

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Towards Domain-Inspired Solutions Tackling Limited and Diverse Data Problems:
Applications in Bioinformatics, Astrobiology, and Biometrics

Permalink

<https://escholarship.org/uc/item/4xq364nk>

Author

Jonnalagedda, Saisri Padmaja

Publication Date

2024

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Towards Domain-Inspired Solutions Tackling Limited and Diverse Data Problems:
Applications in Bioinformatics, Astrobiology, and Biometrics

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Saisri Padmaja Jonnalagedda

December 2024

Dissertation Committee:

Dr. Bir Bhanu, Chairperson

Dr. Xiaoping Hu

Dr. Vagelis Papalexakis

Copyright by
Saisri Padmaja Jonnalagedda
2024

The Dissertation of Saisri Padmaja Jonnalagedda is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

This work and my research journey would not have been possible without the unwavering support and invaluable guidance of many incredible people, all of whom have my heartfelt gratitude. First and foremost, I am deeply grateful to my advisor, Prof. Bir Bhanu, for his support, wisdom, and consistent guidance throughout my PhD journey, which have been instrumental to my growth as a researcher. I am also thankful to him for lending his immense expertise and perspective, and for providing me with incredible opportunities to grow, explore, and collaborate within the research community. I want to extend my heartfelt gratitude to my thesis committee members, Prof. Xiaoping Hu and Prof. Vagelis Papalexakis, for their time, thoughtful feedback, and invaluable guidance that have enriched my work.

I have had the privilege to work on some wonderful projects, alongside incredible collaborators who have profoundly shaped my research. I want to extend my sincere thanks to Prof. Mary Droser, Prof. Chia-en Chang, Dr. Brent Weinberg, Dr. Daniel Schmolze, Michelle Jou, Michelle Rybak, Rachel Surprenant, and Ta (David) I Hung for their exceptional expertise and guidance, whose contribution to this research has been nothing short of invaluable and deeply inspiring. I am also immensely grateful for the constructive feedback, collaborative spirit, and significant contributions of my co-authors, Dr. Taejin L. Min, Dr. Jason Allen, and Dr. Shiv Bhanu, whose insights have greatly enhanced the quality of this work. During these formative years, I was fortunate to work with exceptional researchers in the industry during my internships, which were truly transformative experiences. I am especially thankful for the mentorship of Dr. Xiao Xiao, Dr. Kumar Sricharan, Dr. Ka-

malika Das, and Dr. Anton Schwaighofer, who have taught me valuable lessons not only in research but also in professional growth.

I want to extend my sincere gratitude to Prof. Amit Roy-Chowdhury, Prof. Tofigh Heidarzadeh, and Kim Underhill for their immeasurable support, kindness, and encouragement throughout my life at UCR, which has been a source of strength and motivation. A heartfelt thanks to my amazing lab members, Dr. Adam Witmer, Dr. Runze Li, Dr. Henry Liu, Dr. Rajkumar Theagarajan, Xiu Zhang, Ankit Jain, Rakesh Kumar, and Pratyay Dutta, who have been an incredible source of support, camaraderie, and invaluable feedback through the years, making this journey a truly collaborative and enriching experience.

My research journey has been possible due to the inspiration of many, and I would like to acknowledge the impact of my previous thesis advisors Prof. Ashwin Kothari and Dr. Harinath Garudadri, for encouraging the researcher in me. A special token of thanks to the many individuals whose constant support and encouragement have been instrumental in my journey - Ajay and Puja Panchal, Anuja Deshpande, and my wonderful group of friends from DI, UCR, UCSD, and VNIT. You all made my journey so much brighter and I could not be more grateful to be surrounded by such an incredible group of people.

I owe my deepest gratitude to my parents Dr. Venkatesh Jonnalagedda, and Dr. Megha Jonnalagedda, for being my first teachers, for instilling a sense of research in me, and for being my pillars through thick and thin. I am grateful to my sister Priya Jonnalagedda for her constant unwavering support. I would like to thank my in-laws Sarla Gupta, Adv. Ashok Gupta and my confidante Dr. Amol Gupta for always being there for me. I would like to thank my grandparents and extended family for their continuous encouragement.

I am very grateful to my co-author, best friend, and husband, Dr. Akash Gupta - who has been a steadfast and constant companion. This would not be possible without your love and support.

I would like to thank UCR for their resources. NASA Astrobiology, IARPA, Accenture Federal Services, and NSF for funding my research. This work is fully or in part, a reprint of the following published works:

1. Jonnalagedda, Padmaja, et al. "A Comprehensive Radiogenomic Feature Characterization of 19/20 co-gain in Glioblastoma." IEEE Transactions on Artificial Intelligence (2024).
2. Jonnalagedda, Padmaja, et al. "Computational modeling of tumor invasion from limited and diverse data in Glioblastoma." Computerized Medical Imaging and Graphics (2024)
3. Jonnalagedda, Padmaja, et al. "Searching for Life: End-to-end Automated Detection and Characterization of Ediacaran Biosignatures." IEEE Transactions on Geoscience and Remote Sensing (2024).
4. Jonnalagedda, Padmaja, and Bir Bhanu. "Novel Body Biometric for Long-Range Recognition Under Extreme Conditions." 2023 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2023.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Towards Domain-Inspired Solutions Tackling Limited and Diverse Data Problems:
Applications in Bioinformatics, Astrobiology, and Biometrics

by

Saisri Padmaja Jonnalagedda

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, December 2024
Dr. Bir Bhanu, Chairperson

As the use of machine and deep learning becomes increasingly ubiquitous, there is also a need to address the most common challenges in real-world applications of deep learning. This thesis presents a study of four applications, for building solutions with informed design choices addressing major challenges such as limited and diverse data problems, tackling artifacts, and data imbalance - to name a few.

The first application deals with developing generative models to reproduce and characterize the visual manifestation of unapparent radiogenomic features in MR imaging for a rare mutation in Glioblastoma patients. This research reports three contributions in this study: (1) a novel generative approach that models tumor macro-features in this limited, diverse, and unbalanced dataset, (2) a novel approach to computationally model tumor invasion properties, and (3) a new metric to evaluate GANs.

The second application has an astrobiological utility to study one of the most definitive biogenic signs of early life on Earth, the double-rippled bedforms (DRBs) from the Ediacaran Member (550–555 Ma) of South Australia. We develop automated tools

for this analysis while tackling the challenges of an uncured, limited, and artifact-ridden dataset. We develop a new automated tool for astrobiological investigation that robustly detects miniature biosignatures while tackling these challenges.

The third application addresses the challenge of video-based human recognition using a limited feature set under extreme imaging distortions, imaging range, lack of frames, arbitrary pose, occlusions, air turbulence, and changing clothes. We propose two novel body-based biometrics, one using 2D binary silhouette, and one using 2D RGB images, that can tackle the aforementioned challenges while using minimal information at a time. Extensive evaluation reveals the efficacy of each of these biometrics under varying imaging conditions.

The fourth application is to develop novel generative methods to synthesize full-atom 3D protein conformations for the $A\beta_{42}$ monomer. An intrinsically disordered protein, $A\beta_{42}$ has a highly dynamic nature and a complex feature space. We propose a novel sampling technique to synthesize novel conformations that can capture the high variability in the protein features and has shown to be highly effective in sampling stable, diverse, unique, and biologically useful conformations under varying force fields.

TABLE OF CONTENTS

LIST OF FIGURES	xiii
LIST OF TABLES	xx
1 Introduction	1
2 Radiogenomic Feature Characterization in Glioblastoma MRI using GANs	7
2.1 Introduction	7
2.2 Related Work and Contributions	12
2.2.1 Related Work	12
2.2.2 Contributions	14
2.3 Technical Approach	15
2.3.1 Quantification of Tumor Macro-Features	15
2.3.2 Realistic Radiogenomic Design using Disentanglement in GAN	16
2.3.3 Synthetic Image Fidelity (SIF) Metric	21
2.4 Experiments and Results	25
2.4.1 Description of Datasets	25
2.4.2 Experimental Setup	26
2.4.3 Results: Quantification of Macro-features	30
2.4.4 Results: Qualitative Evaluation	31
2.4.5 Results: Quantitative Evaluation	37
2.4.6 Results: SIF-based Qualitative Evaluation	39
2.4.7 Failure Cases and Limitations	40
3 Computational Modeling of Tumor Invasion from Limited and Diverse Data in Glioblastoma	42
3.1 Introduction	42
3.2 Related Work and Contributions	44
3.2.1 Related Work	44
3.2.2 Contributions	48
3.3 Technical Approach	49
3.4 Experiments and Results	63

3.4.1	Description of Datasets	63
3.4.2	Experimental setup	66
3.4.3	Discriminatory Power for Mutation Detection	67
3.4.4	Generated Image Quality and Diversity	68
3.4.5	Realism of Generated Images	77
3.4.6	Cross-Dataset Testing	80
3.4.7	Ablation Studies	83
3.4.8	Data Augmentation using TI-GAN	88
3.4.9	Limitations of TI-GAN	89
4	Searching for Life: End-to-end Automated Detection and Characterization of Ediacaran Biosignatures	92
4.1	Introduction	92
4.2	Related Work and Contributions	99
4.2.1	Related Work	99
4.2.2	Contributions	101
4.3	Technical Approach	102
4.3.1	Problem Formulation	103
4.3.2	Auxiliary Representation-based Optimization for Region Activation (ARORA).	104
4.3.3	Boundary-Aware Saliency-Encoded Network (BASE-Net)	109
4.3.4	DRB Quantifier	114
4.4	Experiments and Results	116
4.4.1	Dataset: Description and Pre-processing	116
4.4.2	Experimental Setup	117
4.4.3	Qualitative Results	119
4.4.4	Quantitative Results	123
4.4.5	DRB Quantifier	128
4.4.6	Ablation Study	130
5	Novel Body-based Biometrics for Human Recognition under Extreme Imaging Conditions	133
5.1	Related Work and Contributions	138
5.1.1	Related Work	138
5.1.2	Contributions	140
5.2	Technical Approach	141
5.2.1	Outfit Regularizing Biometric (ORB)	142
5.2.2	Recognition Backbone	144
5.3	Experiments and Results	148
5.3.1	Datasets Used	148
5.3.2	Experimental Setup	150
5.3.3	Ablation Studies	158
5.3.4	Training and Inference Details	160
5.3.5	Failure Cases	163

6	Novel Sampling Strategy for Diverse and New Protein Conformation Synthesis using Generative Deep Networks	165
6.1	Introduction	165
6.1.1	Contributions	168
6.2	Materials and Methods	169
6.2.1	Dataset	169
6.3	Technical Approach	172
6.3.1	Problem Formulation	172
6.4	Results	176
6.4.1	Model Validation	176
6.4.2	Model Synthesis	176
6.4.3	Quantify the efficiency of Synthesizing Conformations	178
6.4.4	Biologically important structures in synthetic A β 42 conformations .	180
6.4.5	Relation between the Salt-bridge and toxic conformations	183
7	Conclusions	185
	Bibliography	190

LIST OF FIGURES

2.1	Disentanglement of various macro features in a brain scan. (a) Whole scan image (WSI), (b) WSI showing tumor boundary (visual only, not a separate representation), (c) Tumor shape + texture + location (Shp + Tex + Loc), (d) Tumor shape + texture + tumor microenvironment (Shp + Tex + TME), (e) Delineated tumor mass (Shp + Tex), and (f) Binarized tumor mass (Shp).	11
2.2	Overview of R2D2-GAN. It generates a tumor crop after learning from the shape and texture of a real tumor.	16
2.3	Pipeline to compute the Synthetic Image Fidelity Metric. The process from (*) onwards is separately done for each of the R, G, and B channels – then aggregated to obtain the SIF score. R1 and S1 are shown here for the Red (first) channel. For the GradCam heatmap, the redder color implies a higher significance of that feature, and bluer is less.	23
2.4	Characterizing tumor manifestation in the brain using the distribution of tumor location along the axial, coronal, and sagittal planes conditioned upon mutation status for both datasets. Each color corresponds to a different patient.	29
2.5	Generated Image Quality of R2D2-GAN (vs real data and other models) for Datasets A and B for both classes. We see that most models, including DDPM, suffer from a lack of detail in some tumors in each class. Real (green, solid) vs [a] MSG-GAN (orange, short-dashed), [b] SAGE (dark yellow/long-dashed), [c] FeaD-GAN (purple, dot-dash), [d] StyleGAN2 (red, dot-dot-dash), [e] DDPM (blue, dashed), [f] R2D2-GAN (green, dotted).	29
2.6	Magnified examples of generated images from all GANs. Real: Solid/green, R2D2-GAN: green/dotted, DDPM: blue/dashed, StyleGAN2: red/dot-dot-dash, FeaD-GAN: purple/dot-dash, SAGE: dark yellow/long-dashed, MSG-GAN: orange/short-dashed. These magnified images clearly show the sharpness artifacts in SAGE (orange arrows in Images 1, 3, 4), loss of detail in StyleGAN2 (Images 2, 3, 4), and checkered artifacts in MSG-GAN (images 3 and 4). We note some blurring in DDPM (Images 2 and 4). We note that for all images, R2D2-GAN overcomes the persistent blurriness problem of FeaD-GAN.	32

2.7	Diversity of images generated by R2D2-GAN using Dataset B (vs DDPM, StyleGAN2, and MSG-GAN). Best viewed zoomed in. MSG-GAN shows high-frequency artifacts and a lack of tumor details. StyleGAN2 has better quality but lower diversity. DDPM has some diversity but also has some very bright or very dark images generated due to instability in training. R2D2-GAN: green/dotted, DDPM: blue/dashed, StyleGAN2: red/dot-dot-dash, MSG-GAN: orange/short-dashed.	33
2.8	Visualizing GradCAM outputs in the computation of SIF score. Similar activation maps demonstrate a GAN’s ability to learn high-level complex feature associations that correlate with the unapparent visual manifestations of the presence or absence of a mutation. We note that R2D2-GAN highlights similar regions as real data, thereby establishing its efficacy in learning crucial features. Colorbar for heatmaps is provided in Figure 2.3.	36
2.9	Cases where R2D2-GAN fails to learn the same features as real data. The images were selected based on low SIF scores, demonstrating an additional benefit of the SIF score in identifying failure cases in GANs. Colorbar for heatmaps is provided in Figure 2.3.	41
3.1	Overview of the Tumor Invasion GAN (TI-GAN) Model. The detailed architecture of each TI-GAN module is given in Figure 3.3. The definition of each variable is provided in Table 3.1. The visual representation of the major variables is provided in Figure 3.2.	45
3.2	Visual representation of the variables used in TI-GAN. For the context images, i_+ means the brain scan above the current slide i in the direction of the top of the head and i_- means the slice below, towards the neck. The white arrows represent where the tumor invasion is being represented by the aggregation described in Figure 3.1. The blue diffused region in i_+ shows the overall tumor invasion as we go in the direction of the neck. Similarly, the red area in i_- represents the same as we go towards the top of the head. As pointed out by the yellow arrow in i_- , the red region also represents the eyes we encounter when traveling in that direction. Thus, this context provides an estimate of the location of the tumor in this direction.	47
3.3	(left) Detailed Architecture of the Encoders, Decoders, and Discriminators of the proposed TI-GAN model. (right) Details of the Self-Attention (SA) and Short-Long Term Attention (SLTA) modules. For the Encoder ($\mathcal{E}_\phi$), input 1 is the Encoder feature, and input 2 is the context from $\mathcal{E}_\chi$ (described in Section 3.3). For the decoder ($\mathcal{D}_\phi$), Input 1 is the feature corresponding to ϕ , and Input 2 corresponds to ψ , which helps in joint estimation. Vice versa for $\mathcal{D}_\psi$, as described in Section 3.3	54
3.4	(a) Computation of the volumetric localization context (χ). (b) Visual representation of how the χ varies with varying 2D brain slices for the same patient. Images (1) and (8) represent the topmost and bottom-most slices of tumor occurrence. We note that the extent of the red/blue hue surrounding the tumor is indicative of the relative position and global growth patterns of tumor invasion.	55

3.5	Qualitative Analysis of the performance of TI-GAN. The images represent the intra-class diversity of the real data and demonstrate the capability of the proposed approach in capturing the diversity during synthesis. The results are obtained using the models trained on an 80-20 patient-wise data split. For both datasets and classes, the results of the proposed approach are very similar to real data, thus establishing the efficacy of TI-GAN modeling. . . .	65
3.6	Qualitative evaluation of the quality of generated tumor invasion. Results are compared against real data as well as WGAN-GP, PGGAN, MSG-GAN, and Generative Inpainting models. We zoom into the site of the tumor region to highlight the efficacy in reproducing the tumor invasion. Row-wise: [a] WGAN-GP, [b] PGGAN, [c] MSG-GAN, [d] Generative Inpainting, and [e] TI-GAN. We note blurriness in many of the WGAN-GP and PGGAN images whereas MSG-GAN suffers from high frequency artifacts. We also note some unrealistic skull patterns in GI results.	70
3.7	Visual Confusion Matrix for Radiologist Blind Test. True Positives are when the blind image was real and was predicted by radiologists as real. False Positives are when the blind image was synthetic but was predicted as real, False Negatives are when the blind image was real but predicted as synthetic and True Negatives are when the blind image was synthetic, and radiologists predicted as synthetic. The white text indicates the predictions by three radiologists (R1, R2, R3). * represents the image that was in the blind set.	79
3.8	Cross-dataset or cross-domain testing performance of TI-GAN. We synthesize the tumor invasion of a dataset using the trained model from the other dataset. We report these results adjacent to the same-domain evaluation for comparison. For cross-dataset testing, models are trained on all the data of the dataset in training, and tested on all images of the other dataset.	81
3.9	Ablation Study: Comparing the efficacy of TI-GAN (row 1, green, solid) in modeling the tumor microenvironment without local context (no tumor growth information in the form of ψ_{tumor}), (shown row 2, orange, dotted) and without global context or Volumetric Localization Context χ (shown row 3, blue, dashed). The first set of three rows shows the overall brain scan with synthetic tumor invasion, whereas the second set of three rows zooms into the tumor invasion region for a closer look at the generated results.	85
3.10	Failure cases of TI-GAN, identified using various factors such as SSIM, PSNR, MSE values, and visual inspection. We selected examples of various types of errors seen during TI-GAN synthesis. Some examples show an incorrect estimation of the skull boundary, whereas some show incorrect CSF or brain tissue estimation. In some cases, we note that the tumor growth boundary is incorrectly estimated. We see that in many of the cases, we do not see much local context due to relative tumor size being large and the masking being extensive. We note from ablation studies that the local context significantly affects the overall TI-GAN performance - which is also reflected in the failure cases here.	90

4.1	Depositional processes and resulting sedimentary packaging in mat-dominated systems (e.g., the Ediacaran Period) and in depositional systems with no organic mat growth. A) Deposition of sands on top of mat-bound sediment wherein the original topography of the seafloor (1) is not eroded by newly deposited sand (2). Resulting sedimentary packaging consists of stacked sandstone beds, with no fine-grained interbeds, preserving rippled tops and bases (i.e., Double Rippled Bedforms). B) Deposition of sands on top of sediment that is not stabilized by organic mats wherein the original topography of the seafloor (1) is eroded by newly deposited sand (2); eroded ripple crests are demarcated by red shading. The absence of an organic mat leads to erosion of ripples and amalgamation of successive event beds, resulting in amalgamated sandstone event beds bound by erosional surfaces.	94
4.2	Top (dashed): Annotated cross-sectional view of a stack of double-rippled bedforms (DRBs) from the Ediacara Member. DRBs are labeled by the green bracket. Individual rippled bed junctions are denoted by blue highlights. DRB thickness, reflecting the thickness of a single bed bound by non-erosional rippled bed junctions, is denoted by the red arrow. The yellow bracket denotes length of a single ripple (half-wavelength in DRB Quantifier) and is defined as the length encompassed by two local low points (trough = t) surrounding a local high point (crest = c). Scale: 8 cm. Bottom (a-j): Terrain diversity in our dataset showing variations in lighting conditions, fields-of-view, and artifacts. Scales a-j: 13 cm.	96
4.3	Overview of SPACESeg2.0 with three plug-and-play modules - the “Auxiliary Representation-based Optimization for Region Activation” (ARORA) module for Region Activation and (adaptive) Data Augmentation, the segmentation model “Boundary-Aware Saliency-Encoded Network” (BASE-Net) for robust, terrain-aware segmentation and the “DRB Quantifier Tool” (DRBQ) module for characterization of detected potential biosignatures. In DRBQ “Characterization”, δ (orange): distance between ripples, A (green): amplitude, λ (red): half-wavelength and yellow markers: number of ripples. . . .	103
4.4	Overview of the ARORA module. The dual-tasking ARORA utilizes a pseudo-image representation to generate data-driven context and perform adaptive data augmentation.	105
4.5	Visual demonstration of each step of ARORA (top row), each auxiliary channel of ARORA (bottom row). We show the output of ARORA (σ_0) overlaid on the original image for better visualization. This is labeled $I + \sigma_0$. We follow this representation throughout, as a binary mask is difficult to interpret.107	
4.6	Overview of BASE-Net Architecture (left). Architectures of individual BASE-Net modules (right).	109
4.7	Overview of the DRB Quantifier Tool	115

4.8	Performance of SPACESeg2.0 in detecting DRBs, compared with the performance of PSPNet, DeepCrack, ESPNet, and SPACESeg. Results reported for Cohort A and B. The images are chosen to demonstrate a range of number and size of rippled bed junctions. It can be seen that SPACESeg2.0 is robust to over- and under-segmentation compared with other models. . . .	120
4.9	Performance of SPACESeg2.0 in detecting DRBs from difficult terrains, vs comparative models. The difficulty of terrain is empirically chosen via the presence of higher number of artifacts, varied imaging conditions, low relative DRB thickness, high number and/or closely-spaced rippled bed junctions. .	121
4.10	Cross-testing performance of SPACESeg2.0 in detecting DRBs from Cohort B, when trained using Cohort A. Comparison with PSPNet, DeepCrack, ESPNet, SPACESeg.	123
4.11	Results showing metrics computed using the DRBQ module. The results using ground-truth ripples are also shown as reference to evaluate the precision of segmentation results of SPACESeg2.0. A high degree of similarity with ground-truth statistics means that SPACESeg2.0 can be used for DRB characterization. For cases with complex DRB such as (c), we note that the plots don't match exactly, but the range of values for each metric is consistent - which is key in characterization.	125
4.12	Ablation study demonstrating the impact of the Auxiliary Image Representation (AIR) in detecting difficult rippled bed junctions. The results are compared against other unsupervised techniques. ARORA employs HLFRA method with Gradient Optimization (GO). The impact of AIR can be seen when HLFRA+GO uses an RGB image input instead of AIR input. Meanwhile, CORA [202] uses a different technique to compute AIR and a different active contour optimization. It can be seen that ARORA is more robust against complex imaging conditions than CORA, such as in image 3, where CORA cannot identify many regions of interest due to the complexity of the terrain. In images 4 and 5, where the rippled bed junctions are finer and closer together, there is clumping in CORA, whereas ARORA can tackle it. In isolated beds such as in images 1 and 2, we note that CORA highlights the marker artifact as well, which ARORA almost fully mitigates.	129
4.13	Ablation study demonstrating the effect of Hierarchical Context Fusion and Boundary or Context Aware Pooling (BCAP) on detecting rippled bed junctions, compared against the full model. The results without context are obtained by training BASE-Net without context input and the results without BCAP are obtained by training BASE-Net without both boundary-aware and pyramid pooling. Images are chosen for cases with higher number of bed junctions that are close together and/or images with bed junctions that look like the terrain.	130

5.1	Examples depicting the robustness of ORB across various challenging imaging conditions. (a-c) Indoor examples, with pose variations, articulation, and camera elevation. (d-e) Similar poses, with and without articulation at 200m imaging distance. (f) Similar articulation, extreme imaging condition for imaging range of 500m. (g-h) Close-range outdoor imaging, with walking, articulation, and occlusion. (i-j) Pose variation, with lighting and turbulence challenges at a 100m distance. (k-l) Similar pose, with extreme imaging challenges such as turbulence and occlusion at varying imaging distances of 400m and 200m, respectively. (m-n) Examples of controlled indoor walking video frames used in the gallery. Images (a-c) and (m,n) are indoor, set 1 clothing gallery images. Images (d-l) are outdoor set 2 clothing probe images. All images are taken from the BRIAR dataset. The subject has authorized the use of their images.	135
5.2	Left: Overview of the computational pipeline for the Outfit Regularizing Biometric (ORB). A description for each step is provided in the technical section. Right: Details on the computation of shape-based biometric used in Step (b) "Distance Transform + Difference of Gaussians" of ORB computation. While D_4 is used in ORB computation, B_{skel} , a fusion of D_0 and D_8 is used as an additional modality in some experiments during score fusion in image matching.	137
5.3	Overview of the TransFuse recognition pipeline. Each recognition backbone is made up of a Swin transformer (which can be replaced with any SOTA recognition network). The optimization function details are provided in the technical section. For image matching, image-level matching is done for all public datasets (image-based datasets). For the video-based BRIAR dataset, the image matching protocol is provided in Figure 5.4. Each recognition channel (until cosine score computation) can be added or removed as needed, making the overall pipeline plug-and-play, and modular.	141
5.4	Gallery aggregation and probe video aggregation process in image matching step for the BRIAR data. For the public datasets, simple image-to-image matching is done as they are image-based datasets.	148
5.5	Plots and trends for the performance of TransFuse on BRIAR's standard Evaluation Protocols (EP) EP3.1 (for Phase 1) and EP4.2 (for Phase 2). They represent the plots for verification and closed-set performance on the face-restricted subset of BRIAR. The first two plots are results for the EP3.1 protocol, while the remaining two represent performance on EP4.2.	155
5.6	Failure Cases in BRIAR predictions. We use examples where one or two modalities fail. A detailed explanation is provided in Section 5.3.5.	161
6.1	(A) synthetic conformation of A β 42 with all 4 turns A-D; (B) Classical internal bond-angle-torsion (BAT) coordinate representation (Z-matrix) for a small molecule. Atoms 1-3 are termed root atoms for this molecule. Atoms $i > 3$ are defined by (b_i, θ_i, ϕ_i) , where b, a and θ are bond, angle, and torsions, respectively. For example, atom 4 is presented by (b_4, θ_4, ϕ_4)	170

6.2	Overview of ProFormX architecture. The input X is real data and $\hat{X}$ is noise. The structure label consists of information on structure type and forcefield. This information can be used to generate desired structure types during synthesis.	172
6.3	Model validation using the reconstruction mode in ProFormX. The left side consists of distributions of all sidechain torsion angles “phi”, and the right side shows the distribution of “phi”. The blue plot is predicted structures, and the orange is “real”. The results are reported for 1Z0Q in the 14SB force field.	177
6.4	Model validation using the reconstruction mode in ProFormX. The left side consists of distributions of all sidechain torsion angles “phi”, and the right side shows the distribution of “phi”. The blue plot is predicted structures, and the orange is “real”. The results are reported for 2NAO in the IDPS force field.	177
6.5	Free energy profile (color bar) of RMSD (y-axis) vs radius of gyration (R_g , x-axis) for A β 42. The units of the free-energy landscapes are kcal/mol. . .	180
6.6	(A) Free energy profile of MD run4 and Model 4U. (B) Most dominant conformations select from the free energy profile (purple circle) (C-E)	181
6.7	(A) Free energy profile of MD run2 and Model 4 with ffIDPSFF force field (B) Crystal Structure A β 42 fibril (PDB: 2BEG) (C) Novel Synthetic conformations from Model 4 has similar conformation as (B) found in red square (D) crystal structure of hairpin conformation that form A β 42 fibril from human brain (PDB: 7Q4M). (E) Novel Synthetic conformations from Model 4 have similar conformation as (D) found in the red square.	182
6.8	(A) Crystal structure of hair pin conformation that form A β 42 fibril from human brain (PDB: 7Q4M). (B-D) Novel synthetic structures are obtained from Model 1 with ff14IDPSFF force field. (B) Type II A β 42 fibril with F20 rotate inward and F19 rotate outward. (C-D) Intermediate states that require minimal conformational rearrangement to form type I or type II A β 42 fibril.	183
6.9	Novel synthetic conformations from Model 1 and Model 2. (A) D23-K28 salt-bridge without hydrophobic interaction. (B-C) Tightly pack hydrophobic interaction with D23-K28 salt-bridge forming inward of Turn C (D-E) Tightly pack hydrophobic interaction with D23-K28 salt-bridge forming outward of Turn C (F) Distinct hydrophobic interaction formin at C-terminus with D23-K28 forming outward of Turn C.	184

LIST OF TABLES

2.1	Mutation detection in Dataset A and B using the macro feature representations as shown in Figure 2.1. ResNet18 is used with a learning rate of 0.0001, batch size 16, and Adam optimizer with step decay. The train/test split is 80/20, all images are resized to 256x256, and metrics are reported on 10-fold cross-validation as mean (standard deviation). Metrics are - ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. IL results are calculated for every individual slice and PL results are calculated by a majority vote of IL predictions of all patient scans.	28
2.2	Fréchet Inception Distance (FID) Score, modified Inception score (mIS), and Synthetic Image Fidelity (SIF) score for generated images using R2D2-GAN, DDPM, StyleGAN2, FeaD-GAN, SAGE, and MSG-GAN using 500 synthetic images for each.	34
2.3	Mutation detection in Dataset A and B using only synthetic tumor crops generated using R2D2-GAN, DDPM, StyleGAN2, FeaD-GAN, SAGE, and MSG-GAN. All GANs are trained using the same real training sets and tested on the same corresponding real test sets for all 10-folds. The metrics are reported as mean (standard deviation) over 10 folds. ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. The model trained using real data is used as the baseline.	35
3.1	List of Symbols used	49

3.2	Mutation detection in Dataset A and B using only real data generated using whole images, tumor regions, and only tumor crops. Metrics are - ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. WI: Whole Images, TR: Tumor Region, TC: Tumor Crop. The models are trained using an 80-20 patient-wise split of the data, using a ResNet18 classifier. The metrics reported are mean (standard deviation) over 10 folds. WI outperforms TR since it also contains location information. Comparing the three accuracies provides us with a quantification of the impact of tumor invasion/microenvironment on mutation detection.	64
3.3	Novel Synthesis Metrics: Fréchet Inception Distance (FID) Score and modified Inception score (mIS) for generated images using TI-GAN, Generative Inpainting (GI) [494], MSG-GAN [81], PGGAN [213] and WGAN-GP [143]. The number of generated images per class is 500 for each GAN. For our model, Bold : Best performance, <u>Underline</u> : Second best. <i>Con</i> stands for “Control” class, <i>Mut</i> for Mutated class, and <i>Avg</i> is the average across both classes. Results on Whole Images and Tumor Regions are provided. Whole Images refer to the full 2D brain scan, whereas Tumor Region refers to the tumor and surrounding tissue. Whole Images and Tumor Regions are shown in Figure 3.6.	71
3.4	Exact Synthesis Metrics: Learned Perceptual Image Patch Similarity (LPIPS), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Metric (SSIM) for generated images using TI-GAN, Generative Inpainting (GI) [494], MSG-GAN [81], PGGAN [213] and WGAN-GP [143]. The number of generated images per class is 500 for each GAN.	72
3.5	Mutation detection in Dataset A and B using only synthetic data generated using TI-GAN, Generative Inpainting (GI) [494], MSG-GAN [81], PGGAN [213], and WGAN-GP [143] All GANs are trained using the same real training sets and tested on the same corresponding real test sets for all 10-folds (80-20 splits). The classifier used is ResNet18. The metrics are reported as mean (standard deviation) over 10 folds. Performance using real data is used as the baseline. For real data in the unbalanced Dataset B, the minority class is oversampled using data augmentation techniques such as rotation, flipping, etc. All metrics are in percentages. ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. Results reported for classifiers trained on WI: Whole Images, and TR: Tumor Region. We note that the TI-GAN-generated dataset is a good representation of the real data, which is reflected in the classification performance.	73
3.6	Blind test by radiologists to distinguish between real and TI-GAN generated tumor microenvironment. Metrics shown are ACC - Accuracy, PRE - Precision, REC - Recall, FPR - False Positive Rate, and FNR - False Negative Rate. The text in bold is the mean value across all radiologists.	78

3.7	Cross-dataset Testing performance metrics: FID Score, mIS, SSIM, LPIPS, and PSNR for generated images using TI-GAN. Results are shown for the model trained on Dataset B, and tested on Dataset A (Training: B, Testing: A) and vice versa (Training: A, Testing: B). We report results on Whole Images and Tumor Regions, as shown in Figure 3.6. It can be noted that cross-dataset testing results are comparable to same-dataset testing results, demonstrating the generalizability of TI-GAN. Results from same-dataset testing are used for baseline comparison.	82
3.8	Quantitative metrics for Ablation experiments. Performance is evaluated when TI-GAN is trained without local context (Planar Localization Context) and global context (Volumetric Localization Context). We report results on Whole Images and Tumor Regions, as shown in Figure 3.6. While performance degrades when either context is removed, it can be seen that local context (tumor properties) has a stronger impact on TI estimation.	86
3.9	Mutation detection in Dataset A and B using only synthetic data generated using standard TI-GAN (Model Std), using cross-dataset testing (Model CT), without local context (Model L), and without global context (Model G). The classifier used is ResNet18. The metrics are reported as mean (standard deviation) over 5-fold cross-validation, with train/test split to be 80/20. Performance using the standard setup of TI-GAN (Model Std) is used as the baseline. All metrics are in percentages. ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses.	87
3.10	Mutation detection in Dataset B using synthetic data from TI-GAN to balance the unbalanced real data. We do this by oversampling the minority mutated class. We compare these results against Standard Data Augmentation (SDA) techniques such as rotation, flipping, etc., and data balancing using comparative GANs. Metrics are - ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. The models are trained using an 80-20 patient-wise split of the data, using a ResNet18 classifier. The metrics reported are mean (standard deviation) over 10 folds.	89
4.1	Quantitative performance of our proposed SPACESeg2.0 vs other deep segmentation networks for Cohort A. Metrics - ACC: Accuracy, BAC: Balanced Accuracy, PRE: Precision, REC: Recall, F1: F1/Dice Score, JSC: Jaccard or IoU score, BFSC: Boundary F1-score, and HD: Hausdorff Distance. Metrics are in % and reported as <i>mean (std)</i>	124
4.2	Quantitative performance of our proposed SPACESeg2.0 vs other deep segmentation networks for Cohort B. Metrics - ACC: Accuracy, BAC: Balanced Accuracy, PRE: Precision, REC: Recall, F1: F1/Dice Score, JSC: Jaccard or IoU score, BFSC: Boundary F1-score, and HD: Hausdorff Distance. Metrics are in % and reported as <i>mean (std)</i>	126

5.1	Comparative analysis of the top state-of-the-art models' performance on the PRCC dataset, with our proposed model. We report the results of our model using the Swin backbone and the MGN backbone. Bold represents the best overall result, and <u>Underlined</u> represents the second best result. Results are reported for general (including similar clothing), and clothing change setups. Results are divided based on RGB-only models (listed in chronological order) in rows 1-4; followed by multi-modal models (listed in chronological order) in rows 5-14.	151
5.2	Comparative analysis of the state-of-the-art models' performance on the VC-Clothes and Celeb-ReID datasets, with our proposed model. We report the results of our model using the Swin backbone, as well as the MGN backbone. Bold represents the best overall result, and <u>Underlined</u> represents the second best result. Results are reported for general, similar clothing, and clothing change setups for VC-Clothes, and standard clothing change setup for Celeb-ReID. Results are divided based on RGB-only models (listed in chronological order) in rows 1-5; followed by multimodal models (listed in chronological order) in rows 6-7.	152
5.3	Performance of individual modalities and their fusion across all public datasets using Swin backbone (TransFuse). Bold represents the key performance trend to observe, whereas <i>Italics</i> represent the overall TransFuse results. . .	153
5.4	Performance of individual modalities and their fusion on the most recent BRIAR subsets (BRS 3 and 4). This consists of 363 subjects, with 273 subjects in training and 90 in testing. Bold represents the best overall performance, and <u>Underline</u> represents best standalone modality. SL: Score-level Fusion. Matching is done on the standard BRIAR task, with set 1 clothing indoor data as gallery and set 2 clothing outdoor data as probe.	154
5.5	BRIAR Evaluation Protocol Results for EP3.1 and EP4.2. The metrics of interest are closed-set Rank-20 performance, and verification performances - TAR @ 1% for EP3.1, and TAR @ 0.1% FAR for EP4.2. The target metric for the Face Restricted Treatment (FRT) subset for verification is provided in (bold, parenthesis).	155
5.6	Impact of changing clothes on the recognition task. Experiment done on a subset of 363 subjects from BRIAR, indoor-indoor matching, with changing clothes between gallery and probe. Train/test split: 273/90 subjects. SL: Score-level fusion. Bold represents the best overall performance, and <u>Underline</u> represents the best standalone modality.	156
5.7	Impact of a changing environment (indoor gallery to outdoor probes) with similar set clothing for RGB images. Experiments are done on a subset of 363 subjects from BRIAR data, with 273 subjects in training and 90 subjects in evaluation. Bold represents the best overall performance, and <u>Underline</u> represents the best standalone modality.	157

5.8	Results with varying ranges using TransFuse. The model is trained on all ranges and evaluated for varying ranges of probes. The number of probe frames in close-range were 5663, in 100-300m were 11,732, in 300-600m were 10,727, and in long-range/UAV were 6015. Bold represents the best performance.	157
5.9	Performance of Gait Algorithms using limited number of frames, for Gait3D and OU-MVLP datasets vs SSRNet that uses only single frame. Results for comparative models are from [514] for Gait3D and [490] for OU-MVLP. SL: Score-level fusion. Top results are <u>Underlined</u> , our best results are in Bold	159
5.10	Experiments to determine the optimal backbone. Experiments were conducted on RGB modality, using 363 Briar subjects (273 in training, and 90 in evaluation). Bold represents the best performance.	159
5.11	Experiments to determine the optimal fusion technique. Experiments were conducted on the fusion of ORB, RGB, and B_{skel} , using 363 Briar subjects (273 in training, and 90 in evaluation). Bold represents the best performance.	160
6.1	Details of MD simulation and Training/Validation of ProFormX. Structure Type: Force field and Initial Structure; Tr/Val: Training/Validation; Avg RMSD: Validation of reconstructed conformations, average torsion RMSD (All heavy atoms /Backbone)	171
6.2	Refining the novel synthetic conformations of different models. Type of synthesis: C=Controlled/Perturbation-based Synthesis; U=Uncontrolled Synthesis. Step 1: Number of conformations from initial interpolation in the latent space. Step 2: Number of conformations within energy cutoff. Step 3: Number of distinct conformations (removing self-repeats). Step 4: Number of novel synthetic conformations (repeat with MD confs eliminated).	179

Chapter 1

Introduction

The recent and rapid advancements in machine and deep learning have revolutionized numerous fields, including but not limited to healthcare, astronomy, finance, autonomous systems, etc. As these technologies become increasingly ubiquitous, their deployment in real-world applications requires careful consideration of a myriad of challenges and subsequent design choices and solutions. Unlike controlled experimental setups, real-world scenarios often present complex issues such as limited data availability, diverse and heterogeneous datasets, the presence of artifacts, and significant data imbalance. Addressing these challenges is critical to ensuring that machine learning models are not only accurate but also robust and generalizable in their applications. This dissertation delves into these challenges by studying four distinct applications, each tackling issues encountered in practical deployments of deep learning. By focusing on informed design choices, the research emphasizes the importance of tailoring solutions to the unique demands of each application. Key areas of exploration include strategies for overcoming limited and diverse data prob-

lems, designing mechanisms to detect and mitigate the influence of artifacts, and addressing the pervasive issue of data imbalance. These challenges, while prevalent across domains, often require domain-specific insights and innovative methodologies to overcome effectively.

The first chapter details the work in characterizing tumor macro-features in the radiogenomic analysis of Glioblastoma using axial FLAIR MRIs. For diseases with high morbidity rates such as Glioblastoma Multiforme, the prognostic and treatment planning pipeline requires a comprehensive analysis of imaging, clinical, and molecular data. Many mutations have been shown to correlate strongly with the median survival rate and response to therapy of patients. Studies have demonstrated that these mutations manifest as specific visual biomarkers in tumor imaging modalities such as MRI. To minimize the number of invasive procedures on a patient and for the overall resource optimization for the prognostic and treatment planning process, the correlation of imaging and molecular features has garnered much interest. This work detects and characterizes the impact of tumor biomarkers (such as shape, texture, and location) in detecting a prognostic mutation - the concurrent gain of 19 and 20 chromosomes, and proposes two novel ideas for this analysis. First, to address the challenges associated with the limited, diverse, and complex nature of medical data, this work proposes a novel generative model - the Realistic Radiogenomic Design using Disentanglement in Generative Adversarial Networks (R2D2-GAN), designed to recreate highly subtle, unapparent manifestations of mutations in Magnetic Resonance Imaging. It generates high-resolution, diverse data that captures the discriminatory visual features of the molecular markers while tackling the high diversity, unbalanced, and limited GBM data with rare mutations correlating with patient survival such as 19/20 co-gain. Second,

this study proposes a quantitative metric called the Synthetic Image Fidelity (SIF) Score to evaluate the performance of GANs in learning visually unapparent prognostic features through the use of gradient-based model explanations.

The second chapter extends the radiogenomic analysis and feature characterization to develop a novel generative model to perform computational modeling of tumor growth properties. While the tumor mass is the most significant feature, the impacted tissue surrounding the tumor is also a significant biomarker contributing to the visual manifestation of mutations — which has not been studied as extensively. The pattern of tumor growth impacts the surrounding tissue accordingly, which is a reflection of tumor properties as well. Modeling how the tumor growth impacts the surrounding tissue can reveal important information about the patterns of tumor enhancement, which in turn has significant diagnostic and prognostic value. This work presents the first work to automate the computational modeling of the impacted tissue surrounding the tumor using generative deep learning. The work isolates and quantifies the impact of the Tumor Invasion (TI) on surrounding tissue based on change in mutation status, subsequently assessing its prognostic value. Furthermore, a TI Generative Adversarial Network (TI-GAN) is proposed to model the tumor invasion properties. Extensive qualitative and quantitative analyses, cross-dataset testing, and radiologist blind tests are carried out to demonstrate that TI-GAN can realistically model the tumor invasion under practical challenges of medical datasets such as limited data and high intra-class heterogeneity.

The third chapter presents a novel application of deep learning in astrobiology. With state-of-the-art imaging and analytical tools onboard the NASA Perseverance Rover

Mission, geological information for remote astrobiological analysis is readily available and more widespread than ever. For analysis of such data, there is a need for automated, remote assessment tools, capable of translational and scalable research. To search for evidence of life in the universe, one of the goals of astrobiological research, we must first recognize robust indicators of life on Earth. One method is the study of Earth’s early fossil record, dominated by simple organisms similar to that predicted to have existed on Mars - if life ever did develop. We identify a potential morphological biosignature, known as double-rippled bedforms (DRBs), from the Ediacaran Member (Rawnsley Quartzite, 550–555 Ma) of South Australia. We utilize these DRBs to develop a new tool for astrobiological investigation via an improved end-to-end remote, automated biosignature assessment tool, Scene-aware Perception Automation using Composite Embedding for Segmentation (SPACESeg2.0). The proposed tool accurately detects and quantitatively characterizes DRBs in imagery from varied fields of view, with practical artifact-ridden data, even when the DRBs constitute less than 1% of the total image. We establish the efficacy of SPACESeg2.0 in detecting desired structures compared with other deep-learning techniques.

In the fourth chapter, we address the practical imaging challenges faced in human recognition. Body-based Human recognition faces increased complexity due to a myriad of challenging factors such as variations in imaging platforms, sensors, imaging distance, changes in clothing, arbitrary human poses, occlusion and articulation of the human body, and air turbulence. These conditions often make it difficult to discern facial features, particularly at long distances, where changes in attire also significantly contribute to recognition errors. For videos, the commonly used features such as gait are subject to challenges such as

distortions, occlusions, and lack of frames. For images, the most commonly used RGB and shape-based features, suffer significant drawbacks under such conditions as well. To overcome these challenges, this work introduces two novel biometrics - a shape-based biometric (B_{skel}), and a unified Outfit Regularizing Biometric (ORB). For B_{skel} , this work presents the first study that utilizes single-frame binary silhouettes and their auxiliary representations for human recognition under extreme distortions. The proposed representation is compact, modular, and robust to distortions, allowing for easy deployability for long-range recognition. For ORB, the biometric integrates shape, appearance, and soft biometric features into a unified representation called the Outfit Regularizing Biometric (ORB). Through extensive experimentation, we demonstrate that the ORB representation effectively addresses some of the most significant obstacles in human recognition. We report our findings using the rigorous evaluation protocols of the BRIAR datasets (Phase 1 and 2), alongside widely used clothes-changing datasets such as PRCC, VC-Clothes, and Celeb-ReID. Our analysis using the BRIAR dataset enables us to isolate and examine the impact of various artifacts, including changes in clothing, environmental conditions, and varying imaging distances. We achieve superior performance by employing the ORB representation, in combination with appearance and shape-based features, and leveraging a transformer-based backbone (TransFuse) across different challenging scenarios, consistently outperforming traditional RGB-based systems, delivering nearly a 15% improvement in accuracy.

The fifth chapter discusses the applicability of generative models in synthesizing full-atom 3D protein structures. Proteins exhibit dynamic conformational ensembles that are critical for their biological function. Intrinsically disordered proteins (IDPs), such

as amyloid-beta 42 ($A\beta_{42}$), present a particular challenge due to their transient but stable conformations, which are often implicated in disease-related processes like aggregation. Understanding and sampling these ensembles remains a significant experimental and computational challenge. In this work, we introduce a novel generative framework, ProFormX, which combines principles from Autoencoders (AE), Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs) to explore the conformational landscape of $A\beta_{42}$. Our model incorporates a novel sampling technique that enhances the generation of diverse and biologically relevant conformations. By leveraging molecular dynamics (MD) simulation data, the model learns the latent conformational space of $A\beta_{42}$, enabling the generation of novel synthetic structures with sophisticated backbone and sidechain arrangements. The application of this method to the highly dynamic $A\beta_{42}$ monomer revealed conformational clusters distinct from the ones seen in training data, providing insights into its aggregation-prone states and potential regulatory mechanisms. The generated synthetic conformations also uncovered distinct structural features that have been shown to be biologically relevant. The proposed framework is highly transferable and can be applied to diverse protein systems, making it a powerful tool for studying protein conformational dynamics. The proposed sampling technique can be used with any type of generative model, making it highly generalizable as well. This work underscores the potential of deep learning in advancing protein design and offers a robust approach for generating biologically significant protein conformations with direct relevance to therapeutic development.

Chapter 2

Radiogenomic Feature

Characterization in Glioblastoma

MRI using GANs

2.1 Introduction

Glioblastoma Multiforme (GBM) is a common and aggressive form of malignant brain tumor in adults. A comprehensive assessment of GBM typically involves a combination of clinical, genomic, and multi-modal imaging data. While Magnetic Resonance Imaging (MRI) is often the first step and gold standard in the diagnosis of brain tumors, gene sequencing tests are the most trusted and accurate in detecting mutations that correlate with patient prognosis and response to therapy [278]. The field of individualized therapies targets these anomalies to optimize response and patient survival, but tumor het-

erogeneity presents many roadblocks. The World Health Organization (WHO) emphasizes molecular indicators in deciding the prognosis and response to therapy of GBM [290]. These indicators include rare mutations such as the concurrent gain of 19/20 chromosomes, which correlate with the median survival of the patient. However, not only does molecular testing have limited availability and is expensive, but also invasive, non-localized biopsies could give incorrect assessments due to tumor under-sampling while levying a resource overhead. Compounding this with the increasing number of cases, limited clinical resources and expertise, external factors, processing times, human errors, etc., this task becomes more difficult. These factors can delay the treatment planning process. On the other hand, MRI data is easily available since it is often the first step of diagnosis. It is also non-invasive and provides a complete view of the tumor.

To reduce the number of invasive procedures that a patient undergoes, as well as to optimize the time, expense, and other resources involved in the assessment process, it is highly desirable to have automated techniques that can model and detect the mutations corresponding to prognosis and response to therapy from only imaging data. Radiogenomics, the study of correlating imaging and molecular features of a tumor, is therefore, garnering much attention, due to its potential to streamline the prognostic and treatment planning pipeline [336]. A molecular aberration (or mutation) affects tumor biology and its radiological appearance. This change in the appearance of the tumor mass and brain tissue due to a change in mutation status is indicative of the radiogenomic features. Therefore, the tumor mass properties such as the tumor’s shape, enhancement, necrosis, edema, location, laterality, etc. are considered the most significant in radiogenomics [20, 112, 134]. From an

automated radiogenomic analysis point-of-view, these translate to image analysis features such as tumor shape, texture, location, and the tumor microenvironment. This work is organized from the perspective of analyzing these features. We isolate and quantify the impact of each of these features on the mutation status. Furthermore, we assess the reproducibility of the shape, texture, and location features. GBM, being a strongly infiltrative tumor, has a complex tumor microenvironment [385]. While some efforts have been made to model it mathematically [470, 230], further investigation is required. Specifically, its automated modeling is largely an unexplored area of research. Therefore, we only isolate and quantify the impact of the tumor microenvironment to understand its radiogenomic contribution, and its reproducibility study is reported in the next chapter.

GANs in Medical Imaging: Due to their versatile utility, Generative Adversarial Networks (GANs) have been popular in developing automated digital radiomics tasks such as segmentation, data augmentation, enhancement, etc., and there is now an interest in automating radiogenomic tasks using GANs. From a computational point of view, a GAN must be able to successfully handle some of the most important issues of medical data like tumor heterogeneity, lack of data, data imbalance, etc. Medical data are often restricted and controlled in their distribution. They are also uncured, meaning getting properly annotated images for training by using automated tools is difficult. With limited, highly heterogeneous data like GBM, we show that these models face mode collapse or generate low-quality and/or artifact-ridden image samples. The visual manifestation of a tumor is often unapparent, and therefore, it involves a complex feature space. It is not always viable to exactly quantify the features and the relations among them or to reach a scalable, replicable

consensus over a complete set of factors. Thus, there is a need to develop automated tools that can model the desired visually unapparent biomarkers from limited and/or unbalanced data, while maintaining the realism, quality, and diversity of the original dataset. Thus, we propose *R2D2-GAN (Realistic Radiogenomic Design using Disentanglement in GAN)*.

Utility of R2D2-GAN: R2D2-GAN uses disentanglement and semi-deterministic sampling to reproduce the shape and texture of the visual mutation manifestation. It generates diverse, realistic, and high-quality samples of tumors that reflect the unapparent radiogenomic features while tackling the issues of limited, unbalanced, and heterogeneous training data. The purpose of disentanglement is two-fold: firstly, to isolate and quantify the impact of the individual features in determining mutation status - allowing for interpretability [204]. Secondly, learning statistical models of disentangled features allows us to generate high-quality synthetic brain tumor slices even when the training set is incredibly diverse and contains as few as 80 images. In this work, we establish the qualitative and quantitative impact of disentanglement in radiogenomic modeling while tackling the most common issues of medical data. We perform extensive multi-metric analysis, use multiple datasets, and carry out (condition adherence) testing on downstream tasks to evaluate the reliability of synthetic features. To further demonstrate the significance of the proposed model on limited and highly unbalanced data, we evaluate it on the rarer mutation - 19/20 co-gain. Such an approach can further be utilized to study other rare mutations in GBM as well as tumors of other organs. The proposed R2D2-GAN can not only be used for radiogenomic modeling, but also for other tasks such as data augmentation, data anonymization, disease progression modeling, and more.

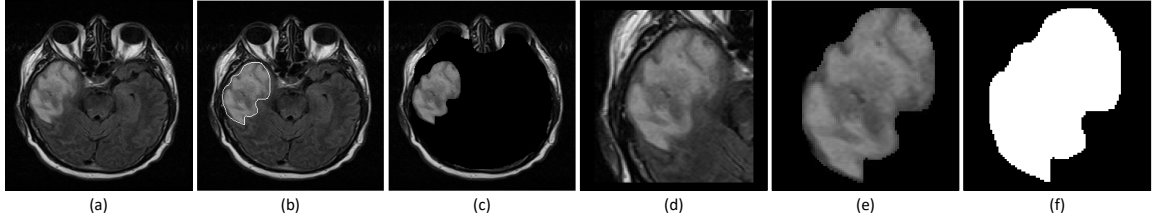


Figure 2.1: Disentanglement of various macro features in a brain scan. (a) Whole scan image (WSI), (b) WSI showing tumor boundary (visual only, not a separate representation), (c) Tumor shape + texture + location (Shp + Tex + Loc), (d) Tumor shape + texture + tumor microenvironment (Shp + Tex + TME), (e) Delineated tumor mass (Shp + Tex), and (f) Binarized tumor mass (Shp).

Model Explanation-based GAN Metric: In this work, we also propose the use of model explanations in evaluating the efficacy of a GAN. While many latest GANs can visually generate similar images, many factors need to be considered in deciding if the desired features are being reproduced. This is particularly important in medical applications of Artificial Intelligence (AI), where the “desired” features often correspond to critical information such as the accurate diagnosis and survival analysis of a patient. When data to the GAN is limited and highly heterogeneous, some GANs may produce good-quality images, but they may suffer from a lack of diversity in their output. On the other hand, some GANs may accurately capture the diversity of data, but fall short of reproducing the desired resolution that can capture prognostic or diagnostic information. Current GAN evaluation metrics are sensitive to challenges such as mode collapse [34]. We, therefore, utilize gradient-based model visualization techniques and quantify the similarity in their activation maps for classification using real vs synthetic data. The extent of similarity of features between the activation maps provides crucial information regarding how effective the GANs are at learning unapparent features, such as cases in which the model may predict the right class outcome but at the expense of learning the wrong features. These may be significant not only in medical

tasks but other tasks that require a high extent of data fidelity. This metric also helps in identifying the failure cases of GAN-based generation.

2.2 Related Work and Contributions

2.2.1 Related Work

Radiogenomics: Researchers have conducted radiogenomic analysis of GBM using various imaging modalities [250, 331, 65]. Some developed models for the more common mutations IDH1 and MGMT methylation [331, 53]. 19/20 co-gain, among others, is a rarer mutation – yet important. Its rarity makes it more difficult to analyze. Techniques that address such mutations [204, 205] pave a path to study other rare mutations, in achieving a comprehensive assessment of any cancer. Furthermore, the majority of research focuses on mutation detection [194]. While significant, it is crucial for clinical settings that mutations not only be detected, but also their discriminatory feature space be interpretable. Additionally, for strongly infiltrative tumors such as GBM, any tumor analysis must include the tumor microenvironment, which has been shown to be a significant biomarker [230].

Generative Deep Learning in Medical Imaging: Generative models have gained popularity in medical image analysis due to their applications in tasks like data augmentation [299], image enhancement [156], segmentation [244], and more [511, 72]. Recently, some researchers have also performed GAN-based radiogenomic analysis [478, 186, 204, 205]. Some of the most common GANs [8, 421] used in medical images for single modality synthesis are DCGAN [351], WGAN-GP [453], PGGAN [149, 151], and MSG-GAN [81]. Some recent models such as StyleGAN2 [214] are also known for their high-quality synthesis. Due

to the unapparent discriminatory feature set and limited, heterogeneous, and unbalanced data, GANs that generate whole-brain images face challenges such as mode collapse and lack of quality. Additionally, since GANs learn from the underlying distribution, there can be a trade-off between the realism of generated images vs. the diversity [388]. To tackle these issues and generate desired high-resolution, realistic, and diverse images, some works have explored generating tumors separately [478, 204, 205]. Recently, diffusion models [170, 219] have risen in popularity due to their strong generative properties. They are capable of generating high-quality images while tackling mode collapse. However, it has been shown that diffusion models have slow sampling, and have high data requirements [219]. This may cause a bottleneck for limited and diverse datasets such as ours.

GAN Evaluation Techniques: Among the most common metrics for evaluating GANs are the Fréchet Inception Distance (FID) score [167] and the Inception Score (IS) [19]. These metrics compute the diversity and quality of GAN image generation. Another metric not as frequently used is the Adversarial Accuracy (AA) [482]. This metric compares the accuracy of classifiers trained on real data versus synthetic data. These metrics have been shown to have a few drawbacks [34] such as limited data size or mode collapse. Another concern with fidelity measures is that they measure realism compared with the closest real sample. A small difference between a real and synthetic sample does not necessarily mean that the generated sample is better, and such a metric is also at odds with computing diversity. A GAN could generate realistic diverse images that are very different from the real set, and that would result in a lower metric for that GAN. There is currently no metric that can establish the extent of “correct” features being learned by a GAN. Interpreting

deep learning models has been an active field of research in the past few years, with many models being proposed that interpret the network outcomes. In high-stakes tasks such as medical imaging [35, 440], a lot of research has been directed towards understanding why models make certain decisions and how. With an increasing number of generative models in medical imaging, it becomes crucial to effectively evaluate the reliability of GANs in tasks with critical outcomes. To that end, we propose a measure of how similar the features behind model predictions are for classifiers trained on real and synthetic data, as opposed to just predictions. In other words, we measure the extent of similarity in the features learned by real data and synthetic data using model explanations.

2.2.2 Contributions

- The proposed research characterizes and quantifies the visual correlation between tumor macro-features and the presence of prognostic mutations. It includes the effects from lesser-analyzed features such as the tumor microenvironment. To the best of our knowledge, this is the first research of its kind that quantifies the visual correlation and features from the tumor microenvironment.
- The proposed R2D2-GAN assesses the reproducibility of tumor shape and texture features and incorporates a novel semi-deterministic sampling and an enhanced training protocol for increased robustness. This ensures diversity in tumor generation and improves the overall quality and realism of the generated tumor mass while tackling the significant problems associated with practical datasets such as limited data, data imbalance, and intra-class heterogeneity to learn visually unapparent features.

We conduct experiments on multiple datasets and report performance with extensive qualitative and quantitative metrics for evaluating the quality, realism, and diversity of generated tumors. We also compare our results with state-of-the-art generative models and outperform them.

- The proposed novel metric Synthetic Image Fidelity (SIF) evaluates the performance of GANs. It quantifies the extent to which a GAN learns the same features as real data when predicting the same sample. For our task, this metric measures how much a GAN can reproduce unapparent visual biomarkers that contribute towards the overall prognosis. We believe this is an important measure for high-stakes, high-fidelity tasks such as medical imaging, and to the best of our knowledge, this is the first metric of its kind.

2.3 Technical Approach

2.3.1 Quantification of Tumor Macro-Features

The tumor mass is the most significant in its contribution to the visual manifestation of the mutation. Previous work [204, 205] has established the existence of a discriminatory feature space in MRI data in detecting the mutation status and quantified the importance of some tumor macro-features. However, a significant biomarker of high clinical importance is the tumor microenvironment, which has not been widely studied in machine learning. In this work, we extend the analysis to the tumor microenvironment in detecting mutation status. The features considered are as shown in Figure 2.1.

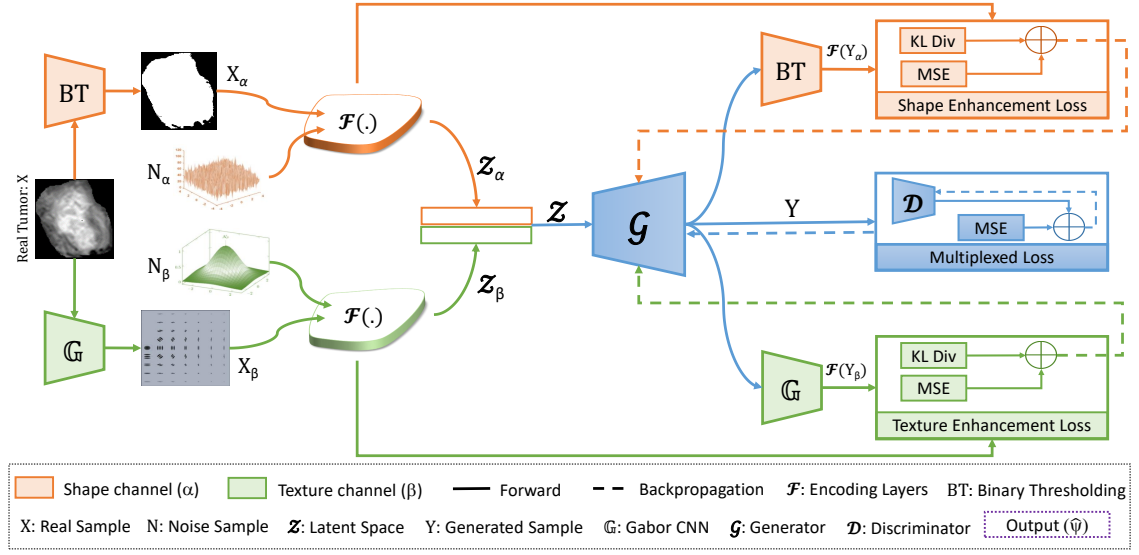


Figure 2.2: Overview of R2D2-GAN. It generates a tumor crop after learning from the shape and texture of a real tumor.

2.3.2 Realistic Radiogenomic Design using Disentanglement in GAN

To study the reproducibility of shape and texture features, we propose R2D2-GAN. Generative models create synthetic images from the underlying statistics of the data. Thus, typically, more data trains a better model. This is a challenge for medical data because of the limited data size, restrictive nature of the data, non-uniform collection protocols, limited access to annotated data, etc. Most proposed approaches aim to synthesize the whole MRI scan at once. This is undesirable for two reasons. The first is that the tumor resolution is compromised, as GANs have limited capability for generating high-resolution images [213]. Specifically for radiogenomic tasks such as ours, the tumor features must be generated at high resolution and with finesse. The second is that most GANs that generate whole images have an inherent feature entanglement that causes the model to lose the capability to learn specific features [441]. R2D2-GAN, as shown in Figure 2.2, involves disentanglement and subsequent generation of the tumor shape and texture, allowing us to increase the apparent

size of the dataset by disentangling and semi-deterministic sampling. This is an effective way to learn from limited training data and provides a lot of control over the diversity of output via individualized re-sampling. The framework of R2D2-GAN is shown in Figure 2.2. We first describe the problem formulation, and methodology to provide technical details.

Problem Statement: Given a tumor crop sample (X), the goal of R2D2-GAN is to generate a synthetic tumor (Y). First, real tumors are disentangled into two channels - the shape (α) and the texture (β). The shape and texture features (X_α) and (X_β) are generated using soft binary thresholding ($\mathbb{BT}(\cdot)$) and Gabor CNN ($\mathbb{G}(\cdot)$), respectively as shown in Eq. (2.1a).

$$X_\alpha = \mathbb{BT}(X_\beta) \quad X_\beta = \mathbb{G}(X) \quad (2.1a)$$

$$Z = \{Z_\alpha, Z_\beta\} = \{\mathcal{F}(X_\alpha), \mathcal{F}(X_\beta)\} \quad (2.1b)$$

$$Y = \mathcal{G}(Z) \quad Y_\alpha = \mathbb{BT}(Y) \quad Y_\beta = \mathbb{G}(Y) \quad (2.1c)$$

X_α and X_β are cast onto the latent space for encoding using $\mathcal{F}$, resulting in Z_α and Z_β . A multiplexed latent space sampling protocol is employed that samples from the latent shape (Z_α) and texture (Z_β) subspaces to generate a semi-deterministic input (Z). Z is the input to the generator ($\mathcal{G}$) which outputs the synthetic tumor crop Y (Eq. (2.1b, c)).

Methodology: To generate synthetic tumors, R2D2-GAN performs feature disentanglement followed by semi-deterministic re-sampling. As shown in Figure 2.2, the model extracts the shape and texture of the tumor, X_α and X_β , respectively. The shape representation is achieved using soft binary thresholding. For the texture representation, we employ Gabor CNNs (GCNNs) [277], which are designed with an inductive bias that ensures a texture

representation of the output. In addition to GCNNs having superior texture learning capabilities, the steerable and scalable kernels of GCNN have pre-defined orientations and scales, reducing the number of learnable parameters. The Gabor orientation Filter (GoF) equations are standard and the back-propagation happens only through filter coefficients - as in any CNN, making the GCNN trainable. We use the constraints as in [204] to steer the GCNN using GoF equations (Eq (2.2)).

$$g(u, v|f, \omega, \delta, \sigma) = \exp\left(-\frac{(u^2/\sigma_u^2 + v^2/\sigma_v^2)}{2}\right) \cos(2\pi fu + \delta) \quad (2.2)$$

where x and y are spatial positions, ω is the orientation of the filters, δ is the phase-shift, $u = x\cos\omega + y\sin\omega$ and $v = -x\sin\omega + y\cos\omega$, σ_u and σ_v are the spread of the filters and f is the frequency of the sinusoid.

Semi-deterministic sampling: To encode the shape and texture features, we employ two linear layers followed by ReLU activation. The encoding layers are collectively represented as $\mathcal{F}$. Appropriate noise is added to each of the channels (denoted by N_α and N_β) and the input to $\mathcal{F}$ is obtained via semi-deterministic sampling. This means that if the proportion of noise in the input is n , then the proportion of data is $(1-n)$. Since the shape is binary, we randomly sample values from $[0, 1]$ to add as noise (salt and pepper noise). For the texture, we add standard Gaussian noise (mean=0, std=1). The encoded Z_α and Z_β are the distributions from which the Generator picks its seed latent vector (Z). Due to the limited nature of our data, we want to ensure maximum diversity and quality of the generated tumors. To this end, we randomly vary the extent of noise added to each channel. The network randomly switches between varying extents of real input and noisy input, making

it a semi-deterministic sampling. Parallels can also be drawn that the sampling in this network lies between the Autoencoder (AE) and standard GAN, and this is done to tackle issues such as mode collapse and limited training samples. This makes the tumor generation robust to the issues of the dataset while generating high-quality diverse images. Let the weights for noise be added to shape and texture by α_n and β_n , respectively. The latent space representation will be obtained as:

$$Z_\alpha = \mathcal{F}[(1 - \alpha_n)X_\alpha + \alpha_n\mathcal{N}_\alpha] \quad (2.3a)$$

$$Z_\beta = \mathcal{F}[(1 - \beta_n)X_\beta + \beta_n\mathcal{N}_\beta] \quad (2.3b)$$

where Z_α/Z_β are the latent space representations of shape/texture, α_n/β_n are the weights for shape/texture noise (changes randomly for robust training), $\mathcal{F}$ denotes the encoding layers, X_α/X_β are input shape/texture, $\mathcal{N}_\alpha/\mathcal{N}_\beta$ are shape/texture noise.

Since the sampling is semi-deterministic, the loss function should reflect that. For a system with generalized input a and output b , if the sampling is done entirely from real data (a_{real}), the network acts like an AE and is trained using a reconstruction loss like Mean Squared Error (MSE) loss (Eq (2.4a)). If the samples are completely noisy (a_{noise}), the network acts as a GAN and loss is Binary Cross-Entropy (BCE) loss (Eq (2.4b)). For a generic multiplexed system, the loss term adapts between these two extremes (Eq (2.5)).

$$\mathcal{L}_{a \sim p(real)} : \mathbb{E}_{a \sim p_{real}(a)} \left[||a_{real} - b||_2 \right] \quad (2.4a)$$

$$\mathcal{L}_{a \sim p(noise)} : \mathbb{E}_{a \sim p_{noise}(a)} \left[\log \left(1 - \mathcal{D}(\mathcal{G}(a_{noise})) \right) \right] \quad (2.4b)$$

$$\mathcal{L}_D = n \times \mathcal{L}_{a \sim p(\text{noise})} + (1 - n) \times \mathcal{L}_{a \sim p(\text{real})} \quad (2.5)$$

where n is fraction of noise, $\mathcal{L}_{a \sim p(\text{real})}$ is generalized reconstruction (L2 Loss) given real input, $\mathcal{L}_{a \sim p(\text{noise})}$ is generalized adversarial loss given noisy input.

Translating this to our system, the overall loss function is the combined loss from the shape and texture channels and the loss from the discriminator (Figure 2.2). We observed that employing three discriminators made the model unstable, and therefore, we use a regular discriminator for our output channel and leverage the encoding layers $\mathcal{F}$ to compute the KL divergence loss between the generated vs real feature for the shape and texture channels. At the core of it, encoding layers followed by KL divergence loss achieves a very similar effect as a discriminator, without the instability caused by more training parameters. Since our model also trains a GCNN simultaneously, it is prone to instability. Wasserstein GANs (WGANs) have been shown to add stability to the GAN training [15] via a Lipschitz constraint, which enforces maximum gradient on the discriminator for improved generator training. This can sometimes result in overshoot, which is an instability in itself. Thus, we apply gradient penalty function to the network [143] to improve stability. This modifies the generalized discriminator loss in Eq (2.5) to the one shown in Eq (2.6). Substituting discriminator loss with KL divergence loss for each of the shape and texture channels gives us Eq (2.7) and (2.8) respectively.

$$\begin{aligned} \mathcal{L}_D = & n \times \left(\mathbb{E}_{Y \sim p_Y} [\mathbf{D}(Y)] - \mathbb{E}_{X \sim p_X} [\mathbf{D}(X)] \right. \\ & \left. + \mathbb{E}_{\hat{X} \sim p_{\hat{X}}} [\|\nabla_{\hat{X}} \mathbf{D}(\hat{X})\|_2] \right) \\ & + (1 - n) \times \mathbb{E}_{X \sim p_X} [\|X - Y\|_2] \end{aligned} \quad (2.6)$$

$$\begin{aligned}\mathcal{L}_\alpha &= n_\alpha \times \left(\mathbf{D}_{\mathbf{KL}}(\mathcal{F}(Y_\alpha) || \mathcal{F}(X_\alpha)) \right) \\ &\quad + (1 - n_\alpha) \times \mathbb{E}_{X_\alpha \sim p_{X_\alpha}} \left[||X_\alpha - Y_\alpha||_2 \right]\end{aligned}\tag{2.7}$$

$$\begin{aligned}\mathcal{L}_\beta &= n_\beta \times \left(\mathbf{D}_{\mathbf{KL}}(\mathcal{F}(Y_\beta) || \mathcal{F}(X_\beta)) \right) \\ &\quad + (1 - n_\beta) \times \mathbb{E}_{X_\beta \sim p_{X_\beta}} \left[||X_\beta - Y_\beta||_2 \right]\end{aligned}\tag{2.8}$$

where encoding layers are denoted by $\mathcal{F}$, generated tumor Y is $\mathcal{G}(Z)$, (Z) is the multiplexed latent vector computed in Eq (2.3), Y_α/Y_β are shape/texture features of Y ,

$\hat{X}$ terms are the weighted average between real and synthetic images (subscript α for shape and β for texture), n_α and n_β are weights of noise added to respective channel and n is the average of the two, $\mathcal{G}$ is the Generator, $\mathcal{L}_D$ is overall multiplexed loss, $\mathcal{L}_\alpha$ and $\mathcal{L}_\beta$ are the multiplexed shape and texture losses, respectively. The overall loss function of the model will thus be:

$$\mathcal{L} = \mathcal{L}_D + 0.5 \times (\mathcal{L}_\alpha + \mathcal{L}_\beta)\tag{2.9}$$

2.3.3 Synthetic Image Fidelity (SIF) Metric

The hypothesis for developing SIF is that many GANs may generate similar-looking features or images, resulting in high Fréchet Inception Distance (FID) or Inception scores. However, it has been shown in the literature that these metrics are not always robust [34]. FID score measures the fidelity and diversity of synthetic data but is sensitive to sample size and uses limited statistics [34]. IS, on the other hand, is sensitive to mode collapse [34]. Additionally, both FID and IS lump quality and diversity into a single

metric which may be ineffective in computing either. AA compares the accuracy of classifiers trained on real and GAN data on a test set, but this does not confirm whether the GAN-trained classifier is simply predicting the right samples or if it is actually learning the correct features. In other words, if GANs generate a few images that look like real data, they can have good quantitative metrics. However, they may not adequately learn the complex feature associations for the whole dataset. This is particularly a key aspect for limited and heterogeneous data applications such as ours. The visual biomarkers of mutation may need insights from all data points at high resolution to be able to model appropriate feature associations that reflect the presence or absence of mutations. Failure to capture either the high level of detail in any image or the extent of heterogeneity or diversity in the dataset will lead to incorrect features being emphasized during synthesis.

The proposed SIF Metric utilizes model interpretability tools to compute the extent of similarity between decisions made by classifiers trained on real and synthetic data. A model such as GradCAM reflects the magnitude of gradient backpropagation, which strongly correlates with the features that the classifier has used to predict the mutation status [380]. SIF aims to compare the extent of similarity between the features being learned between real and synthetic data. In essence, SIF investigates how closely the underlying statistical features learned from synthetic data mirror those learned from real data. Therefore, if synthetic data considers the same features during classification as real data, it indicates that the generative model has performed comprehensive mode coverage and effective learning of the diverse yet sparse input distribution of real data. For example, if a GAN fails to capture the data heterogeneity, then during classification, it will have learned feature

associations from a limited subset than what the real data represents. This will be reflected in the visualization of output images when it highlights regions different from what the real data model does. Similarly, if the GAN fails to capture data quality, it may not generate images that accurately represent the finer details such as a necrotic core, for example. This could be the difference in the presence or absence of a mutation, since a more pronounced necrotic core can reflect the presence of 19/20 co-gain, i.e., better prognosis of a patient. This analysis serves as a qualitative assessment of the efficacy of a GAN in learning highly unapparent yet critical feature associations.

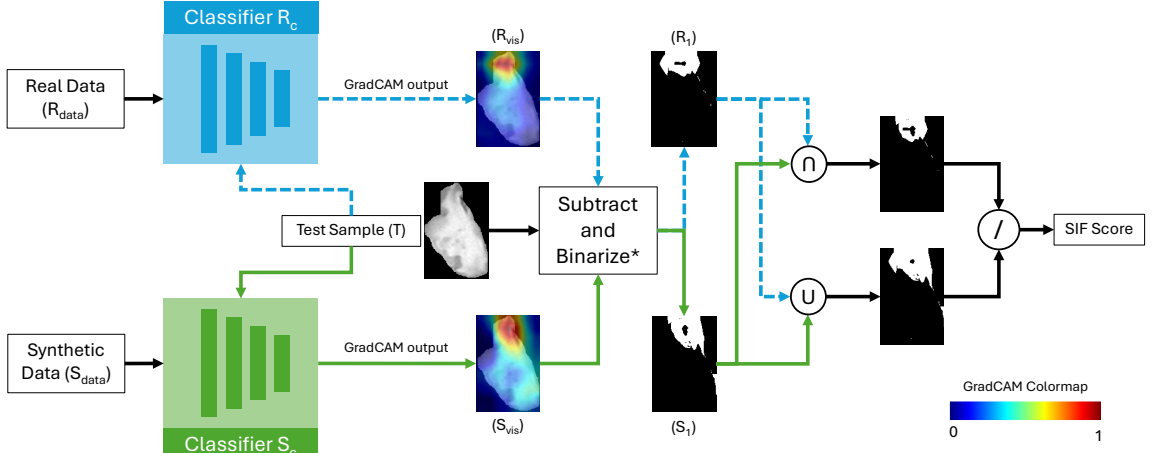


Figure 2.3: Pipeline to compute the Synthetic Image Fidelity Metric. The process from (*) onwards is separately done for each of the R, G, and B channels – then aggregated to obtain the SIF score. R_1 and S_1 are shown here for the Red (first) channel. For the GradCam heatmap, the redder color implies a higher significance of that feature, and bluer is less.

To compute this metric, we train two classifiers, one on GAN-generated data (S_c) and one on real data (R_c), as shown in Figure 2.3. Both classifiers are used to classify real samples (T). Then, a gradient-based visualization model (here, GradCAM [380]) is used to visualize which features of the real data get activated during classification using the real

classifier and the GAN classifier. We binarize the heatmaps from GradCAM for each of the channels of RGB. The threshold for binarization is computed using Otsu’s method [325]. These binary heatmaps are then subjected to an “AND” operation and an “OR” operation. The ratio is an Intersection over Union (IoU)-like metric that reports the extent of similarity for each channel. The more similar these heatmaps from GradCAM are, the more effective a GAN is at learning desired features. The three similarity measures are then aggregated to obtain SIF. It is to be noted that the heatmap could be binarized without splitting into R, G, and B channels. However, the color on the regions highlighted by GradCAM carries meaning on how significant a feature is (Ref. Figure 2.3). The redder a region is, the more significant it is for prediction. The bluer regions carry less significance. Binarizing all three channels at once compresses all three channels into one, thereby losing this information. Consequently, by not isolating the channels, SIF can only capture if the GAN data can capture similar features, but not how significant the features are according to the GAN. Thus, we isolate and later aggregate the channels to make SIF a stronger, more holistic metric. The algorithm for SIF computation is provided in Algorithm 1. We also note an added benefit of SIF is that it can help identify failure cases in any GAN for its analysis. If a sample has low SIF, it means that even if it looks real, the sample has not captured the discriminatory features of that image class. Since it is an IoU-based metric, SIF always lies between 0 to 1, and the higher the value, the better the GAN.

Algorithm 1 Algorithm for SIF Computation

Input: Real image test sample (T in Figure 2.3)

Output: SIF score

Step 1: Train classifier R_c on real data, and classifier S_c using synthetic data (Figure 2.3)

Step 2: Use the trained classifiers to test a real sample T

Step 3: Perform GradCAM computation to obtain heatmaps R_{vis} and S_{vis} corresponding to T using classifiers R_c and S_c , respectively

Step 4: Get bare heatmaps using $R_{vis} - T$ and $S_{vis} - T$ (remove tumor)

Step 5: For each channel (Red, Green, and Blue) of image’s bare heatmaps, binarize them using Otsu’s method to compute the binarization threshold to obtain R_1 and S_1 (for red channel), R_2 and S_2 (for green channel), and R_3 and S_3 (for blue channel)

Step 6: Compute a component of SIF metric for each channel via $SIF_i = \sum(R_i \cap S_i) / \sum(R_i \cup S_i)$

Step 7: $SIF_T = (SIF_1 + SIF_2 + SIF_3) / 3$

Step 8: Repeat this for all images in the testing set, and aggregate them to find SIF

2.4 Experiments and Results

2.4.1 Description of Datasets

We report our results on two datasets, a dataset collected at Emory University (Dataset A) and the publicly available The Cancer Imaging Archive Glioblastoma (TCIA-GBM) [71] Dataset (Dataset B). The use of human data in Dataset A was approved by the Institutional Review Board (IRB) of Emory University, Protocol #95120. Dataset A is very limited and diverse whereas Dataset B is highly imbalanced. The evaluation of Dataset A highlights the effectiveness of R2D2-GAN in learning discriminative features from limited data, whereas, the evaluation of Dataset B demonstrates its utility in combating data imbalance. Both datasets exhibit high intra-class diversity and tumor heterogeneity. Together, they represent common, critical issues faced by real-world medical data.

Dataset A. Dataset A contains FLAIR MRI for 25 patients with known mutation status. This cohort has 14 control and 11 mutated patients. The control class has cases

without 19/20 co-gain and mutated class, with co-gain. Both classes have about 100-130 images each with a tumor, making it a limited dataset.

Dataset B. The public Dataset B (TCIA-GBM) has a cohort of 164 patients with FLAIR MR images, divided into 133 control and 31 mutated patients. The corresponding molecular data for all patients were obtained from The Cancer Genome Archive (TCGA) website [430]. Extracting the 19/20 co-gain status was done by taking gistic data and thresholding for each arm of the chromosome, and then checking if $>25\%$ of the genes are duplicated on both the long and short arms of chromosomes 19 and 20 [204].

2.4.2 Experimental Setup

To characterize the features, we perform mutation classification for both datasets using the representations shown in Figure 2.1. A quantitative comparison shows the impact of each feature. To analyze the location feature, we plot tumor distribution along the axial, coronal, and sagittal planes for both classes of both the datasets. For shape and texture, we report qualitative and quantitative evaluations of R2D2-GAN and compare them with other generative models. While some radiogenomic GAN works [204, 205] are included in the comparison, some [478, 407] are not included as their code-base are unavailable. We extend our analysis to the state-of-the-art image generation models such as StyleGAN2 [214], BigGAN [39], and Denoising Diffusion Probability Model (DDPM) [170]. In our experiments, BigGAN had trouble converging on the limited Dataset A. StyleGAN2 is more appropriate for our analysis since it uses the disentanglement principle as well. Most medical imaging tasks use some version of DCGAN [351, 148], WGAN or WGAN-GP [148, 233, 453], PGGAN [149, 152, 151] or MSG-GAN [81]. We report a comparison

with DDPM, StyleGAN2, MSG-GAN, FeaD-GAN, and SAGE in our results. Among these, SAGE is a disentanglement-based multi-step approach, while FeaD-GAN and StyleGAN2 are disentanglement-based single-step GANs. MSG-GAN is a single-step GAN without disentanglement. While single-step GANs are easier to train, training them on limited and heterogeneous data leads to mode collapse, under-learning, low-quality, or artifact-ridden synthesis. We demonstrate via extensive qualitative and quantitative results that they are an oversimplification of the task at hand. Multi-step GANs such as SAGE require more setup and longer training. Thus, comparing these models emphasizes the importance of the disentanglement and improved sampling and training protocol in R2D2-GAN. While DDPM outperforms most of the comparative GANs, we note that for limited and diverse datasets, R2D2-GAN has the most consistent performance, outperforming DDPM for most metrics. Furthermore, each image using DDPM takes almost 20X longer to generate as compared to R2D2-GAN ($>17s$ vs $<1s$). Based on their reference models in medical imaging tasks, we performed fine-tuning of MSG-GAN [81] to perform better with our data. StyleGAN2 was used as is, by reducing augmentation probability to 0.25 as suggested in [214] when dealing with limited data. For DDPM, we use Algorithm 1 as proposed in [170].

The desired features of a radiogenomic GAN are that the generated images be high quality, diverse, realistic, and replicate the (visually unapparent) discriminatory feature set. We report quantitative results demonstrating all of these properties of R2D2-GAN compared with other models. We report qualitative results evaluating the quality, diversity, and realism of R2D2-GAN compared with other models.

Table 2.1: Mutation detection in Dataset A and B using the macro feature representations as shown in Figure 2.1. ResNet18 is used with a learning rate of 0.0001, batch size 16, and Adam optimizer with step decay. The train/test split is 80/20, all images are resized to 256x256, and metrics are reported on 10-fold cross-validation as mean (standard deviation). Metrics are - ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL; Image Level and PL: Patient Level analyses. IL results are calculated for every individual slice and PL results are calculated by a majority vote of IL predictions of all patient scans.

Input Source	Dataset A					Dataset B				
	ACC (PL)	ACC (IL)	SENS	SPEC	DICE	ACC (PL)	ACC (IL)	SENS	SPEC	DICE
WSI (All)	92 (08)	90 (06)	85 (07)	95 (05)	86 (08)	81 (09)	78 (12)	73 (08)	86 (08)	70 (10)
Shp + Tex + Loc	95 (03)	92 (05)	88 (06)	98 (02)	88 (08)	84 (08)	78 (10)	72 (06)	88 (07)	71 (10)
Shp + Tex + TME	88 (10)	84 (13)	77 (13)	88 (08)	77 (12)	77 (11)	76 (09)	71 (11)	83 (09)	68 (12)
Shp + Tex	85 (12)	80 (09)	67 (09)	84 (06)	69 (07)	74 (09)	73 (12)	66 (10)	77 (06)	64 (13)
Shp	70 (06)	68 (07)	65 (09)	75 (05)	66 (08)	64 (10)	62 (08)	58 (06)	70 (05)	60 (08)

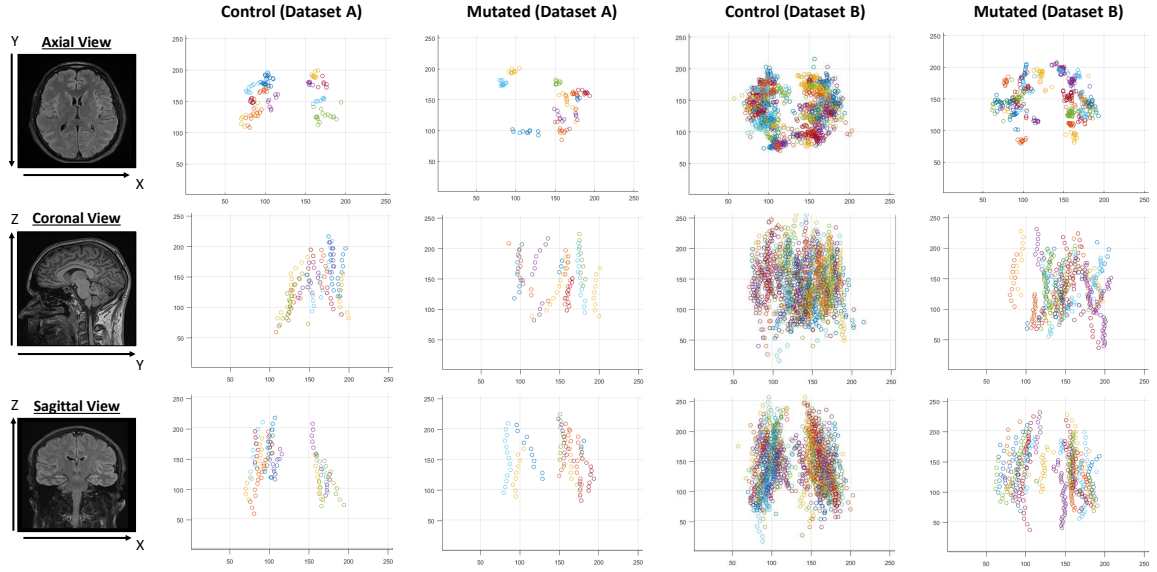


Figure 2.4: Characterizing tumor manifestation in the brain using the distribution of tumor location along the axial, coronal, and sagittal planes conditioned upon mutation status for both datasets. Each color corresponds to a different patient.

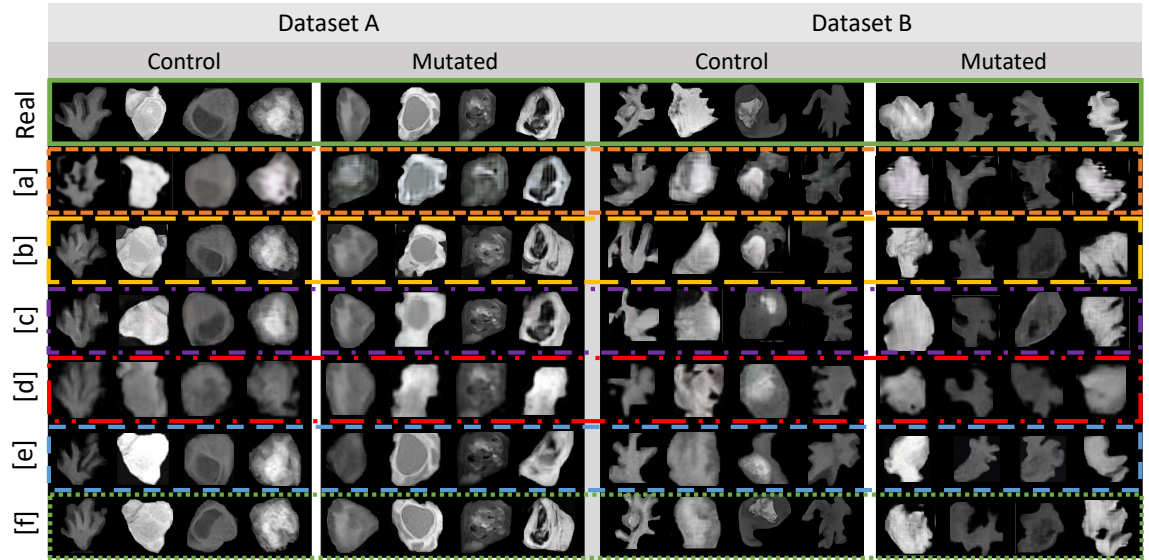


Figure 2.5: Generated Image Quality of R2D2-GAN (vs real data and other models) for Datasets A and B for both classes. We see that most models, including DDPM, suffer from a lack of detail in some tumors in each class. Real (green, solid) vs [a] MSG-GAN (orange, short-dashed), [b] SAGE (dark yellow/long-dashed), [c] FeaD-GAN (purple, dot-dash), [d] StyleGAN2 (red, dot-dot-dash), [e] DDPM (blue, dashed), [f] R2D2-GAN (green, dotted).

2.4.3 Results: Quantification of Macro-features

We report classification performance in detecting mutations for various feature representations (shown in Figure 2.1) in Table 2.1. Here, accuracy (ACC) is the proportion of total correct predictions, sensitivity (SENS) is the true positive rate, specificity (SPEC) is the true negative rate, dice score (DICE) is the harmonic mean of precision (proportion of true positives over all the positives), and recall (sensitivity). We note that the inclusion of each of these macro-features augments the overall classification performance, thus indicating their impact on the visual manifestation of the mutation. We note that the newly included tumor microenvironment is indeed a significant biomarker in itself, as it improves the classification performance by at least 3% for both datasets. We note that the location feature is also a significant biomarker. Towards characterizing the location feature, we plot the locations of control and mutated tumors for both datasets in Figure 2.4. We see that there is a notable difference in where the tumor is localized based on the presence or absence of mutation. This variation manifests more strongly in the axial view but shows some difference in the other views as well. One of the factors this can affect is how easily a tumor can be surgically resected, which subsequently impacts the overall survival of the patient. For the mutated tumors, which correlate positively with median survival, we note that the tumors tend to be less internal to the overall brain structure. In other words, many tumors appear to be more easily accessible.

2.4.4 Results: Qualitative Evaluation

Generated Image Quality

Figure 2.5 shows a comparison of the images generated by R2D2-GAN vs other GANs and includes the real data as reference for Dataset A and B, respectively. This serves as a visual demonstration of the efficacy of R2D2-GAN in learning the underlying data distribution of the mutation biomarkers. We note from the results of Dataset A particularly, which is a very limited dataset, that R2D2-GAN generates the desired features accurately. MSG-GAN performance gets better for larger sets of data such as the Control group of Dataset B, but has a persistent checkered artifact for others, which affects the tumor texture and possibly any mutation biomarkers. DDPM outperforms all these models, but loses some detail (Figure 2.5) or has a blurriness issue (Figure 2.6). Some example images of heterogeneous tumors are shown in Figure 2.6 for a detailed analysis. These magnified images clearly show the checkered artifacts in MSG-GAN (images 3 and 4), sharpness artifacts in SAGE (arrows in Images 1, 3, and 4), and loss of detail in StyleGAN2 (Images 2, 3, 4). For DDPM, we see a blurriness issue in all images, specifically images 2 and 4. We note that for all images, R2D2-GAN overcomes the persistent blurriness problem of FeaD-GAN. An interesting example is Image 2, which seems low in detail but has a lot of minute details in its texture, which is captured accurately only by R2D2-GAN.

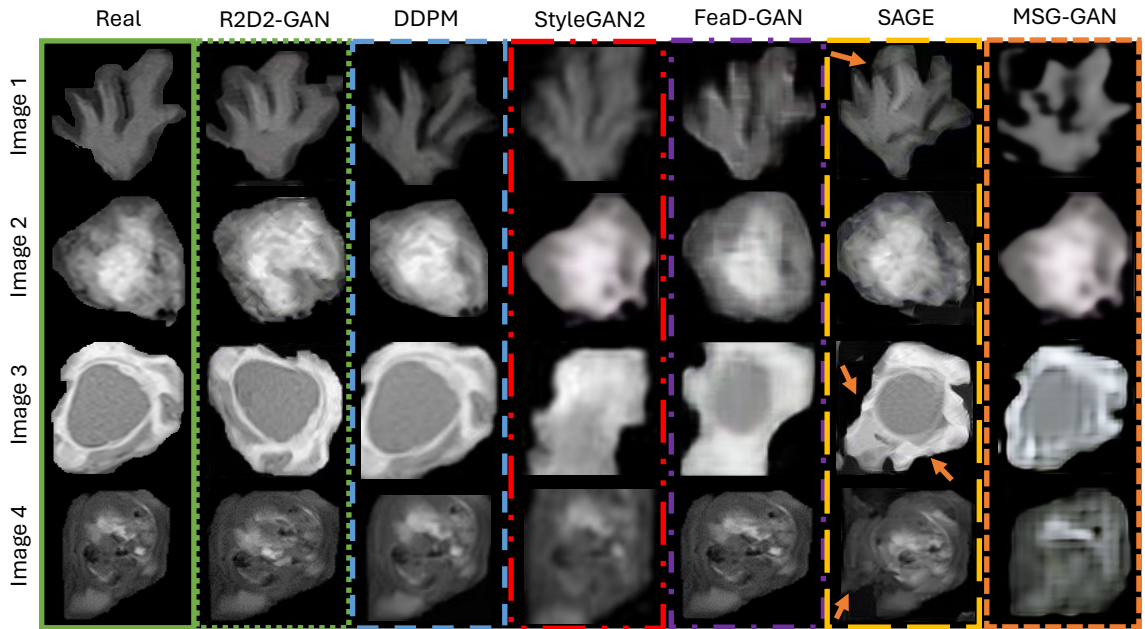


Figure 2.6: Magnified examples of generated images from all GANs. Real: Solid/green, R2D2-GAN: green/dotted, DDPM: blue/dashed, StyleGAN2: red/dot-dot-dash, FeaD-GAN: purple/dot-dash, SAGE: dark yellow/long-dashed, MSG-GAN: orange/short-dashed. These magnified images clearly show the sharpness artifacts in SAGE (orange arrows in Images 1, 3, 4), loss of detail in StyleGAN2 (Images 2, 3, 4), and checkered artifacts in MSG-GAN (images 3 and 4). We note some blurring in DDPM (Images 2 and 4). We note that for all images, R2D2-GAN overcomes the persistent blurriness problem of FeaD-GAN.

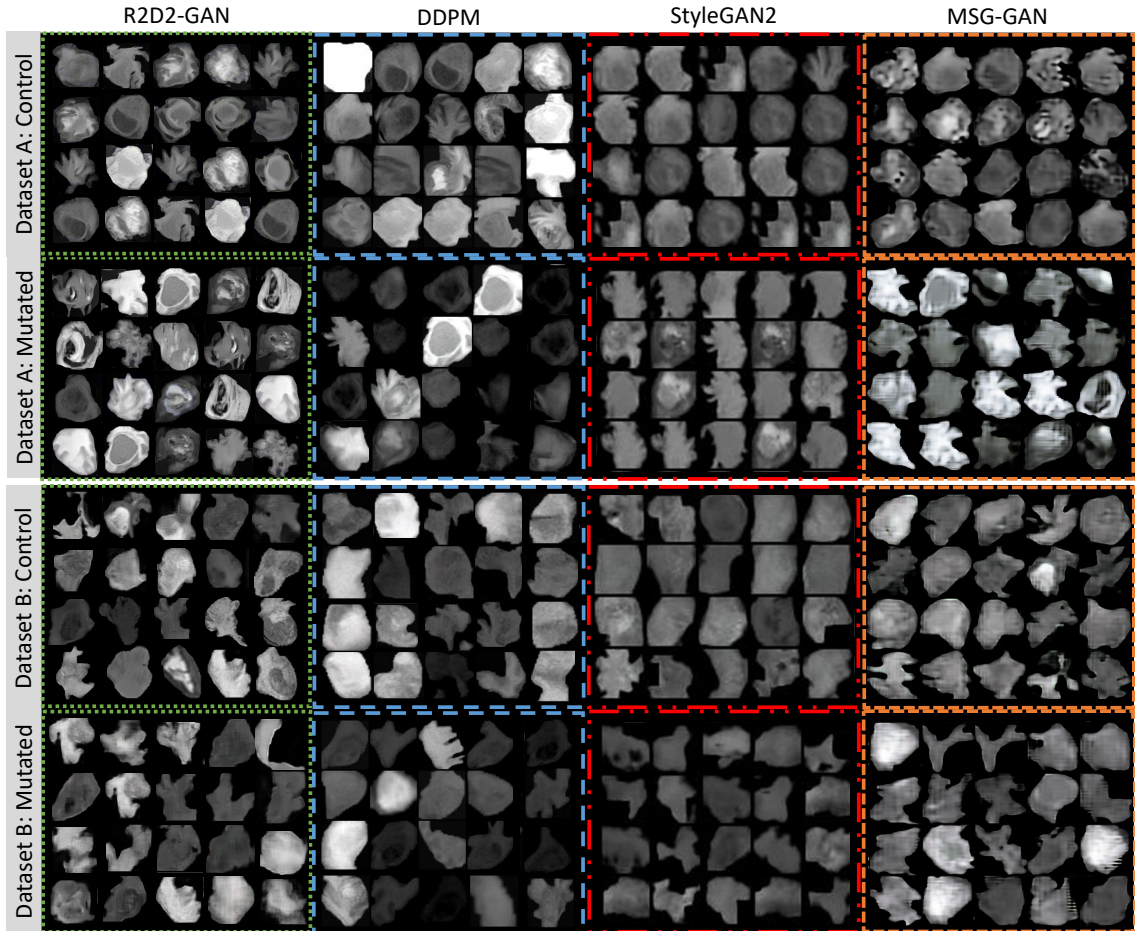


Figure 2.7: Diversity of images generated by R2D2-GAN using Dataset B (vs DDPM, StyleGAN2, and MSG-GAN). Best viewed zoomed in. MSG-GAN shows high-frequency artifacts and a lack of tumor details. StyleGAN2 has better quality but lower diversity. DDPM has some diversity but also has some very bright or very dark images generated due to instability in training. R2D2-GAN: green/dotted, DDPM: blue/dashed, StyleGAN2: red/dot-dot-dash, MSG-GAN: orange/short-dashed.

Table 2.2: Fréchet Inception Distance (FID) Score, modified Inception score (mIS), and Synthetic Image Fidelity (SIF) score for generated images using R2D2-GAN, DDPM, StyleGAN2, FeaD-GAN, SAGE, and MSG-GAN using 500 synthetic images for each.

Model	Dataset A						Dataset B					
	SIF $\uparrow$			FID $\downarrow$			mIS $\uparrow$			SIF $\uparrow$		
	Con	Mut	Avg	Con	Mut	Avg	Con	Mut	Avg	Con	Mut	Avg
R2D2-GAN (ours)	0.31	0.37	0.34	198.23	202.45	200.34	1.42	0.24	0.28	0.26	171.13	188.87
DDPM [170]	0.30	0.31	0.30	193.32	212.26	202.79	1.40	0.21	0.27	0.24	174.63	176.23
StyleGAN2 [214]	0.24	0.34	0.30	236.74	251.86	244.30	1.35	0.30	0.20	0.25	193.52	172.67
FeaD-GAN [204]	0.26	0.35	0.31	207.22	211.09	209.15	1.40	0.23	0.25	0.24	183.64	200.96
SAGE [205]	0.28	0.36	0.32	204.83	210.21	207.52	1.34	0.24	0.25	0.24	189.56	195.43
MSG-GAN [81]	0.22	0.32	0.27	256.90	248.01	252.45	1.12	0.25	0.25	0.25	233.70	220.54
												227.12

Table 2.3: Mutation detection in Dataset A and B using only synthetic tumor crops generated using R2D2-GAN, DDPM, StyleGAN2, FeaD-GAN, SAGE, and MSG-GAN. All GANs are trained using the same real training sets and tested on the same corresponding real test sets for all 10-folds. The metrics are reported as mean (standard deviation) over 10 folds. ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. The model trained using real data is used as the baseline.

Data	Input Source	Dataset A						Dataset B					
		ACC(PL)	ACC(IL)	SENS	SPEC	DICE		ACC(PL)	ACC(IL)	SENS	SPEC	DICE	
<i>Tumor Crops</i>	Real Data	85 (12)	80 (09)	69 (09)	84 (06)	70 (07)		74 (09)	73 (12)	66 (10)	77 (06)	64 (13)	
	R2D2-GAN	85 (10)	82 (08)	72 (10)	85 (06)	72 (08)		83 (10)	82 (09)	76 (11)	90 (09)	78 (12)	
	DDPM [170]	83 (12)	82 (10)	72 (13)	84 (09)	71 (11)		82 (11)	82 (13)	76 (14)	88 (10)	77 (13)	
	StyleGAN2 [214]	76 (16)	75 (18)	65 (16)	74 (19)	68 (17)		80 (09)	78 (10)	74 (09)	86 (08)	75 (08)	
	FeaD-GAN [204]	85 (08)	80 (08)	67 (09)	84 (06)	69 (09)		82 (06)	79 (08)	75 (09)	90 (04)	74 (06)	
	SAGE [205]	85 (10)	81 (07)	72 (07)	87 (06)	72 (08)		83 (07)	81 (10)	76 (12)	91 (06)	77 (08)	
	MSG-GAN [81]	79 (13)	73 (15)	70 (12)	79 (13)	67 (15)		78 (12)	79 (11)	73 (15)	84 (12)	72 (12)	

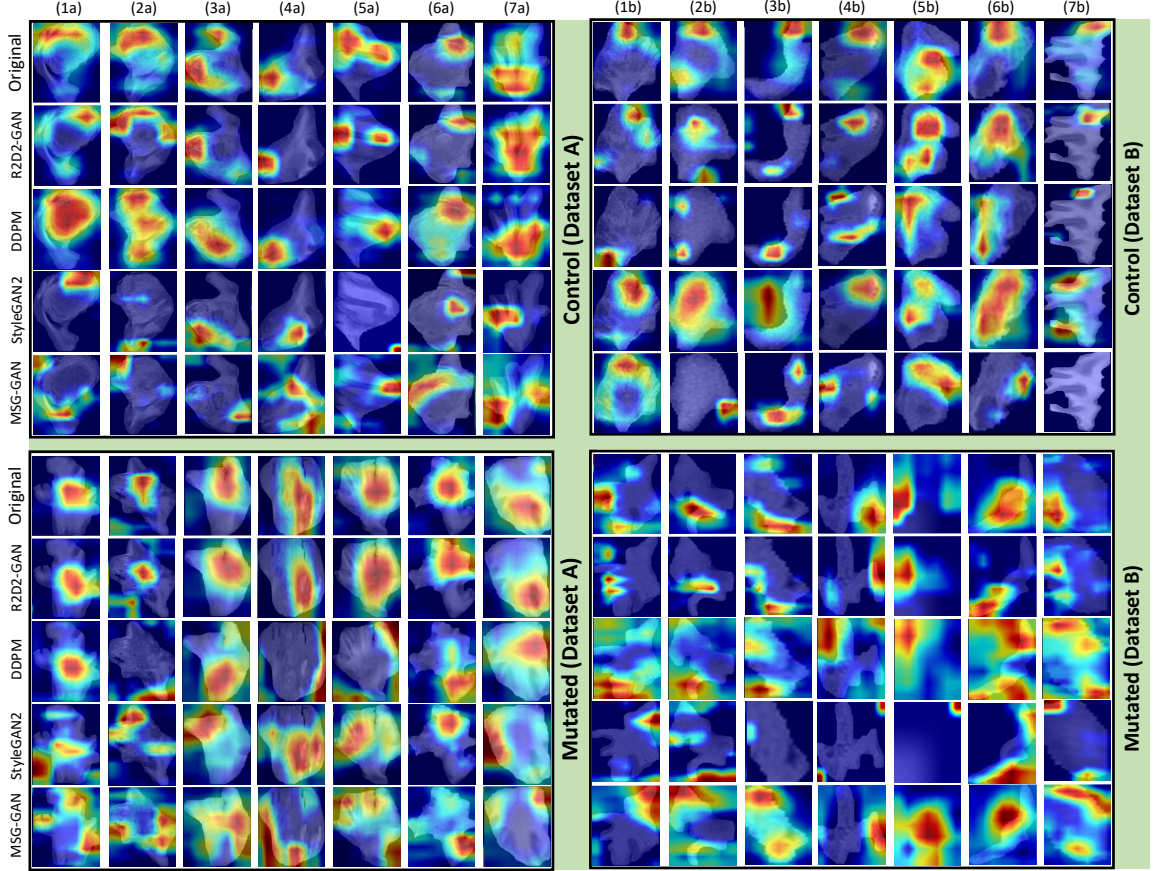


Figure 2.8: Visualizing GradCAM outputs in the computation of SIF score. Similar activation maps demonstrate a GAN’s ability to learn high-level complex feature associations that correlate with the unapparent visual manifestations of the presence or absence of a mutation. We note that R2D2-GAN highlights similar regions as real data, thereby establishing its efficacy in learning crucial features. Colorbar for heatmaps is provided in Figure 2.3.

Diversity of Generated Images

We compare the diversity of R2D2-GAN generated images with images generated by single-step GANs (MSG-GAN, and StyleGAN2) and DDPM in Figure 2.7. We show that compared to R2D2-GAN, there is a trade-off in other GANs between quality and diversity. It also represents the robustness of R2D2-GAN as compared to other GANs to survive mode collapse. For the limited yet highly diverse Dataset A, we note that StyleGAN2 has fair quality but a lack of diversity, and MSG-GAN has a lack of diversity and quality. In terms of mode coverage and diverse generation of images, while DDPM outperforms most of the other GANs, it sometimes ends up generating too bright or completely dark tumors due to instability in training. With increasing data size (Dataset B), DDPM, StyleGAN2, and MSG-GAN show a performance improvement, but have similar concerns. It is to be noted that R2D2-GAN effectively tackles mode collapse while maintaining image quality and diversity for both datasets.

2.4.5 Results: Quantitative Evaluation

Image Quality and Diversity Metrics

We report the Fréchet Inception Distance (FID) score [167], modified Inception Score (mIS), and the proposed SIF score to evaluate generated images. These results are reported in Table 2.2. FID score represents the feature distance between real and generated images and measures synthetic image quality and diversity. It is also shown to be robust to many common image perturbations and can detect mode collapse and intra-class diversity, making it highly useful to our task. Lower FID indicates a better model. However, a

drawback of FID is that it cannot correctly estimate when the sample size is limited. IS [19] is a measure of image quality and diversity, whose value lies between 1 and the number of classes (here, 2). Higher IS represents better performance. Unlike traditional IS, which uses ImageNet weights, we report modified IS (mIS) by finetuning the last few layers of the Inception model for 10 epochs before computing IS to address the properties of the data. A drawback of IS/mIS is that it is sensitive to mode collapse. We report the SIF score for all GANs as well. For FID score, DDPM outperforms R2D2-GAN slightly for Dataset A (Control) and Dataset B (Avg) (Table 2.2). We note for SIF score that except for the largest set (Dataset B, Control), where StyleGAN2 has the best performance - R2D2-GAN outperforms all GANs across all data subsets. This highlights the significance of R2D2-GAN in tackling limited and diverse datasets. We demonstrate the utility of SIF in visualization as well as the identification of failure cases in future subsections (F and G). We also report Adversarial Accuracy (AA) [482], which computes the classification accuracies achieved by two classifiers on the same validation set, one trained on real data and one on synthetic data. If both accuracies are close, it implies that the GAN has effectively learned the underlying distribution of the real data. We reported the accuracy values for various GANs and real data baseline in the ACC (IL) column of Table 2.3. To compute these metrics, we used 500 synthetic images per class for both datasets for all GANs. Performance variation due to data limitation is more apparent in MSG-GAN and StyleGAN2. We note that R2D2-GAN has the most consistent performance of all these models. For AA in Table 2.3, we also note that R2D2-GAN has the closest performance to real data for the balanced Dataset A. For highly imbalanced Dataset B, we note that the performance of many GANs is better

than real data, which can be attributed to the data augmentation (over-sampling) using synthetic samples. We note that R2D2-GAN has the best overall performance.

Learning Discriminatory Features

The goal of employing GANs in automated radiogenomic analysis is to assess if the generated images capture the distribution representing mutation information and learn the unapparent feature space. For this, we perform classification using only real data and then using only synthetic data from various GANs. The hypothesis is that if synthetic data can accurately distinguish between the presence and absence of mutation in real data, then it has captured the desired features. We compare the classification results using R2D2-GAN images with classification using real data and other models. We use the same classifier and report the same metrics as Table 2.1. The results are reported in Table 2.3. The results are reported for both Datasets A and B. It can be seen in Table 2.3 that R2D2-GAN performs just as well as real data (sometimes better) while outperforming other GANs, indicating that it can capture the underlying statistical distribution and thereby able to learn and model the discriminatory features. The improvement in performance compared to real data can be attributed to the limited and unbalanced nature of Datasets A and B, respectively. Thus, we can see that for varying data size and intrinsic mutation properties, R2D2-GAN consistently has the best performance.

2.4.6 Results: SIF-based Qualitative Evaluation

An added benefit of SIF computation is the ability to visualize and compare GAN prediction performance compared with real data prediction. We note from previous results

that among state-of-the-art approaches, DDPM performs the best after R2D2-GAN. In this analysis, we demonstrate that improved quantitative performance does not necessarily result in more accurate features being learned. As established in previous sections, accounting for complex unapparent feature associations that can only be learned by accurately modeling for the limited and diverse data during training are key to effective GANs in high-stakes, high-fidelity tasks such as medical imaging. We note here from Figure 2.8 that when visualizing the features that the model is seeing to make decisions on mutation detection, R2D2-GAN consistently activates the regions that the real data also activates. We note that some GANs also are correct in images such as 4a, 6a (control), 7b (control), 2b (mutated), etc., but are unable to perform so consistently across the board - specifically for the highly limited and diverse classes of Dataset A. This is also reflected in the SIF scores in Table 2.2. This shows us that R2D2-GAN, due to its effective latent space sampling, can accurately learn unapparent features in the visual manifestation of mutations. Furthermore, these insights also help us in studying which features of a tumor contribute to mutation detection. For example, most control class images focus on the enhancing region (which is the growing part of the tumor) whereas most mutated class images focus on the necrotic core. This makes sense as the control class has the worse prognosis, thereby the tumor growth is the feature to observe.

2.4.7 Failure Cases and Limitations

Limitation: R2D2-GAN, in its current state, can only be applied to the tumor crops, and not the tumor microenvironment. Thus, to generate whole MRI images using R2D2-GAN, there needs to be an additional step of learning the tumor microenvironment.

Failure Cases: It is difficult to discern the errors visually if the generated images are of good quality, especially when the desired features to be learned are unapparent. To identify the failure cases in these cases, we employed SIF-based filtering. When the SIF score of an image is too low, the image can be filtered and observed for insights. In Figure 2.9, we show some results where the SIF score is less than 0.2 (this can be varied as needed, and can be chosen empirically based on application). This shows there is some room for improvement in R2D2-GAN for learning images with higher fidelity.

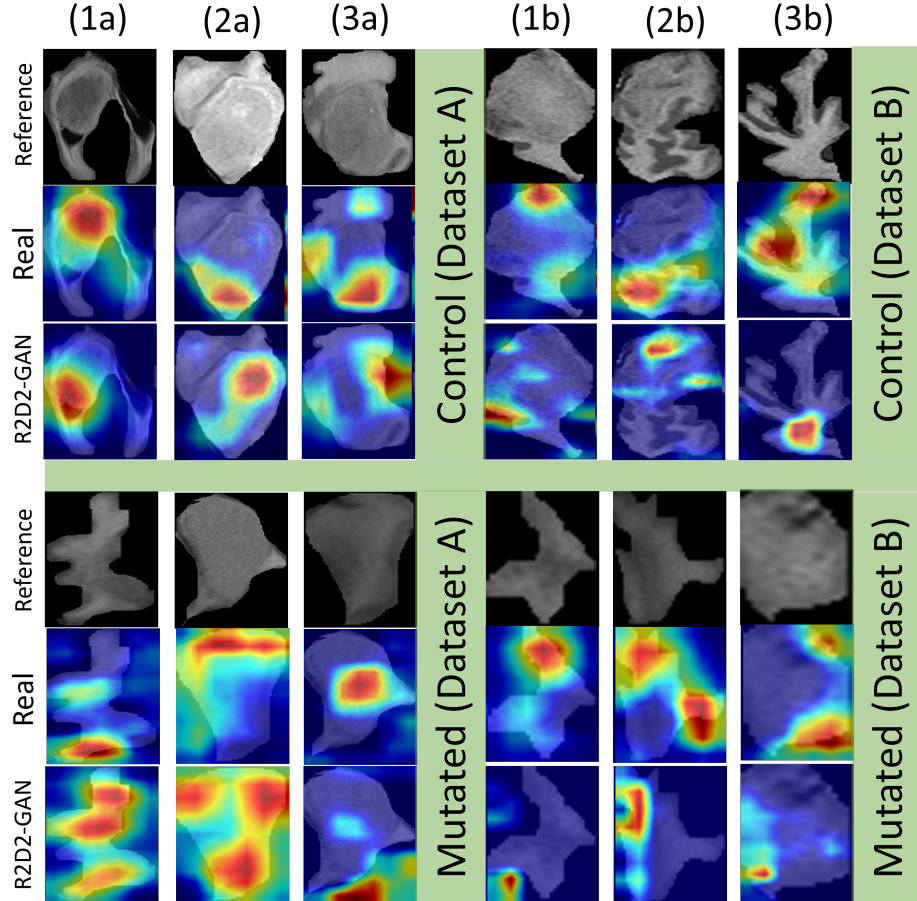


Figure 2.9: Cases where R2D2-GAN fails to learn the same features as real data. The images were selected based on low SIF scores, demonstrating an additional benefit of the SIF score in identifying failure cases in GANs. Colorbar for heatmaps is provided in Figure 2.3.

Chapter 3

Computational Modeling of Tumor Invasion from Limited and Diverse Data in Glioblastoma

3.1 Introduction

A mutation in a tumor affects tumor biology and its radiological appearance. Therefore, the tumor mass properties such as the tumor's location, laterality, shape, necrosis, edema, enhancement, etc. are rightfully considered the most significant in automated mutation detection. As such, many proposed automated methods either focus on the overall brain scan or the quantifiable properties of the tumor mass [134, 112, 172]. Meanwhile, the non-enhancing parenchyma (impacted brain tissue around the enhancing tumor) which affects the assessment by radiologists [527, 230] is yet to receive due importance in machine

learning. A mutation correlates with the tumor properties such as the enhancement of the tumor and its location, and consequently, the impacted tissue surrounding the tumor. Therefore, the extent of this tumor invasion into the surrounding tissue is often associated with disease aggressiveness and patient prognosis [109, 20]. For example, a high extent of tumor invasion may suggest a higher resistance to treatment. The patterns in tumor invasion cause alterations in tissue density, shape irregularities, and disruptions in anatomical structures, which can be extracted as features using image processing or machine learning techniques and can aid in mutation detection.

In this work, we focus on the less explored 19/20 co-gain as our mutation of interest. We analyze and characterize the imaging features that correlate with the visual manifestation of 19/20 co-gain, aiding in its detection using MRI data. Specifically, we isolate and quantify the discriminatory power of the impacted brain tissue surrounding the tumor, in detecting a mutation. Our aims are twofold: firstly, to quantitatively assess the impact of tumor invasion on prognostic analysis, and secondly, to develop a generative model capable of faithfully reproducing the properties of the affected tissue. This approach not only enriches our understanding of tumor behavior but also facilitates the characterization of features essential for comprehensive radiogenomic analysis.

From a computational point of view, the second goal presents many challenges. The correlation of diverse patterns of tumor infiltration with mutation status can be subtle, complex, and often unapparent to the naked eye. Furthermore, being a rarer mutation that manifests in extremely diverse ways, the dataset for 19/20 co-gain is highly unbalanced and limited, with high inter-class similarity and intra-class diversity. A generative model that

can tackle these challenges should be able to learn complex feature associations from limited and highly heterogeneous data. To this end, we propose a Generative Adversarial Network to model Tumor Invasion (TI) called TI-GAN. TI-GAN learns the unconditional statistical distribution of TI using the limited data and improves the accuracy of this estimation via conditional joint learning using context from tumor features such as tumor shape, texture, and location.

We demonstrate that unconditional modeling helps us estimate the generalizable properties of TI, whereas conditional learning using tumor properties helps in generating tumor invasion that is realistic for a given brain scan. Through quantitative and qualitative analysis, cross-dataset testing, radiologist blind tests, and application to a downstream classification task, we demonstrate that our proposed approach can consistently and realistically model tumor invasion properties. While we do not claim that TI-GAN is deployable in clinical settings in its current form, we believe that the extensive qualitative and quantitative analyses support its efficacy in modeling tumor invasion.

3.2 Related Work and Contributions

3.2.1 Related Work

To streamline the prognostic and treatment planning pipeline of GBM, researchers have attempted radiogenomic analysis using various imaging modalities [331, 65]. Many have developed models automating the radiogenomic modeling of the more common mutations IDH1 mutation and MGMT methylation [331, 53, 172, 112]. Specifically, research

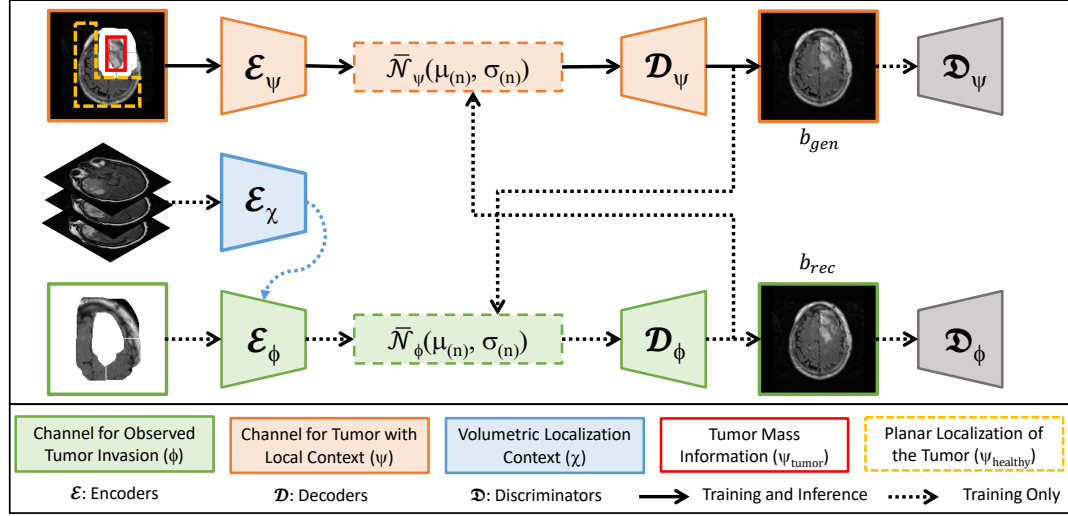


Figure 3.1: Overview of the Tumor Invasion GAN (TI-GAN) Model. The detailed architecture of each TI-GAN module is given in Figure 3.3. The definition of each variable is provided in Table 3.1. The visual representation of the major variables is provided in Figure 3.2.

has been directed at analyzing various properties of the tumor mass and their radiogenomic contributions [134, 46, 526]. 19/20 co-gain, among many others, is a rarer mutation – yet important in its prognostic value. Its rarity makes it more difficult to analyze. Techniques that address such mutations [204, 205] pave a path to study other rare mutations, in achieving a comprehensive assessment of any cancer. Traditional feature descriptors focus heavily on the properties of the tumor such as the tumor location, shape, enhancement, necrosis, etc. [138, 134, 187, 146, 292], while many deep learning techniques focus on the whole 2D or 3D brain scan [103, 283]. While these perform the critical task of establishing that visual biomarkers of mutations exist, it is crucial for clinical settings that mutations not only be detected, but also their discriminatory feature space be interpretable. To that end, there has been research effort toward characterizing some tumor macro-features and their impact in identifying mutations from MRI [204, 205].

Strongly infiltrative tumors such as GBM have a diverse and complex tumor microenvironment [385]. There is a significant therapeutic benefit in learning the properties of this microenvironment since it has been shown to contain important prognostic information [337, 413, 20, 109]. Due to its significance in diagnostic, prognostic, and therapeutic pipelines, there have been efforts to perform mathematical modeling of how the tumor growth impacts the surrounding brain tissue [470, 230]. However, there are many complexities in modeling tumor invasion that way. While deep networks are capable of learning such non-linear and high-level feature associations, there has been little effort in automating these techniques that model the tumor microenvironment. In this work, we quantifiably establish the impact of the tumor microenvironment on the status of 19/20 co-gain mutation. Furthermore, we propose a generative model with the capability to model and reproduce the tumor microenvironment properties.

Generative deep learning models, especially GANs, have received strong interest from researchers due to their immense capability to learn underlying feature associations in images and synthesize high-quality images. In medical imaging applications, they are used in data augmentation [299, 31], data anonymization [387], image enhancement [503, 437], anomaly detection [153], and more [125, 72, 86, 329, 516]. For single modality image generation tasks such as ours, the most commonly used GANs are some variations of Deep Convolution GAN (DC-GAN), Wasserstein GAN (WGAN-GP), Progressively Growing GANs (PGGAN), and Multi-Scale Gradients for GANs (MSG-GAN) [8, 240, 454, 421, 81]. These have shown promise in the synthesis of good-quality images for tasks like data augmentation.

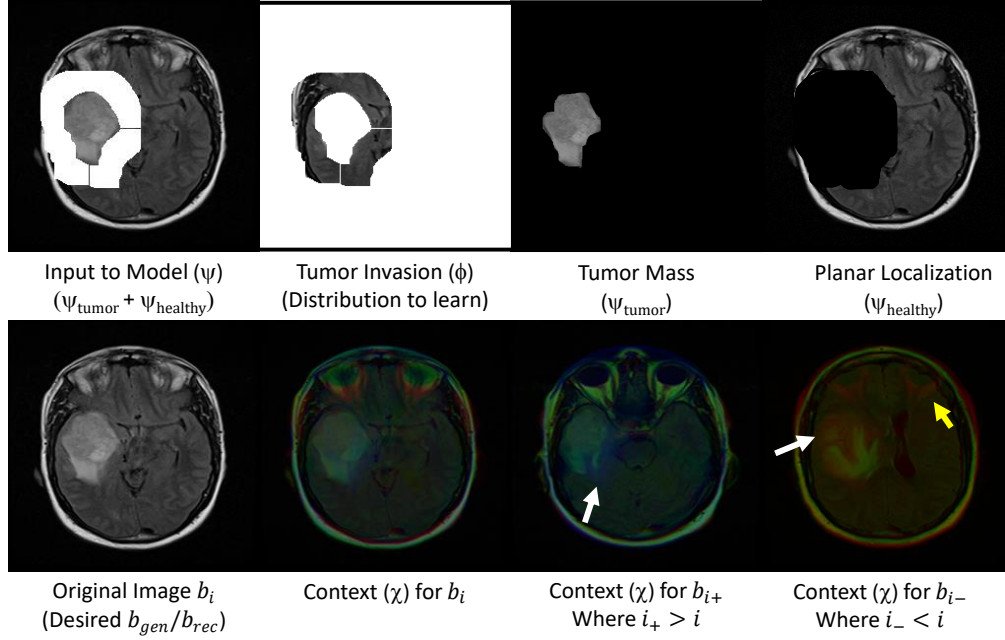


Figure 3.2: Visual representation of the variables used in TI-GAN. For the context images, i_+ means the brain scan above the current slide i in the direction of the top of the head and i_- means the slice below, towards the neck. The white arrows represent where the tumor invasion is being represented by the aggregation described in Figure 3.1. The blue diffused region in i_+ shows the overall tumor invasion as we go in the direction of the neck. Similarly, the red area in i_- represents the same as we go towards the top of the head. As pointed out by the yellow arrow in i_- , the red region also represents the eyes we encounter when traveling in that direction. Thus, this context provides an estimate of the location of the tumor in this direction.

Recently, diffusion models have also risen in popularity due to their ability to generate high-quality images, and have been used in image-to-image translation [280, 229], segmentation [468], and synthesis [90, 91, 328] tasks, among others [108, 219]. These are computationally intensive and have very high data requirements. For rare mutations, the data available is limited, which causes a bottleneck. Some studies have shown that diffusion models have better mode coverage than GANs, whereas GANs have a more efficient sampling strategy [219]. For our task, they represent methods that can unconditionally generate images. However, as previously established, tumor invasion is not an independent variable and its synthesis depends on tumor properties. In other words, methods that perform synthesis with provided context (conditions) are also useful. These consist of methods such as generative inpainting [494] which have been quite popular in image completion tasks. In this work, we analyze both unconditional generation using GANs and conditional generation using generative inpainting to establish the foundation for our proposed model - which performs joint conditional and unconditional estimation for improved performance.

3.2.2 Contributions

- To the best of our knowledge, this is the first deep learning approach to model the impact of mutations on the visual manifestation of tumor microenvironment, a very significant biomarker. This can further be utilized for radiogenomic analysis of other cancers as well as other medical applications such as cancer pseudo-progression analysis.
- We employ joint (conditional and unconditional) estimation and asymmetric masking to tackle challenges of practical datasets such as data diversity and limited dataset, respec-

tively, to learn visually unapparent features realistically. We conduct experiments on multiple datasets and report performance with extensive qualitative analysis and quantitative metrics to demonstrate the robustness of our approach.

- We carry out blind tests performed by three US board-certified radiologists to establish the realism of images of the generated tissue, achieving almost 50% accuracy, i.e., chance-level classification (between real or synthetic images), which is the most desirable.

Table 3.1: List of Symbols used

Symbol	Description
B	3D brain scan
b_i	2D brain slice of interest
N	Number of slices in B
$\phi, \hat{\phi}$	Real and estimated affected tissue
$\psi, \hat{\psi}$	Real and estimated brain sans affected tissue
ψ_{tumor}	Tumor Mass or Tumor Crop
$\psi_{healthy}$	Planar Localization of the Tumor
χ	Volumetric Localization of the Tumor
$\mathcal{E}_{\phi/\psi/\chi}$	Encoders for each parameter
$\mathcal{D}_{\phi/\psi}$	Decoder/Generator for each parameter
$\mathbb{D}_{\phi/\psi}$	Discriminator for each parameter
$\bar{\mathcal{N}}_{\phi/\psi}$	Estimated PDF for each parameter
θ_{cond}	Conditional likelihood estimator
θ_{unc}	Unconditional likelihood estimator
$\hat{\theta}$	Overall estimate of tumor invasion
b_{rec}, b_{gen}	Reconstructed and generated images (training)
$\hat{b}_i$	Novel synthetic 2D scans (inference)

3.3 Technical Approach

Intuitive Explanation: From a computational point of view, our goal is to estimate the underlying statistical distribution reflecting the tumor microenvironment properties using

generative modeling. We treat this as a maximum likelihood estimation (MLE) problem. The challenge, however, is that the parameter to be estimated has a high variance and there are only a small number of data points. This means the distribution of training data is sparse. If we were to train a general GAN, it would perform unconditional statistical modeling of the parameter to learn the general trends. However, in this case, it is simply learning what tumor invasion will generally look like for data from various subjects, with no knowledge of what the tumor size, shape, or location is. This means that it could generate random tumor invasion for a new data sample during test time, which might not look realistic. On the other hand, if we trained a conditional model, by giving the context of tumor shape, size, and location - then the best performance of the estimator will be achieved by overfitting on the limited samples to account for diversity. In other words, since the goal is to perform MLE, the model could theoretically get to the lower bound by doing exact reconstruction on all data samples (since they are limited). Thus, by only performing unconditional estimation, the network will be too generic due to the sparse underlying distribution to learn from. On the other hand, it will not generalize enough to new data if we perform only conditional estimation. To combat this, we perform joint conditional and unconditional estimation. We describe the technical approach of TI-GAN in detail in this section. To demonstrate the performance of fully unconditional or conditional generations, we compare our model with state-of-the-art (SOTA) GANs (for unconditional generation) and a generative inpainting model (for conditional generation).

Problem Formulation: The proposed tumor invasion model is shown in Figure 3.1. The 3D scan of the brain (B) consists of 2D slices $\{b_1, b_2, \dots, b_N\}$. Each slice b_i can then be

disentangled into two subspaces: the impacted tissue (ϕ_i) immediately surrounding the tumor and the rest of the brain tissue (ψ_i), as shown in Figure 3.2 (top row, first two images). The two things to note here are that (1) ϕ and ψ are mutually exclusive, such that $b_i = \phi_i \cup \psi_i$ and (2) ψ_i can further be considered to be a combination of two mutually exclusive subspaces: the tumor mass ($\psi_{i,tumor}$) and the un-impacted brain tissue ($\psi_{i,healthy}$), as shown in Figure 3.2 (top row, 3rd and 4th images). All the variables used are summarized in Table 3.1 for reference.

$$B = \left[b_i \right]_{i=1}^N \text{ s.t. } \phi_i \cup \psi_i = b_i, \phi_i \cap \psi_i = \emptyset, \quad (3.1a)$$

$$\hat{b}_i \rightarrow f\left([\hat{\phi}_i, \hat{\psi}_i, B]\right), \quad (3.1b)$$

$$i \rightarrow f(\chi), \chi \rightarrow f(B), \quad (3.1c)$$

$$\psi_i = \psi_{i,healthy} + \psi_{i,tumor}. \quad (3.1d)$$

Hereafter, for ease of formulation, ϕ_i and ψ_i will be represented in the generalized form, ϕ and ψ .

The goal of TI-GAN is to estimate the tumor-impacted tissue (ϕ_i). As established previously, the impacted tissue depends on the tumor mass properties such as its shape, texture, and location. Without these features, TI-GAN will perform unconditional generation which is not realistic. Thus, we provide the tumor features as input in the form of local context. In other words, we provide $\psi_{i,tumor}$ and $\psi_{i,healthy}$ as inputs to the network. To provide tumor information on a global scale, we use a Volumetric Localization (VL, χ)

that provides context regarding the overall location of the tumor, and overall tumor diffusion patterns in the 3D scan. This allows the model to learn the feature associations of a particular 2D slice and its neighboring slices. Thus, the network inputs and outputs are:

During training: **Inputs:** ϕ , ψ , and χ ; **Outputs:** $\hat{\phi}$ and $\hat{\psi}$,

During inference: **Input:** ψ ; **Output:** $\hat{\phi}$,

where ψ input contains information about the tumor mass and the relative 2D location of the tumor through $\psi_{i,tumor}$ and $\psi_{i,healthy}$, respectively (i.e., the local context). We do not provide χ as input during inference to allow for the practical utility of TI-GAN so that during the test time it should not require full 3D scans for generating tumor microenvironment. Instead, the model relies on the underlying conditional distributions learned during training to make relevant inferences. In other words, during training, we use the input ψ_i and Volumetric Localization (χ) obtained from the 3D scan (B) to develop a statistical model that can estimate the surrounding tissue $\hat{\phi}_i$. During inference, a brain scan with no tumor invasion $\hat{\psi}_i$ is given as the input, and the tumor invasion $\hat{\phi}_i$ is estimated using the learned statistical distribution of ψ , ϕ , χ . Since χ is unknown during inference, we learn it as an implicit variable, conditioned upon ϕ . This means that every unconditional estimate of $P_{\hat{\phi}}$ essentially comes from a more realistically learned distribution $P_{\phi|\chi}$. Subsequently, a synthetic brain scan ($\hat{b}_i$) can be estimated by using $\hat{\phi}_i$ and $\hat{\psi}_i$, which is the desired output of this model. This problem formulation is given by Eq (3.1). It is to be noted that since i is a representation of the volumetric location of a slice, it can be estimated using χ .

We can model the subspace representation of ϕ_i and ψ_i separately and learn to generate the impacted brain tissue surrounding the tumor, similar to an inpainting task.

Without affecting the formulation, during disentanglement, we can include the edge of the tumor in the surrounding region, so that it is now a part of ψ_i . We do this to avoid spurious edge artifacts and generate a smoother transition between the tumor edge and the impacted tissue. We vary the radius of the mask in ψ_i to account for varying extents of tumor enhancement and thereby tackle the limited nature of the dataset.

Estimating Tumor Invasion (Encoding): As mentioned, the two important features we focus on are the tumor mass properties and the relative location of the tumor in the brain. We want to estimate $\hat{\phi}$ (impacted tissue) given ψ (tumor properties). To build an estimate of the tumor invasion, our model has to learn the distributions of ϕ and ψ relative to one another. To this end, we use convolutional encoders $\mathcal{E}_\phi$, $\mathcal{E}_\psi$ for ϕ and ψ , respectively. For the two disentangled paths of ϕ and ψ , the encoders generate subspace representations based on their distributions P_ϕ and P_ψ . It should be noted that ϕ and ψ are mutually exclusive, but not independent. To obtain estimates of underlying distributions, we design $\mathcal{E}_\phi$ and $\mathcal{E}_\psi$ to be Conditional Variational Autoencoders (CVAE) [394]. The detailed architecture of each module is shown in Figure 3.3. The encoders $\mathcal{E}_\phi$ and $\mathcal{E}_\psi$ are responsible for estimating the parametric representation of ϕ and ψ with respect to their relative subspaces. For modeling desired properties from diverse input, we take inspiration from [512, 513]. However, a typical pluralistic task [512] aims to generate realistic and diverse output given high variance prior (diverse input), whereas in our case it has to do with realistic and semi-deterministic generation given diverse input. In other words, the tumor invasion estimated by TI-GAN cannot be unconditionally creative - it has to make biological sense, conditioned upon the tumor’s properties. At its core, this means enforcing a tighter lower variational bound on

conditional log-likelihood than shown in [512, 513]. We do this by incorporating global and local context via χ and ψ , respectively. The local context is readily available as ψ is given as input to the encoder directly. To incorporate global context χ , we compute a context flow and cast the information of the relative tumor location onto a 3D scan which can then be encoded during training.

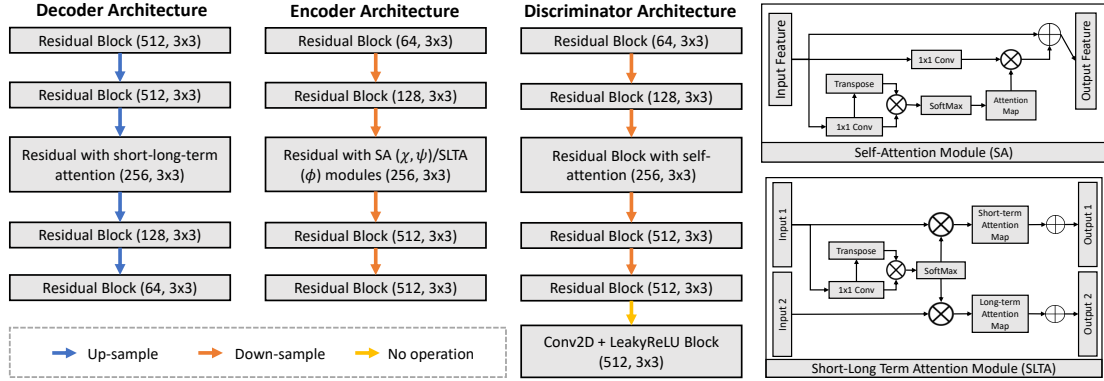


Figure 3.3: (left) Detailed Architecture of the Encoders, Decoders, and Discriminators of the proposed TI-GAN model. (right) Details of the Self-Attention (SA) and Short-Long Term Attention (SLTA) modules. For the Encoder ($\mathcal{E}_\phi$), input 1 is the Encoder feature, and input 2 is the context from $\mathcal{E}_\chi$ (described in Section 3.3). For the decoder ($\mathcal{D}_\phi$), Input 1 is the feature corresponding to ϕ , and Input 2 corresponds to ψ , which helps in joint estimation. Vice versa for $\mathcal{D}_\psi$, as described in Section 3.3

Computing Global Context Flow: The process of computing χ is shown in Figure 3.4.

The idea is to generate an auxiliary 2D brain scan that incorporates the information on the relative 3D location of the tumor. Thus, we exploit the three channels of any RGB image, to encode this information. Given a 2D brain slice b_i , the R-channel of χ represents a cumulative flow of all the slices starting from the top of the head (b_1) until b_i . The G-channel is b_i . The B-channel is the cumulative flow of all the slices starting from b_i until the neck (b_N). This is represented by the equations in Figure 3.4.

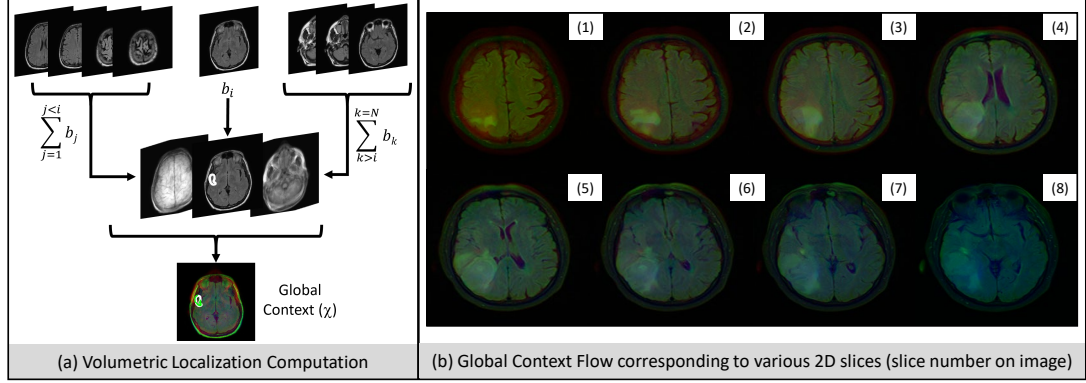


Figure 3.4: (a) Computation of the volumetric localization context (χ). (b) Visual representation of how the χ varies with varying 2D brain slices for the same patient. Images (1) and (8) represent the topmost and bottom-most slices of tumor occurrence. We note that the extent of the red/blue hue surrounding the tumor is indicative of the relative position and global growth patterns of tumor invasion.

The algorithm for computation of χ , and a visual representation of how χ varies with various slices of the same brain scan are shown in Figure 3.4. We see from the intensity of red or blue hues surrounding the tumor region in χ , where the tumor lies, as well as a global representation of the extent of tumor invasion and diffusion patterns. For an isolated view into the R- and B-channel of χ , we refer to Figure 3.2 (bottom row, last two images) which shows the red and blue hues of various intensities reflecting the patterns of brain tissue variations (white/yellow arrows, bones, eyes, etc.).

Incorporating Context: Accuracy in estimated tumor invasion ($\hat{\phi}$) requires knowledge of tumor properties. Thus, we use the key tumor properties as context. The most significant feature, the tumor mass, is incorporated via ψ_{tumor} in the input. $\psi_{healthy}$ is the Planar Localization Context (PLC), which augments the estimation of tumor invasion by providing the relative tumor location in the 2D slice. Meanwhile, χ provides the Volumetric Localization Context (VLC), a cumulative flow representation of a 3D scan onto one 2D

scan. We incorporate the local context via self-attention modules as shown in Figure 3.3. To incorporate the global context, we perform cross-attention using the Short-Long Term Attention (SLTA) module. The encoders $\mathcal{E}_\psi$ and $\mathcal{E}_\chi$ are regular encoders with self-attention modules (Figure 3.3). However, as previously mentioned, $\mathcal{E}_\phi$ incorporates the global context implicitly - and thereby has to learn from $\mathcal{E}_\chi$. Thus, we use SLTA module (Figure 3.3), where Input 1 and Input 2 are features from $\mathcal{E}_\phi$ and $\mathcal{E}_\chi$, respectively.

We can remove the local context information (without affecting the formulation) and still get some estimate of tumor invasion. However, the network does not have enough information on the tumor growth properties without this to realistically learn tumor invasion patterns. Similarly, without 3D context, the encoders lose contextual knowledge of diffusion trends in the tumor along the rest of the brain, causing unrealistic infiltration patterns. The results with and without context are reported as an ablation study. For a slice b_i , the three channels of VLC are computed as shown in Figure 3.1 (right). If the z-channel goes from the neck to the top of the head (ref. Figure 3.1 and Eq (3.1)), the first channel of χ is an aggregation along the z-channel prior to b_i (neck to current 2D slice being processed), the second channel is ψ , and the third channel is a flow along z-channel post b_i until N (from the current slice to the top of the head). During training, both contexts act as known priors to determine the conditional estimation of the tumor invasion.

Ensuring Realism in Output (Conditional Estimation): While we now have a variety of input distributions being modeled, we still need to generate the tumor microenvironment ($\hat{\phi}$) that makes biological sense. Since there is a limit to how creative the generated output can be for a medical task (the estimated tumor invasion has to be realistic), we need

a tighter lower bound on the estimation than used in [512]. To achieve this, given the localization context of a brain and tumor properties, realistic generation of surrounding tissue is made possible using additional conditions on the global context χ (VLC). The CVAE $\mathcal{E}_\chi$ reparameterizes the global context (χ) for this task.

During training, the goal is to get the best estimate of the generalized distribution of ϕ using ψ and χ such that during inference, the best estimate $\hat{\phi}$ (or implicitly, $\phi|\chi$) can be efficiently inferred from only $\hat{\psi}$. Therefore, during training, the network uses the known posterior P_ϕ conditioned upon χ to find the best estimate of the prior P_ψ . At the same time, P_ϕ tries to best estimate P_ψ . In the training steps, ϕ , ψ and χ are known. At every iteration, this means that the goal is to minimize the KL divergence between the known prior P_ψ and the ϕ channel's estimate $P_{\hat{\psi}|\chi}$ and that between $P_{\phi|\chi}$ and the ψ channel's estimate $P_{\hat{\phi}}$. During inference, only the input ψ is available, which means the final objective of the network is to find the best estimate of the posterior $P_{\phi|\psi}$, assuming χ as the implicit variable being learned jointly with ϕ . Thus, during training, the conditional likelihood function has a lower bound as shown in Eq (3.2).

$$\begin{aligned} \log P(\phi | \psi, \chi) \geq & -\text{KL}[q_\phi(z_\phi | \phi, \psi, \chi) \| P_\psi(z_\phi | \psi, \chi)] \\ & + \mathbb{E}_{q_\phi(z_\phi | \phi, \psi, \chi)} [\log P_\theta(\phi | z_\phi, \psi, \chi)]. \end{aligned} \quad (3.2)$$

where z_ϕ is the latent vector (conditioned upon the implicit variable χ), q_ϕ is the posterior sampling function, P_ψ is the conditional prior, P_θ is the likelihood function. Since the tissue ϕ and tumor ψ are exclusive sets, the posterior sampling can only depend on ϕ . Therefore, our estimator function now reduces to:

$$\begin{aligned} \log P(\phi | \psi, \chi) &\geq -\text{KL}[q_\phi(z_\phi | \phi, \chi) \| P_\psi(z_\phi | \psi, \chi)] \\ &\quad + \mathbb{E}_{q_\phi(z_\phi | \phi, \chi)} [\log P_\theta(\phi | z_\phi, \psi, \chi)]. \end{aligned} \quad (3.3)$$

This equation represents the conditional estimation of the tumor invasion when all the inputs are known. Using this estimation alone in a generative model will cause the network to overfit on the training data (learn the exact distribution), resulting in a lack of generalization to unknown inputs.

Ensuring Generalization to Unknown Inputs (Unconditional Estimation): The current data has limited yet highly diverse samples. Applying this estimator to high dimensional, high Degree of Freedom (DoF) tasks such as ours leads to difficulty in convergence [512]. For high variance priors such as ours, since ϕ and ψ are coupled during training, the model can approach the desired lower bound by overfitting, i.e., not learning the overall distribution but the exact solutions for the limited number of samples. For our data, this is highly possible because the model prefers to do exact reconstruction given tumor properties than to do sparse modeling of heterogeneous samples to learn generalized trends of tumor invasion. In other words, the randomized z_ϕ from the learned distribution will get ignored since the network exactly knows $p(\phi|\psi)$. To tackle this issue, [512] included an unconditional likelihood estimation of ϕ . The unconditional estimation using a standalone, fixed prior and random posterior sampling is:

$$\begin{aligned}\log P(\phi) &\geq -\text{KL}[q_\phi(z_\phi | \phi) \| P(z_\phi)] \\ &\quad + \mathbb{E}_{q_\phi(z_\phi | \phi)}[\log P_\theta(\phi | z_\phi)].\end{aligned}\tag{3.4}$$

However, in our case, we have an implicit variable being learned coupled with ϕ , which is only available during training. Thus, in our unconditional estimation, we learn the conditional distribution of the coupled variables, which can be directly sampled during inference as follows:

$$\begin{aligned}\log P(\phi | \chi) &\geq -\text{KL}[q_\phi(z_\phi | \phi, \chi) \| P(z_\phi)] \\ &\quad + \mathbb{E}_{q_\phi(z_\phi | \phi, \chi)}[\log P_\theta(\phi | z_\phi, \chi)].\end{aligned}\tag{3.5}$$

This is called unconditional estimation because it does not know ψ . Thus, it acts similarly to a standard generative model. Not knowing the information about tumor mass and the 2D location of the tumor means that the network will generate tumor invasion that is not realistic to the given input.

To summarize, the estimator in Eq (3.3) learns the tumor invasion when the tumor mass and the 2D location are known. This will be realistic to known inputs but can not be generalized to unknown data. On the other hand, the estimator in Eq (3.5) will always generate diverse outputs but will not be realistic to the given input. Thus, for realistic estimation on diverse inputs, we perform a joint estimation using the conditional and unconditional estimator. In other words, the model will predict random tumor invasion

using Eq (3.5) and will narrow down to predict the realistic invasion when paired with the additional information provided by the tumor in Eq (3.3). This is shown in Eq (3.7c).

Joint Estimation (Decoding): The crux of joint estimation is that both estimators are learning from each other. This is made possible by SLTA modules in the decoders $\mathcal{D}_\psi$ and $\mathcal{D}_\phi$. For SLTA in $\mathcal{D}_\psi$, Input 1 is the feature vector sampled from the estimated distribution of ψ , and Input 2 is from ϕ . For $\mathcal{D}_\phi$, Input 1 comes from ϕ and Input 2 comes from ψ . This is represented as the dotted lines in Figure 3.1. Essentially, the learned representations of ϕ and ψ are aiding each other via SLTA to bolster the estimates $\hat{\psi}$ and $\hat{\phi}$, respectively.

Model Convergence: This estimator in Eq (3.3) cannot be solved definitively since the sampling function $q_\phi(\cdot|\phi, \chi)$ and context input (χ) are only available during training. In other words, during inference, our sampling using the learned distribution of P_ϕ conditioned upon ψ and χ must efficiently synthesize desired brain tissue $\hat{\phi}$, independent of any prior information on ϕ and χ . Thus, we should be able to obtain an estimate of $\hat{\phi}$ without $q_\phi(\cdot|\phi, \chi)$, using only observed prior $\hat{\psi}$ and the sampled latent variable z_ϕ . Combining this with our original estimate for training, we get a new lower bound as in Eq (3.6).

$$\begin{aligned} \log P(\phi | \psi, \chi) &\geq \lambda \times \left(-\text{KL}[q_\phi(z_\phi | \phi, \chi) \| P_\psi(z_\phi | \psi, \chi)] \right. \\ &\quad \left. + \mathbb{E}_{q_\phi(z_\phi | \phi, \chi)} [\log P_\theta(\phi | z_\phi, \psi, \chi)] \right) \\ &\quad + (1 - \lambda) \times \left(\mathbb{E}_{P_\psi} [\log P_\theta(\phi | z_\phi, \psi)] \right). \end{aligned} \quad (3.6)$$

where λ is a number between 0 and 1. We vary λ randomly between epochs for robust estimation with and without q_ϕ .

Objective Function: While training, the network needs to simultaneously maximize the likelihood and the unconditional prior. The new objective function has both the conditional likelihood function Eq (3.6) and unconditional prior estimator Eq (3.5). When using the posterior sampling function, the network favors reconstruction (b_{rec}), whereas when using the learned conditional prior, the network generates output from estimated distribution (b_{gen}). The unconditional estimator in Eq (3.4) and the second term in Eq (3.2) both facilitate reconstruction of the brain scan (θ_{rec} corresponding to b_{rec}). Meanwhile, the modified term in Eq (3.6) facilitates the novel generation (θ_{gen} corresponding to b_{gen}). We can combine those terms to get the final estimator function as shown in Eq (3.7c). This allows for improved generalizability of the model.

$$\begin{aligned}\theta_{unc} = & -\text{KL}[q_\phi(\cdot, \chi) \| P_\psi(z_\phi)] \\ & + \mathbb{E}_{q_\phi(z_\phi | \phi, \chi)} [\log P_{\theta_{rec}}(\phi | z_\phi)],\end{aligned}\tag{3.7a}$$

$$\begin{aligned}\theta_{cond} = & \lambda \times \left(-\text{KL}[q_\phi(z_\phi | \phi, \chi) \| P_\psi(z_\phi | \psi, \chi)] \right. \\ & \left. + \mathbb{E}_{q_\phi(z_\phi | \phi, \chi)} [\log P_{\theta_{rec}}(\phi | z_\phi, \psi, \chi)] \right) \\ & + (1 - \lambda) \times \left(\mathbb{E}_{P_\psi} [\log P_{\theta_{gen}}(\phi | z_\phi, \psi)] \right),\end{aligned}\tag{3.7b}$$

$$\hat{\theta} = \arg \max [\theta_{unc} + \theta_{cond}].\tag{3.7c}$$

where θ_{unc} , θ_{cond} are the unconditional and conditional maximum likelihood estimates, and $\hat{\theta}$ is the overall estimate.

To obtain a differentiable loss for the network in Figure 3.1, we grouped terms similar to standard loss functions. The KL terms in Eq (3.7a and 3.7b) for unconditional

and conditional priors are the KL divergence losses for image reconstructed given ϕ (b_{rec} , output of decoder D_ϕ) and the image generated given ψ (b_{gen} , output of decoder D_ψ), respectively (Figure 3.1). To maximize the likelihood during training (relating to b_{rec}), we can directly employ an L1 loss between the input training image and b_{rec} . When learning the conditional prior, the generated image's $\hat{\psi}$ (relating to b_{gen}) should be similar to the deterministic ψ , so we employ L1 loss between those, not the whole images. The two KL divergence losses and two reconstruction losses make up the terms of the estimator. We must also ensure the generated images do not diverge from the underlying distributions. We, therefore, have two discriminators $\mathcal{D}_\phi$ and $\mathcal{D}_\psi$ to minimize KL divergence between real and generated distributions, adding two adversarial loss terms. Thus, the final loss:

$$\mathcal{L} = (\mathcal{L}_{KL}^{rec} + \mathcal{L}_{L1}^{rec} + \mathcal{L}_{adv}^{rec} + \mathcal{L}_{KL}^{gen} + \mathcal{L}_{L1}^{gen} + \mathcal{L}_{adv}^{gen})/6. \quad (3.8)$$

During inference, we perform the likelihood estimation as in Eq (3.7c) to obtain the best estimate of $\hat{\phi}$ given $\hat{\psi}$. The final output $\hat{b}_i$ is obtained by combining the given $\hat{\psi}$ and estimated $\hat{\phi}$. The overall algorithm is described in Algorithm 2.

Addressing varying extents of tumor invasion: While we tackle diversity in small datasets using joint estimation as shown in Eq (3.7c), we also employ an additional step to ensure the model trains well on limited data. As previously mentioned, the first step of TI-GAN is to disentangle the tumor invasion and the rest of the brain by masking out the desired region. However, for diverse tumors such as GBM, there is no definitive radius of impacted tissue. To address varying extents of tumor enhancement, as well as to ensure robustness in training and subsequent inference, the masks are made irregular,

and of varying radii. Based on the average relative size of the tumor (for a brain scan of 256x256), we perform asymmetric masking. We randomly pick the outward mask radius (r_{mo}) from 10-30 pixels, masking outward of the tumor-tissue boundary, and vary inward radius (r_{mi}) 5-10 pixels, masking inwards. This property is particularly useful because it does not require the tumor delineation to be very precise, and it caters to varying degrees of how aggressive a tumor can be. Additionally, this step also doubles as a data augmentation step - effectively providing more samples for robust training of TI-GAN for limited datasets.

Algorithm 2 Algorithm for TI-GAN Training

Input : ψ (varying asymmetric masking), ϕ , χ

Output: Estimated $\hat{\phi}$ given $\hat{\psi}$

At Epoch 0: $\mathbb{E}(\phi)$ is random (e.g., all gray pixels for estimated tumor invasion)

while $loss > threshold$ **do**

Step 1: Compute χ using VLC method (Figure 3.1)

Step 2: Encode ϕ and ψ using $\mathcal{E}_\phi$ and $\mathcal{E}_\psi$ with $\mathcal{E}_\chi$ as context

Step 3: Sample input to decoders from $\tilde{\mathcal{N}}_\phi$ and $\tilde{\mathcal{N}}_\psi$ to generate estimates of $\hat{\phi}$ and $\hat{\psi}$

Step 4: Estimate ψ given ϕ and χ . This will be deterministic since all inputs are known

Step 5: Maximize $\mathbb{E}(\hat{\phi}|\chi)$ using current best estimate of ψ (Eq 3.6)

Step 6: Minimize $\mathcal{L}$ and update $\mathcal{E}$, $\mathcal{D}$ and $\mathcal{D}$ for both ϕ (reconstruction) and ψ (synthesis) channels

end

3.4 Experiments and Results

3.4.1 Description of Datasets

We report our results on two datasets, a dataset collected at Emory University (Dataset A) and the publicly available The Cancer Imaging Archive Glioblastoma (TCIA-GBM) [71] Dataset (Dataset B). Dataset A is very limited and diverse whereas Dataset B is highly imbalanced. The evaluation using Dataset A highlights the effectiveness of TI-GAN

Table 3.2: Mutation detection in Dataset A and B using only real data generated using whole images, tumor regions, and only tumor crops. Metrics are - ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. WI: Whole Images, TR: Tumor Region, TC: Tumor Crop. The models are trained using an 80-20 patient-wise split of the data, using a ResNet18 classifier. The metrics reported are mean (standard deviation) over 10 folds. WI outperforms TR since it also contains location information. Comparing the three accuracies provides us with a quantification of the impact of tumor invasion/microenvironment on mutation detection.

		Dataset A					Dataset B				
	Input	ACC(PL)	ACC(IL)	SENS	SPEC	DICE	ACC(PL)	ACC(IL)	SENS	SPEC	DICE
	WI	92 (08)	90 (06)	85 (07)	95 (05)	86 (08)	81 (09)	78 (12)	73 (08)	86 (08)	70 (10)
	TR	88 (10)	84 (13)	77 (13)	88 (08)	77 (12)	77 (11)	76 (09)	71 (11)	83 (09)	68 (12)
	TC	85 (12)	80 (09)	67 (09)	84 (06)	69 (07)	74 (09)	73 (12)	66 (10)	77 (06)	64 (13)

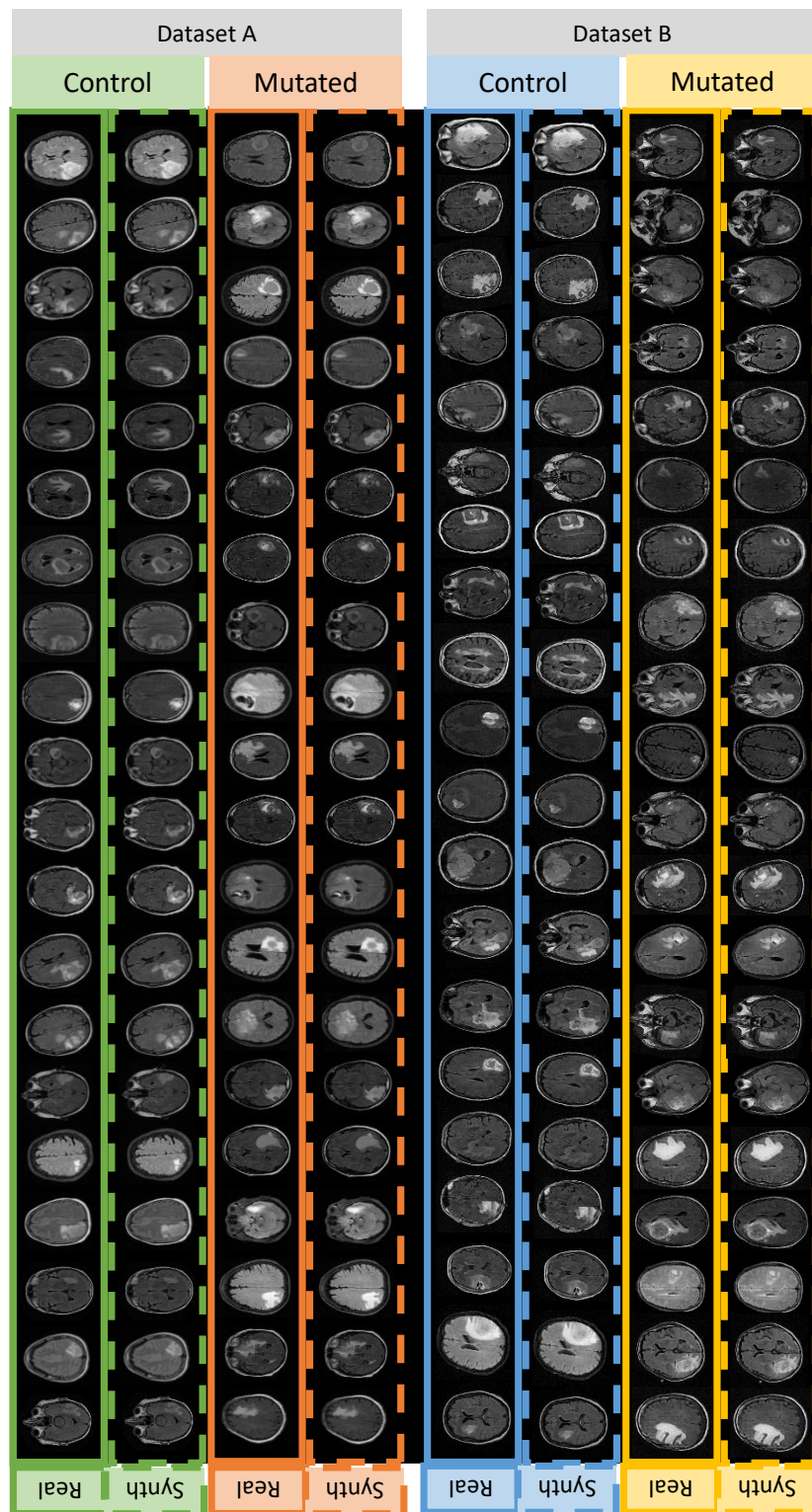


Figure 3.5: Qualitative Analysis of the performance of TI-GAN. The images represent the intra-class diversity of the real data and demonstrate the capability of the proposed approach in capturing the diversity during synthesis. The results are obtained using the models trained on an 80-20 patient-wise data split. For both datasets and classes, the results of the proposed approach are very similar to real data, thus establishing the efficacy of TI-GAN modeling.

in learning discriminative features from limited data, whereas, the evaluation for Dataset B demonstrates its utility in combating data imbalance. Both datasets exhibit high diversity. Together, they represent two common, critical issues faced by real-world medical data and the efficacy of TI-GAN in handling diversity and lack of data in one or both classes.

Dataset A. Dataset A is a private dataset that consists of FLAIR MRI data for 25 patients with known mutation status. This cohort has 14 control and 11 mutated patients. The control class has cases without 19/20 co-gain and mutated class, with co-gain. Both classes have about 100-130 images each containing a tumor, making it a limited dataset. The use of human data in Dataset A was approved by the Institutional Review Board (IRB) of Emory University, Protocol #95120.

Dataset B. The public Dataset B has a cohort of 164 patients with FLAIR MR images, which is divided into 133 control and 31 mutated patients. To extract the 19/20 co-gain status, the corresponding molecular data for all patients was extracted from The Cancer Genome Archive (TCGA) website [430]. It was done by taking gistic data from TCGA and thresholding for each arm of the chromosome, checking if more than 25% of the genes are duplicated on both the long and short arms of chromosomes 19 and 20 [204].

3.4.2 Experimental setup

TI-GAN uses jointly conditional and unconditional estimations for synthesizing the tumor microenvironment. In the following subsections, we first establish the mutation detection power of tumor invasion, followed by qualitative and quantitative results on the three aspects of generated images: image quality, diversity, and realism. We compare these results with commonly used unconditional GANs (WGAN-GP, PGGAN, and MSG-GAN)

and conditional generation using inpainting (Generative Inpainting (GI)). Furthermore, to establish realism, we perform a blind test by 3 US board-certified radiologists and report the qualitative and quantitative results. To demonstrate the generalizability of TI-GAN, we perform cross-dataset testing. To establish the importance of local and global context - in other words, to establish the importance of knowing tumor shape, texture, and location information in tumor invasion modeling - we report quantitative and qualitative results of ablation studies without using local context (primarily tumor shape and texture) and global context (primarily tumor location), respectively. For TI-GAN-generated images (vs other models), cross-testing results, and ablation studies, we also report the impact of using synthetic images in mutation detection. Finally, to demonstrate the efficacy of TI-GAN in data augmentation tasks, we use TI-GAN-generated images to balance the minority mutated class in the unbalanced Dataset B - and report the performance improvement.

3.4.3 Discriminatory Power for Mutation Detection

We begin our analysis by isolating and quantifying the impact of the tumor microenvironment in predicting the mutation status. This establishes the extent to which the mutation manifests in the surrounding tissue. For this, we perform classification of the presence or absence of mutation using real data at various levels of abstraction of the feature set. We perform three classification tasks, one using whole brain images, the second using the tumor regions (tumor mass plus the surrounding tissue) and the third using only tumor mass. The comparison of the classification performance of Tumor Region (TR) and Tumor Crop (TC) in Table 3.2 allows us to isolate the impact of only tumor invasion on the presence or absence of mutation. This is because, referring to Figure 3.2, $TC = \psi_{tumor}$, and

$TR = \phi + \psi_{tumor}$. For the tumor region, we extract rectangular patches of size 25 pixels around each direction of the tumor boundary, based on 256x256 brain slices. We evaluate the classification performance for both Dataset A and B, and the results are reported in Table 3.2. For the highly unbalanced Dataset B, we perform oversampling of the minority class to perform classification using standard data augmentation techniques like rotation and flipping. The classifier used is ResNet18, with a learning rate of 0.0001, Adam optimizer and trained for 30 epochs with early stopping criteria. The classifier uses a 10-fold cross-validation technique - and the results for each metric are reported as mean (standard deviation). It can be seen that including the tumor region increases the overall performance of the classifier by at least 3%, indicating that it contains information reflecting mutation status. We note that the best performance is achieved using whole images. This is because whole images have an additional feature, the location of the tumor, the impact of which has been quantified in [204].

3.4.4 Generated Image Quality and Diversity

Qualitative Results

Diversity: In Figure 3.5, we report the qualitative results demonstrating the diversity of images generated by TI-GAN for Datasets A and B. Note specifically the intra-class heterogeneity of the tumors. For the limited Dataset A, this diversity leads to increased complexity in training. However, a comparison with real data demonstrates that the proposed model can model the desired properties effectively. We note that paired estimation using two channels and spatial context (as shown in Figure 3.1) tackles the data diversity

challenges. In Figure 3.5, we show examples of real data with high intra-class heterogeneity and complex tumor properties. We compare these images with corresponding images generated by TI-GAN and note that TI-GAN is consistently able to reproduce the desired features with high fidelity and perceptual quality (also demonstrated using quantitative metrics) even in the most limited classes such as the mutated class of Dataset A. We note that the diversity of the generated images while maintaining quality represents that the underlying sparse and complex distribution of tumor invasion is being effectively learned using TI-GAN. Specifically, it means that TI-GAN overcomes common issues faced by GANs such as mode collapse when dealing with limited and diverse data.

Quality: It is important that the modeled tumor invasion is of good quality, and reflects the subtle features of mutations being studied. To provide a qualitative analysis, we zoom into the region surrounding the tumor and compare the results with the real data. We also provide a comparison with other generative models. For comparative models, we report three types of results: (a) WGAN-GP to generate whole images, (b) PGGAN and MSG-GAN for the high-resolution synthesis of whole images, and (c) Generative Inpainting (GI) [494] to synthesize only the surrounding tissue (similar setup to TI-GAN). These results are shown in Figure 3.6 for Datasets A and B. We note that using WGAN-GP to generate whole images does not offer the same level of realism, as it is not able to replicate the quality or details, and shows a lot of blurriness. Two of the most commonly used medical GANs, PGGAN, and MSG-GAN, are known for their high-resolution generation- and we have, thus, included them in our comparisons. We note from the results that when the dataset is

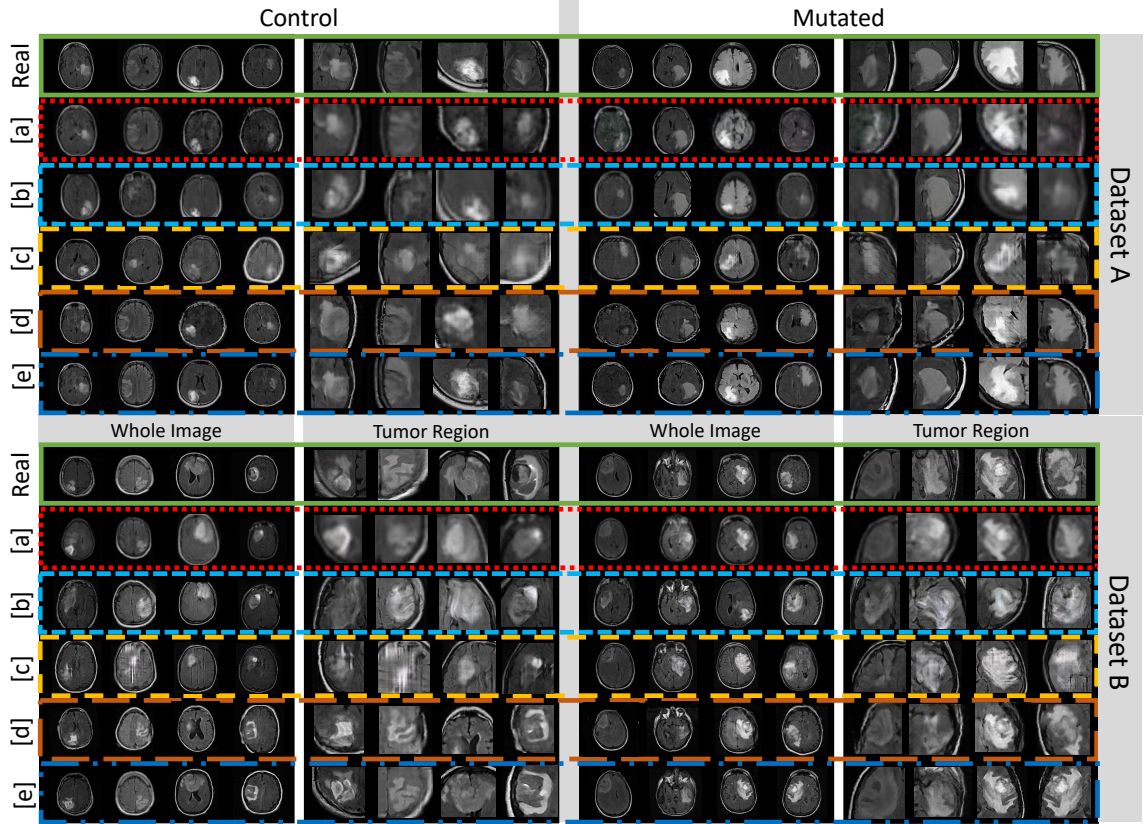


Figure 3.6: Qualitative evaluation of the quality of generated tumor invasion. Results are compared against real data as well as WGAN-GP, PGGAN, MSG-GAN, and Generative Inpainting models. We zoom into the site of the tumor region to highlight the efficacy in reproducing the tumor invasion. Row-wise: [a] WGAN-GP, [b] PGGAN, [c] MSG-GAN, [d] Generative Inpainting, and [e] TI-GAN. We note blurriness in many of the WGAN-GP and PGGAN images whereas MSG-GAN suffers from high frequency artifacts. We also note some unrealistic skull patterns in GI results.

Table 3.4: Exact Synthesis Metrics: Learned Perceptual Image Patch Similarity (LPIPS), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Metric (SSIM) for generated images using TI-GAN, Generative Inpainting (GI) [494], MSG-GAN [81], PGGAN [213] and WGAN-GP [143]. The number of generated images per class is 500 for each GAN.

<i>Input: Whole Images</i>		<i>Dataset A</i>			<i>Dataset B</i>		
Model		LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑
TI-GAN (ours)	GI	0.07	25.02	0.82	0.08	24.51	0.79
		0.17	23.18	0.75	0.18	22.68	0.70
	MSG-GAN	0.29	20.06	0.63	0.21	21.15	0.72
	PGGAN	0.35	19.32	0.66	0.33	19.80	0.68
	WGAN-GP	0.33	18.60	0.60	0.34	19.46	0.68
<i>Input: Tumor Region</i>		<i>Dataset A</i>			<i>Dataset B</i>		
Model		LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑
TI-GAN (ours)	GI	0.17	23.31	0.77	0.21	23.12	0.74
		0.25	19.56	0.69	0.26	20.42	0.68
	MSG-GAN	0.36	16.12	0.51	0.34	18.92	0.70
	PGGAN	0.44	15.82	0.55	0.40	16.78	0.63
	WGAN-GP	0.40	17.32	0.57	0.38	17.54	0.60

Table 3.5: Mutation detection in Dataset A and B using only synthetic data generated using TI-GAN, Generative Inpainting (GI) [494], MSG-GAN [81], PGGAN [213], and WGAN-GP [143] All GANs are trained using the same real training sets and tested on the same corresponding real test sets for all 10-folds (80-20 splits). The classifier used is ResNet18. The metrics are reported as mean (standard deviation) over 10 folds. Performance using real data is used as the baseline. For real data in the unbalanced Dataset B, the minority class is oversampled using data augmentation techniques such as rotation, flipping, etc. All metrics are in percentages. ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. Results reported for classifiers trained on WI: Whole Images, and TR: Tumor Region. We note that the TI-GAN-generated dataset is a good representation of the real data, which is reflected in the classification performance.

		Dataset A						Dataset B					
		ACC(PL)	ACC(IL)	SENS	SPEC	DIC		ACC(PL)	ACC(IL)	SENS	SPEC	DIC	
WI	Real Data	92 (08)	90 (06)	85 (07)	95 (05)	86 (08)		81 (09)	78 (12)	73 (08)	86 (08)	70 (10)	
	TI-GAN	93 (08)	92 (09)	91 (07)	94 (08)	89 (10)		85 (10)	84 (13)	81 (10)	90 (07)	78 (09)	
	GI	90 (09)	88 (11)	84 (12)	91 (07)	84 (10)		82 (12)	82 (14)	78 (12)	82 (10)	74 (13)	
	MSG-GAN	80 (15)	80 (14)	74 (12)	83 (09)	74 (11)		81 (13)	80 (10)	74 (12)	81 (10)	70 (09)	
	PGGAN	90 (09)	86 (09)	80 (12)	90 (06)	82 (09)		75 (11)	77 (09)	70 (12)	78 (15)	67 (12)	
	WGAN-GP	85 (06)	80 (07)	76 (09)	88 (05)	75 (08)		72 (13)	75 (10)	66 (11)	75 (09)	64 (14)	
TR	Real Data	88 (10)	84 (13)	77 (13)	88 (08)	77 (12)		77 (11)	76 (09)	71 (11)	84 (09)	68 (12)	
	TI-GAN	88 (09)	85 (11)	78 (08)	87 (07)	76 (10)		80 (12)	80 (13)	76 (14)	85 (12)	75 (15)	
	GI	84 (14)	82 (11)	72 (10)	85 (12)	73 (12)		77 (14)	75 (13)	71 (11)	82 (15)	72 (14)	
	MSG-GAN	76 (10)	74 (13)	69 (12)	78 (09)	69 (10)		76 (11)	76 (12)	70 (10)	79 (08)	70 (10)	
	PGGAN	79 (12)	76 (13)	70 (11)	80 (10)	76 (13)		76 (14)	76 (15)	72 (10)	76 (11)	70 (12)	
	WGAN-GP	76 (13)	73 (10)	68 (14)	79 (11)	68 (12)		73 (10)	72 (13)	74 (15)	72 (11)	72 (16)	

limited, their performance drops drastically, resulting in checkered or other high-frequency artifacts. Compared with a generative inpainting model with the same setup as TI-GAN, we see that the former suffers from many distortions, specifically for Dataset A. Increasing the size of the dataset improves the performance of all models (see Control class of Dataset B), which is not a limitation for TI-GAN - indicating its robustness against data size and diversity.

Quantitative Results

To perform quantitative evaluation, we present a two-fold analysis: (a) Image quality and diversity metrics, and (b) Using only synthetic data to perform downstream classification tasks. Details for each and their results are provided in this subsection.

(a) Image Quality and Diversity Metrics: To quantify the quality and diversity of images generated, we report various novel synthesis and exact synthesis metrics for the tumor region as well as the whole images as shown in Figure 3.6 (zoomed). The results are reported in Table 3.3 for the novel synthesis metrics and in Table 3.4 for the exact synthesis metrics. We computed these metrics using 500 images per class for both datasets. The metrics are described below. Novel synthesis metrics refer to metrics such as modified Inception Score (mIS) [19] and Fréchet Inception Distance (FID) score [167]. These are useful in evaluating the performance of GANs because they evaluate the similarity/dissimilarity of images generated using deep feature associations, and do not perform exact pixel-wise comparison of generated images with respect to real images. Exact synthesis metrics consist of metrics such as Structural Similarity Index Metric (SSIM) [174], Peak Signal-to-Noise Ratio (PSNR), and Learned Perceptual Image Patch Similarity (LPIPS) [502] measure the

similarity/dissimilarity based on the synthetic image being an identical version of the real data. In our case, novel synthesis metrics are more suitable for GANs such as WGAN, PG-GAN, and MSG-GAN, and exact synthesis metrics are more suitable for GI. The description of each of the metrics is given below.

Fréchet Inception Distance (FID) Score: FID score computes the feature distance between real and generated images. It is a metric of synthetic image quality and diversity that is robust against common image perturbations and can detect intra-class diversity and mode collapse. To compute the FID score, we used PyTorch’s official implementation [379]. The lower the FID score, the higher the quality and diversity of generated images. Since our dataset suffers from a high degree of inter-class similarity and intra-class variance, we also report FID scores for each class, in addition to the average. We note that across both datasets and classes, TI-GAN consistently outperforms all the models.

Modified Inception Score (mIS): Inception Score (IS) evaluates Image Quality and Image Diversity. The lowest value of IS is 1 and the highest is the number of classes (in our case, 2). The higher the IS, the better. Traditional IS is calculated using ImageNet weights. Since our data are different in appearance, we finetuned the last two layers of the Inception model for 20 epochs before calculating IS, which we call the modified Inception Score (mIS). Note that IS has an additional disadvantage in that it cannot detect intra-class diversity or mode collapse. This can be seen in the performance of MSG-GAN in Dataset B, where it outperforms TI-GAN but has a lower FID score.

Learned Perceptual Image Patch Similarity (LPIPS): LPIPS measures the perceptual similarity between two images. It measures this similarity by factoring in texture, color, and

structure that influence how humans perceive visual content. It is typically computed using deep networks that have been trained on large datasets of images. Here, we use AlexNet-based implementation as reported in [502]. The lower the value of LPIPS, the better the quality of generated images.

Peak Signal-to-Noise Ratio (PSNR): PSNR evaluates the quality of a reconstructed image compared to its original version. It is calculated by comparing the maximum power of an image to the power of corrupting noise that affects the fidelity of the image. Here, we use the implementation as shown in [221] which uses the peak value of an image and the MSE between real and synthetic image. The higher the PSNR value, the better the quality of the reconstructed image.

Structural Similarity Index Measure (SSIM): SSIM is also measure of perceptual similarity between real and synthetic images based on structural properties. SSIM ranges between 0 and 1, where 0 means the images are completely dissimilar, and 1 means they are identical. For SSIM and PSNR of unconditional models (WGAN-GP, PGGAN, and MSGGAN) which do not always have exact matches with real data after synthesis, we compute pairwise values between real and generated image sets and pick the generated image with the highest similarity with a given real image. We report the average value of these metrics.

Across all the metrics, we note that our model has consistently superior performance for both datasets (Tables 3.3 and 3.4). For most comparative methods, we see their performance improve for the larger Dataset B. For example, MSG-GAN sees a sharp improvement in the FID score for the control class of Dataset B, which is the largest subset among our data. This trend, however, does not carry over to the rest of the classes.

(b) Application to Downstream Classification In addition to image quality and diversity metrics, we demonstrate the quality of TI-GAN-generated images by evaluating how accurately TI-GAN learns the underlying statistical distribution of the desired features. In its essence, we use only the synthetic data to train the classifier to detect the mutation status on the same test sets as used in Table 3.2. The idea is, if the synthetic training data can detect mutation status without using real data, it has captured the desired underlying feature associations. To measure this, we use a metric called Adversarial Accuracy (AA) as defined in [482]. AA computes the classification accuracies achieved by two classifiers on the same validation set, one trained on real data and one on synthetic data. If both accuracies are close, then the GAN has learned the underlying distribution of the real data accurately. We report accuracies for synthetic data generated by various GANs and real data in the ACC (IL) column of Table 3.5. Comparing real vs synthetic classifier accuracies, it can be seen that TI-GAN outperforms all GANs. TI-GAN for Dataset B slightly outperforms real data as well. Accounting for random seeds in training classifiers and higher standard deviation in TI-GAN, we do not claim that TI-GAN learns any more features than the real data itself. However, we can safely say that it effectively learns features similar to real data.

3.4.5 Realism of Generated Images

Quantitative Results:

We conducted a blind test with three radiologists to determine the realism of our generated images modeling tumor invasion. The task for radiologists was to look at a set of data and determine which images were real and which images were synthetic. This blind

set included images trained on both Datasets A and B. The control set had 44 synthetic images and 40 real images, whereas the mutated set had 45 synthetic images and 39 real images. The number of images was picked at random so as to not have balanced sets. For the quantitative evaluation of these results, we report accuracy, precision, recall, false positive rate, and false negative rate for each radiologist, as well as average performance in Table 3.6. For a realistic generation, there should be no distinction between real and synthetic images in a set. Ideally, our goal is that a balanced blind set performs equal to or worse than chance. We notice that the average blind test accuracy for control and mutated classes was 0.50 and 0.54, which is close to chance. The most significant statistic to be noted is the False Positive Rate, which is 0.46 for control and 0.48 for mutated. This is a very encouraging result as it indicates that the generated tumor microenvironment is realistic enough for radiologists almost half the time.

Table 3.6: Blind test by radiologists to distinguish between real and TI-GAN generated tumor microenvironment. Metrics shown are ACC - Accuracy, PRE - Precision, REC - Recall, FPR - False Positive Rate, and FNR - False Negative Rate. The text in bold is the mean value across all radiologists.

Class	Rad.	ACC	PRE	REC	FPR	FNR
<i>Con</i>	1	0.50	0.47	0.45	0.45	0.55
	2	0.57	0.55	0.52	0.39	0.48
	3	0.43	0.40	0.40	0.54	0.60
	Mean	0.50	0.47	0.46	0.46	0.54
<i>Mut</i>	1	0.61	0.59	0.49	0.29	0.51
	2	0.51	0.48	0.49	0.47	0.51
	3	0.50	0.47	0.69	0.67	0.31
	Mean	0.54	0.51	0.56	0.48	0.44

Qualitative Results:

For the qualitative evaluation of the results of the radiologist blind test, we provide a Visual Confusion Matrix (VCM) in Figure 3.7. In a 2x2 VCM, true positives are examples

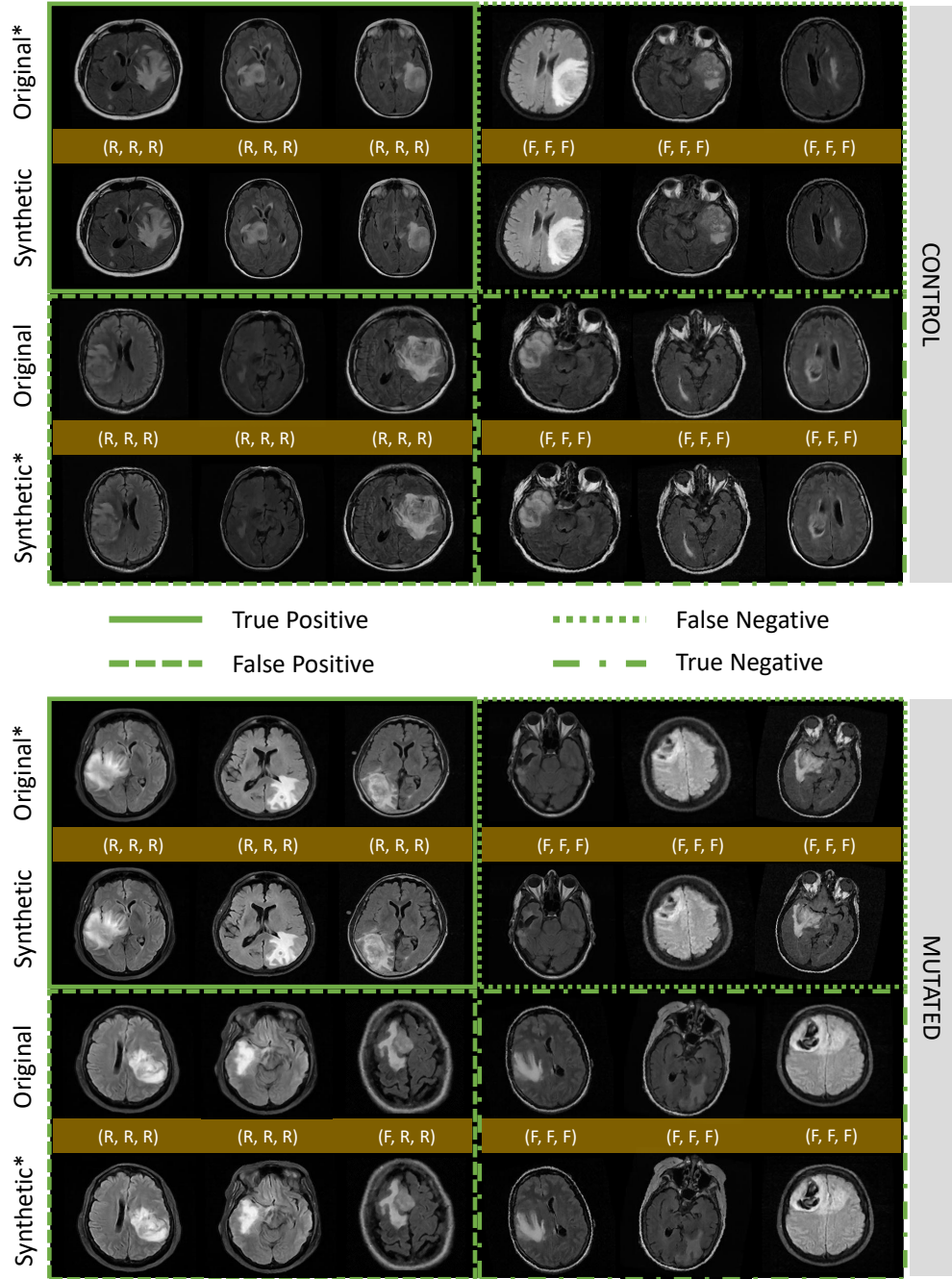


Figure 3.7: Visual Confusion Matrix for Radiologist Blind Test. True Positives are when the blind image was real and was predicted by radiologists as real. False Positives are when the blind image was synthetic but was predicted as real, False Negatives are when the blind image was real but predicted as synthetic and True Negatives are when the blind image was synthetic, and radiologists predicted as synthetic. The white text indicates the predictions by three radiologists (R1, R2, R3). * represents the image that was in the blind set.

of images that were real in the blind set, which the radiologists classified as real, false positives are synthetic images in the blind set, that were classified as real, false negatives are images that were real but predicted as synthetic, and true negatives are synthetic images that were correctly predicted to be synthetic. For ease of reference and comparison, we also show the synthetic/real images corresponding to the real/synthetic images used in the blind set in Figure 3.7. In this figure, we emphasized the cases where the radiologists’ predictions were unanimous (except for one example of the mutated false positive).

3.4.6 Cross-Dataset Testing

Qualitative Results

Medical data are often restrictive, and various clinical centers have different data collection protocols. Therefore, deep-learning models trained on a certain dataset typically perform badly with other datasets [381, 80]. Thus, for clinical viability, machine learning models should be generalizable. To address this, we show synthetic tissue generated by TI-GAN via training on Dataset A and testing on Dataset B and vice-versa. We compare the cross-dataset testing with intra-dataset evaluation results in Figure 3.8. To demonstrate the generalizability of TI-GAN generation, we show the results using cross-dataset testing in Figure 3.8. We note that by training on Dataset A/B and testing on Dataset B/A, TI-GAN is still able to generate tumor invasion properties on a diverse set of images. In this figure, we also show images with high intra-class heterogeneity to show that TI-GAN synthesis is accurate under a wide array of input conditions. We can see that the quality and diversity via cross-domain testing are almost at par with the same-domain testing.

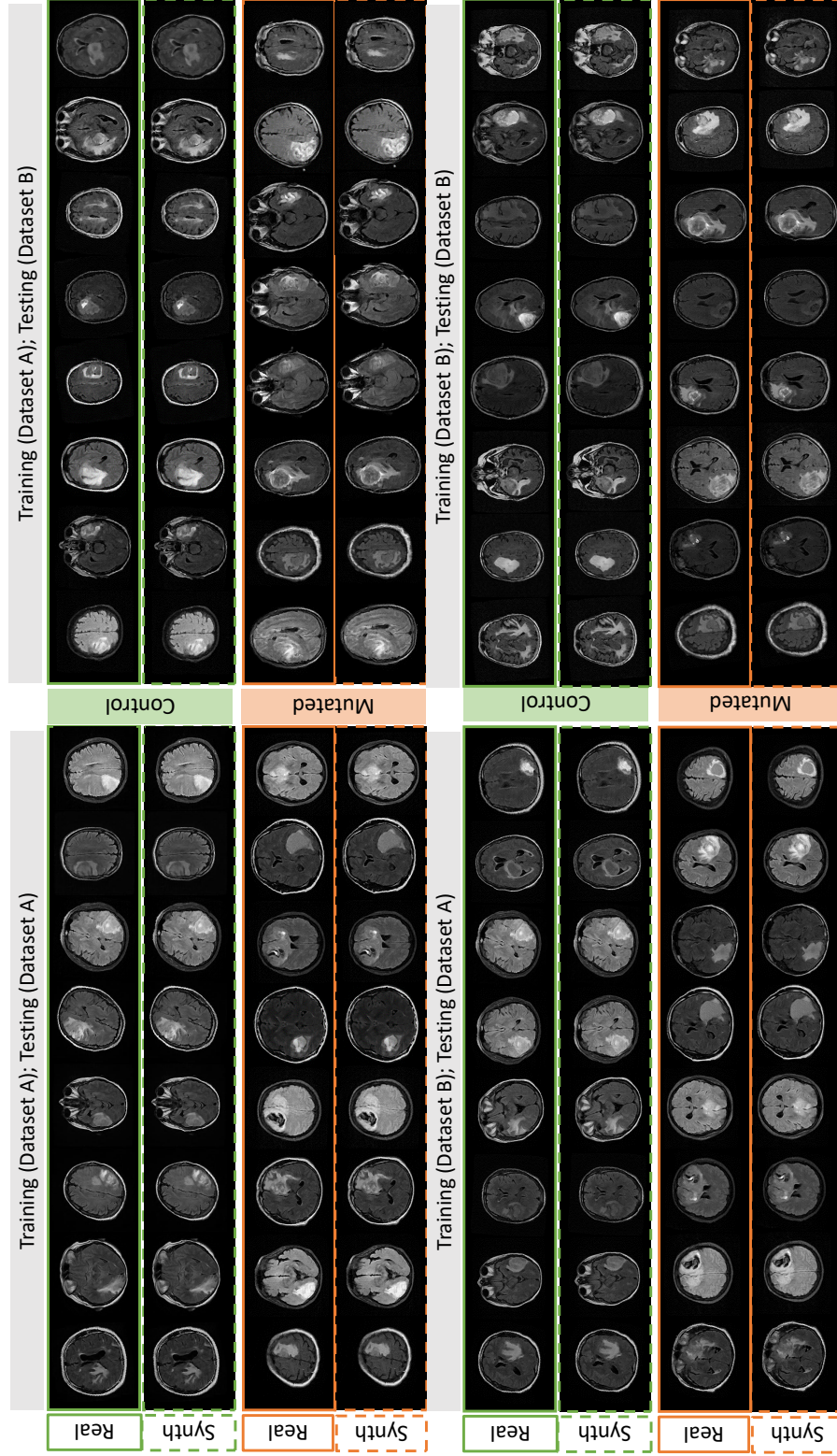


Figure 3.8: Cross-dataset or cross-domain testing performance of TI-GAN. We synthesize the tumor invasion of a dataset using the trained model from the other dataset. We report these results adjacent to the same-domain evaluation for comparison. For cross-dataset testing, models are trained on all the data of the dataset in training, and tested on all images of the other dataset.

Table 3.7: Cross-dataset Testing performance metrics: FID Score, mIS, SSIM, LPIPS, and PSNR for generated images using TI-GAN. Results are shown for the model trained on Dataset B, and tested on Dataset A (Training: B, Testing: A) and vice versa (Training: A, Testing: B). We report results on Whole Images and Tumor Regions, as shown in Figure 3.6. It can be noted that cross-dataset testing results are comparable to same-dataset testing results, demonstrating the generalizability of TI-GAN. Results from same-dataset testing are used for baseline comparison.

<i>Input: Whole Images</i>		<i>Testing Set: A</i>					<i>Testing Set: B</i>				
Training Set		FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑
Same-dataset		86.22	1.76	0.07	25.02	0.82	67.04	1.60	0.08	24.51	0.79
Cross-dataset		90.54	1.74	0.09	24.25	0.81	82.34	1.54	0.11	22.85	0.76
<i>Input: Tumor Region</i>		<i>Testing Set: A</i>					<i>Testing Set: B</i>				
Training Set		FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑
Same-dataset		176.08	1.41	0.17	23.31	0.77	132.17	1.40	0.21	23.12	0.74
Cross-dataset		182.25	1.38	0.19	22.88	0.76	150.47	1.31	0.23	20.21	0.71

Quantitative Results

For quantitative evaluation of cross-dataset testing results, we provide the image quality metrics mIS, average FID score, LPIPS, PSNR, and SSIM in Table 3.7 for the whole image as well as the tumor region, similar to the results shown in Tables 3.3 and 3.4. We report results for the model trained on Dataset B and tested on Dataset A, and vice versa. It can be seen that since Dataset A is much smaller in size than Dataset B, the generalization results are slightly lower. However, the metrics are comparable to same-dataset testing results, demonstrating the capability of TI-GAN in learning and generalizing from a very limited dataset. When trained on a larger Dataset B, and tested on Dataset A, the performance of TI-GAN holds well. We also apply the results of cross-dataset testing on the downstream task of mutation detection. We perform 5-fold cross-validation mutation detection where the training data is obtained from the output of cross-dataset testing. We note that TI-GAN has comparable performance for both datasets as compared to the same domain testing results (standard TI-GAN setup, Model Std in Table 3.9). These results are shown in Table 3.9 (Model CT).

3.4.7 Ablation Studies

Qualitative Results

To assess the importance of using local and global context in estimating the tumor invasion into the surrounding tissue, we demonstrate the results of TI-GAN in Figure 3.9 with both contexts (row 1), without local context (row 2), and without global context (row 3). For no local context, we withhold tumor growth information from the network by

inputting only $\psi_{healthy}$ of the tumor, without ψ_{tumor} . For no global context, we train the model without χ . Zooming in, in Figure 3.9 (zoomed), we see that there are deformities in the skull structure and artifacts surrounding the tumor. This is because the model is unable to locate from which part of the brain the tumor has been generated - thereby facing a lack of critical information about the surrounding tissue, the diffusion parameters at that location, or the relative location of the skull. It can be noted that local context is particularly critical in estimating the skull region, which makes sense as it contains information about the tumor core properties and its position with respect to known brain tissue.

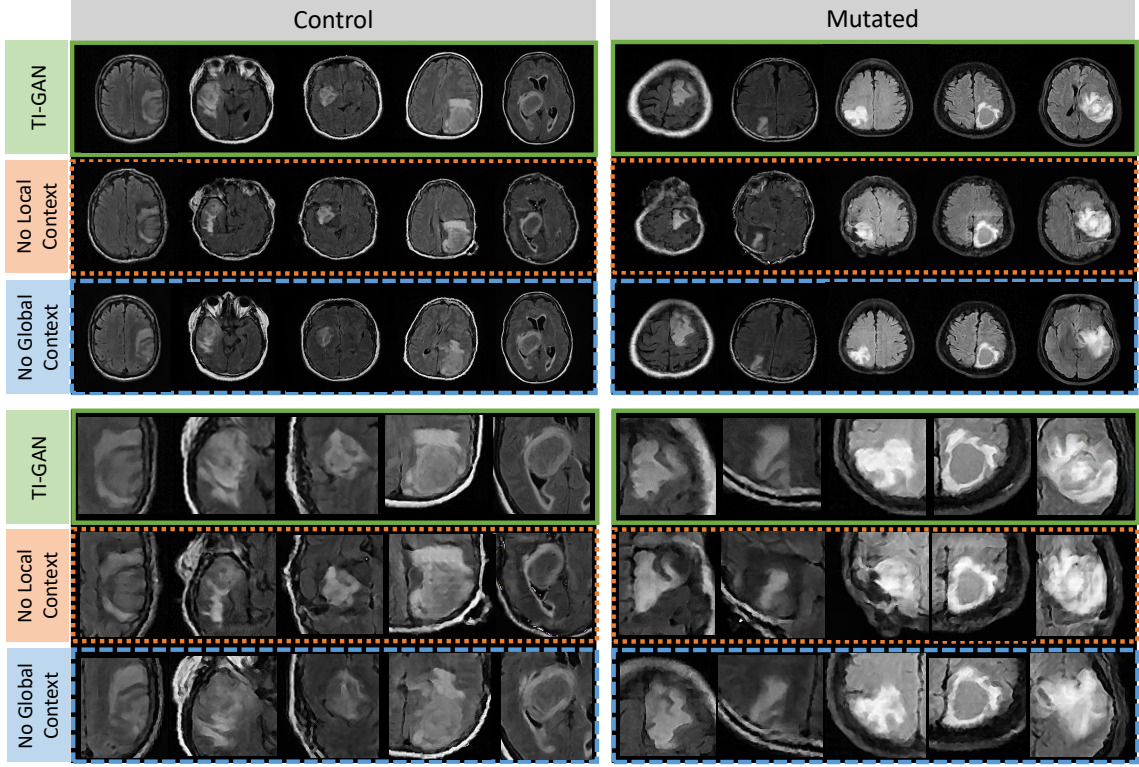


Figure 3.9: Ablation Study: Comparing the efficacy of TI-GAN (row 1, green, solid) in modeling the tumor microenvironment without local context (no tumor growth information in the form of ψ_{tumor}), (shown row 2, orange, dotted) and without global context or Volumetric Localization Context χ (shown row 3, blue, dashed). The first set of three rows shows the overall brain scan with synthetic tumor invasion, whereas the second set of three rows zooms into the tumor invasion region for a closer look at the generated results.

Quantitative Results

Similar to the quantitative results presented for previous experiments (image quality and cross-dataset testing), we report the image quality and diversity metrics for the ablation studies. We report results on both datasets for training without local and global context, respectively. Similarly to previous results, we report these results for whole images and tumor regions. These results are shown in Table 3.8. While results drop when either

Table 3.8: Quantitative metrics for Ablation experiments. Performance is evaluated when TI-GAN is trained without local context (Planar Localization Context) and global context (Volumetric Localization Context). We report results on Whole Images and Tumor Regions, as shown in Figure 3.6. While performance degrades when either context is removed, it can be seen that local context (tumor properties) has a stronger impact on TI estimation.

<i>Input: Whole Images</i>		<i>Dataset A</i>					<i>Dataset B</i>				
Model		FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑
TI-GAN		86.22	1.76	0.07	25.02	0.82	67.04	1.60	0.08	24.51	0.79
	No Local Context	102.54	1.66	0.14	20.45	0.73	96.20	1.52	0.15	21.90	0.70
	No Global Context	94.78	1.72	0.10	23.22	0.78	84.56	1.58	0.13	22.78	0.74
<i>Input: Tumor Region</i>		<i>Dataset A</i>					<i>Dataset B</i>				
Model		FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	mIS ↑	LPIPS ↓	PSNR ↑	SSIM ↑
TI-GAN		176.08	1.41	0.17	23.31	0.77	132.17	1.40	0.21	23.12	0.74
	No Local Context	198.66	1.33	0.25	19.64	0.68	140.24	1.34	0.26	19.02	0.68
	No Global Context	188.11	1.36	0.20	21.06	0.71	138.65	1.37	0.23	20.36	0.70

Table 3.9: Mutation detection in Dataset A and B using only synthetic data generated using standard TI-GAN (Model Std), using cross-dataset testing (Model CT), without local context (Model L), and without global context (Model G). The classifier used is ResNet18. The metrics are reported as mean (standard deviation) over 5-fold cross-validation, with train/test split to be 80/20. Performance using the standard setup of TI-GAN (Model Std) is used as the baseline. All metrics are in percentages. ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses.

Input	Dataset A					Dataset B				
	ACC(PL)	ACC(IL)	SENS	SPEC	DIC	ACC(PL)	ACC(IL)	SENS	SPEC	DIC
Model Std	93 (08)	92 (09)	91 (07)	94 (08)	89 (10)	85 (10)	84 (13)	81 (10)	90 (07)	78 (09)
Model CT	91 (09)	90 (10)	90 (08)	93 (08)	87 (10)	83 (11)	81 (12)	79 (13)	88 (10)	75 (13)
Model L	84 (13)	84 (12)	83 (13)	88 (10)	83 (10)	80 (10)	77 (12)	75 (11)	82 (13)	73 (12)
Model G	88 (10)	87 (12)	87 (12)	90 (09)	84 (12)	82 (12)	79 (14)	78 (13)	86 (09)	75 (10)

of the contexts are removed, we note that the local context (tumor shape and texture) has a higher impact than global context since “without local context” results are worse. Thus, the tumor properties, as well as the volumetric context of tumor play a role in estimation of tumor invasion. Similar to cross-dataset testing results, we perform 5-fold cross-validation mutation detection where the training data is obtained from the ablation study models. These results are shown in Table 3.9. Model L refers to the ablation model with no local context, and Model G refers to no global context. We note that as compared to standard TI-GAN mutation detection performance, both Models L and G have lower performance. Specifically, we note that Model L has the lowest performance as the network does not know the tumor shape or texture, and therefore has no context about tumor growth patterns.

3.4.8 Data Augmentation using TI-GAN

TI-GAN can also be used as a data augmentation tool. In this case, specifically for Dataset B as it is highly imbalanced. We use the generated images from TI-GAN to oversample the minority mutated class of Dataset B and perform mutation detection under similar conditions to the one reported in Table 3.2. The results for comparative GANs used as data augmentation instead of TI-GAN are reported for comparison. The baseline is training on real data with data balance achieved using standard data augmentation tools such as rotation and flipping. These results are reported in Table 3.10. We see that comparative GANs have better performance as data augmentation tools than standalone detection as shown in Table 3.5 since the data augmentation test contains real data in training which offsets any impact of unrealistic images when trained using only synthetic images. On the other hand, for TI-GAN, we notice a minor drop in performance. While

it may not be statistically significant, a small drop may imply that the diversity in tumor invasion synthesis may be helping detection in Table 3.5 as compared to Table 3.9, where only a limited number of synthetic samples are used.

Table 3.10: Mutation detection in Dataset B using synthetic data from TI-GAN to balance the unbalanced real data. We do this by oversampling the minority mutated class. We compare these results against Standard Data Augmentation (SDA) techniques such as rotation, flipping, etc., and data balancing using comparative GANs. Metrics are - ACC: Accuracy, SEN: Sensitivity, SPEC: Specificity, and DIC: Dice Score for IL: Image Level and PL: Patient Level analyses. The models are trained using an 80-20 patient-wise split of the data, using a ResNet18 classifier. The metrics reported are mean (standard deviation) over 10 folds.

Input	ACC (PL)	ACC (IL)	SENS	SPEC	DICE
Unbalanced Data	78 (13)	72 (15)	68 (12)	82 (08)	68 (11)
Balanced Data (SDA)	81 (09)	78 (12)	73 (08)	86 (08)	70 (10)
TI-GAN	84 (09)	82 (11)	82 (12)	88 (08)	80 (11)
GI	83 (10)	82 (13)	80 (10)	85 (13)	78 (13)
MSG-GAN	82 (11)	78 (12)	72 (14)	84 (09)	76 (11)
PGGAN	78 (09)	73 (11)	74 (13)	80 (10)	73 (12)
WGAN-GP	78 (13)	75 (10)	72 (12)	82 (08)	74 (10)

3.4.9 Limitations of TI-GAN

To analyze the limitations of TI-GAN, we filtered the generated images based on their SSIM, PSNR, and MSE values. Then, we handpicked examples that showed different types of error cases. These examples are shown in Figure 3.10. For reference, we also showed the corresponding training image and the reference real image. From these examples, we note that in cases where the relative size tumor is very large compared to the overall brain scan, and/or when the tumor region masked input to TI-GAN has a very large mask - the TI-GAN does not have sufficient local context to generate a highly accurate and realistic

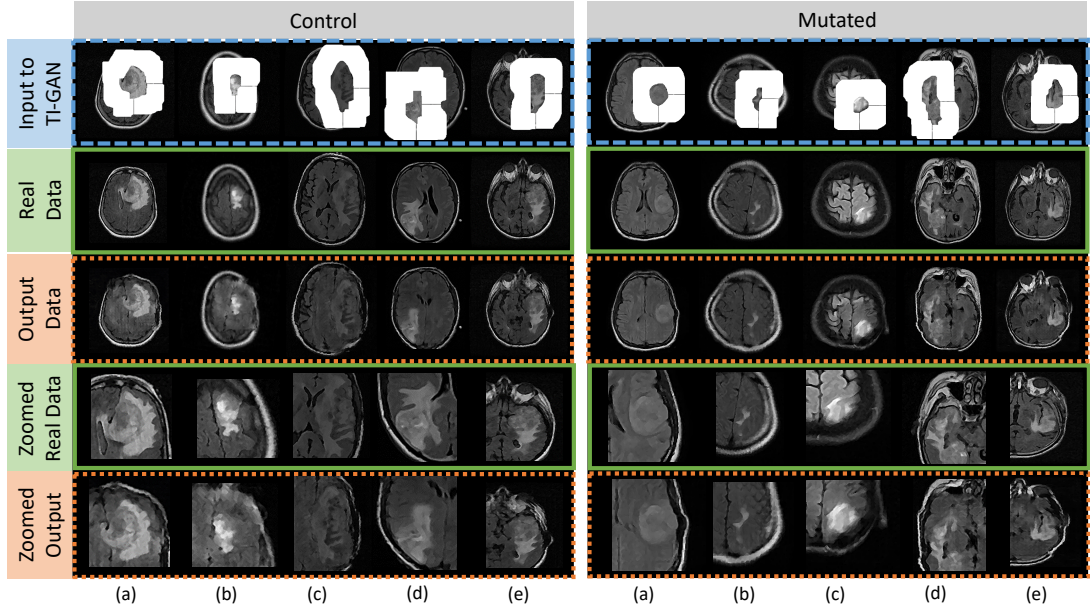


Figure 3.10: Failure cases of TI-GAN, identified using various factors such as SSIM, PSNR, MSE values, and visual inspection. We selected examples of various types of errors seen during TI-GAN synthesis. Some examples show an incorrect estimation of the skull boundary, whereas some show incorrect CSF or brain tissue estimation. In some cases, we note that the tumor growth boundary is incorrectly estimated. We see that in many of the cases, we do not see much local context due to relative tumor size being large and the masking being extensive. We note from ablation studies that the local context significantly affects the overall TI-GAN performance - which is also reflected in the failure cases here.

output. In these cases, since the conditional estimate is not accurate, TI-GAN relies on the unconditional estimate which can sometimes be a random sampling from all training data and it does not correlate realistically to the given sample. In Figure 3.10 we also report the generated images zoomed into the tumor microenvironment to show the lack of realistic detail in synthesis. As can be seen in Control: images (a), (b), and (e) - the shape of the skull is completely unrealistic due to the lack of context. The same goes for Mutated: images (c), (d), (e). For Control image (d), we notice that the tumor growth is incorrectly estimated, as is the CSF. For the Mutated class, we also note a checkered artifact in the image (b). A combination of these, in addition to incorrect brain tissue estimation, can be seen in Control image (c) and Mutated image (a).

Chapter 4

Searching for Life: End-to-end Automated Detection and Characterization of Ediacaran Biosignatures

4.1 Introduction

The study of the origin, evolution, and distribution of life in the universe is a central scientific motivation for extraterrestrial exploration and is a top research priority in NASA's most recent Decadal Strategy for Planetary Science and Astrobiology [309]. To successfully conduct astrobiological research and to develop frameworks for the robust detection of signs of life in the rock records of other planets, one must first exhaustively

study the evolution of life on Earth. Accounting for variability in atmospheric conditions, gravity, etc., that may influence geological features on other planets, this research could be translated to other settings (translational research). The first ~ 3 billion years of Earth's history were dominated by simple, single-celled organisms. This early evolution of life on Earth is preserved within the Precambrian rock record, making it an ideal natural laboratory for developing search images to guide the recognition of past life on Earth and other planets and for developing strategies for the detection of potential biosignatures [82, 462, 5, 463, 87, 293, 193, 386]. While body fossils of the single-celled organisms that dominated the Precambrian Eon provide essential insight into what those organisms were and how they were preserved [11, 192, 373], the record of the organic mats formed by a consortia of these simple organisms, namely cyanobacteria, provide a comparatively large number of sedimentary structures and anactualistic sedimentary features (i.e., features that are not observed in environmentally analogous successions of the Phanerozoic Eon when organic mats are rare), such as wrinkle structures, mat chips, and shrinkage cracks, that form via the interaction of organic mats with physical processes (e.g., erosion, deposition, etc.) [318, 315, 316, 317]. These sedimentological signatures of past microbial life, known as microbially induced sedimentary structures and textured organic surfaces [315, 131], are common targets for the identification of potential morphological (i.e., structural) biosignatures that can be detected in situ [462, 309].

The Adelaide Superbasin of South Australia is well known for its nearly continuous record of mid-Tonian (ca. 850 Ma) to late Ediacaran (635-539 Ma) marine sediments [273] that preserve extensive evidence of organic mats [131, 132], providing a useful dataset for

the identification of potential in situ biosignatures in the form of organic mat induced sedimentary structures. Within the Adelaide Superbasin, the highest diversity and complexity of organic mat signatures is found within the Ediacara Member of the Rawnsley Quartzite [131, 130, 93], which consists of 10-300 meters of shallow marine feldspathic sandstone [129, 420]. Through sedimentological characterization of the Ediacara Member (555-550 Ma) in Ikara-Flinders Ranges National Park and Nilpena Ediacara National Park we identify a potential morphological biosignature that is readily identified in macroscopic imagery and easily tested for biogenicity.

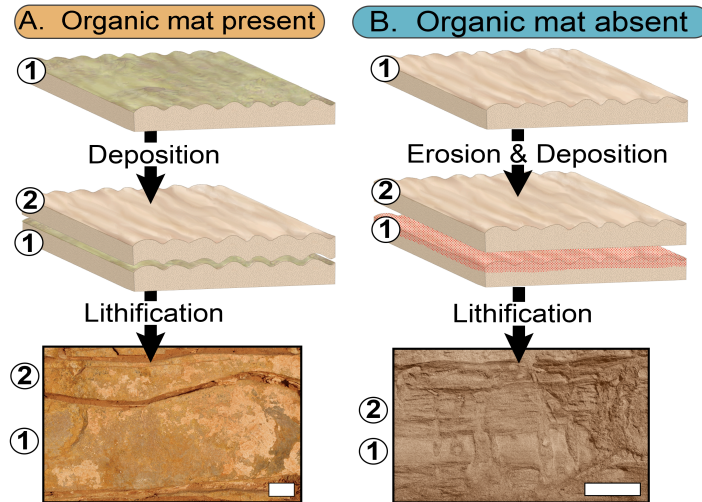


Figure 4.1: Depositional processes and resulting sedimentary packaging in mat-dominated systems (e.g., the Ediacaran Period) and in depositional systems with no organic mat growth. A) Deposition of sands on top of mat-bound sediment wherein the original topography of the seafloor (1) is not eroded by newly deposited sand (2). Resulting sedimentary packaging consists of stacked sandstone beds, with no fine-grained interbeds, preserving rippled tops and bases (i.e., Double Rippled Bedforms). B) Deposition of sands on top of sediment that is not stabilized by organic mats wherein the original topography of the seafloor (1) is eroded by newly deposited sand (2); eroded ripple crests are demarcated by red shading. The absence of an organic mat leads to erosion of ripples and amalgamation of successive event beds, resulting in amalgamated sandstone event beds bound by erosional surfaces.

This potential biosignature, here referred to as Double Rippled Bedforms (DRBs), is the cross-sectional expression of stacked palimpsest ripples, which occur in successions of sandstone storm deposits when a younger set of ripples forms on top of an older set of ripples without eroding or reworking them (Figure 4.1A) [334]. This results in stacks of bedding planes, representing consecutive storm events, that have rippled bed tops and bottoms that appear as double rippled bedforms in cross-section [420]. For palimpsest ripples to form, pre-existing ripples must be stabilized to prevent their erosion upon deposition of new sand bodies [378]. In the absence of a stabilizing agent, pre-existing ripples are eroded by newly deposited sand, resulting in amalgamated beds bound by erosional junctions with no preserved ripple crests (Figure 4.1B) [420]. Thus, the presence of palimpsest ripples and their cross-sectional manifestation as DRBs, necessitates that the sediment was stabilized, either by organic mats or through low-energy draping of fine-grained sediment on top of ripples before deposition of subsequent beds [420]. This presents a biogenic and an abiogenic mechanism for the formation of palimpsest ripples and DRBs. However, abiogenic formation of palimpsest ripples through the low-energy draping of ripples by fine-grained siliciclastics is necessarily characterized by the presence of thin layers of fine-grained sediment between doubly rippled sandstone beds. In contrast, palimpsest ripples formed via the biological stabilization of sediment by organic mats (Figure 4.1A) are characterized by stacks of discrete, doubly rippled beds of the same lithology (i.e., synlithological) and no fine-grained interbeds (e.g., Figure 4.2 (top)) [420]. Therefore, the characterization of DRB grain size can be used as a straightforward method for discerning their biogenicity.

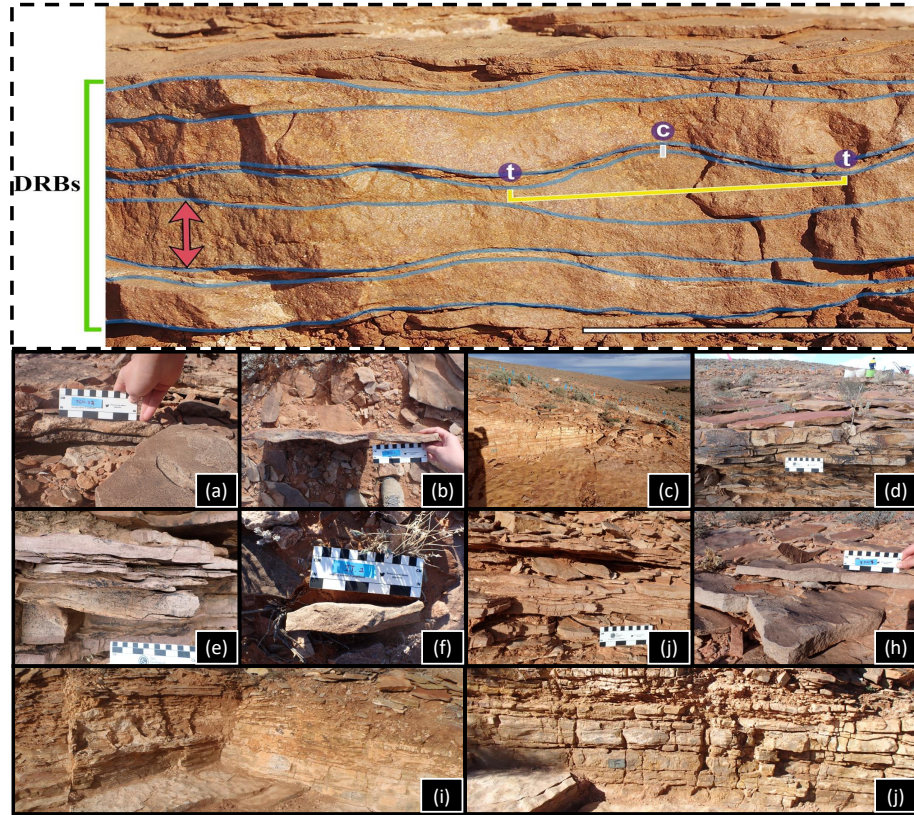


Figure 4.2: **Top (dashed)**: Annotated cross-sectional view of a stack of double-rippled bedforms (DRBs) from the Ediacara Member. DRBs are labeled by the green bracket. Individual rippled bed junctions are denoted by blue highlights. DRB thickness, reflecting the thickness of a single bed bound by non-erosional rippled bed junctions, is denoted by the red arrow. The yellow bracket denotes length of a single ripple (half-wavelength in DRB Quantifier) and is defined as the length encompassed by two local low points (trough = t) surrounding a local high point (crest = c). Scale: 8 cm. **Bottom (a-j)**: Terrain diversity in our dataset showing variations in lighting conditions, fields-of-view, and artifacts. Scales a-j: 13 cm.

DRBs are here identified as a useful potential biosignature due to the clear distinction between abiogenic and biogenic, and because they are easily recognized as complex and distinct macroscopic features in standard imagery of geological cross-sections, aligning with criteria established for highly probable biosignatures [309, 307] and allowing for rapid and low cost in situ detection of outcrop that has high potential to preserve evidence of past life. In the Ediacara Member of South Australia, DRBs occur commonly as synlithological stacks of sandstone beds that are associated with fossilized mat textures on the bases of bedding planes [130, 420]. The absence of fine-grained sediment between DRBs and their close association with fossilized mat textures in the Ediacara Member establishes these DRBs as potential definitive signatures of past life [420]. The definitively biogenic DRBs of the Ediacara Member can thus be used to establish a search image for this potential biosignature in other geological units and to develop novel methods for the remote detection of potential signatures of past life in geologic cross-section. The ability to detect DRBs in imagery of geological cross-section is ideal for scalable and translational research because this type of data can be easily collected for any other region on Earth and is similar to the imagery data collected by the Mastcam-Z on the Mars 2020 Perseverance Rover [121, 40]. Thus, we propose an end-to-end approach, “Scene-aware Perception Automation using Composite Embedding for Segmentation” (SPACESeg2.0), for remote detection and characterization of DRBs using images of biogenic DRBs taken from the Ediacara Member at the Ikara-Flinders Ranges National Park and the Nilpena Ediacara National Park.

Motivation for SPACESeg2.0: From a computational point of view, automating the detection of DRBs poses significant challenges. The terrain surrounding and containing

these DRBs is very diverse, and the relative DRB thickness varies significantly (Figure 4.2). This issue is compounded by the fact that the images collected from the Ediacara Member have unknown or widely varying fields-of-view. Furthermore, the dataset is uncured, i.e., it has no control parameters over imaging conditions. Thus, it consists of challenging artifacts such as scale, humans, foliage, and shadows. A few examples of the dataset depicting the diversity of images and artifacts are shown in Figure 4.2 (bottom, a-j). While this compounds the complexity of the task, building a robust methodology that tackles these very practical issues brings us significantly closer to performing translational research on other geologic cross-sections on Earth, a critical tool for the search for past life on other rocky planets such as Mars. We developed “Scene-aware Perception Automation using Composite Embedding for Segmentation” (SPACESeg) [202], a computer vision method for the automated detection of DRBs while tackling the aforementioned challenges.

SPACESeg2.0 is designed to improve upon the drawbacks of the original SPACESeg model. The goal for SPACESeg and SPACESeg2.0 is to be “scene-aware”, i.e., detect rippled bed junctions while tackling the issues of varying fields-of-view and imaging conditions. “Perception automation” means the model’s predictions are human perceptible. For example, computationally, the shadows look a lot like ripples, but a human can distinguish between them - and so should SPACESeg2.0. SPACESeg2.0, like its predecessor, has two complementary modules - a rule-based unsupervised technique “ARORA” (Auxiliary Representation-based Optimization for Region Activation), that adapts to local terrain properties to detect perceptually relevant regions and the supervised deep learning model “BASE-Net” (Boundary-Aware Saliency Encoded segmentation Network), a novel model

that performs terrain-aware segmentation. The two modules work together to perform highly accurate bed junction segmentation. Each module of SPACESeg2.0 is optimized for increased performance over SPACESeg. Compared to SPACESeg, SPACESeg2.0 has three major modifications: (a) an improved region activation method with adaptive data augmentation technique incorporated for dual functionality, (b) BASE-Net, a boundary-aware saliency-encoded segmentation network, and (c) a DRB Quantifier (DRBQ) tool that performs bed junction characterization via the number of ripples, ripple amplitude, half-wavelength and distance between successive ripples on each surface, at any resolution.

4.2 Related Work and Contributions

4.2.1 Related Work

Given the increased interest in search for life beyond Earth [309, 5, 307, 111, 465], a suite of geological features from the Precambrian rock record have been proposed as potentially astrobiologically significant biosignatures to be prioritized in the search for life on Mars [386]. Many of these proposed biosignatures require microscopy, spectroscopy, or chemical analyses to be identified in the rock record (i.e., amino acids, carbon isotopes; biogenic micro-textures) [386]. Additionally, most of these biosignatures can be produced via both abiotic and biotic processes, and discerning between an abiogenic and biogenic origin can require intensive tests including chemical analyses and microscopy [294]. This indicates the need for the identification of a biosignature that is macroscopic, easily detected in low-cost formats (i.e., imagery), and can be established as biogenic through straightforward tests (e.g., grain size analysis). DRBs fulfill these criteria. Thus, development of automated

methods for detection of DRBs is an important next step for the utilization of this form of sedimentary packaging as a potential biosignature and has the potential to substantially contribute to the search for life in the universe. In addition to the ability to detect DRBs in imagery, utilizing Machine Learning (ML) and Computer Vision (CV) methodologies allows for the remote detection and characterization of the rippled bed junction morphologies that comprise DRBs. This, supplemented with additional collection of data on the lithology of the beds, provides a rapid and low-cost mechanism for processing large datasets and for identifying extraterrestrial outcrops with high potential astrobiological significance.

In astrobiology and geology, ML and CV have a myriad of applications. For example, research efforts have been dedicated to applying deep learning to characterize various aspects of Mars topography [92, 45, 303] such as identifying Barchan Dunes [361] or Martian dust storms [322]. Some researchers have also performed crater detection [102, 395], rock classification [484] and rock segmentation [267] on Martian surface. In the search for potential biosignatures, Convolutional Neural Networks (CNNs) have been employed to distinguish between abiotic and biotic sediment cracks to identify microbially induced sedimentary structures in imagery in [73]. Semantic segmentation models were proposed to aid in characterizing sediment core descriptions in [105], and to perform segmentation of Scanning Electron Microscope (SEM) images of mud rocks in [29]. Characterizing grain size is a critical part of assessing if DRBs have biotic origins, and some researchers have proposed deep learning techniques for this task outside the context of DRB detection [398, 285]. However, the task of objectively identifying DRBs remains a problem yet to be solved. We aim to build a robust methodology that can precisely segment DRBs with varying

thicknesses and relative sizes while tackling artifacts and other imaging problems. Our goal is a model that can handle limited datasets while using a low-cost, scalable input modality.

Automated detection of DRBs falls under the semantic segmentation task, a widely researched area in ML and CV. Aligning most closely are the tasks of boundary detection [232, 18], contour optimization [171, 61, 339] and scene-aware segmentation [492, 427, 476, 509]. Some applications for boundary or contour detection include nerve detection [362, 459], crack detection on roads [107, 507, 175, 529], detecting defects in structures [357, 352], etc. Meanwhile, many real-world applications, specifically in remote sensing, require complex-scene-aware segmentation [247, 496, 195, 84, 256, 450]. Performing scene-aware segmentation and boundary detection in our task is complicated by the fact that our regions of interest are of varying sizes in images that have no set field-of-view or control over imaging conditions. To this end, we compare our proposed model with some of the state-of-the-art techniques in deep learning in both scene-aware segmentation and boundary detection [202, 509, 529, 360, 296].

4.2.2 Contributions

Astrobiological Contributions:

- SPACESeg2.0 performs precise segmentation of DRBs, a highly probable signature of past life observable in geologic cross-section. To the best of our knowledge, this is the first end-to-end automated system to remotely detect and characterize DRBs.
- Our proposed model is tested against a varied set of practical imaging conditions, artifacts, and dataset sizes for improved deployability and scalability.

- We demonstrate that our model is a robust computational tool that has the potential to be scaled to other terrains for geological assessment, on and beyond Earth.

Machine Learning Contributions:

- We propose a novel deep network, BASENet, for scene-aware boundary segmentation with domain-inspired context, loss function and overall design curated to tackle the challenges of an unbalanced segmentation problem (DRB vs background) on a highly diverse, limited dataset.
- We develop a plug-and-play terrain-invariant, dual-task module (ARORA) that generates domain-specific context and performs adaptive, data-driven augmentation for robust training of the BASENet segmentation model.
- SPACESeg2.0 performs accurate minute-structure segmentation and tackles a plethora of practical artifacts - outperforming many deep-learning segmentation models for a varying set of imaging conditions and sample sizes.

4.3 Technical Approach

The task of detecting rippled bed junctions is complicated severely by the variations in relative ripple size, bed thickness, field-of-view, artifacts, imaging conditions, etc. Subsequently, the discriminatory feature space required to isolate these rippled bed junctions is high-dimensional. While deep learning techniques excel in modeling non-linear, complex feature associations - they also fall prey to over- or under-learning due to the challenging nature of our data. One such challenge is the shadow artifact, which looks very

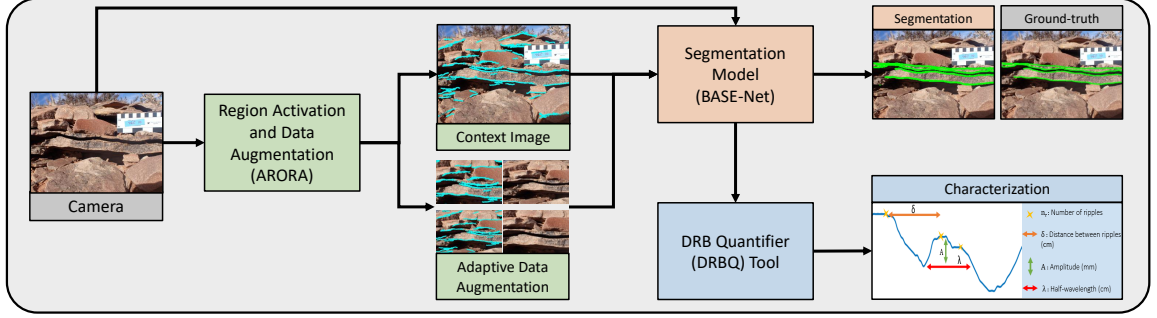


Figure 4.3: Overview of SPACESeg2.0 with three plug-and-play modules - the “Auxiliary Representation-based Optimization for Region Activation” (ARORA) module for Region Activation and (adaptive) Data Augmentation, the segmentation model “Boundary-Aware Saliency-Encoded Network” (BASE-Net) for robust, terrain-aware segmentation and the “DRB Quantifier Tool” (DRBQ) module for characterization of detected potential biosignatures. In DRBQ “Characterization”, δ (orange): distance between ripples, A (green): amplitude, λ (red): half-wavelength and yellow markers: number of ripples.

similar to rippled bed junctions. Another challenge is the high class imbalance, where the relative ripple size can be as low as $< 1\%$ of the image width. To combat this issue, we bootstrap the deep learning segmentation with an unsupervised augmentation technique. The unsupervised algorithm ARORA activates the regions with ripples as saliency (similar to an attention map) - highlighting regions-of-interest, while also performing adaptive data augmentation. The output of ARORA is given as context to BASE-Net, a deep learning-based segmentation model, along with the augmented dataset. Together, they make a scene-aware model that rejects perceptually irrelevant features and detects the desired DRBs with high precision, which are then characterized using the DRBQ. The overview of SPACESeg2.0 is shown in Figure 4.3.

4.3.1 Problem Formulation

The detection of DRBs is done in two steps: the unsupervised ARORA for training data augmentation and the segmentation network BASE-Net. Let X_0 be the cross-sectional

view of stacked DRBs. Let the ground-truth be Y_{gt} . X_0 is given as an input to the unsupervised ARORA model. The output of this step is the data-driven context (σ_0) and the adaptive, information-rich patches ($\tilde{x}$). The original image samples (X_0), ripple-rich patches ($\tilde{x}$) and context (σ_0) are then given as input to the deep learning model BASE-Net. While σ_0 acts as external attention or context to the BASE-Net, $\tilde{x}$ are used for adaptive data augmentation. The output of BASE-Net is Y_0 , i.e., predicted ripples using SPACESeg2.0.

To characterize the DRBs, we give Y_0 as input to the DRB Quantifier (DRBQ) tool, which then outputs various statistics such as the number of ripples in a sample (n_r), average half-wavelength of the ripples (λ), amplitude (A) and distance between ripples (δ). The detailed technical approach of each step is provided in the following subsections.

4.3.2 Auxiliary Representation-based Optimization for Region Activation (ARORA).

The first step of SPACESeg2.0 is the unsupervised ARORA module, which augments the training of the subsequent segmentation module in two ways: (a) Generating a saliency or attention map (σ_0) and, (b) Adaptive training data augmentation through information-rich (ripple-rich) patches ($\tilde{x}$). ARORA generates an auxiliary image representation followed by gradient optimization to highlight the potential regions where the rippled bed junctions might be present. Subsequently, the regions with the most ripples are utilized for informed data augmentation of the training set. The inputs to the BASE-Net model are the ARORA augmented auxiliary image set and attention maps, in addition to the original training set. The details of the model development of ARORA are provided in this section. The overview of ARORA is shown in Figure 4.4. While following the same principle as the

Composite image-based Optimization for Region Activation (CORA) in [202], the auxiliary representation in ARORA is less resource-intensive and more effective. We demonstrate via qualitative analysis that ARORA is more robust to artifacts than CORA [202].

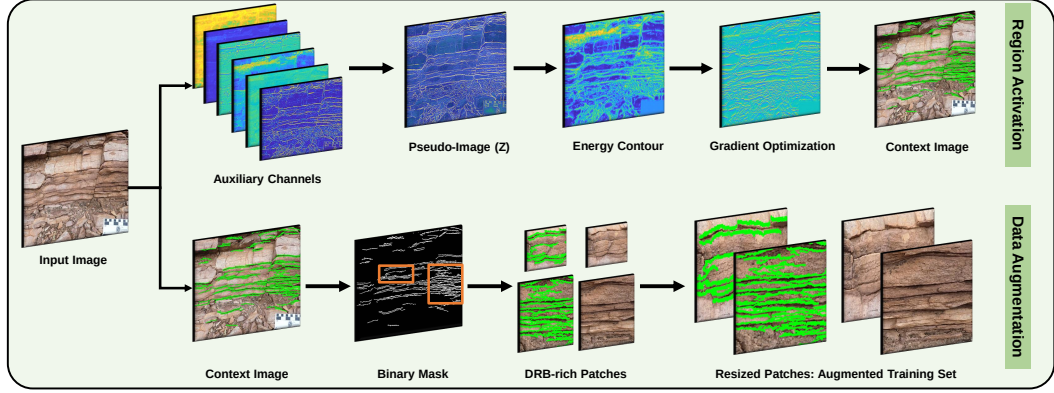


Figure 4.4: Overview of the ARORA module. The dual-tasking ARORA utilizes a pseudo-image representation to generate data-driven context and perform adaptive data augmentation.

Auxiliary Space Generation: The auxiliary representation is generated to address some of the biggest concerns of a non-curated imaging set. These issues are variations in lighting, shadows, atmospheric conditions, human errors, varying fields-of-view and diversity in relative ripple size in an image. Exploratory data analysis revealed that some image representations highlight or suppress one or more of the features of interest. Unlike CORA, which utilizes Gabor features (scale and orientation) and Fast Fourier Transform (FFT) features [202], we limit the auxiliary features in ARORA to the spatial domain for faster processing. For this, we leverage other image spaces and filtering techniques to synthesize a six-channel Auxiliary Image Representation (AIR). These channels are computed as:

- *Channel 1:* The first channel is designed to highlight DRB-like structures. A Hessian is a second-order representation of intensity variations and describes the local curvature.

Eigenvector decomposition of this can be used in highlighting tubular (i.e., thin, elongated) structures [483, 364]. We perform multi-scale Hessian filtering using the technique in [364], which shows that if the first and second eigenvalues have low and high positive value respectively, we can extract the darker boundary-like structures (i.e., rippled bed junctions).

- *Channel 2:* For the second channel, we aggregate the red-difference and blue-difference chroma components. These are obtained by subtracting the gamma-compressed Luma channel from the YCbCr representation. This channel brightens the shadows, normalizing the terrain lighting.
- *Channel 3:* The third channel is obtained by normalizing the hue of an image with the saturation (from the HSV color space), which isolates the artifacts of an image.
- *Channel 4:* The fourth channel addresses the intensity inconsistencies in an image by performing a flat-field correction using a Gaussian kernel with a standard deviation of 5 to estimate the dark regions.
- *Channel 5:* The fifth channel normalizes the dark channel of an image with its saturation, rejecting the high saturation, high luminescence regions such as measuring artifacts.
- *Channel 6:* The final channel is an entropy filtered image of the flat field normalized image, with a 9x9 size kernel.

The channels are then subjected to PCA and the first three principal components are cast onto a 3-channel space to obtain a pseudo-image Z . It should be noted that since

the AIR channels are obtained using non-linear transformations, the subsequent pseudo-image Z resides in a mock-RGB color space and is not expected to look realistic. All the AIR channels and Z are shown in Figure 4.5. These are shown as colorized heatmaps to aid in visualization. It can be seen that each channel isolates a different attribute of the image, and the subsequent images cast onto a mock-RGB space (Z) reject the most common artifacts such as markers, sky, human artifacts, etc. and highlights desired contours.

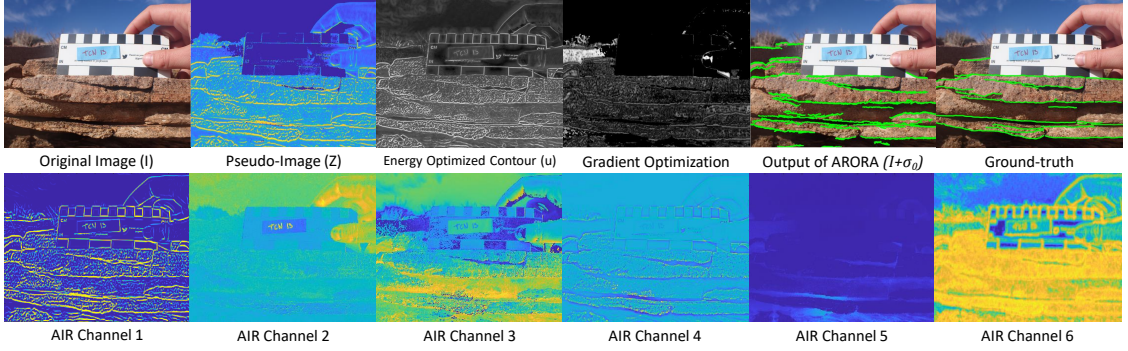


Figure 4.5: Visual demonstration of each step of ARORA (top row), each auxiliary channel of ARORA (bottom row). We show the output of ARORA (σ_0) overlaid on the original image for better visualization. This is labeled $I + \sigma_0$. We follow this representation throughout, as a binary mask is difficult to interpret.

Fuzzy energy and gradient optimization: To “activate” the regions of interest, we compute the active contours followed by gradient optimization. Active contour models (ACM) [216, 359] are widely popular in both region and boundary segmentation tasks. However, a few drawbacks of ACM models, specifically in extracting object boundaries, are their sensitivity to initial level-set function and image inhomogeneity, as well as the non-convexity of the energy function. We address the issue of image inhomogeneity by utilizing pseudo-image Z instead of the RGB image. To navigate the sensitivity towards initial level-set functions, we employ the Hybrid and Local Fuzzy Region-edge based Active

(HLFRA) Contour model [110]. The hybrid and local fuzzy energy terms of HLFRA have also been shown to be strictly convex, ensuring optimization [110].

We perform fuzzy energy minimization for Z to obtain a region contour u . The rippled bed junctions are boundary-like structures, residing at the local minima of the contour. Thus, we compute the second-order gradient for u and the lowest 10% values are extracted as the predicted DRB regions. The threshold is chosen to deliberately allow over-activation, to accommodate the wide variations of relative size of the rippled bed junctions in our images. Here, “relative size” collectively refers to the thickness and cross-sectional length of individual rippled bed junctions, relative to the height and width of the image. The region contour (u), gradient optimized image and the final output of ARORA (σ_0) are shown in Figure 4.5.

Adaptive data-augmentation: The region activated image (σ_0) serves a dual purpose. In addition to being the context for better training of the segmentation model, it is also used for informed data augmentation. In many of our images, a large area can be either covered by artifacts or have no rippled bed junctions at all. Thus, data augmentation techniques such as cropping, zooming or scaling prove ineffective as they may feed noisy or irrelevant samples to the segmentation model. To address this, we perform data-driven or informed augmentation. Such adaptive augmentation has proven highly effective for complex tasks in our previous work [201]. Since the region-activated image (σ_0) is specifically designed to address the primary data concerns, we use it to extract information-rich patches ($\tilde{x}$). Using the binary mask of σ_0 , we find the regions which have the highest density of structures of interest. We extract patches of random sizes and aspect ratios from these regions and

re-scale them to 256x256. The field terrains have a high heterogeneity in relative ripple sizes. Thus, an added benefit of extracting (and re-scaling) random sizes and aspect ratios of patches is that they reflect this property of the dataset and bolster the training set with samples of varying magnification and relative ripple size. This makes the subsequent segmentation model more robust to varying imaging conditions, which is a significant hurdle in dataset such as ours.

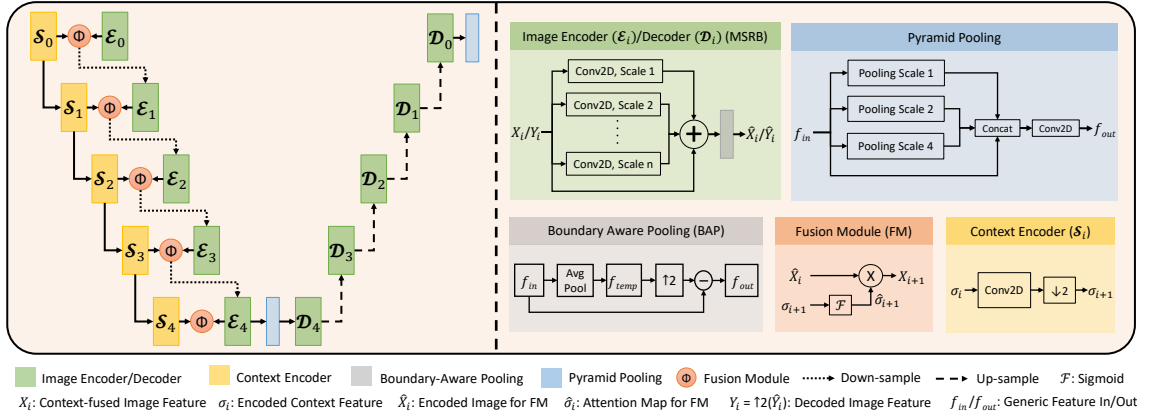


Figure 4.6: Overview of BASE-Net Architecture (left). Architectures of individual BASE-Net modules (right).

4.3.3 Boundary-Aware Saliency-Encoded Network (BASE-Net)

The augmented data from ARORA is used to train the Boundary-Aware Saliency-Encoded segmentation Network (BASE-Net). The underlying model of BASE-Net is the ResUNet-a architecture [84]. To leverage the distinct properties of the data, while navigating extreme imaging conditions, we enhance this model through (a) Hierarchical context fusion for domain-aware segmentation, (b) Boundary Aware Pooling (BAP) to utilize the structural properties of ripples for improved segmentation, and (c) A loss function that is

effective for highly unbalanced and edge-dependent segmentation. The overview of BASE-Net is shown in Figure 4.6. The design choice and workflow of each BASE-Net module are provided in the following subsections.

Hierarchical Context Fusion: The inputs to the network are the image (X_0) and context (σ_0). The output is the predicted segmentation map (Y_0). At each level of abstraction, the image features (X_i) are enhanced with the context features (σ_i). The structures of image encoder ($\mathcal{E}_i$) and context encoder ($\mathcal{S}_i$) are provided in Figure 4.6. Each $\mathcal{E}_i$ is a multi-scale residual encoder, which inputs X_i and outputs the intermediate image encoding $\hat{X}_i$. Similarly, $\mathcal{S}_i$ takes input σ_i and outputs the context feature map σ_{i+1} . For context fusion with image features, σ_{i+1} is passed through a linear layer followed by a sigmoid layer, to obtain an attention map. A channel-wise product results in a context-embedded image feature X_{i+1} . This process is carried out for all layers, as shown in Eq (4.1) and (4.2).

$$X_{i+1} = \downarrow (\Phi(\hat{X}_i, \hat{\sigma}_{i+1})) = \hat{X}_i \odot \hat{\sigma}_{i+1} \quad (4.1)$$

$$\hat{X}_i = \mathcal{E}_i(X_i), \quad \hat{\sigma}_{i+1} = \mathcal{F}(\mathcal{S}_i(\sigma_i)) \quad (4.2)$$

where, i represents the layer number, $\odot$ is the channel-wise product, $\downarrow$ is the down-sampling operation and $\hat{X}_i$ is the intermediate image feature (output of $\mathcal{E}_i$), $\hat{\sigma}_{i+1}$ is the attention map, σ_{i+1} is the context feature (output of $\mathcal{S}_i$), X_{i+1} is context-encoded image feature (input to the $(i+1)^{th}$ encoding layer) and $\mathcal{F}(\cdot)$ is a linear layer followed by sigmoid function. Thus, each level of abstraction in the encoding process of BASE-Net has inherent context embedding, which is subsequently propagated through the decoding layers as well. During decoding,

the context embedded features from the encoding layers (X_{i+1}) are decoded using a parallel residual decoder $\mathcal{D}_i$. This process is depicted in Eq (4.3).

$$Y_i = \begin{cases} \uparrow (\mathcal{D}_i(X_i)), & \text{if } i == n. \\ \uparrow (\mathcal{D}_i(Y_{i-1})), & \text{otherwise.} \end{cases} \quad (4.3)$$

where $\mathcal{D}_i$ represents the i^{th} layer of decoder and X_n is image encoded representation from the final encoding layer and Y_{i-1} is the $(i-1)^{th}$ -level decoder output.

Hierarchical context fusion provides strong feature associations for boundary-like segmentation tasks. Augmenting this with multi-scale residual blocks for feature extraction gives us superior performance in DRB detection.

Multi-scale Residual Blocks (MSRB): Each encoder and decoder module is made of a multi-scale residual convolution unit. This is the basic unit of ResUNet-a [84], which we adopt for BASE-Net. Multi-scale features have shown great promise in efficient and fine-grained segmentation tasks [509, 84, 426, 519, 455]. Here, the multi-scale features are obtained by varying the dilation rates on convolutional filters, thereby varying the receptive fields of the subsequent filter bank. For BASE-Net, we use three scales, and an additional identity path (residual block). Using an identity path helps us avoid the diminishing gradient problem or training degradation [162]. This property becomes particularly useful when training on more limited sample sets. The general structure of MSRB is shown in the Encoders/Decoders module of Figure 4.6 (green). For each encoder/decoder module, we replace average or max-pooling steps with boundary-aware pooling to improve the detection of thin, edge-like structures, i.e., rippled bed junctions.

Boundary-aware pooling: We replace the average or max pooling with a pooling method that leverages the nature of our structure of interest. In developing boundary-aware models, inspiration is taken from traditional image sharpening or edge detection in image processing [455, 343, 528]. In our case, since the rippled bed junctions are thin, boundary-like structures, we mimic the process of edge-boost filtering. To sharpen an image or highlight its edges, the image is blurred and then subtracted from the original image. We replicate this process by employing average pooling, followed by upsampling, and subtracting this feature map from the original feature map. The average pooling downsamples the feature map by half. By upsampling this, we essentially “blur” the feature map. Subtracting the “blurred” map from the original effectively “highlights” the edges, which in our case highlights the rippled bed junctions. This process is depicted as “Boundary Aware Pooling” in Figure 4.6 and is computed as shown in Eq (4.4).

$$f_{out} = f_{in} - \uparrow_{\times 2}(f_{in,avg}) \quad (4.4)$$

where, f_{in} is a generic term for any feature input, $f_{in,avg}$ (f_{temp} in Figure 4.6) is obtained by average pooling f_{in} and f_{out} is the boundary-aware pooling output.

Pyramid Pooling: Our dataset suffers from highly variable field-of-view terrains. Thus, global context is a key property, as it can aid in effective scene-parsing. The empirical receptive field of regular CNNs (even for residual blocks such as ours) is shown to be low [518]. While global average pooling is most commonly used for incorporating global context, it does not suffice for complex scenes [509]. To this end, a hierarchical global contextual prior called the pyramid pooling layer was proposed in [161], which improved

the scene-parsing capabilities of a model. This layer computes and concatenates the highest-level feature maps at various scales, then fed to a 3x3 convolution layer. Two such pooling layers are used in BASE-Net, one each at the end of encoding and decoding hierarchies. The general architecture of this layer is shown as the “Pyramid Pooling” block in Figure 4.6.

Loss Function: The desired structures to be segmented often occupy a small fraction of the image. The loss function of BASENet should address this and adapt to our highly unbalanced (DRB to terrain ratio) segmentation task. Losses such as weighted cross-entropy and Dice loss are often used, but they suffer from optimization difficulty and gradient explosion, respectively [4]. To this end, we utilize the Focal-Tversky Loss (FTL) [4]. The Tversky index is a generalization of the Dice and Jaccard coefficients, as shown in Eq (4.5), where TP are True Positives, FP are False Positives, FN are False Negatives and TN are True Negatives. By tuning the parameters α and β , we can either weigh FPs or FNs more, thereby forcing the model to focus on more hard-to-segment examples. The FTL is a generalization of this metric, defined in Eq (4.6). This loss term helps us in achieving fine-grained segmentation in a highly unbalanced segmentation task [4, 188].

$$TI = \frac{TP}{TP + \alpha \times FN + \beta \times FP} \quad (4.5)$$

$$\mathcal{L}_{FT} = \sum (1 - TI)^\gamma \quad (4.6)$$

where the value of α is 0.7, β is 0.3 and γ is 1.33. We set the value of α higher than β to penalize the FNs more, which has been shown to be effective in tackling high class imbalance [4]. The value of $\gamma > 1$ is chosen so that the loss function has higher gradient for

the harder examples, i.e., more minute ripples. The impact of these parameters is reflected in our performance (Table 4.1 and 4.2 in Results section).

4.3.4 DRB Quantifier

The DRB Quantifier is a tool built to extract rippled bed junctions from an image and compute their characterizing statistics. We compute four metrics for each rippled surface detected in each cross-sectional image. The description of each metric is provided below, and a visual depiction of each metric is shown in Figure 4.7. Per Figure 4.7, the DRB consist of two rippled bed junctions, and the metrics are computed for each of the two. These are exported to an XLS file. The statistics provided by the DRBQ allows for the quantification of *mat-related ripple muting*. This phenomenon has been described from the Ediacara Member of South Australia wherein the extent of mat development corresponds to the height and width of preserved ripples [420, 93]. Well-developed mats “mute” the rippled surfaces by growing over originally formed ripples and filling in the ripple troughs. This results in ripples with a shorter crest-to-trough height and corresponding shorter half-wavelengths and shorter average amplitude values. Distances between ripple crests and the number of ripples on a surface relate to the spacing of ripples on a bedding plane which broadly correspond to the energy level of the system at the time of ripple formation. The four metrics provided by the DRBQ are, therefore, useful for estimating extent of mat development and system energy from cross-sectional views alone. Thus, these statistics provide additional information beyond just the identification of the potential presence of fossilized evidence of life which are generally not able to be ascertained without excavation of the bedding planes, serving as an essential tool for remote investigations.

With increasing number of data samples, these characteristics could be quantified with increasing confidence.

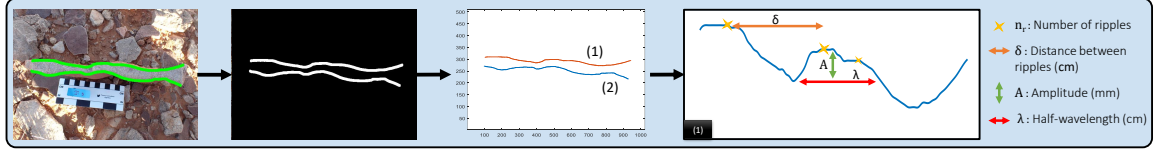


Figure 4.7: Overview of the DRB Quantifier Tool

(a) Half-wavelength (λ): We compute the half-wavelength as the distance between two consecutive local minima (ripple trough), encompassing a local maxima (ripple crest). This is shown as the red measure in Figure 4.7. The local maxima and minima are computed using first and second-order gradients of the ripple curve and computed for each individual ripple.

(b) Amplitude of each ripple (A): The amplitude of each ripple crest is computed as the height of the crest from the average height of the ripple. Since there is no definitive universal reference point, we report amplitude as a relative measure. This is shown as the green measure in Figure 4.7.

(c) Distance between ripple crests (δ): This value is computed between consecutive crests on a single rippled bed junction. This is shown as the orange measure in Figure 4.7. Since the ripple size is variable, even on the same rippled surface, we report this metric separately from the half-wavelength. If all the ripples were perfectly symmetric/sinusoidal, the distance between crests would be the same as the half-wavelength.

(d) Number of ripples per surface: This shows how many ripple crests are present on every bed junction.

4.4 Experiments and Results

4.4.1 Dataset: Description and Pre-processing

We used two cohorts of data for our experiments in this work. The images were collected by our team from outcroppings of the Ediacara Member in the Ikara-Flinders Ranges National Park and the Nilpena Ediacara National Park in South Australia. The only control parameter on the imaging was that the pictures be cross-sectional views. Beyond that, no control parameters were assigned to mitigate any natural artifacts and noise in the data. Thus, we ensure that the data is realistic and developing our model on such extensive and robust conditions only elevates the practicality of SPACESeg2.0. For analysis on the DRBQ, only images with known field-of-view were used - so that the ripple characterization can be done in standardized measurement units. Assuming that the horizontal field-of-view is V_h cm, and the original size of the image is $h \times w \times 3$, the scaling factor to convert a metric to standard units (cms) will be V_h/w .

The first cohort consists of 657 images. Of these images, 594 images had known field-of-view. The resolution of these images ranges from 24 pixels/cm to 225 pixels/cm. The field-of-view of these images ranges from 17.8cm to 166.7cm. The second cohort is a much smaller cohort of 27 images, which was used in [202]. Of these images, 23 images had known field-of-view. The resolution of these images ranges from 0.7pixels/cm to 265pixels/cm. The field-of-view of these images ranges from 6.6cm to 75.3cm. Cohort B is included for fair comparison between SPACESeg2.0 and [202] as well as to demonstrate the robustness and efficacy of SPACESeg2.0 in tackling a highly diverse and very limited dataset.

Pre-processing: To focus on sedimentological features that are geologically significant, we excluded bed junctions in which $<2\%$ of the length of the entire cross-section had ripple-like features, preventing the inclusion of bed plane irregularities that are not ripples. Owing to the varying fields-of-view, we did not levy such restriction upon the relative width of the ripple. Images are resized to 256x256 before training. For the comparative models, data augmentation as performed in their respective papers, if any, were used [360, 509, 529, 296, 202] for Cohort A. Irrespective of the original implementation, oversampling via random cropping and scaling were performed for Cohort B for all comparative models due to the limited size of the dataset, to improve performance. For SPACESeg2.0, we used ARORA for adaptive data augmentation.

4.4.2 Experimental Setup

Brief technical details and training parameters for SPACESeg2.0 are included in this section. The results are compared with various deep-learning segmentation models.

Auxiliary Representation-based Optimization for Region Activation: To compute the energy contours, the weight for hybrid energy is 0.5 and local energy is 1 [110]. The second-order gradient is minimized with a threshold allowing the lowest 10% of the values. The subsequent binary masks of the contours are then filtered based on their relative area (less than 1% are removed), aspect ratio (excluded if the width of each individual detected structure is less than five times its length), and orientation (any structure over 45° is excluded). The final detected structures, as binary masks, are used as context for BASE-Net and also used for the data augmentation step of ARORA. For the data augmentation path of ARORA, we vary the aspect ratio (width/height) of patches between 0.25 to 4 in

intervals of 0.25. The scale is varied from 0.5 to 2 in intervals of 0.5. The size of the patches can vary anywhere from a fifth of the width of the image to half the width. All the above thresholds and ratios were empirically chosen to reflect realistic fields-of-view and relative ripple size for a majority of images. Thus, the resultant patches are ripple-rich, but with random relative size and field-of-view. This adds to the robustness of the model, allowing it to learn to detect ripples irrespective of aspect ratio, scale, and resolution. Four patches are extracted per training image for Cohort A and 10 patches are extracted per training image for Cohort B since it is a limited set.

Boundary-Aware Saliency-Encoded Network: All the images, augmentation patches, and context masks are resized to 256x256 before training. The dataset is divided into 80% data for training and validation, and the remaining 20% is used for blind testing. We also report the results of the model trained on Cohort A and tested on Cohort B, to demonstrate the generalizability of SPACESeg2.0. For training, we employ the Adam optimizer with a learning rate of 0.0001. The training is done for 200 epochs, and 5-fold cross-validation is performed.

Comparative methods: We compare our results with UNet [360], PSPNet [509], DeepCrack [529], ESPNet [296] and SPACESeg [202]. These models were chosen for comparison for three reasons: (a) due to their extensive use in remote sensing applications, (b) they are all suited for boundary-like and highly unbalanced segmentation, and, (c) all of these methods (except UNet) use some form of multi-scale processing, which allows for a fair comparison with our model. Here, segmentation is unbalanced because the DRBs occupy a small fraction as opposed to the background. We use the same training and testing data

for training all these models for consistency in comparison. We use the configurations from the respective publications, fine-tuning where necessary to improve performance.

For DRBQ: The horizontal field-of-view is used to convert the metrics from pixel space to centimeters (cm). The lowest detectable height of an individual ripple (A_{min}) is 0.1mm and the peaks are computed over a sliding window encompassing 10% of the image at a time. These values are used for reporting the results in this work, but can be modified based on the granularity of analysis desired in an application.

4.4.3 Qualitative Results

Overall Performance: For the qualitative results, we compared the results of SPACE-Seg2.0 with PSPNet [509], DeepCrack [529], ESPNet [296], and SPACESeg [202] (Figure 4.8). From our experiments, we noted that UNet demonstrated low performance (see quantitative results in Section 4.4.4), and it is, therefore, excluded from the qualitative results. For Cohort A, we note that ESPNet suffers from errors in a lot of the more difficult-to-recognize DRBs, such as the ones that look similar to the terrain. Meanwhile, DeepCrack has high precision but misses several rippled bed junctions. For Cohort B, we note that the lack of data affects DeepCrack and ESPNet, and they show the same behavior of low recall (in DeepCrack) and over-segmentation (in ESPNet). For both cohorts, PSPNet outperforms ESPNet and DeepCrack in general detection of DRBs, but does not perform well for more difficult terrains, as can be seen in Figure 4.8. We note that SPACESeg2.0 outperforms all the models for both cohorts, including our previous SPACESeg [202].

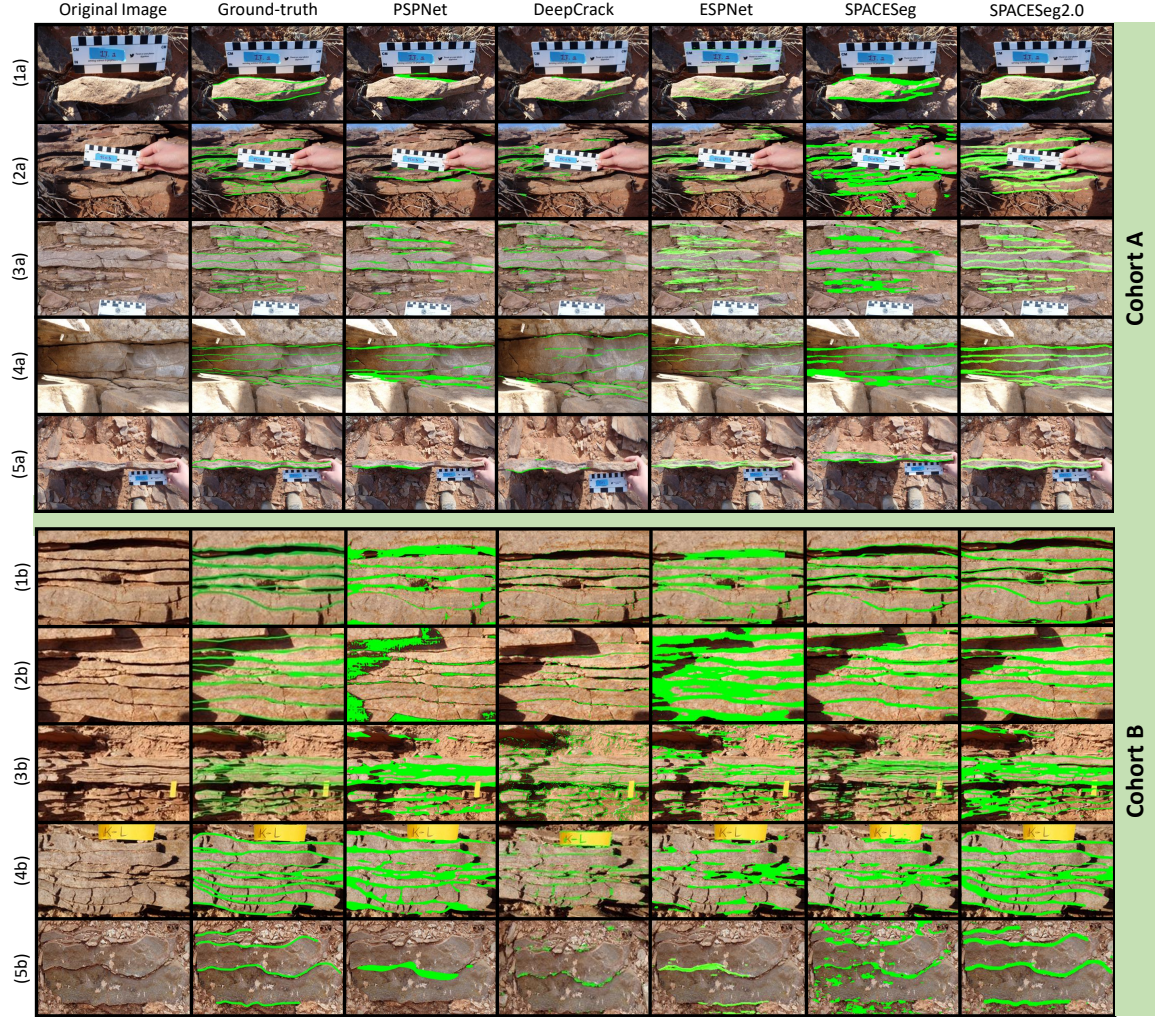


Figure 4.8: Performance of SPACESeg2.0 in detecting DRBs, compared with the performance of PSPNet, DeepCrack, ESPNet, and SPACESeg. Results reported for Cohort A and B. The images are chosen to demonstrate a range of number and size of rippled bed junctions. It can be seen that SPACESeg2.0 is robust to over- and under-segmentation compared with other models.

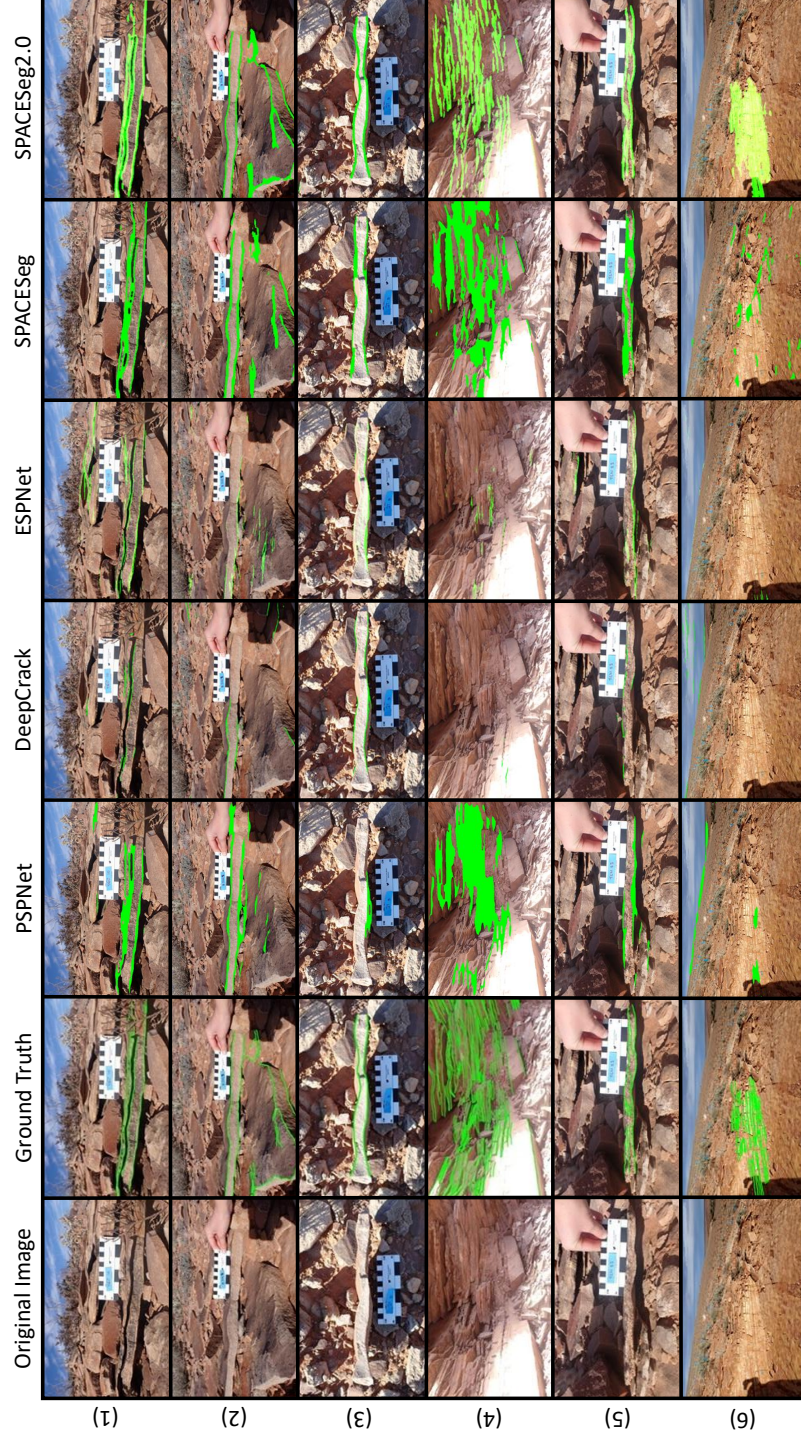


Figure 4.9: Performance of SPACESeg2.0 in detecting DRBs from difficult terrains, vs comparative models. The difficulty of terrain is empirically chosen via the presence of higher number of artifacts, varied imaging conditions, low relative DRB thickness, high number and/or closely-spaced rippled bed junctions.

Difficult Terrains: As established, our data faces a lot of practical challenges in terms of terrain variability. The artifacts include (but are not limited to) shadows, measuring markers, humans, lighting conditions, and landscape artifacts such as shrubs and sky. To demonstrate the robustness of SPACESeg2.0 in tackling these issues, we report the performance on difficult terrains and compare our results with PSPNet, DeepCrack, ESPNet, and SPACESeg. The difficulty of a terrain is subjective, and in this case, we evaluate difficulty empirically based on the length of rippled bed junctions, relative DRB thickness, varied lighting, and/or extent of artifacts. Some examples are shown in Figure 4.9. Image (1) consists of thin DRBs and is affected by shrubs, debris and the sky as artifacts. Image (2) has DRBs that are surrounded by very similar looking terrain, which makes it difficult to see them. Images (3) and (4) have very thin DRBs which are strongly affected by lighting conditions as well as shadows. Image (5) has shadows, debris and human artifact. Image (6) does not have a traditional cross-sectional view, while consisting of very thin DRBs, sky, shrubs and debris. It can be seen that even in cases of less distinct junctions, SPACESeg2.0 outperforms all the other methods.

Cross-cohort Testing: As a test of generalizability and scalability, we report the performance of SPACESeg2.0 trained on Cohort A and tested on Cohort B. We compare these results with all the other models and the results are shown in Figure 4.10. It can be seen that SPACESeg2.0 is still able to extract desired DRBs. SPACESeg [202] performs second best, with some failure cases in difficult terrains like Images (3), (4) and (6). We do not report the results with training on Cohort B and testing on A due to the limited size of Cohort B.

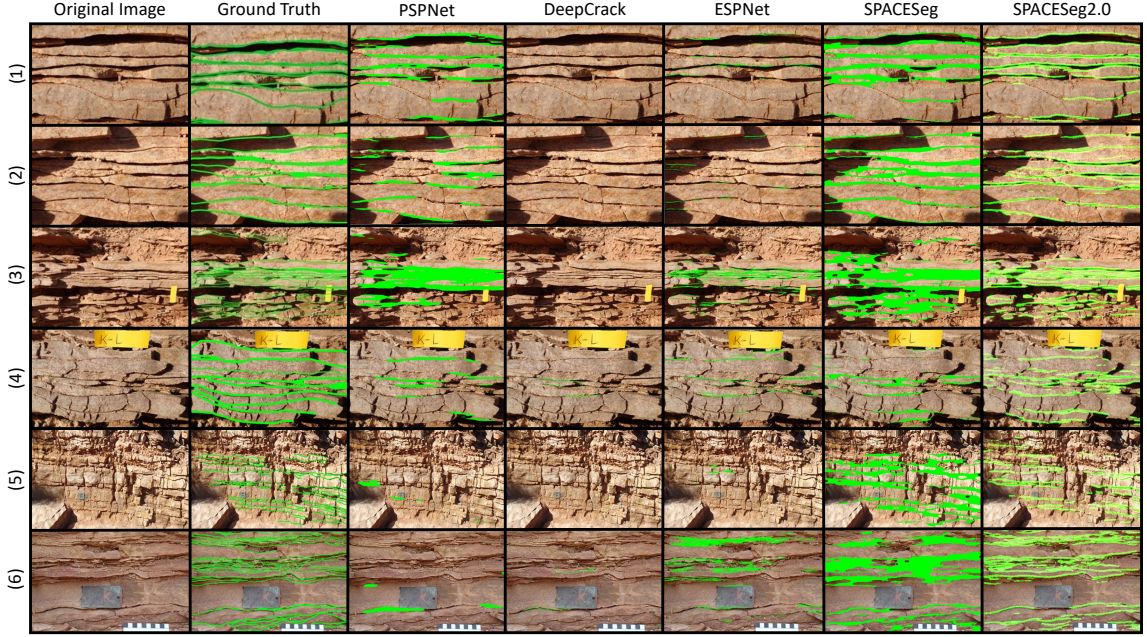


Figure 4.10: Cross-testing performance of SPACESeg2.0 in detecting DRBs from Cohort B, when trained using Cohort A. Comparison with PSPNet, DeepCrack, ESPNet, SPACESeg.

4.4.4 Quantitative Results

For all of the algorithms compared, we report the accuracy (ACC), Balanced Accuracy (BAC), Precision (PREC), Recall (REC), Dice/F1-score (F1), IoU/Jaccard Score (JSC), BFSC Score (BFSC), and Hausdorff Distance (HD). We compute these metrics to evaluate SPACESeg2.0 and compare with UNet [360], PSPNet [509], ESPNet [296], DeepCrack [529] and SPACESeg [202]. Table 4.1 and Table 4.2 include these results for Cohort A and B, respectively. The metrics are defined as follows, where TP is number of True Positives (rippled bed junctions classified as rippled bed junctions), FP is number of False Positives (background classified as rippled bed junctions), TN is number of True Negatives (background classified as background), FN is number of False Negatives (rippled bed junctions classified as background).

Table 4.1: Quantitative performance of our proposed SPACESeg2.0 vs other deep segmentation networks for Cohort A. Metrics
- ACC: Accuracy, BAC: Balanced Accuracy, PRE: Precision, REC: Recall, F1: F1/Dice Score, JSC: Jaccard or IoU score,
BFSC: Boundary F1-score, and HD: Hausdorff Distance. Metrics are in % and reported as *mean (std)*.

Algorithm	ACC $\uparrow$	BAC $\uparrow$	PRE $\uparrow$	REC $\uparrow$	F1 $\uparrow$	JSC $\uparrow$	BFSC $\uparrow$	HD $\downarrow$
UNet [360]	84.75 (06.36)	68.72 (05.99)	58.03 (18.32)	53.64 (13.46)	50.56 (11.45)	34.59 (10.96)	63.63 (14.51)	14.45 (01.92)
PSPNet [509]	87.13 (09.88)	73.35 (07.44)	60.36 (12.47)	49.84 (20.64)	51.78 (13.64)	36.02 (12.00)	62.18 (15.50)	13.15 (01.35)
ESPNet [296]	86.96 (07.30)	69.05 (06.78)	63.02 (13.78)	40.36 (14.95)	46.89 (13.11)	31.53 (10.65)	73.77 (13.95)	13.28 (01.64)
DeepCrack [529]	85.32 (09.25)	55.98 (03.95)	89.68 (15.91)	31.31 (09.22)	43.27 (13.15)	29.73 (09.61)	61.88 (22.80)	15.17 (02.06)
SPACESeg [202]	86.09 (08.15)	77.90 (16.49)	67.75 (10.09)	63.03 (10.76)	63.12 (13.11)	50.23 (12.70)	78.71 (12.42)	13.01 (02.10)
SPACESeg2.0	89.84 (07.48)	84.62 (10.16)	72.23 (08.10)	72.43 (19.68)	70.82 (12.12)	54.86 (13.30)	85.57 (10.63)	12.23 (02.23)

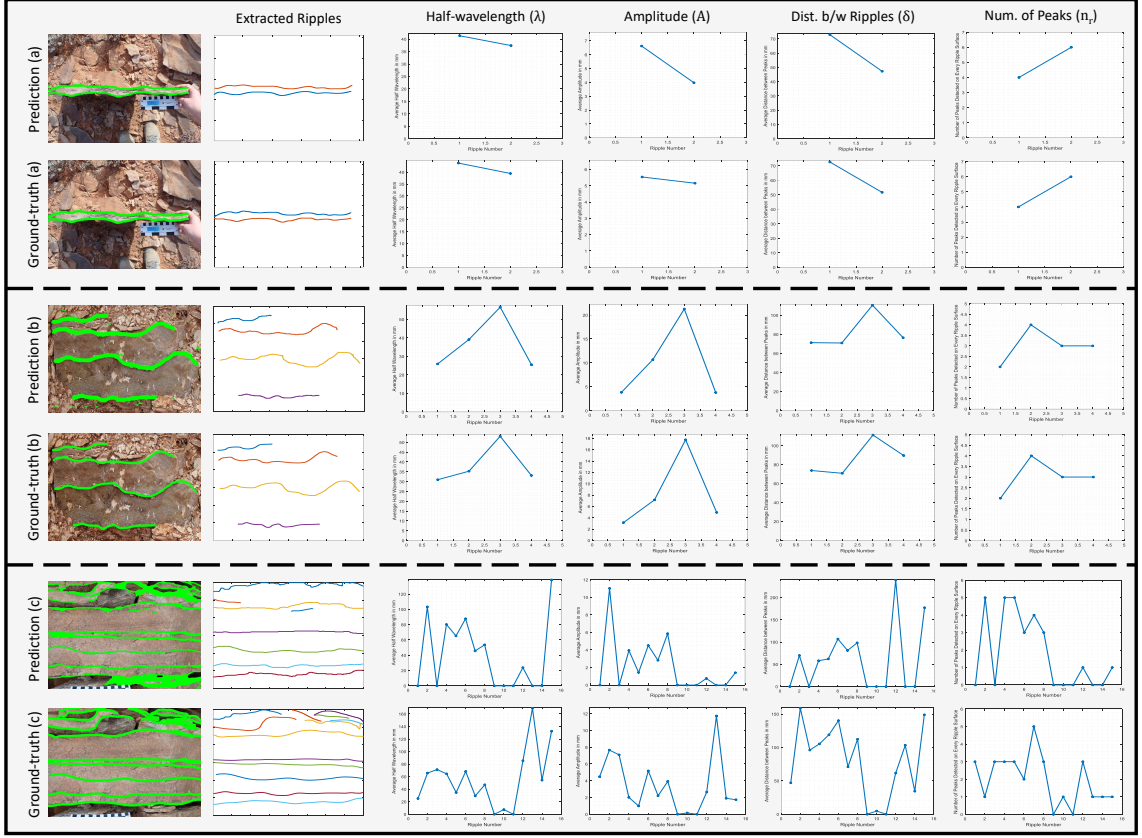


Figure 4.11: Results showing metrics computed using the DRBQ module. The results using ground-truth ripples are also shown as reference to evaluate the precision of segmentation results of SPACESeg2.0. A high degree of similarity with ground-truth statistics means that SPACESeg2.0 can be used for DRB characterization. For cases with complex DRB such as (c), we note that the plots don't match exactly, but the range of values for each metric is consistent - which is key in characterization.

Accuracy (ACC): Measures the proportion of all pixels that were correctly segmented, and is computed as $(TP+TN)/(TP+TN+FP+FN)$.

Balanced Accuracy (BAC): An arithmetic mean of the true positive rate and true negative rate, this computes the average performance of each class. We report this metric since our classes are imbalanced (more background than DRBs), and accuracy can be high even when fewer DRBs are detected.

Table 4.2: Quantitative performance of our proposed SPACESeg2.0 vs other deep segmentation networks for Cohort B. Metrics - ACC: Accuracy, BAC: Balanced Accuracy, PRE: Precision, REC: Recall, F1: F1/Dice Score, JSC: Jaccard or IoU score, BFSC: Boundary F1-score, and HD: Hausdorff Distance. Metrics are in % and reported as *mean (std)*.

Algorithm	ACC $\uparrow$	BAC $\uparrow$	PRE $\uparrow$	REC $\uparrow$	F1 $\uparrow$	JSC $\uparrow$	BFSC $\uparrow$	HD $\downarrow$
UNet [360]	78.33 (15.12)	68.43 (09.40)	54.24 (14.64)	82.87 (06.54)	63.93 (09.78)	47.39 (07.84)	65.37 (14.86)	15.02 (01.31)
PSPNet [509]	80.23 (08.31)	73.61 (12.34)	65.19 (10.36)	74.43 (06.16)	66.31 (09.38)	51.13 (10.65)	71.34 (11.13)	14.53 (01.92)
ESPNet [296]	84.14 (08.12)	65.07 (07.95)	55.94 (09.80)	79.48 (11.60)	62.60 (13.32)	48.16 (09.79)	61.15 (08.48)	15.13 (01.82)
DeepCrack [529]	78.54 (11.87)	54.02 (06.82)	72.14 (08.36)	60.50 (13.29)	66.78 (09.64)	49.61 (08.94)	70.71 (08.53)	15.87 (02.57)
SPACESeg [202]	88.47 (06.35)	75.27 (10.84)	70.75 (07.39)	82.03 (08.26)	76.51 (07.92)	61.79 (07.19)	81.41 (07.22)	13.84 (01.93)
SPACESeg2.0	92.32 (08.61)	80.12 (09.54)	75.33 (10.71)	85.82 (10.43)	80.78 (09.87)	64.86 (13.30)	84.73 (09.88)	12.95 (02.04)

Precision (PRE): The proportion of total positive predictions that were correctly predicted; calculated as $(TP)/(TP+FP)$. This metric reflects how many pixels segmented as rippled bed junctions are actually rippled bed junctions.

Recall (REC): Measures what proportion of actual positives were correctly predicted, computed as $(TP)/(TP+FN)$. It means of all rippled bed junctions, how many were detected correctly.

Dice or F1 Score (F1): Useful in evaluating unbalanced segmentation, it is the harmonic mean of precision and recall, computed as $(2*TP)/(2*TP+FP+FN)$. It is a combined measure of how many rippled bed junctions were segmented, and how precisely.

IoU or Jaccard Similarity Score (JSC): A similarity metric, which measures the proportion of true positives over all the true and predicted positives, computed as the intersection over union between ground-truth and predicted for each class, $TP/(TP+FP+FN)$. Here, it means of all the real and predicted rippled bed junctions, how many are actually present.

Boundary F1 Score (BFSC): Proposed in [75], BFSC is a type of F1-score, computed as a contour matching score by finding the boundary of the binary maps. We use the same implementation as in [75]. This metric is important here because the rippled bed junctions are thin boundary-like structures and a contour matching score better reflects the model performance.

Hausdorff Distance (HD): This measures the distance between two point sets and is reported as an indicator of the highest error in segmentation [23]. The formula for HD is given as:

$$HD(X,Y) = \max\left(\max_{x \in X} \min_{y \in Y} \|x - y\|_2, \max_{y \in Y} \max_{x \in X} \|x - y\|_2\right)$$

For the larger Cohort A (Table 4.2), most models have similar accuracy, and the difference in performance can be seen from the balanced accuracy, F1 and Jaccard scores. For the limited Cohort B (Table 4.2), most models have higher recall. This means they sacrifice “exact” detection in favor of “more” detection. This is because the number of samples are limited, and the models do not have enough data to learn from, causing over-segmentation. DeepCrack has high precision even for Cohort B, but suffers from low recall, which means that it detects rippled bed junctions without over-segmentation, but often misses some due to this phenomenon - causing lower recall in the process. For all these models, it can be seen that balanced accuracy is lower than accuracy, implying that one class is favored over the other. Most models report better performance for the larger Cohort A - indicating the importance of dataset size. We note that SPACESeg2.0 has the most consistent performance among all models for both cohort sizes. Compared with SPACESeg2.0, DeepCrack has higher precision in Cohort A and UNet has similar recall in Cohort B. However, SPACESeg2.0 outperforms both models in other metrics, suggesting that the high precision in DeepCrack is due to under-segmentation whereas high recall in UNet is due to over-segmentation.

4.4.5 DRB Quantifier

The DRBQ inputs the output of BASE-Net and computes various metrics to characterize the ripples comprising the DRBs. Currently, we report the half-wavelength, amplitude, distance between ripple crests and number of crests per rippled bed junction. The output of DRBQ is shown in Figure 4.11. For increasing number of rippled bed junctions (images a to c), we show the performance of DRBQ in extracting the various ripple characteristics. For reference, we also report the metrics computed using the ground-truths of

the same images. It can be noted that these metrics are similar for SPACESeg2.0 predicted and ground-truthed surfaces. For the more complex Image (c), even though the distribution is not exactly the same for all statistics, we note that the prediction and ground-truth lie within similar measurement ranges (in mm) for all the metrics. If the predicted metrics are similar to the ground-truth statistics, this implies that SPACESeg2.0 can be employed for DRB characterization. The extracted ripples and the metrics for each ripple are shown in Figure 4.11.

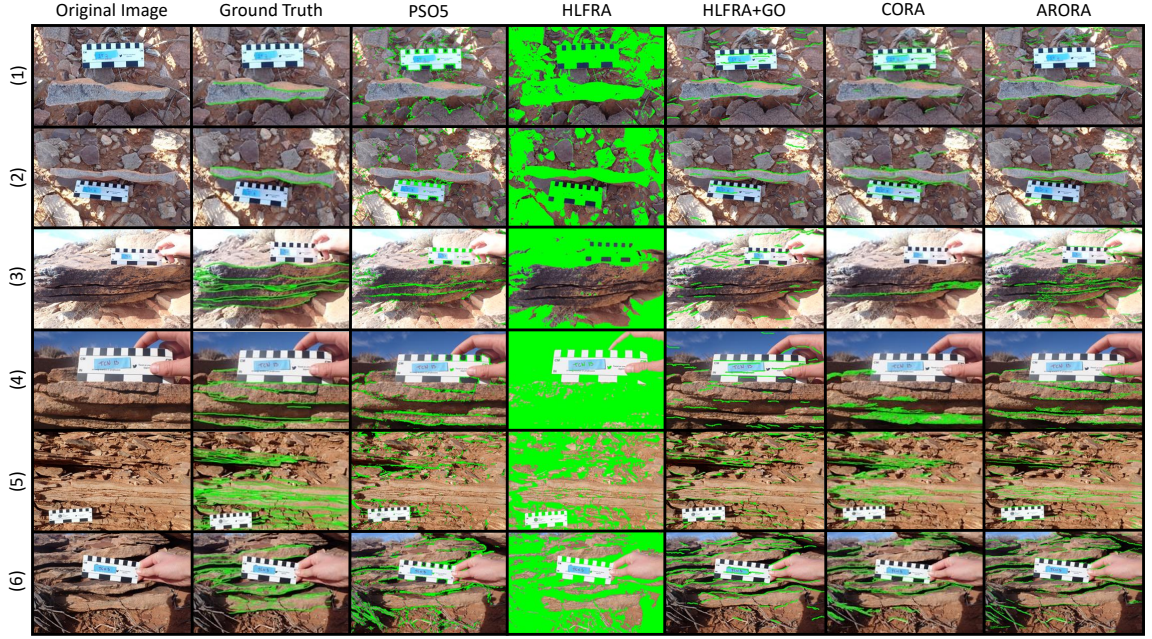


Figure 4.12: Ablation study demonstrating the impact of the Auxiliary Image Representation (AIR) in detecting difficult rippled bed junctions. The results are compared against other unsupervised techniques. ARORA employs HLFRA method with Gradient Optimization (GO). The impact of AIR can be seen when HLFRA+GO uses an RGB image input instead of AIR input. Meanwhile, CORA [202] uses a different technique to compute AIR and a different active contour optimization. It can be seen that ARORA is more robust against complex imaging conditions than CORA, such as in image 3, where CORA cannot identify many regions of interest due to the complexity of the terrain. In images 4 and 5, where the rippled bed junctions are finer and closer together, there is clumping in CORA, whereas ARORA can tackle it. In isolated beds such as in images 1 and 2, we note that CORA highlights the marker artifact as well, which ARORA almost fully mitigates.

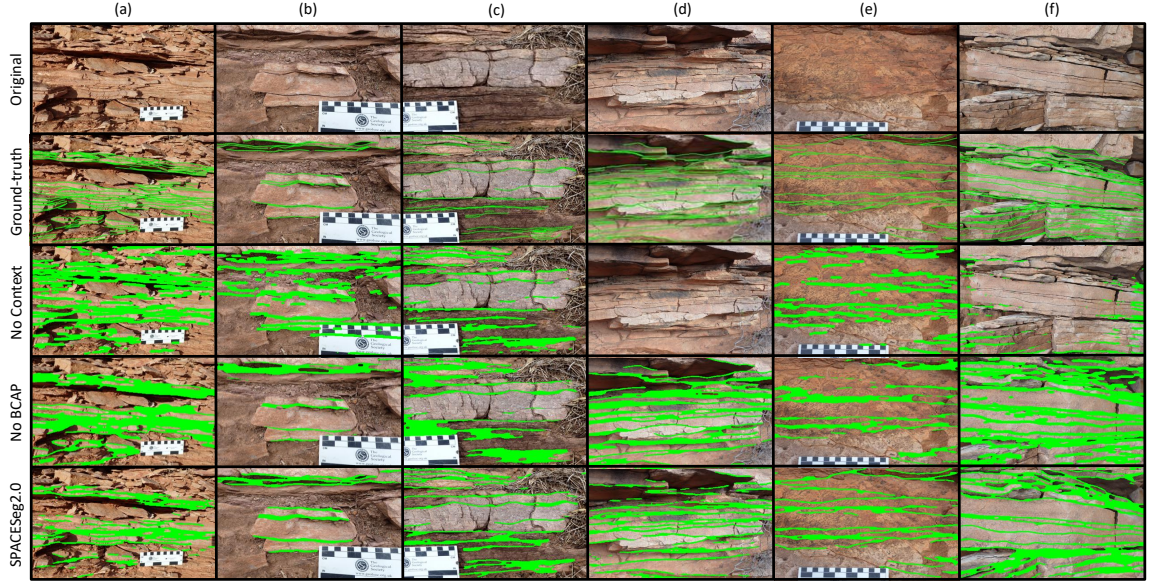


Figure 4.13: Ablation study demonstrating the effect of Hierarchical Context Fusion and Boundary or Context Aware Pooling (BCAP) on detecting rippled bed junctions, compared against the full model. The results without context are obtained by training BASE-Net without context input and the results without BCAP are obtained by training BASE-Net without both boundary-aware and pyramid pooling. Images are chosen for cases with higher number of bed junctions that are close together and/or images with bed junctions that look like the terrain.

4.4.6 Ablation Study

Impact of Auxiliary Image Representation (AIR): It can be argued that contour optimization followed by gradient optimization should activate desired regions without the use of the AIR. In this section, we report the results of ARORA compared with standalone contour optimization using HLFRA, region detection using HLFRA followed by Gradient Optimization (HLFRA+GO), the SPACESeg counterpart of ARORA (CORA), and the ARORA module. Additionally, we report a baseline performance using clustering using Particle Swarm Optimization (PSO) technique. Unlike [202], we perform a five-class clustering as opposed to three for PSO, as this five-class cluster showed better performance

among various number of clusters. The parameters for optimization for HLFRA and HLFRA+GO are the same as for ARORA, the only difference being the AIR. The parameters for CORA are the same as used in [202]. Figure 4.12 shows the results for each method, and examples are chosen from both cohorts. It can be seen that for non-AIR samples, PSO and HLFRA methods show high dependency on imaging conditions such as intensity difference between DRB-rich cross-sections and background, lighting conditions, etc. Without the AIR, when the DRB-rich cross-sections are the center of focus of the image (images (c) and (d)), it can be seen that HLFRA, HLFRA+GO, and CORA all fail whereas ARORA can still reject most artifacts and focus on the correct region. While PSO performs well here, we can see that it does poorly for images (a), (b) and (e). We note that ARORA (i.e., AIR+HLFRA+GO) is the only consistent technique to extract the ripples in difficult terrains, of which AIR is a significant step.

Impact of Context Embedding in BASE-Net: We report an ablation study demonstrating the impact of context fusion on the detection results of BASE-Net. For this analysis, the BASE-Net is trained using the same configuration as Figure 4.6, only without context. The results are reported in the “No Context” column of Figure 4.13. We note that for difficult terrains, the lack of context causes one of two kinds of errors: over-segmentation in the presence of too much debris (images a and b), and under-segmentation with different imaging (focus) settings (images d and f). For all images a-f, we also note that there is an error in segmenting rippled bed junctions that look similar to the surrounding terrain. BASE-Net performs better under all these conditions, establishing the importance of including context.

Impact of Boundary and Context Aware Pooling: For the same set of images, we also report results demonstrating the effect of Boundary and Context Aware Pooling (BCAP) in BASE-Net. The “No BCAP” row in Figure 4.13 shows the results when BASE-Net is trained with all the same configurations except the BAP and pyramid pooling layers, both of which are replaced by global average pooling. It can be seen that for all images, the segmentation is coarser when we do not use BAP and pyramid pooling. Additionally, some of the rippled bed junctions that look similar to the rock sample (images b and e) are not segmented. Thus, these pooling layers help us achieve more fine-grained segmentation.

Chapter 5

Novel Body-based Biometrics for Human Recognition under Extreme Imaging Conditions

Human recognition plays a crucial role in various applications, including surveillance, security, and retail analytics. It involves matching individuals across different camera views and times, despite variations in their appearance due to factors such as lighting, pose, and occlusions. One particularly challenging aspect of human recognition is the variability in clothing, as individuals often change their attire between observations. The task is further complicated by many challenges such as imaging distance, sensing platform and camera orientation, articulation of the human body and its pose variations, occlusion, atmospheric turbulence, and many other distortions.

For image-based person ReID, color is by far the most commonly used modality. It is a low-cost, ubiquitous modality that is rich in visual cues. It can also be used easily with existing recognition backbones. While it has the significant advantage of being an appearance-based biometric, it can also be its biggest disadvantage. In images with high distortions, RGB modality has significant drawbacks. In images, where the face is not visible, RGB images have to rely heavily on identifying a human from shape or clothing. When there is a change in clothing, RGB modality suffers from errors. Even for similar clothing, yet high extent of imaging distortions, the lack of reliable facial attributes causes a lot of errors.

To mitigate these, shape-based features such as silhouettes or a person’s skeletal information are commonly used. Independent of color and texture information, they provide robust performance in clothing-changing scenarios and are fairly robust to lighting and environmental changes. However, due to the same lack of color and texture information, they suffer from lower discriminatory power. Furthermore, they do not provide accurate pose estimation on their own and are affected by camera orientation, as the image perspective changes.

Pose-based features estimate body joints to identify a person. These features have benefits over shape-based features as they are robust to clothing and environmental changes and can tackle partial occlusions. However, in cases of articulation, atmospheric turbulence, or imaging elevation, these methods suffer greatly.

Recently, 3D body-based biometrics has been gaining traction. It offers robustness to clothing changes, camera angle variations, and environmental distortions. It is well-suited

for dynamic environments with pose variations and provides enhanced discriminative power for structural features. Its ability to estimate depth and spatial relationships makes it more reliable in complex multi-view systems and in scenarios where traditional 2D methods struggle. However, current models are computationally intensive, while also being heavily affected by occlusion, articulation, the weight of the person, and low image resolution. As these are very common challenges of any human recognition task, 3D models still have a ways to go before becoming a commonly used component of a biometrics system.

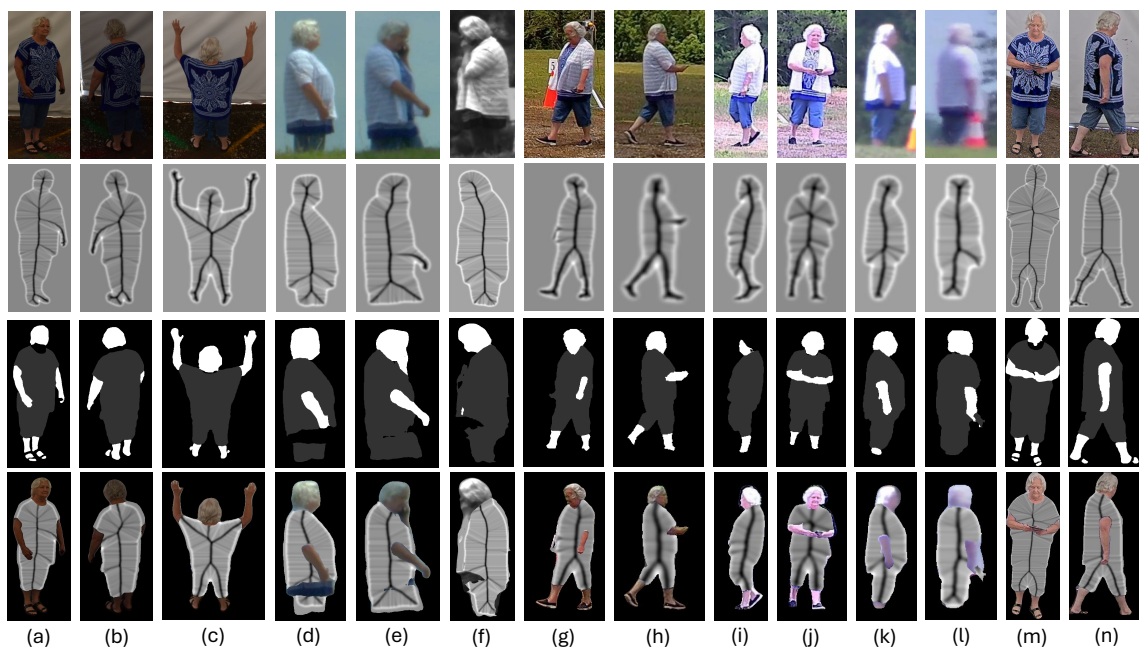


Figure 5.1: Examples depicting the robustness of ORB across various challenging imaging conditions. (a-c) Indoor examples, with pose variations, articulation, and camera elevation. (d-e) Similar poses, with and without articulation at 200m imaging distance. (f) Similar articulation, extreme imaging condition for imaging range of 500m. (g-h) Close-range outdoor imaging, with walking, articulation, and occlusion. (i-j) Pose variation, with lighting and turbulence challenges at a 100m distance. (k-l) Similar pose, with extreme imaging challenges such as turbulence and occlusion at varying imaging distances of 400m and 200m, respectively. (m-n) Examples of controlled indoor walking video frames used in the gallery. Images (a-c) and (m,n) are indoor, set 1 clothing gallery images. Images (d-l) are outdoor set 2 clothing probe images. All images are taken from the BRIAR dataset. The subject has authorized the use of their images.

Most human recognition or person reID methods focus on developing sophisticated backbone recognition models, with adaptive attention mechanisms, transformer-based feature extraction, hierarchical processing, multi-task learning, self-supervision, etc. to improve performance. While these are effective, they also have an upper limit on how discriminatory their feature space is, limiting their performance. Only RGB or pose input has a limit on how identity-preserving it can be, and a recognition model can leverage that upper limit only. To that end, many researchers focus on multi-modal fusion, where multiple input modalities are used to perform recognition. On the other hand, some researchers employ super-resolution or generative models to improve the input to the recognition model. While these are sometimes effective, they also levy a resource overhead for a limited improvement in performance. We propose developing a better input modality that outperforms any appearance or shape-based biometric, that is compatible with any existing recognition model. Furthermore, this modality can be used in a plug-and-play fashion to improve the performance of any existing multi-modal fusion model.

In this work, we propose a unified body-based biometric, called the Outfit Regularizing Biometric (ORB), built as an extension to our previously proposed B_{skel} representation [200]. ORB is designed to leverage the benefits of the aforementioned representations to generate a singular, robust body-based biometric. ORB uses the appearance information from RGB, as well as the soft biometric information such as hair and skin color. It also incorporates a novel shape-based representation, B_{skel} proposed in [200] that estimates the human skeleton from single-frame binary silhouettes of a person. ORB performs human parsing to localize and group the various body parts and attire, and subsequently devel-

ops an auxiliary image that consists of solely identity-preserving features of a person. We demonstrate that ORB is robust to significant imaging challenges such as clothing change, environment change (indoor-to-outdoor), imaging range variation, etc. ORB is fast to compute, and we demonstrate using standard transformer-based recognition backbones that it significantly outperforms the state-of-the-art appearance-based (RGB), and shape-based (B_{skel}) modalities. While not only being the best standalone modality, we also demonstrate that ORB fused with RGB and B_{skel} (TransFuse) achieves state-of-the-art (or near SOTA) performance on datasets such as BRIAR, PRCC, VC Clothes, and Celeb ReID. By focusing on inherent and enduring physical attributes, our method aims to provide a more reliable and robust solution to the problem of person re-identification, even when there are significant clothing changes between observations. Some examples of the robustness of ORB representation are shown in Figure 5.1.

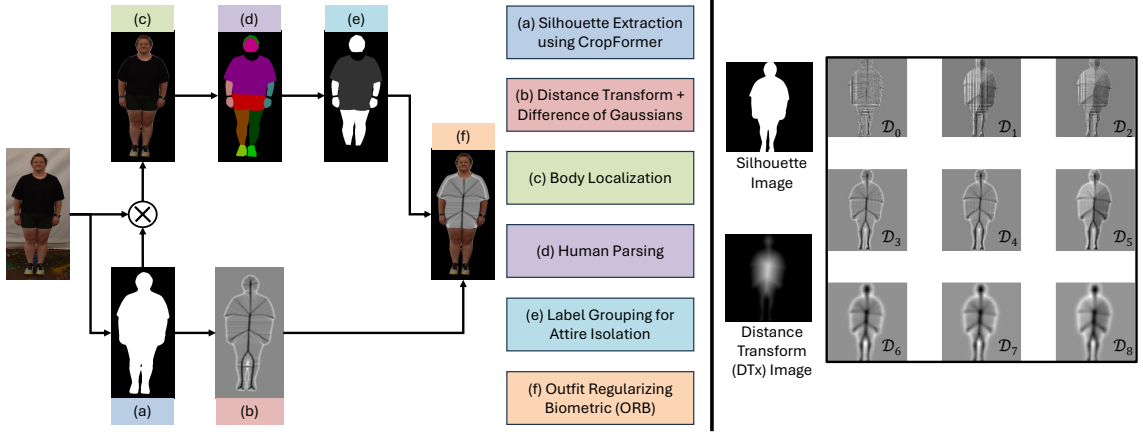


Figure 5.2: Left: Overview of the computational pipeline for the Outfit Regularizing Biometric (ORB). A description for each step is provided in the technical section. Right: Details on the computation of shape-based biometric used in Step (b) “Distance Transform + Difference of Gaussians” of ORB computation. While $\mathcal{D}_4$ is used in ORB computation, B_{skel} , a fusion of $\mathcal{D}_0$ and $\mathcal{D}_8$ is used as an additional modality in some experiments during score fusion in image matching.

5.1 Related Work and Contributions

5.1.1 Related Work

Imaging Challenges: In addition to the significant challenge of changing clothes, the task of human recognition and person re-ID is heavily affected by challenges such as imaging range, camera elevation, turbulence, articulation, varying pose, etc. Many approaches have been proposed, that tackle these challenges separately. For example, some work has been proposed to perform activity recognition [392, 140, 347, 16], to tackle the challenge of articulation and aid in recognition. Recognition under occlusion has been of interest to many researchers recently [314, 458, 448, 282]. Many approaches have been proposed towards pose invariant recognition [410, 504, 349]. Approaches such as [177, 264, 522] tackle the challenge of whole-body recognition under imaging distance and elevation, while research has also been directed towards emulating atmospheric conditions to tackle this challenge [50, 66, 304]. The BRIAR challenge [74] has developed a dataset that consists of all these challenges, including changing clothes. The task is to perform indoor (gallery) to outdoor (probe) matching - another significant challenge. Approaches have been proposed in this challenge that use pose information [410, 504, 349, 405, 370], 3D models [504, 523], gait-based models [145, 523], appearance-based models [145, 522], and shape-based models [410, 301, 145, 266, 522, 200]. There has been work towards the use of generative models to augment recognition performance [504, 436, 489].

We note that in most of the above cases, the approaches focus on developing highly sophisticated backbone architectures. Alternatively, we propose that generating a more effective, robust, input feature representation that can replace or augment RGB

inputs can be another way forward in effective clothes-changing (and distortion-invariant) recognition. We demonstrate that existing recognition backbones can provide state-of-the-art performance by augmentation via a better input modality. To that end, we propose an Outfit Regularizing Biometric (ORB) that is a unified representation consisting of the most identity-preserving attributes of appearance-based RGB, shape-based measures from silhouettes [200], and soft biometrics.

Clothing-change ReID: Traditional ReID approaches primarily rely on appearance-based features from RGB images. The use of deep networks (mainly, CNNs) has become increasingly popular in recent years, making them the current primary setup for human recognition [479]. More recently, transformers have been used for the same [371]. While these methods demonstrate high accuracy when the clothing remains consistent, their performance dramatically deteriorates in scenarios involving clothes-changing [252]. To that end, there has been a great deal of research towards extracting discriminative features from clothing, accessories, and background cues and performing efficient human recognition under clothes-changing conditions [77, 265, 154, 85]. While many of these works focus on using RGB input to perform their task [141, 409, 508, 178, 179], several works have used auxiliary information to bolster the results [141, 173, 154, 487].

Using auxiliary information addresses the limitations of only appearance-based methods by utilizing extra information obtained via pose estimation [28, 313, 342, 173]. Some techniques utilize shape-based features [6, 253, 254, 197, 173, 144]. While promising, these methods are often sensitive to inaccuracies in pose or body shape estimation, especially in occluded or crowded scenes. Many approaches use additional information such as clothes

ID to significantly improve recognition performance [141, 154, 487, 77]. However, this is contingent upon the availability of the clothing ID of every subject under every view - which is not always possible.

Another emerging direction in clothes-changing ReID research is the use of 3D features [57, 265]. Some methods also employ fusion with depth-based features [211, 245] or infrared features with RGB [62, 447, 467]. These methods provide greater invariance to appearance changes but are often constrained by the availability of depth or additional sensor data and the high complexity of 3D model reconstruction.

5.1.2 Contributions

- A novel shape-based biometric from single-frame binary silhouettes that roughly estimates the human skeleton. We report experiments that quantitatively demonstrate it to be robust to distortions like limited frames, imaging distance, change in environment, and changing clothes.
- We propose a novel unified body biometric (ORB) that integrates the enduring, identity-preserving attributes of standard shape- and appearance-based body representations into a single, robust, body biometric.
- We report experiments that quantitatively demonstrate that ORB is robust to distortions like changing clothes, imaging distance, change in environment, pose, articulation, camera elevation, and atmospheric turbulence.
- The proposed ORB and TransFuse (fusion of transformers) are fast, modular, and compact, and can be used individually or as plug-and-play modules with other approaches to bootstrap performance of long-range human recognition.

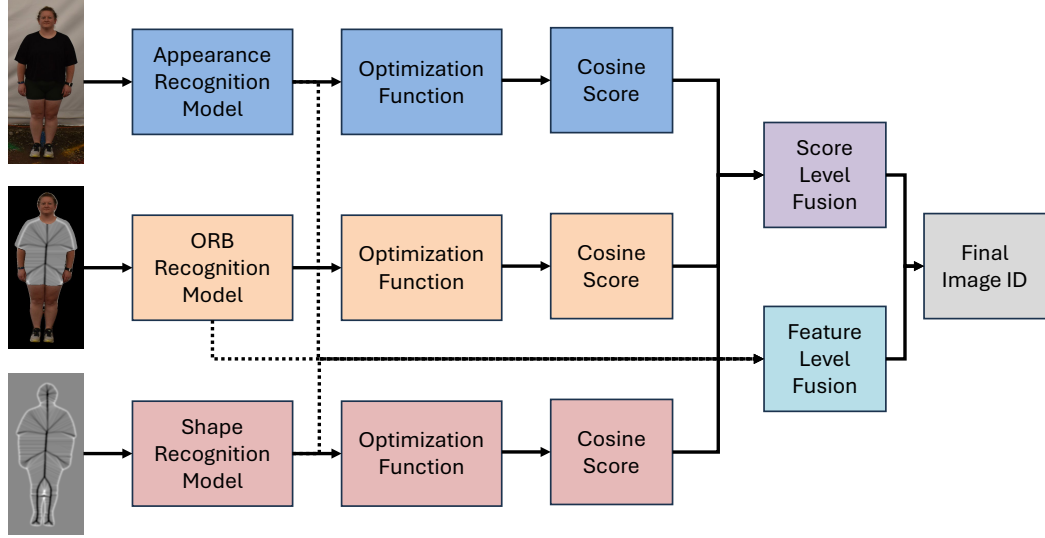


Figure 5.3: Overview of the TransFuse recognition pipeline. Each recognition backbone is made up of a Swin transformer (which can be replaced with any SOTA recognition network). The optimization function details are provided in the technical section. For image matching, image-level matching is done for all public datasets (image-based datasets). For the video-based BRIAR dataset, the image matching protocol is provided in Figure 5.4. Each recognition channel (until cosine score computation) can be added or removed as needed, making the overall pipeline plug-and-play, and modular.

5.2 Technical Approach

Figure 5.2 provides an overview of the biometric computation pipeline, whereas Figure 5.3 provides the overall recognition pipeline. We begin by isolating the parts and using the semantic knowledge to isolate the body biometric information and replacing the noisy information with shape biometric information to generate the new novel biometric. We then use a fusion of these biometrics to perform aggregated recognition. The details of each part are provided below.

5.2.1 Outfit Regularizing Biometric (ORB)

(a) Silhouette Extraction: The first step consists of extracting the silhouette of the human, as shown in Figure 5.2-a. To do this, we first perform human detection using YOLOv5. This is followed by silhouette extraction using CropFormer [341]. We used YOLOv5 to strike a balance between speed and accuracy for larger datasets, although this step can be replaced with newer methods. The extracted silhouette is used for two steps in ORB computation: (1) It is used to localize the body in the RGB frame (ref. Figure 5.2-c). This removes any noise pertaining to the surroundings and only keeps the human body in focus, and (2) it is used to compute the robust shape-based biometric, B_{skel} , which estimates a rough skeleton of the body and is scale-invariant [200]. The computed B_{skel} is subsequently used as the shape attribute in ORB.

(b) Computation of Shape Representation: The shape representation used is B_{skel} as proposed by us in [200], and is shown in Figure 5.2-b. To obtain this, we compute the distance transform (DTx) of the binary silhouette, followed by computing the Difference-of-Gaussians pyramid. This provides coarse information on the human body joints. The computation of the DoG pyramid involves hierarchical scaling and blurring using a set of Gaussian kernels, varied at a preset scaling factor. For the pyramid in this implementation, we employ 3 scales (representing the standard deviation (sigma) value of Gaussian kernels) and 3 octaves (representing the input image resolution) [419]. This results in a 9-image DoG pyramid. It is demonstrated in [200], that the fusion of the lowest (D_8) and highest (D_0) resolution images of this pyramid gave the optimal recognition performance (B_{skel} in [200]). While we use B_{skel} in our fusion model in recognition as one of the standalone

modalities, for the computation of ORB, we pick the image on the pyramid which is the same resolution as the original image (D_4) as shown in Figure 5.2 (right).

(c) *Body Localization:* We isolate the human body from the RGB image or frame by extracting the bounding box of the silhouette, followed by a pixel-wise product of the original image with the binary silhouette.

(d) *Human Parsing:* The next step in ORB computation is identifying the attire of a person. To do that, we employ the SCHP model proposed in [249]. Evaluating the three main datasets for human parsing, LIP [257], ATR [258], and Pascal [59], we noted that SCHP does not isolate clothing on Pascal, which is key to our task. Between ATR and LIP, LIP is geared towards more real-life complex scenarios whereas ATR focuses on fashion AI. Furthermore, LIP has 50000+ images compared to 17000+ for ATR. LIP has 20 labels for ATR’s 18 labels. Thus, we chose the model trained on the LIP dataset for our application. The 20 labels for LIP are: ‘Background: 0’, ‘Hat: 1’, ‘Hair: 2’, ‘Glove: 3’, ‘Sunglasses: 4’, ‘Upper-clothes: 5’, ‘Dress: 6’, ‘Coat: 7’, ‘Socks: 8’, ‘Pants: 9’, ‘Jumpsuits: 10’, ‘Scarf: 11’, ‘Skirt: 12’, ‘Face: 13’, ‘Left-arm: 14’, ‘Right-arm: 15’, ‘Left-leg: 16’, ‘Right-leg: 17’, ‘Left-shoe: 18’, ‘Right-shoe: 19’. This is followed by label grouping for body part isolation.

(e) *Label Grouping:* In this step, we suppress the labels for attire (1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19: group 1) and emphasize the labels for body parts (2, 13, 14, 15, 16, 17: group 2). For this, we set the background pixels to 0, group 1 labels to 128, and group 2 labels to 255. The subsequent image obtained is as shown in Figure 5.2-e.

(f) *ORB Computation:* The suppressed pixels in the label grouping are replaced with the estimated B_{skel} representation as the original resolution (D_4). The highlighted pixels

of the label-grouped images are replaced with the original RGB image. Thus, the ORB representation tackles the clothing change issue by replacing these attributes with a robust, distortion-invariant shape biometric while preserving the original appearance-based attribute of the face, arms, and legs. Furthermore, ORB indirectly integrates soft biometrics via skin and hair color. This leads to ORB being a more optimized feature representation that only preserves the identity-preserving features of any person while suppressing the others. The overall pipeline is shown in Figure 5.2.

5.2.2 Recognition Backbone

(a) Data Augmentation: Data augmentation is a key factor in enhancing the robustness of model training. It is essential that the data augmentations used accurately reflect the challenges posed by the dataset, especially for outdoor datasets such as BRIAR and Celeb ReID. To this end, we employ a variety of augmentation techniques, including linear transformations such as shear and translation, perspective transformations, and random erasing. Additionally, we implement a custom augmentation method inspired by [200], which randomly applies dilation or erosion to specific regions of the body. We hypothesize that each of these augmentations addresses particular challenges identified in the training data. For example, random body part erosion simulates effects like image distortion, while random erasing is intended to mimic occlusion. Similarly, perspective and shear transformations are designed to reflect real-world scenarios, such as camera movement or variations in the camera’s pitch angle. Through these targeted augmentations, we aim to improve the model’s ability to generalize across a variety of real-world scenarios.

(b) *Recognition Backbone:* The next step is feature extraction using the recognition backbone. We employ Swin Transformers [270], with their hierarchical structure and attention mechanisms, as they are particularly adept at modeling both local and global dependencies in images. Compared to traditional CNNs, Swin transformers have many advantages. Compared to the limited receptive fields in CNNs, Swin has a better global contextual understanding due to the self-attention mechanism. The hierarchical approach of Swin cuts down on computational costs while maintaining efficiency. Compared to CNNs and other transformers, Swin has a sliding window approach that allows for a flexible attention mechanism. This provides Swin with the capability to model longer-range feature associations. We report the results of Swin-based performance in ablation studies and compare it with other backbones to establish its efficacy. For our overall recognition pipeline, in addition to the unified ORB representation, we also use the purely appearance-based RGB and shape-based B_{skel} representations. We use separate feature extraction Swin per representation and fuse the scores for overall recognition. This overall pipeline will henceforth be referred to as “TransFuse”. By assigning a separate Swin Transformer to each representation, the pipeline can learn distinct, representation-specific patterns, enhancing its ability to extract relevant features for recognition. This multi-branch approach ensures that the network integrates complementary information from the different representations, leading to more robust and accurate recognition performance. We establish this property of the pipeline in our Experiments and Results section.

(c) *Optimization Function:* Given the complexity of the data, there tends to be significant diversity within the latent space. To address this challenge, we propose a loss function

that employs multi-objective optimization. The first component is an absolute loss, specifically the cross-entropy loss, which helps the classifier distinguish between different classes. This loss is represented by Eq (5.1). For more challenging cases, where embeddings from the same class are dispersed or embeddings from different classes are closer together, we incorporate additional loss terms to address both the absolute distance between pairs of points and the distance between triplets of points. The second component is a pairwise loss, where we use contrastive loss [227], which has proven effective in separating embeddings of different classes and clustering those from the same class. The corresponding equation is shown in Eq (5.2). The third component is an angular loss [451], which we prefer over the traditional triplet loss [166]. Angular loss is both scale- and rotation-invariant, as it optimizes cosine distance between the data points. This makes it particularly suitable for handling the re-scaling of input data features. Additionally, while triplet loss only considers two sides of a triplet triangle, angular loss leverages all three, providing a more comprehensive metric [451]. Angular loss also enables faster convergence, as it narrows the margin search to a smaller angular space, making it more scalable for complex data. The equation for angular loss is provided in Eq (5.3). The complete loss function for our proposed model is outlined in Eq (5.4). During training, the three loss terms are weighed equally. We compute the L2 norm of each embedding before computing the loss.

$$\mathcal{L}_{CE} = \sum_{x_i \in \mathcal{B}} (y_i \cdot \log(\hat{y}_i)) \quad (5.1)$$

$$\mathcal{L}_{con} = \frac{1}{M} \sum \left[I_{i,j} \cdot \|x_i, x_j\|_2 + (1 - I_{i,j}) \cdot [\alpha - \|x_i, x_j\|_2]^+ \right] \quad (5.2)$$

$$\mathcal{L}_{ang} = \frac{1}{N} \sum_{x_a \in \mathcal{B}} \log \left[1 + \sum_{\substack{x_n \in \mathcal{B} \\ y_n \neq y_a, y_p}} \exp(f_{a,p,n}) \right] \quad (5.3)$$

$$\mathcal{L} = \frac{1}{3} \cdot (\mathcal{L}_{CE} + \mathcal{L}_{con} + \mathcal{L}_{ang}) \quad (5.4)$$

where, $f_{a,p,n} = 4\tan^2\theta(x_a + x_p)^T x_n - 2(1 + \tan^2\theta)x_a^T x_p$, y_k is true label, $\hat{y}_k$ is predicted label, x_k is the embedding vector for sample k , samples a, p and n are the anchor, positive and negative respectively, α is the margin for contrastive loss, θ is the margin for angular loss, $\mathcal{B}$ is the batch, $I_{i,j}$ is the indicator function that is 1 if $i = j$, else -1.

(d) Image Matching: We use cosine distance for image search and matching. For all public datasets, which are image-based datasets, we compute a gallery to probe matching for all individual pairs. For the video-based BRIAR data, we perform gallery aggregation before performing matching. We aggregate embeddings across all gallery stills and sequences by averaging to obtain a single gallery embedding for each subject. For probes, we compute image-level embeddings independently. If the probe inputs are videos rather than images, we aggregate all frame-level embeddings to generate one embedding per probe video. However, in this work, we match each probe frame individually. It is important to note that the term “single-frame” in our approach refers to the input and explicitly indicates that no temporal information is utilized. Since our proposed approach at its core is still an image-based pipeline, we treat each frame separately during both training and evaluation. The aggregation technique for both the gallery and probes is generalizable beyond the proposed approach. The image-matching process for BRIAR is shown in Figure 5.4. This process saves significantly on resources.

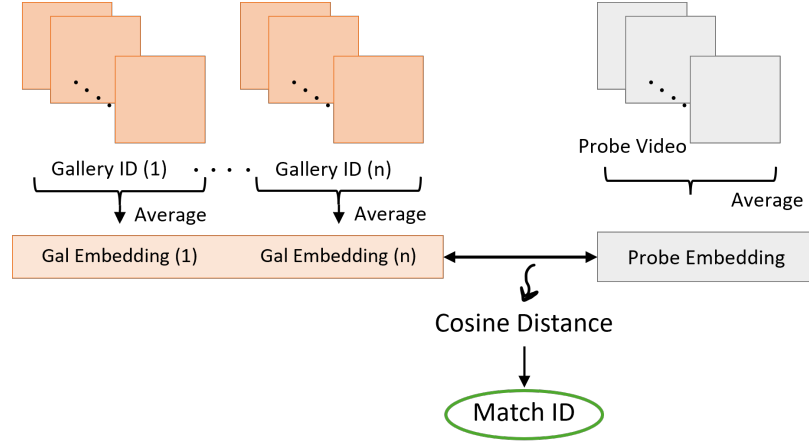


Figure 5.4: Gallery aggregation and probe video aggregation process in image matching step for the BRIAR data. For the public datasets, simple image-to-image matching is done as they are image-based datasets.

5.3 Experiments and Results

5.3.1 Datasets Used

(a) **PRCC**: PRCC is the most commonly used dataset for changing clothes-based person re-identification [485]. The PRCC dataset comprises 33,698 images of 221 distinct identities. The dataset is structured to test Re-ID systems under realistic conditions where clothing changes occur. Specifically, images from Cameras A and B feature the same individuals in identical clothing but captured in different rooms, while Camera C presents the same individuals in different outfits on different days.

(b) **Celeb-ReID**: This dataset consists of street snapshots of celebrities from the internet [180]. It consists of 34,186 images with 1,052 subjects. In addition to clothing change, this dataset also suffers from imaging distortions and pose variations.

(c) **VC-Clothes**: This data is a synthetic clothes-changing dataset generated using GTA5 game [445]. It features 512 identities captured across 4 scenes/cameras, with each identity

having an average of nine images per scene, totaling 19,060 images. The dataset is split evenly between 256 identities for training and 256 for testing, with about 1-3 outfit changes per subject.

(d) *BRIAR*: IARPA’s Biometric Recognition and Identification at Altitude and Range (BRIAR) dataset [74] is aimed towards tackling the challenges of human recognition under extreme imaging conditions. The recognition task of BRIAR is complex since it is an indoor gallery to outdoor probe matching task, with clothes changed between the gallery and probe. The dataset contains videos captured at distances varying from 100m (indoor/outdoor) to 1000m (outdoor). While the dataset is aimed towards identifying individuals from high altitudes or at long ranges, it has significant other challenges such as clothing change, occlusion, pose variation, atmospheric turbulence, UAV viewpoint, and articulation. In this work, we report the results of the BRIAR evaluation protocol EP3.1 and 4.2, while the models are trained on all training subsets released (BRS1, 1.1, 2, 3, and 4). Additionally, we report results on individual imaging modalities using the most recently released subsets (BRS3 and 4), including ablation studies results.

BRIAR Data Preparation: The BRAIR dataset consists of controlled stills and controlled indoor videos for gallery, and unconstrained outdoor videos for probes. In order to achieve optimal performance with the least amount of training data, we perform range-based sampling of the training data. We use all the gallery stills (108 images) and evenly sample 50 frames for controlled gallery videos. For the probe videos, this protocol is particularly effective, as the number of videos is unequal for shorter and longer ranges. Thus, we ensure 150 frames per range per subject. For example, if a subject has 5 videos at the 1000m distance,

we evenly sample 30 frames per video for all 5 videos. This ensures appropriate diversity across all ranges for all subjects. For BRS1-4 subsets, we have a total of 743 subjects, which results in approx. 1000-1400 images per subject.

5.3.2 Experimental Setup

TransFuse Results: Benchmark Datasets

Table 5.1 shows the results of the proposed model for changing clothes and general setups of the PRCC dataset, compared with the state-of-the-art models. We see that the TransFuse model achieves results comparable to SOTA. When the backbone Swin is replaced with MGN [449] backbone, we note that we achieve the best SOTA results. This demonstrates that ORB is compatible with any existing recognition backbone, similar to RGB. Table 5.2 shows the results using TransFuse for the same using the VC-Clothes and Celeb-ReID datasets. For VC-Clothes, we report results using the general setup across all cameras, the same clothing setup (between cameras 2 and 3), and changing clothes setup (between cameras 3 and 4). We see that for both VC-Clothes and Celeb-ReID, we achieve SOTA/near-SOTA comparable performance with TransFuse. We note that across all the datasets, ORB has a very strong standalone performance, and fusion via TransFuse has SOTA or near SOTA performance.

For a more in-depth analysis, in Table 5.3, we show the performance comparison of ORB, RGB, and B_{skel} modalities using the Swin backbone. We report the performance of each standalone modality and its various fusion metrics. We demonstrate that while

Table 5.1: Comparative analysis of the top state-of-the-art models’ performance on the PRCC dataset, with our proposed model. We report the results of our model using the Swin backbone and the MGN backbone. **Bold** represents the best overall result, and Underlined represents the second best result. Results are reported for general (including similar clothing), and clothing change setups. Results are divided based on RGB-only models (listed in chronological order) in rows 1-4; followed by multi-modal models (listed in chronological order) in rows 5-14.

Methods	Modalities	Reference	PRCC Dataset			
			Clothing Change	General		
			Rank-1	mAP	Rank-1	mAP
PCB [409]	RGB	ECCV 2018	41.8	38.7	99.8	97.0
RGAS-C [508]	RGB	CVPR 2020	42.3	-	98.4	-
RCSANet [178]	RGB	ICCV 2021	48.6	50.2	100	97.2
MCSC-CAL [179]	RGB	TIP 2024	57.8	57.3	99.8	99.8
CESD [342]	RGB + Pose	ACC V 2020	-	-	-	-
FSAM [173]	RGB + Pose + Sil	CVPR 2021	54.5	-	98.8	-
3DSL [57]	RGB + 3D pose + Sil	CVPR 2021	51.3	-	-	-
CAL [141]	RGB + Clothes ID	CVPR 2022	55.2	55.8	100	99.8
AIM [487]	RGB + Clothes ID	CVPR 2023	57.9	58.3	100	99.9
CCFA [154]	RGB + Clothes ID	CVPR 2023	61.2	58.4	99.6	98.7
DCR-ReID [77]	RGB + Sil + Clothes ID	TCSVT 2023	57.2	57.4	100	99.7
SC-Net [144]	RGB + Sil	ACM MM 2023	61.3	59.9	100	97.8
CVSL [313]	RGB + Pose	WACV 2024	57.5	56.9	97.5	99.1
Ours (using Swin)	ORB + B_{skel} + RGB		58.6	55.7	100	99.6
Ours (using MGN)	ORB + B_{skel} + RGB		62.0	58.3	100	99.8

Table 5.2: Comparative analysis of the state-of-the-art models’ performance on the VC-Clothes and Celeb-ReID datasets, with our proposed model. We report the results of our model using the Swin backbone, as well as the MGN backbone. **Bold** represents the best overall result, and Underlined represents the second best result. Results are reported for general, similar clothing, and clothing change setups for VC-Clothes, and standard clothing change setup for Celeb-ReID. Results are divided based on RGB-only models (listed in chronological order) in rows 1-5; followed by multimodal models (listed in chronological order) in rows 6-7.

Methods	Modalities	Publication	VC-Clothes						Celeb-ReID	
			General (All cams)		SC (Cams 2 and 3)		CC (Cams 3 and 4)		General CC	
			Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
HACNN [251]	RGB	CVPR 2018	68.6	69.7	-	-	49.6	50.1	47.6	9.5
PCB [409]	RGB	ECCV 2018	87.7	74.6	94.7	94.3	62.0	62.2	37.1	8.2
TransReid [163]	RGB	ICCV 2021	90.5	80.1	95.1	94.5	70.0	71.8	58.9	14.6
FSAM [173]	RGB + Pose + Sil	CVPR 2021	-	-	94.7	94.8	78.6	78.9	-	-
CAL [141]	RGB + Clothes ID	CVPR 2022	92.9	87.2	95.1	95.3	81.4	81.7	-	-
Ours (using Swin)	ORB + B_{skel} + RGB		<u>92.7</u>	<u>84.2</u>	96.07	94.37	88.23	85.79	<u>56.42</u>	<u>12.18</u>

Table 5.3: Performance of individual modalities and their fusion across all public datasets using Swin backbone (TransFuse). **Bold** represents the key performance trend to observe, whereas *Italics* represent the overall TransFuse results.

Modalities	PRCC SC		PRCC CC		VC-Clothes SC		VC-Clothes CC	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
RGB	99.95	99.43	43.3	44.7	96.47	94.88	72.94	74.06
B_{skel}	56.41	29.87	28.3	17.4	40.78	23.69	47.84	26.39
ORB	86.45	75.30	52.0	44.8	94.90	91.48	83.14	79.49
<i>All three</i>	<i>100</i>	<i>99.6</i>	<i>58.6</i>	<i>55.7</i>	<i>96.07</i>	<i>94.37</i>	<i>88.23</i>	<i>85.79</i>

RGB has strong performance for the same clothing setting, ORB significantly outperforms RGB modality for the clothing change setting. This shows that in similar clothing settings, RGB also does color matching on clothing which is unavailable for cloth-changing settings - leaving only biometric features to be analyzed. In such cases, the proposed ORB has a much better performance, outperforming RGB by almost 8-11%.

TransFuse Results: BRIAR Dataset

We set up the evaluation of the BRIAR dataset in two ways. The first setup uses the two latest released subsets of the BRIAR training data (BRS 3 and 4). This consists of 363 subjects, with 273 subjects in training and 90 in testing. We report the results of various input modalities (ORB, B_{skel} , and RGB), the robustness of the proposed method towards changing clothes, changing environment, and across varying imaging ranges. The second setup is the standard BRIAR task, with training performed using all 5 released BRIAR training subsets (BRS 1, 1.1, 2, 3, and 4), and evaluation done on Phase 1 and Phase 2 evaluation protocols by ORNL (EP3.1 and EP4.2, respectively). Since the primary goal of our proposed biometric is recognition under extreme conditions, for both protocols, we focus on the Face-Restricted Treatment (FRT) subsets. The FRT subsets of any BRIAR evaluation protocol are defined by probes where less than 1% of the face is visible in each

frame of a video, thereby severely limiting the dependency on purely appearance-based biometric matching. For EP3.1, the test set consists of data from 4 BRIAR testing subsets (BTS1, 1.1, 2, and 3), whereas EP4.2 also additionally contains data from BTS4. For the FRT subset, verification is the most important metric, with Phase 1 focusing on TAR @ 1% FAR and Phase 2 focusing on TAR @ 0.1 % FAR.

Table 5.4: Performance of individual modalities and their fusion on the most recent BRIAR subsets (BRS 3 and 4). This consists of 363 subjects, with 273 subjects in training and 90 in testing. **Bold** represents the best overall performance, and Underline represents best standalone modality. SL: Score-level Fusion. Matching is done on the standard BRIAR task, with set 1 clothing indoor data as gallery and set 2 clothing outdoor data as probe.

Input	Rank1	Rank5	Rank20
$\hat{\mathcal{D}}_0$	18.97	44.44	77.45
$\hat{\mathcal{D}}_8$	19.17	44.84	78.28
RGB	26.61	58.27	85.39
ORB	<u>44.19</u>	<u>77.13</u>	<u>94.06</u>
B_{skel} (SL)	19.72	45.57	79.03
$B_{skel} + \text{RGB}$ (SL)	32.80	64.58	91.20
RGB + ORB (SL)	44.24	76.18	95.22
$B_{skel} + \text{ORB}$ (SL)	39.52	71.50	93.57
All three (SL)	44.81	76.85	95.90

Table 5.4 reports the results of the first evaluation setup of BRIAR across various imaging modalities and their subsequent fusion. We note that ORB significantly outperforms RGB modality. We also note that unlike public datasets, the fusion of all three modalities performs the best for BRIAR (instead of only ORB and RGB), due to the extremely challenging nature of the data, where B_{skel} learns crucial complementary information. In fact, a key observation here is that the fusion of RGB and B_{skel} gives the biggest boost from their respective individual modalities than any other pair-wise fusion. This is because appearance and shape features are inherently completely complimentary, unlike ORB which has information from both. Table 5.5 contains the results of TransFuse for the closed-set

Table 5.5: BRIAR Evaluation Protocol Results for EP3.1 and EP4.2. The metrics of interest are closed-set Rank-20 performance, and verification performances - TAR @ 1% for EP3.1, and TAR @ 0.1% FAR for EP4.2. The target metric for the Face Restricted Treatment (FRT) subset for verification is provided in (**bold, parenthesis**).

Input	Rank-20	TAR@1% FAR	TAR@0.1% FAR
EP3.1	76.5	53.5 (50)	19.7
EP4.2	77.7	55.2	24.0 (50)

detection and verification tasks under EP3.1 and 4.2 protocols. Figure 5.5 shows the plots for the same. The first half has the verification and closed-set plots for EP3.1, and the second half shows plots for EP4.2. EP3.1 consists of 8136 probes, whereas EP4.2 contains 7642 probes. In this section, for both evaluation setups, the gallery is indoor data with Set 1 clothing, and the probe is outdoor data with Set 2 clothing, the standard BRIAR setup.

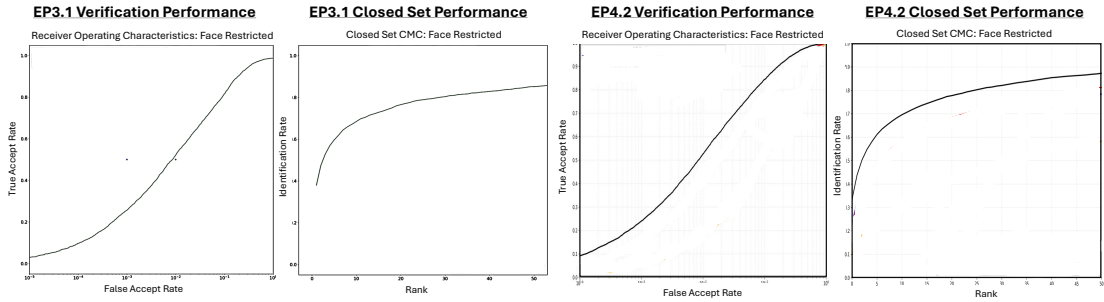


Figure 5.5: Plots and trends for the performance of TransFuse on BRIAR’s standard Evaluation Protocols (EP) EP3.1 (for Phase 1) and EP4.2 (for Phase 2). They represent the plots for verification and closed-set performance on the face-restricted subset of BRIAR. The first two plots are results for the EP3.1 protocol, while the remaining two represent performance on EP4.2.

(a) Robustness to Clothing Change: We use the most recent subsets of BRIAR to isolate the impact of clothing change artifacts. We isolate the clothing change by performing an indoor-to-indoor matching, across both sets of clothing. Thus, we use the Set 1 clothing gallery images as the gallery, and Set 2 clothing gallery images as the probe. Since the indoor images of BRIAR data are captured under similar conditions for both clothing, this

removes any other variables in the imaging such as imaging distortions. The results for this experiment are given in Table 5.6. We note that ORB significantly outperforms RGB in this setup, establishing ORB’s efficacy in clothing change recognition. We note that the fusion of other modalities with ORB still provides a noticeable improvement in the overall performance.

Table 5.6: Impact of changing clothes on the recognition task. Experiment done on a subset of 363 subjects from BRIAR, indoor-indoor matching, with changing clothes between gallery and probe. Train/test split: 273/90 subjects. SL: Score-level fusion. **Bold** represents the best overall performance, and Underline represents the best standalone modality.

Input	Rank1	Rank5	Rank20
$\hat{\mathcal{D}}_0$	38.09	65.40	88.56
$\hat{\mathcal{D}}_8$	38.87	66.86	89.33
RGB	47.91	70.37	91.59
ORB	<u>68.64</u>	<u>87.67</u>	<u>94.99</u>
B_{skel} (SL)	39.79	66.90	89.38
$B_{skel} + \text{RGB}$ (SL)	58.32	81.68	93.99
ORB + B_{skel} (SL)	66.71	85.42	94.64
ORB + RGB (SL)	68.28	86.64	96.06
All three (SL)	71.31	87.95	95.49

(b) Robustness to Environment Change: We isolate the environment change aspect of the BRIAR data, to evaluate the efficacy of the proposed ORB under indoor-to-outdoor matching task. Towards this, we use Set 1 clothing indoor data as the gallery and Set 1 clothing outdoor data as the probe. The results are reported in Table 5.7. We note that in this case, RGB outperforms ORB. This is because there is no clothing change, and so when the face is not visible, RGB can still match pixels on the clothing to ensure a match. We note that this is not a robust method for matching, as it can be seen that under unconstrained conditions (Table 5.4), ORB has a distinct advantage.

(c) Robustness to Imaging Distance: To evaluate the robustness of the proposed representation across varying imaging ranges, we split the probes into ranges of <100m

Table 5.7: Impact of a changing environment (indoor gallery to outdoor probes) with similar set clothing for RGB images. Experiments are done on a subset of 363 subjects from BRIAR data, with 273 subjects in training and 90 subjects in evaluation. **Bold** represents the best overall performance, and Underline represents the best standalone modality.

Input	Rank1	Rank5	Rank20
$\hat{\mathcal{D}}_0$	36.51	63.10	85.48
$\hat{\mathcal{D}}_8$	36.84	63.09	85.25
RGB	90.46	97.61	99.45
ORB	67.49	86.22	95.97
B_{skel} (SL)	37.87	64.31	86.12
B_{skel} + RGB (SL)	87.25	96.48	99.21
ORB + B_{skel} (SL)	65.69	84.53	95.28
ORB + RGB (SL)	58.31	87.56	96.19
All three (SL)	89.25	97.00	99.45

(close-range), 100-300m, and 300-600m, and UAV/long-range (over 600m). We use the same gallery (indoor, set 1 clothing) to evaluate performance on these various probes. The results are reported in Table 5.8. It can be seen that ORB consistently, and significantly outperforms all other modalities. This is a significant characteristic of ORB that makes it a highly effective and robust biometric.

Table 5.8: Results with varying ranges using TransFuse. The model is trained on all ranges and evaluated for varying ranges of probes. The number of probe frames in close-range were 5663, in 100-300m were 11,732, in 300-600m were 10,727, and in long-range/UAV were 6015. **Bold** represents the best performance.

Input	Test Set	Rank1	Rank5	Rank20
RGB	Close-range	29.50	64.93	89.01
	100-300m	29.82	62.95	87.20
	300-600m	22.79	51.15	82.77
	UAV/Long-range	24.44	55.54	82.94
B_{skel}	Close-range	31.52	57.02	82.78
	100-300m	25.23	53.82	83.45
	300-600m	9.52	32.18	72.88
	UAV/Long-range	16.01	42.43	77.87
ORB	Close-range	51.37	83.19	96.02
	100-300m	49.78	80.95	96.26
	300-600m	34.05	69.20	91.45
	UAV/Long-range	44.47	77.82	92.17

(d) **Robustness of B_{skel} to limited video frames:** Here, we report the performance of our single-frame model compared to the performance of state-of-the-art gait recognition models when a limited number of frames are available. For Gait3D dataset [514], per-

formance is reported when limited frames are used in training and testing for three gait recognition algorithms, SMPLGait [514], GaitSet [54] and GaitPart [106]. We report these results when either training or evaluation data used 10 frames each, and compare with our single-frame approach. A similar study was conducted for OU-MVLP dataset [490]. We report this analysis for the authors’ model in [490] and GaitSet for 5 and 10 frames in training. These results are reported in Table 5.9. We report only Rank-1 accuracies for consistency with [514] and [490]. For OU-MVLP dataset, we note that our model has comparable performance to the top method. For the more complex Gait3D dataset, except for SMPLGait trained on 10 frames, our model significantly outperforms all models by nearly 7% and 33% for limited frames in training and testing, respectively, demonstrating the efficacy of our model even with a single frame. These results are reported in [200], where recognition backbone used is HRNet.

5.3.3 Ablation Studies

(a) Varying Recognition Backbone: In this section, we establish the efficacy of using the Swin transformer as our recognition backbone. For a fair comparison, we evaluate the performance on various backbones using RGB, but not ORB. This is to ensure that we compare our results using ORB against the best-performing RGB backbone. We perform comparisons using ResNet50, HRNet30, and Swin. We use the same loss function as reported in Section 5.2.2 (c -optimization function) for all models. We use the BRIAR data for this analysis, as it is the most challenging dataset in our evaluation. We note that Swin model outperforms all the other models.

Table 5.9: Performance of Gait Algorithms using limited number of frames, for Gait3D and OU-MVLP datasets vs SSRNet that uses only single frame. Results for comparative models are from [514] for Gait3D and [490] for OU-MVLP. SL: Score-level fusion. Top results are Underlined, our best results are in **Bold**.

Input	Model	Rank 1
<i>Limited Frames: Gait3D Dataset</i>		
10 frames (Training)	GaitPart	19.10
	GaitSet	32.90
	<u>SMPLGait</u>	<u>40.90</u>
10 frames (Testing)	GaitPart	4.50
	GaitSet	6.40
	SMPLGait	6.80
B_{skel} (SL)	SSRNet	40.05
<i>Limited Frames: OU-MVLP Dataset</i>		
5 frames (Training)	Yao et. al. [490]	78.0
	GaitSet	69.8
10 frames (Training)	<u>Yao et. al. [490]</u>	<u>78.1</u>
	GaitSet	72.2
B_{skel} (SL)	SSRNet	76.88

Table 5.10: Experiments to determine the optimal backbone. Experiments were conducted on RGB modality, using 363 Briar subjects (273 in training, and 90 in evaluation). **Bold** represents the best performance.

Backbone	Rank1	Rank5	Rank20
ResNet50	25.67	57.83	84.83
HRNet-30	25.01	51.53	85.35
Swin Transformer	26.61	58.27	85.39

(b) Varying Aggregation Techniques: We utilized a range of techniques for both feature-level and score-level fusion. In the case of feature-level fusion, we combined the feature embeddings from each modality—namely ORB, RGB, and B_{skel} —right before the matching score was computed. For the score-level fusion, we explored a variety of methods, such as multiplying the individual scores from each modality, selecting the maximum score from the available modalities, and aggregating the scores in different ways. Through experimentation, we found that score-level fusion using the mean of the scores provided the best overall performance.

Table 5.11: Experiments to determine the optimal fusion technique. Experiments were conducted on the fusion of ORB, RGB, and B_{skel} , using 363 Briar subjects (273 in training, and 90 in evaluation). **Bold** represents the best performance.

Fusion Method	Rank1	Rank5	Rank20
Feature-level	44.42	73.88	93.72
Score-level (prod)	39.94	66.81	82.40
Score-level (max)	39.28	73.22	94.80
Score-level (sum)	44.81	76.85	95.90

5.3.4 Training and Inference Details

Resource Utilization: The speed of computing B_{skel} on a CPU is about 12-15 images/sec. When using a NVIDIA GeForce RTX 3080 GPU, the overall time to process a single RGB frame to obtain ORB representation is $<0.5s$. One of the benefits of our proposed approach and the proposed unified biometric (ORB) is that the embedding size is very small that allows for faster computations. The overall speed of template matching and template size are important criteria in the BRIAR evaluation protocol. In the proposed configuration of TransFuse, the embedding size is 1024 per modality (approx. 4KB), and

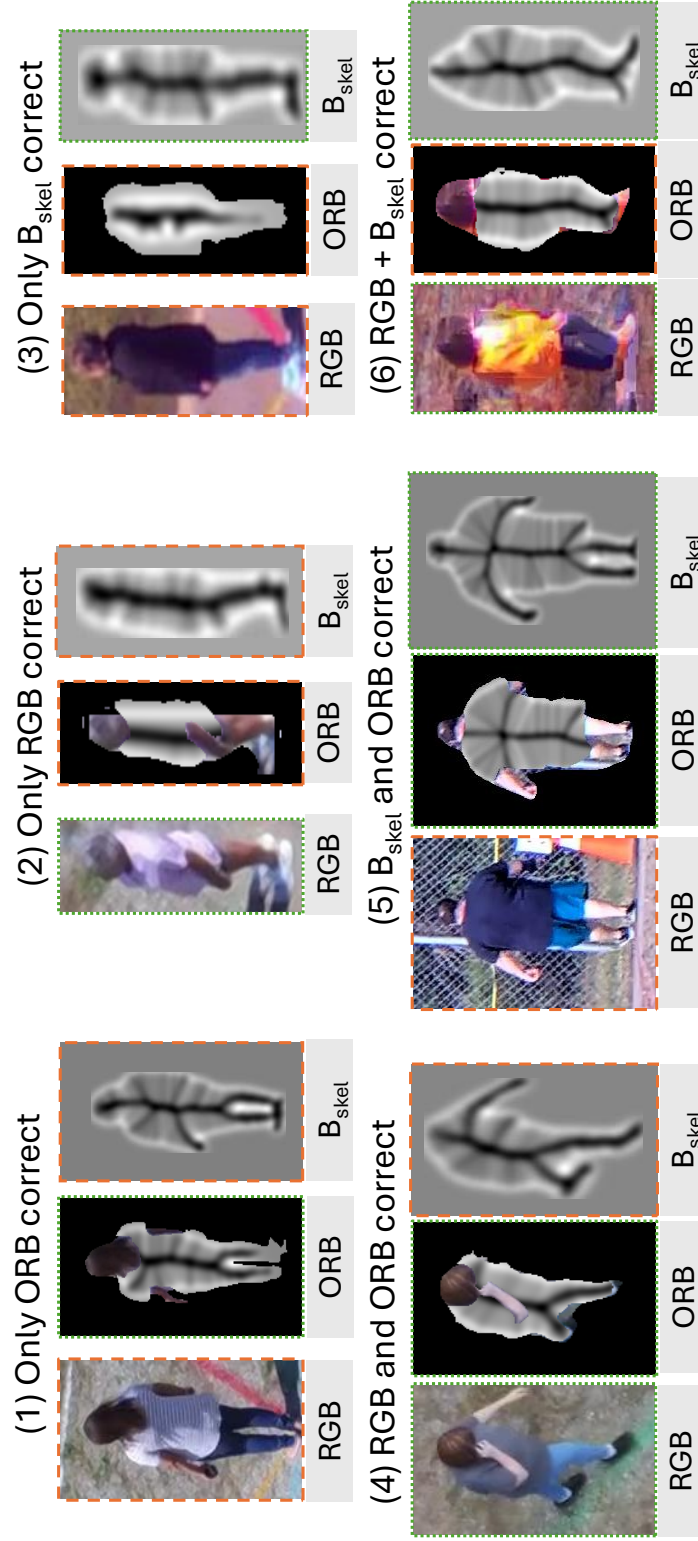


Figure 5.6: Failure Cases in BRIAR predictions. We use examples where one or two modalities fail. A detailed explanation is provided in Section 5.3.5.

each gallery and probe embedding utilizes a size of about 12KB. The BRIAR requirement for a max template size is $< 1\text{MB}$, which is much higher than our embedding size. When computing on NVIDIA GeForce RTX 3080 GPU, the speed to compute embeddings per modality is $< 0.02\text{s}$ and that for image matching is greater than 2500 probes/sec.

Training and Computation Details: To extract silhouettes from RGB videos of the BRIAR dataset, we first use YOLOv5 for human detection to identify bounding boxes. Then, we employ CropFormer to obtain the silhouettes. The BRIAR dataset is large-scale, so we chose YOLOv5 to balance speed and accuracy. For computing the distance transform, we apply a 5×5 mask with the L2 distance metric. To generate the Difference of Gaussians (DoG) images, we start with a Gaussian kernel with $\sigma=1$ for the first of three scales, scaling it by a factor of $1/2^n$ for the subsequent scales. The input size for the Swin Transformer is 224×224 , while for the MGN backbone, it is 128×384 .

To maintain the aspect ratio of subjects within tight bounding boxes, we use resizing with padding instead of standard resizing. A separate recognition backbone is trained for each input image modality. For example, when fusing RGB and ORB, two networks are used. Separate models are trained for each modality because: (1) Modularity: This approach allows for any feature to be applied in a “plug-and-play” fashion as needed without additional dependencies, and (2) Performance: Through experiments with early input-level fusion, late-stage feature fusion, and score-level fusion, we found that score-level fusion yielded the best performance. Each dataset is trained using its own training data, given the different recognition tasks. We use the base Swin model, training it for 120 epochs and selecting the model with the lowest loss. The training process uses the Adam optimizer,

a learning rate of 0.001, and a cosine learning rate scheduler [274]. The loss function and data augmentations are described in previous sections. For data augmentations, we employ random erasing with a 0.4 probability, horizontal flipping with a 0.5 probability, and shear, translation, and perspective transforms with a 0.2 probability each. Random erosion and dilation are applied with a 0.25 probability. We use a batch size of 24 and an embedding size of 1024. The MGN model is trained for 1000 epochs, with a MultiStepLR scheduler with steps at 300 and 600. We set an initial learning rate of 5e-4 with a batch size of 4 per ID. For score-level fusion, we assign equal weights to each score.

5.3.5 Failure Cases

In this subsection, we discuss the examples where each of the modalities fails, using BRIAR data as case points. We show examples where the final prediction is incorrect based on errors due to one or two modalities. The examples are shown in Figure 5.6. For Image (1), we see that the subject is turned away from the camera, which causes the RGB (appearance-based) modality to fail. It also has a camera elevation, which causes the shape-based B_{skel} to fail. However, we note that the ORB modality has the correct prediction. In Image (2), we note that RGB can detect the person correctly but ORB fails. This is potentially due to suboptimal parsing (where the face and hair pixels appear distorted). A similar error due to parsing can be seen in Image (3), where we note B_{skel} has the correct prediction but we can see that ORB does not detect the head of the person at all, thereby preventing it from learning any soft biometric features (red hair). RGB image suffers from illumination and pose issues, thereby having a wrong prediction. For the fourth image, we see that due to the camera elevation, the shape-based B_{skel} suffers adversely, and becomes

the only modality to fail. In the fifth image, we see that RGB is the only modality with incorrect prediction. In this case, except for the face turned around, the pose of the person is not complicated, and the image does not suffer from significant distortions. This implies that the most likely cause of this error is clothing change from gallery to probe. For the last image, we note that the image suffers from a high level of distortion, despite which RGB has the correct prediction. Furthermore, even though the camera has a slight elevation, shape-based B_{skel} has correct prediction too. This image represents a clear failure case of ORB, which can be improved in future works.

Chapter 6

Novel Sampling Strategy for Diverse and New Protein Conformation Synthesis using Generative Deep Networks

6.1 Introduction

Sampling protein conformations is a crucial tool in biology and medicine, with applications in structure prediction, protein-protein interactions, and targeted therapies. Characterizing the biological properties of proteins, as they pertain to their structure and dynamics, is a critical part of many biomedical tasks, such as drug discovery. Sampling from protein conformations allows us to predict protein function and, subsequently, their

utility. Traditionally, computational methods such as molecular dynamic (MD) simulations have been the gold standard for protein conformation sampling. Although effective, these techniques suffer from many drawbacks such as high computational costs, difficulty escaping local energy, and limited temporal resolution [241, 376, 239]. Deep learning has shown remarkable performance in learning nonlinear feature associations on large-scale data, making it ideal for studying protein structures. Recently, many efforts have been dedicated to employing machine and deep learning techniques to understand protein properties [33, 51, 443], protein-protein interactions [397, 396, 47], protein structure generation [124, 10, 432], etc. Some models, such as AlphaFold2 [209], have shown remarkable performance in generating desired protein structures. One limitation of this model is that it does not sufficiently capture the dynamic nature of proteins to generate novel conformations. One particularly challenging target is Amyloid-beta 42 ($A\beta_{42}$), an Intrinsically Disordered Protein (IDP) strongly implicated in the pathogenesis of Alzheimer’s disease. $A\beta_{42}$ ’s propensity to misfold and aggregate into insoluble fibrils makes it a key focus in the study of neurodegenerative diseases. Understanding and generating novel, stable $A\beta_{42}$ conformations could significantly contribute to drug discovery efforts, offering insights into the peptide’s misfolding pathways and identifying new therapeutic strategies. Despite advancements in protein structure prediction and design, the challenge of generating novel $A\beta_{42}$ conformations persists due to the peptide’s high conformational flexibility and aggregation tendencies. Traditional protein design methods, while powerful, often struggle to balance the exploration of the vast conformational space and the generation of energetically stable, physically valid protein structures. Furthermore, existing generative models tend to either fail in captur-

ing the subtlety of A β 42’s conformational diversity or generate redundant samples that provide little insight into novel structural states. Generative deep learning models have a very high capability to learn underlying data associations, making them highly desirable in the quest to synthesize novel protein conformations. Consequently, a number of generative models such as Variational Autoencoders (VAEs) [382, 281], Generative Adversarial Networks (GANs) [225, 261], and Graph Networks [404, 183] have been proposed for protein structure synthesis. These models serve as important milestones in characterizing the underlying computational models of protein structures and dynamics. However, some of these models only generate backbone conformations or study small or structured proteins. More recently, efforts have been dedicated towards MD sampling of IDPs using the C α representation [524, 191], taking a step towards modeling both the structural and dynamic properties of larger IDPs. Although these show great promise, a more granular approach in the feature descriptor for IDPs is needed to better describe their function. To address these challenges, we propose a novel generative framework that combines the strengths of two powerful machine learning paradigms: Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs). Using conformations obtained from MD simulations as our dataset, our approach introduces a novel sampling technique designed to enhance the diversity and stability of the generated A β 42 conformations, enabling the generation of a broad range of novel structures unseen in our dataset. This hybrid generative model, PROtein conFORMATION Xplorer (ProFormX) captures both the continuous nature of the conformational space and the high-dimensional structural relationships within A β 42, making it well-suited for protein design. The model learns from the intrinsic disorder in the

A β 42 dynamics and explores its conformational space to synthesize potentially new and stable structures. A key innovation in our work is a novel sampling technique, which leverages the latent space learned by the model to guide it toward synthesizing energetically stable conformations. This sampling method is designed to overcome common pitfalls in generative models, such as mode collapse and poor sample diversity. By encouraging the exploration of underrepresented regions of the conformational space, the technique ensures the generation of novel conformations that go beyond simple perturbations of known structures. Furthermore, this sampling technique can be applied to existing methods, making it highly generalizable.

In this study, we demonstrate its efficacy in a standard autoencoder-type model, and present a comprehensive evaluation of the proposed framework in the context of A β 42 conformational sampling. We conduct extensive computational experiments to assess the diversity, stability, and physical plausibility of the generated conformations. Our results demonstrate that the model not only successfully generates novel A β 42 structures but also provides valuable insights into the conformational landscape of this peptide. Moreover, the generated conformations exhibit key characteristics associated with A β 42 aggregation, offering potential starting points for the development of therapeutics aimed at modulating its misfolding and aggregation pathways.

6.1.1 Contributions

The findings of this study represent a significant step forward in computational protein sampling, particularly for challenging targets like A β 42. By combining advanced generative modeling techniques with a novel sampling approach, our method opens up new

possibilities for exploring protein conformational landscapes and designing structures that could be used in future drug discovery efforts. The key contributions of our work are:

1. A novel sampling strategy for generating diverse and energetically stable protein conformations from the latent space.
2. A generative network, that is able to faithfully synthesize new protein structures of A β 42, given force field information. With varying force fields, the network is able to adapt to learning the corresponding properties for accurate synthesis.
3. Proposed ProFormX (sampling strategy + generative network) has a powerful capability of a high degree of control over synthesis and is able to perform reconstruction, controlled perturbation-based synthesis, as well as uncontrolled synthesis based on the user’s choice.
4. Extensive analysis of the realism of generated structure, establishing that ProFormX is able to sample A β 42 conformations that are realistic, stable, and different from the original MD nearly 85% of the time. Results demonstrate that our model is able to sample conformations from the regions in the 2D energy plane that did not have structures in the actual MD.

6.2 Materials and Methods

6.2.1 Dataset

MD simulation Protocol

For this study, we study the 42-residue amyloid- β peptide (A β 42). Specifically, we utilize the aqueous solution structure (1Z0Q) and the amyloid fibril (2NAO) from the

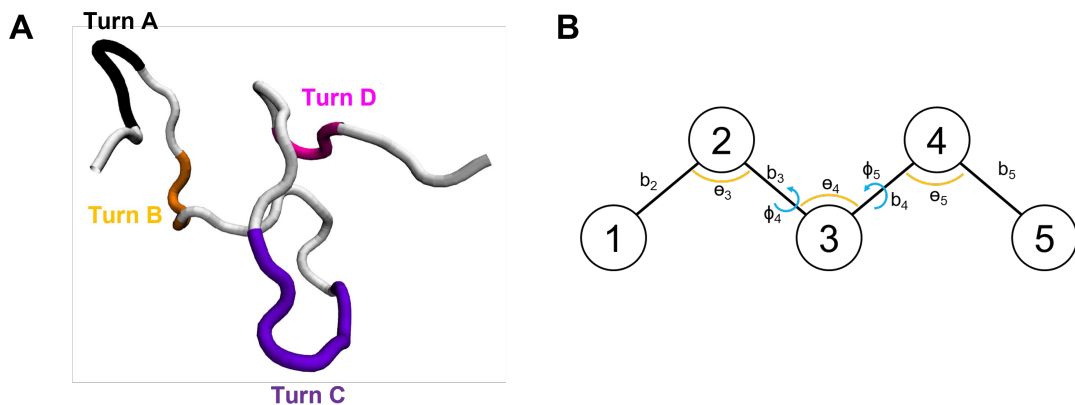


Figure 6.1: (A) synthetic conformation of A β 42 with all 4 turns A-D; (B) Classical internal bond-angle-torsion (BAT) coordinate representation (Z-matrix) for a small molecule. Atoms 1-3 are termed root atoms for this molecule. Atoms $i > 3$ are defined by (b_i, θ_i, ϕ_i) , where b , a and θ are bond, angle, and torsions, respectively. For example, atom 4 is presented by (b_4, θ_4, ϕ_4) .

PDB database. The structures were subjected to two forcefields, 14SBFF and IDPS, and a total of 4 trajectories were obtained at a $1\mu s$ resolution. The description of these trajectories is detailed in Table 6.1. We note that a combination of these trajectories holistically captures the dynamic nature of the A β 42 monomer. To avoid oversampling of conformations that are highly similar to each other, we remove conformations whose backbone RMSD values were within $2\text{\AA}$. MD runs 1-4 were performed using the AMBER20 package using either ff14SB or ff14IDPSFF force fields with different initial conformations (PDB: 1Z0Q or PDB:2NAO) (Table 6.1) [284, 400]. The systems were simulated using the TIP3P explicit solvent model at a temperature of 298K with an NPT ensemble [207]. $12\text{\AA}$ cutoff was used for short-range non-bonded interactions and the long-range electrostatic interactions were computed by the particle mesh Ewald method (PME) [104]. Each frame was saved at 1ps time integral which made up a total of 1,000,000 frames. Other MD trajectories using either PDB 2NAO

or 1Z0Q as the initial structure were obtained from our previous work and detailed in Table 6.1 [300, 429, 444].

Table 6.1: Details of MD simulation and Training/Validation of ProFormX. Structure Type: Force field and Initial Structure; Tr/Val: Training/Validation; Avg RMSD: Validation of reconstructed conformations, average torsion RMSD (All heavy atoms /Backbone)

MD Simulations		Model Construction/Validation	
MD Index	Structure Type	# Frames of Tr/Val	Avg RMSD
MD Run 1	ff14IDPSFF (2NAO)	16,000/4,000	0.339 / 0.244 rad
MD Run 2	ff14IDPSFF (1Z0Q)	16,000/4,000	0.339 / 0.257 rad
MD Run 3	ff14SB (2NAO)	16,000/4,000	0.279 / 0.204 rad
MD Run 4	ff14SB (1Z0Q)	16,000/4,000	0.241 / 0.151 rad

Bond-Angle-Torsions (BAT) Coordinate

Our method is based on the Bond-Angle-Torsion (BAT) internal coordinate system. Unlike Cartesian coordinates, BAT coordinates accurately define the natural circular motion of the dihedral rotation of the molecules [52]. Classical BAT is defined as sets of root atoms and the position of other atoms depends on the previously defined atoms bonded in sequence. For instance, Atom 1 is at the origin; atom 2 is defined by a bond length (b_2) and placed on the z-axis; and atom 3 is defined by a bond length (b_3), bond angle (θ_3) and placed on the x-z plane, and thus we eliminate 6 degrees of freedoms. The total degree of freedom of molecules with N atoms has $3N-6$ internal coordinates. Atoms 1-3 are termed root atoms for this molecule. Atoms $i > 3$ are defined by (b_i, θ_i, ϕ_i) , where ϕ are torsion angles. For example, atom 4 is presented by (b_4, θ_4, ϕ_4) (Figure 6.1-B). Starting from root atoms, all subsequent atoms are defined sequentially. To mitigate the periodicity issue, we use the sin and cos of the angles in our input to ProFormX.

6.3 Technical Approach

The overview of ProFormX is shown in Figure 6.2. The real data input to the model consists of the sin and cos of dihedrals and bond angles, the bond lengths, root-based atom information, and force-field information. The key characteristic of ProFormX is the novel multiplexed sampling protocol (MSP) which adds randomness to the input and adapts the training loss accordingly. To that end, the noise added per feature is standard normal scaled to the standard minima and maxima of that feature. For example, for a dihedral angle, the minimum and maximum are -3.14 to 3.14 ($-\pi$ to π).

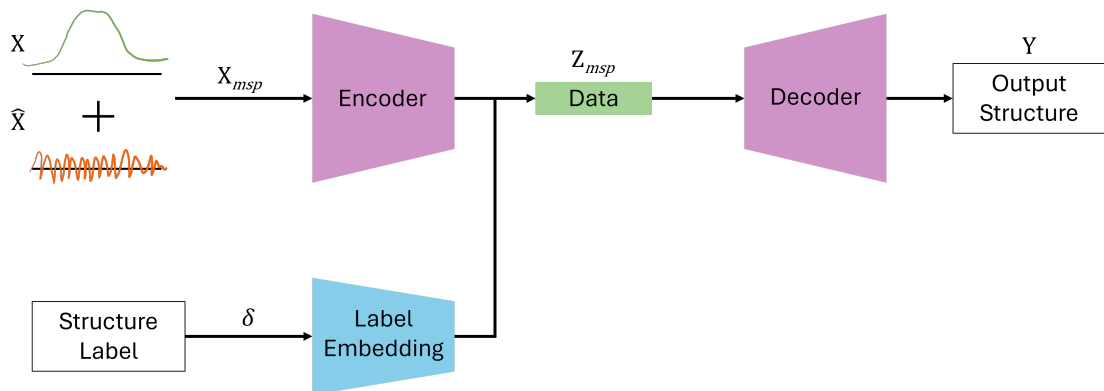


Figure 6.2: Overview of ProFormX architecture. The input X is real data and $\hat{X}$ is noise. The structure label consists of information on structure type and forcefield. This information can be used to generate desired structure types during synthesis.

6.3.1 Problem Formulation

The inputs to ProFormX are real BAT coordinates (X), noise ($\hat{X}$), and the one-hot encoded force-field information (δ). Let the real data be X , comprised of BAT features (bond length, sin and cos of bond angle, sin and cos of torsions), and the corresponding noise

be $\hat{X}$ in a similar format. The proportion of noise added is α , which is chosen randomly between 0-1. The subsequent MSP feature is X_{msp} . The encoded feature vector in the latent space of the real data and multiplexed data are Z , and Z_{msp} , respectively. The desired output is Y (the same format as X , subsequently converted to Cartesian space). Given then encoder ($\mathcal{E}$) and decoder ($\mathcal{D}$):

$$Z_{msp} = \mathcal{E}([X_{msp}, \delta]); Y = \mathcal{D}(Z_{msp}) \quad (6.1)$$

Input Encoding

The input encoding consists of the real data, noise, and force-field information. To reflect the force-field condition on the input, we have one-hot encoded it as label information as follows: 1Z0Q+FF14SB: Class 0, 2NAO+FF14SB: Class 1, 1Z0Q+IDPS: Class 2, 2NAO+IDPS: Class 3. During synthesis, this label can be given as an input to generate a structure of that desired label.

Sampling in ProFormX

The proposed multiplexed sampling protocol (MSP) explicitly adds disorder to the deterministic input to make it semi-deterministic. The extent of noise added to each feature is randomly chosen and also determines the loss function. Given the nature of our data, IDP, the dihedral rotations move much more than the bond length and angle, which have a small deviation. Thus, the α is randomly chosen separately for each feature, allowing us further control over the sampling. The α for dihedrals are sampled randomly between 0-1, whereas based on the standard deviation of the other features, their α is chosen between

0-0.1. If the inputs X and $\hat{X}$ are multiplexed with a factor α , then the semi-deterministic input becomes $X_{msp} = \alpha \times \hat{X} + (1 - \alpha) \times X$. For instance, if no noise is added to the input, then ideally the generative model should do a perfect reconstruction of the input data. This is ensured with a reconstruction loss such as L1 loss. On the other hand, if the input was all noise (like a GAN), then the output can be novel, and the extent of its realism is assessed by a similarity loss function such as KL divergence loss. In our case, since the weight of noise varies at every epoch, the loss adapts accordingly. Furthermore, the synthesis of novel conformations is dependent upon the sampling of the latent space. Therefore, we would like the latent space to be regularized, wherein a slight perturbation of an input feature would reflect accordingly in the synthesized protein. Thus, based on the extent of noise added to the input, we compute another KL divergence loss of the latent space features. If the average proportion of noise added to input is α , then the loss function is:

$$\mathcal{L} = (1 - \alpha) \times |Y - X| + \alpha \times \mathbf{D}_{KL}(\mathcal{Q}(z_{msp})||p(z_r)) + \mathbf{D}_{KL}(p(x)||p(y)) \quad (6.2)$$

where $\mathcal{Q}$ is the sampling function that samples from the MSP latent distribution, $\mathbf{D}_{KL}(\cdot)$ is KL divergence loss, $|\cdot|$ is L1 loss.

Synthesis of structures

During synthesis, we provide real or noisy input to ProFormX, based on the type of output we want it to generate. We also input the label of the desired forcefield (14SB/IDPS) and structure type (1Z0q/2NAO). Owing to the nature of sampling, ProFormX has a key capability to generate structures in one of three modes: reconstruction,

perturbation-based/controlled synthesis, and uncontrolled synthesis. The three modes are described in detail below:

Reconstruction: For this mode, the noise added to the input data is zero. This mode can typically be used to validate the correctness of ProFormX. In other words, it can check how well the model has trained. Since there is no noise, given the label and input BAT coordinates, the model should generate a nearly identical structure.

Perturbation-based Synthesis: This mode uses multiplexed input to synthesize structures. Thus, it does not have fully random noise input like a GAN would, but its input is constrained to a small region around the actual real data point. In other words, it is as if the real data point was perturbed slightly, and the new structures are generated from the impact of this perturbation in the latent space. Thus, this is a controlled synthesis mode.

Uncontrolled Synthesis: This mode uses a completely noisy input with no real information and relies on the distributions learned by the model to synthesize novel structures. Thus, this mode does not have any localized properties being translated, but a fully random combination of dihedral angles being inputted to obtain new conformations.

Model Architecture and Training Details

Given N ($=318$) atoms in a protein structure, the concatenation of dihedrals, bond angles, and bond lengths leads to an input size of 3129. The encoded latent space size is 32, which is concatenated with a one-hot label encoding of 4×1 (corresponding to force field and structure type). The architecture of ProFormX consists of 8 fully connected layers, with the number of channels varying from 3129 to 32, reducing by a factor of 2 per layer. We add a dropout of 0.1 after each layer. The activation function used is tanh, since the

features such as sin and cos lie between -1 to 1. The features are normalized after each layer. The encoded feature vector is 32+4 (label), which is given as input to the decoder. The decoder is the inverse architecture of the encoder, going from 36 to 3129 in factors of 2. The network is trained for 10000 iterations, with a batch size of 2048, learning rate 0.0001, and Adam optimizer.

6.4 Results

6.4.1 Model Validation

To validate if the model is learning correctly, we plot the backbone dihedral angle distributions between real and reconstructed structures. For quantitative analysis of the same, we compute the RMSD between the real and reconstructed structures. The angle plots are shown in Figures 6.3 and 6.4, whereas the RMSD values are reported in Table 6.1. We note that ProFormX can faithfully reconstruct the input structures, suggesting the training is effective. We report RMSD results for all four classes (combinations of 1Z0Q/2NAO and 14SB/IDPS), and the plots for the lowest and highest RMSD values (Fig 6.3: lowest RMSD for 1Z0Q-14SB, Fig 6.4: highest RMSD for 2NAO IDPS).

6.4.2 Model Synthesis

For the controlled and uncontrolled synthesis, we perform an analysis of stability (energy analysis), uniqueness (structures different from input MD), and realism (biological feasibility). For the MD runs described in Table 6.1, we have two sets of results:

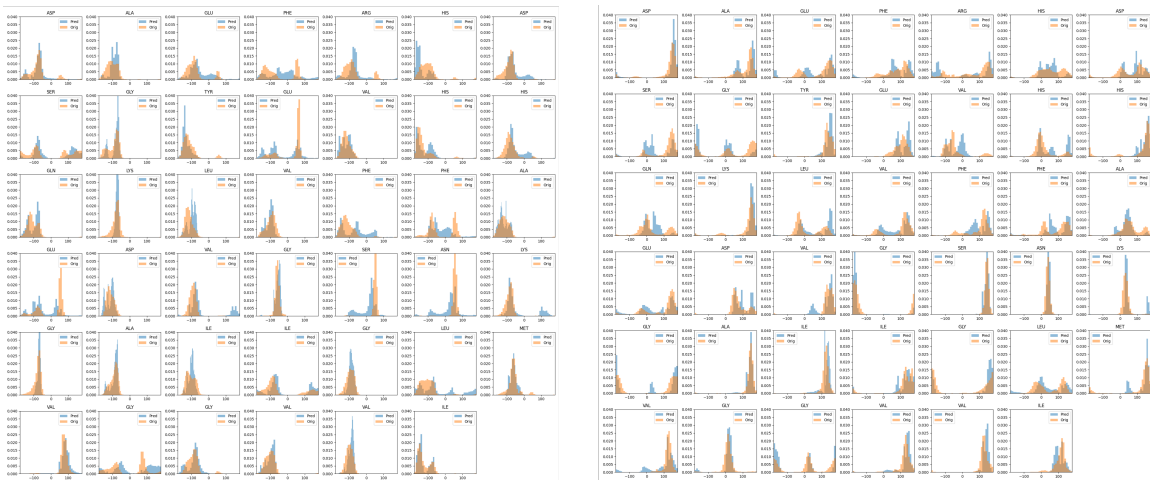


Figure 6.3: Model validation using the reconstruction mode in ProFormX. The left side consists of distributions of all sidechain torsion angles “phi”, and the right side shows the distribution of “phi”. The blue plot is predicted structures, and the orange is “real”. The results are reported for 1Z0Q in the 14SB force field.

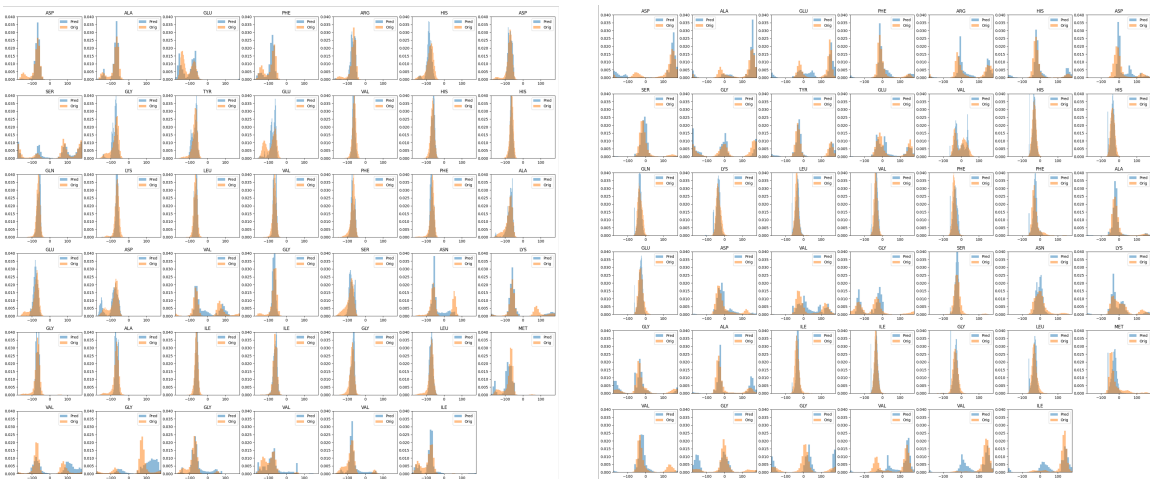


Figure 6.4: Model validation using the reconstruction mode in ProFormX. The left side consists of distributions of all sidechain torsion angles “phi”, and the right side shows the distribution of “phi”. The blue plot is predicted structures, and the orange is “real”. The results are reported for 2NAO in the IDPS force field.

perturbation-based controlled synthesis (Model C), and uncontrolled synthesis (Model U).

So for MD run 1, the results will be Model 1C and 1U.

6.4.3 Quantify the efficiency of Synthesizing Conformations

The 4 MD simulations were initiated using different A β 42 monomer structures and/or force fields, and 20000 conformations were synthesized for each type of model (Model 1-4, C/U). We subjected these conformations to refinement steps, each assessing a different property of the generated structures. Assuming the synthesis of structures is Step 1, for Step 2 (Table 6.2), any unstable structures that were below the energy cutoff (-400kcal/mol) were filtered out. This leads to allowing only stable structures in our final generated structures. For Step 3, we analyze distinctness by removing any self-repeats. We note that a very small percentage of structures were removed as self-repeats - suggesting the model does not succumb to issues such as mode collapse. The final step is to assess uniqueness by removing structures that are similar to raw MD within 2Å RMSD. The raw MD contains Y conformations each. We note that after following these four steps, the models still yield approximately 17,000 distinct conformations per run (Table 6.2). Four local hairpins: Turn A (F4–H6), Turn B (E11–H14), Turn C (S26–K28), and Turn D (V36–G38) were identified (Figure 6.1-A).

To evaluate the synthesis of novel conformations, we plotted the distribution of the Radius of Gyration (R_g) and Root Mean Square Distance (RMSD) for MD Runs 1-4 and synthetic conformations from Models 1-4 (C/U) (Figure 6.5). The broader coverage of the free energy profile in our model demonstrates its ability to sample a wider range of

Table 6.2: Refining the novel synthetic conformations of different models. Type of synthesis: C=Controlled/Perturbation-based Synthesis; U=Uncontrolled Synthesis. Step 1: Number of conformations from initial interpolation in the latent space. Step 2: Number of conformations within energy cutoff. Step 3: Number of distinct conformations (removing self-repeats). Step 4: Number of novel synthetic conformations (repeat with MD confs eliminated).

MD Simulation	Synthesis Method	Synthetic Conformations			
		Step 1	Step 2	Step 3	Step 4
MD Run 1	Model 1C	20000 (100%)	17353 (86.7%)	17045 (85.2%)	17035 (85.2%)
	Model 1U	20000 (100%)	16803 (84.0%)	16701 (83.5%)	16698 (83.5%)
MD Run 2	Model 2C	20000 (100%)	17340 (86.7%)	17228 (86.1%)	17224 (86.1%)
	Model 2U	20000 (100%)	17376 (86.9%)	17225 (86.1%)	17221 (86.1%)
MD Run 3	Model 3C	20000 (100%)	17108 (85.5%)	17063 (85.3%)	17059 (85.3%)
	Model 3U	20000 (100%)	16781 (83.9%)	16626 (83.1%)	16622 (83.1%)
MD Run 4	Model 4C	20000 (100%)	16910 (84.6%)	16828 (84.1%)	16824 (84.1%)
	Model 4U	20000 (100%)	17489 (87.4%)	17214 (86.1%)	17204 (86.0%)

energetically stable conformations. As illustrated in Figure 6.6, MD simulations struggle to sample conformations from one minimum to the other and can easily be stuck in one local minima (Figure 6.6 B). In contrast, our model generates a wide range of conformations with distinct alpha helix interactions and hydrophobic packing absent in MD simulations (Figure 6.6 C-E).

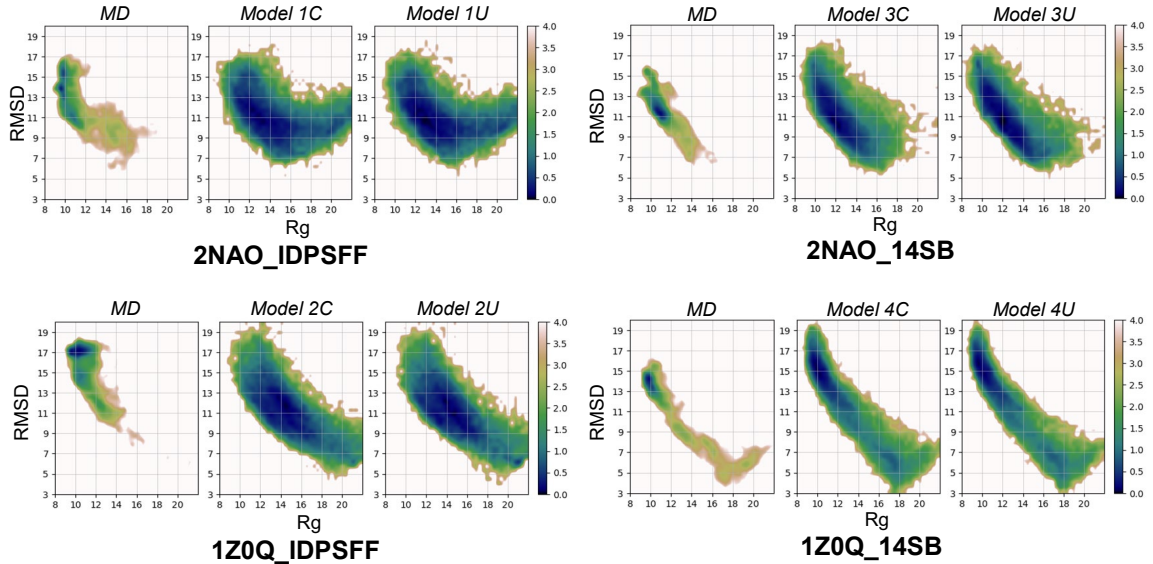


Figure 6.5: Free energy profile (color bar) of RMSD (y-axis) vs radius of gyration (Rg, x-axis) for A β 42. The units of the free-energy landscapes are kcal/mol.

6.4.4 Biologically important structures in synthetic A β 42 conformations

Understanding the conformational ensemble of A β 42 monomers is crucial for elucidating their aggregation mechanism and toxicity. Experimental studies have identified key structures prone to aggregation, focusing on toxic Turn C (S26–K28) and Turn D (V36–G38). Turn C, part of the “ β -turn- β ” motif, is frequently observed in fibrils of both

A β 40 [435] and A β 42 [486], while both Turn C and D forms one of two bends in the S-shaped fold identified in brain-derived A β 42 fibrils via cryo-EM studies [486]. The N-terminal region (D1–K16), associated with metal binding under abnormal conditions, contains Turns A and B and remains highly flexible [259]. Using novel synthetic conformations derived from Model 2U, we observed a partial hairpin structure resembling fibril-derived A β 42 [279] (Figure 6.7B and 6.7C). We also identified an intermediate state featuring both Turn C and Turn D, closely matching fibril conformations from human brain studies [486] (Figure 6.7D and 6.7E).

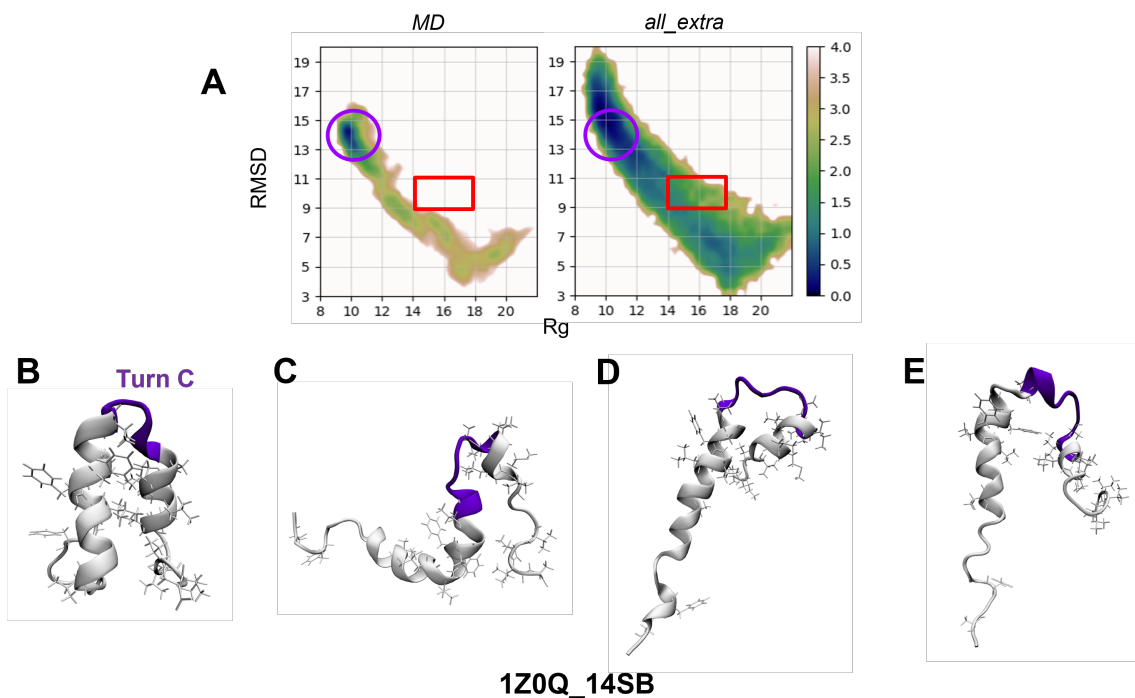


Figure 6.6: (A) Free energy profile of MD run4 and Model 4U. (B) Most dominant conformations select from the free energy profile (purple circle) (C-E)

Different orientations of F19 and F20 contribute to different types of brain-isolated fibril structures [486]. We further investigate the local side-chain rearrangement and observe similar F19 and F20 orientation as type II brain-isolated fibril structures (Figure 6.8 B). Figures 6.8C and 6.8D show intermediate states that require only minimal structural rearrangement for either type I/II brain-isolated fibril structures. Interestingly, all conformations also have toxic Turn C and D with slightly different local arrangements. Without the need for the formation of any salt bridges, the hydrophobic packing through F19 and F20 and surrounding nonpolar residues can induce the toxic Turn C formation.

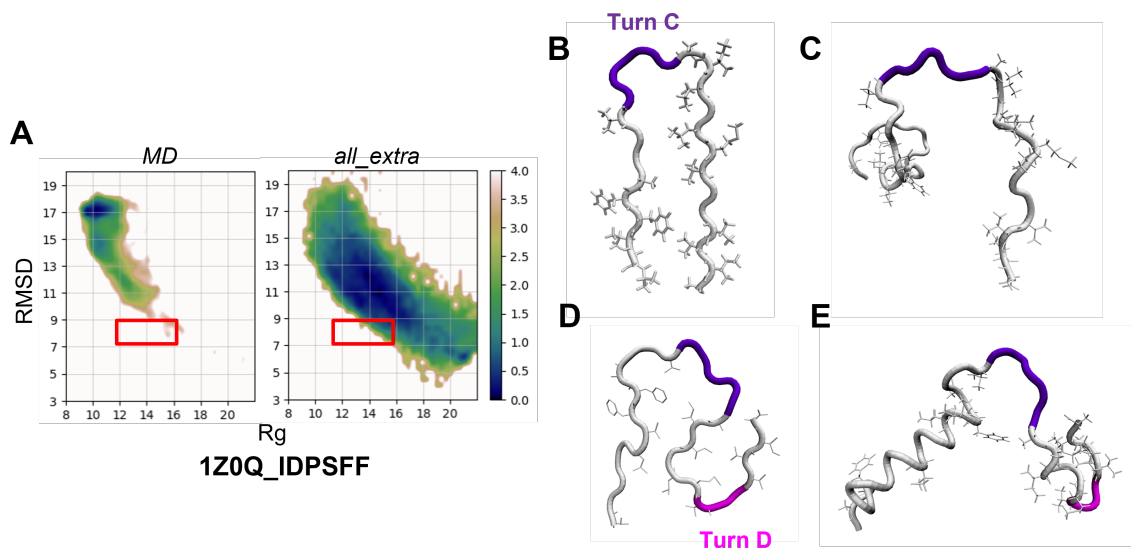


Figure 6.7: (A) Free energy profile of MD run2 and Model 4 with ffIDPSFF force field (B) Crystal Structure A β 42 fibril (PDB: 2BEG) (C) Novel Synthetic conformations from Model 4 has similar conformation as (B) found in red square (D) crystal structure of hairpin conformation that form A β 42 fibril from human brain (PDB: 7Q4M). (E) Novel Synthetic conformations from Model 4 have similar conformation as (D) found in the red square.

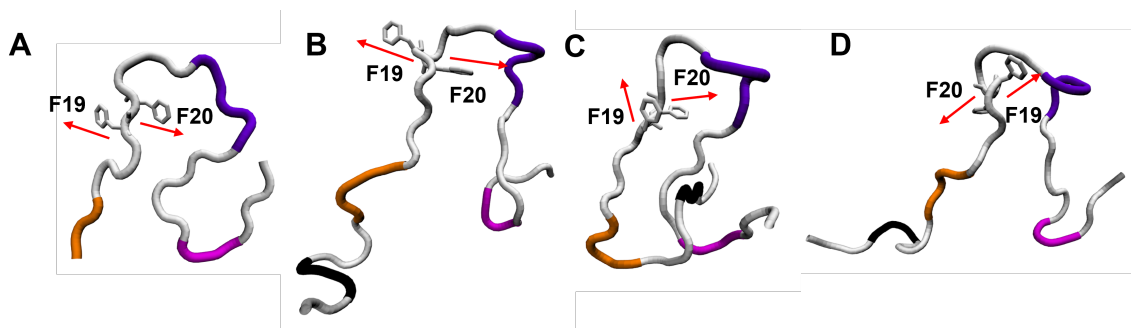


Figure 6.8: (A) Crystal structure of hair pin conformation that form A β 42 fibril from human brain (PDB: 7Q4M). (B-D) Novel synthetic structures are obtained from Model 1 with ff14IDPSFF force field. (B) Type II A β 42 fibril with F20 rotate inward and F19 rotate outward. (C-D) Intermediate states that require minimal conformational rearrangement to form type I or type II A β 42 fibril.

6.4.5 Relation between the Salt-bridge and toxic conformations

We have shown that hydrophobic core induced formation of toxic Turn C and eventually leads to fibril formation and oligomerization. Interestingly, our model also shows an alternative interaction through salt-bridge formation of D23-K28 which can also lead to the formation of Turn C. Salt bridges play an important role in providing intra-molecular attractions and conformational specificity that directly relate to A β 42 aggregation. This non-covalent electrostatic interaction forms when oppositely charged side chains come into proximity, emphasizing the need for detailed atomistic descriptions and accurate sampling of sidechain dynamics in protein studies. Although the D23-K28 salt bridge did not exist in the 1- μ s MD Run1 used in our training set, our Model 1C and 1U samples several conformations featuring D23-K28 interaction alongside various hydrophobic core arrangements (Figure 6.9). For instance, Figure 6.9A depicts weak hydrophobic packing with F19 and F20 oriented outward, making the salt bridge the primary driver for Turn C formation. In Figure 6.9B-E, the F19 is rotated inward and contributes to the hydrophobic packing,

while Figure 6.9F shows both a distinctive D23–K28 salt bridge and hydrophobic packing at the C-terminus. The result suggests that there are multiple pathways either through salt-bridge or hydrophobic interaction of forming toxic Turn C and thus, leads to potential oligomerization. These novel conformations provide diverse insights into atomistic side chain rearrangements, salt-bridge interactions, and hydrophobic packing, offering valuable perspectives for interpreting experimental findings and understanding A β 42 aggregation.

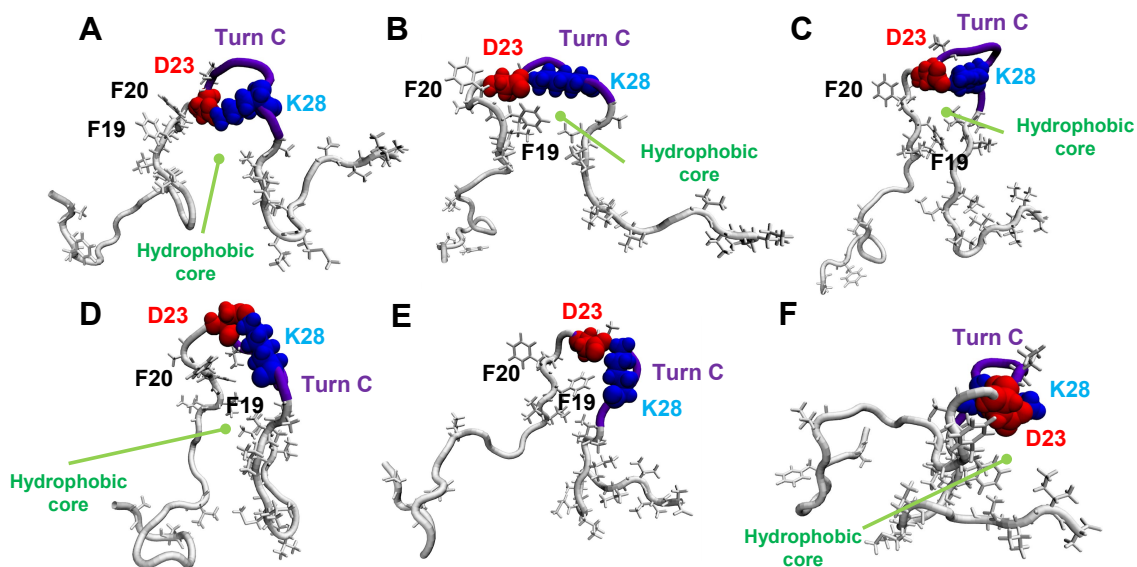


Figure 6.9: Novel synthetic conformations from Model 1 and Model 2. (A) D23–K28 salt-bridge without hydrophobic interaction. (B–C) Tightly pack hydrophobic interaction with D23–K28 salt-bridge forming inward of Turn C (D–E) Tightly pack hydrophobic interaction with D23–K28 salt-bridge forming outward of Turn C (F) Distinct hydrophobic interaction form at C-terminus with D23–K28 forming outward of Turn C.

Chapter 7

Conclusions

This thesis presented novel approaches in the fields of bioinformatics, astrobiology, and human recognition while tackling realistic dataset challenges. Through the study of four diverse applications, this work tackles critical issues such as limited and heterogeneous data, artifacts, and data imbalance—barriers that often hinder the performance and reliability of machine learning models in practical settings.

In the first application, we provided a comprehensive characterization of the visual manifestation of prognostic biomarkers of the mutation 19/20 co-gain. We extended the radiogenomic feature characterization to the tumor-infiltrated tissue - a significant biomarker. We also provided visual analysis of the reproducibility of shape, texture, and location features, conditioned upon mutation status. We proposed R2D2-GAN as a generative deep-learning approach for modeling and reproducing the visual manifestation of mutations in GBM tumor mass. We reported qualitative and quantitative evaluations of the efficacy of the novel sampling technique in R2D2-GAN. We showed that it can efficiently handle

limited-sized heterogeneous data, such as 19/20 co-gain mutation in GBM. We proposed a novel GAN evaluation metric (SIF score) that quantifies whether a GAN has learned the unapparent feature associations that are inherent and crucial to the data. Through this and other metrics, we reported results compared with SOTA models, demonstrating that diverse, high-quality, and realistic tumors are generated using R2D2-GAN. To the best of our knowledge, ours is the first work in the complete characterization of tumor macro-features in radiogenomic analysis, demonstrating the reproducibility of tumor shape, texture, and location conditioned upon mutation status, proposing a novel sampling and training protocol, and using model interpretability tools for a quantitative evaluation of GANs. We believe that with these contributions, our work can be extended to a holistic analysis of radiogenomics of other mutations and cancers, disease progression analysis, and other such crucial applications.

Furthermore, we quantified the impact of the tumor invasion patterns on the presence or absence of 19/20 co-gain. We demonstrated that the tissue impacted by tumor growth contains information regarding the mutation, and thus, contains information of prognostic value. We characterized the properties of the impacted tissue via a novel generative model (TI-GAN), that can accurately and realistically model tumor invasion using joint conditional and unconditional estimation. We reported qualitative and quantitative analyses to demonstrate the performance of the proposed approach in terms of diversity and quality. We demonstrated the realism of generated tumor invasion through a blind test carried out by board-certified radiologists that achieved high false-positive and false-negative rates. We demonstrated the generalizability of the model through cross-domain testing

and reported both qualitative and quantitative results. We performed ablation studies that evaluated the importance of learned context via tumor shape, texture (local context), and tumor location (global context) and demonstrated via qualitative and quantitative analyses that the tumor information is very useful in modeling the patterns of tumor invasion. We demonstrated that TI-GAN can act as an effective data augmentation tool for unbalanced datasets. We evaluated our model against significant and common challenges of medical datasets such as limited quantity of data, unbalanced datasets, and high tumor heterogeneity, and showed that our model is robust. We also compared our model against common medical generative models, outperforming them consistently. Thus, we established the importance of studying tumor invasion and modeled it via a novel computational model. We envision that this model will be useful in other applications beyond radiogenomic analysis such as tumor growth and disease progression analyses.

In the field of astrobiology, in order to leverage DRBs as a potential biosignature for the search for life beyond Earth, we developed SPACESeg2.0 for the automated detection of these features in the imagery of geological cross-sections. We demonstrated that SPACESeg2.0 can address the challenges of practical data such as sample cohort size, varying imaging conditions, a plethora of imaging artifacts, as well as a range of fields-of-view and relative DRB variation. We showed using extensive qualitative and quantitative analyses that SPACESeg2.0 outperformed many state-of-the-art segmentation models in detecting DRBs. Additionally, we showed from ablation studies the impact of the unsupervised ARORA in context generation and data augmentation. We also reported ablation studies to demonstrate the impact of design choices in the novel deep segmentation net-

work BASE-Net. The DRB Quantifier module of the SPACESeg2.0 computes metrics that help characterize the morphology of the ripples that comprise DRBs to gain further insight from these potential biosignatures. Thus, every step of SPACESeg2.0 is carefully designed to allow for maximum information to be analyzed, which vastly increases its potential for scalability. While SPACESeg2.0 can effectively identify and characterize DRBs in imagery, it does not guarantee the biogenicity of the detected morphological features without further grain size analysis. Thus the primary utility of SPACESeg2.0 is the rapid detection of potentially astrobiologically significant geological cross-sections that have a high potential to preserve past life and should be targeted for further human-led analysis.

For human recognition under extreme imaging conditions, we showed that the performance of advanced person re-identification backbones can be further improved by incorporating a robust input modality; one that overcomes the limitations of standard RGB images. We introduced a unified and resilient biometric modality, ORB, which combines the most identity-preserving features of both appearance- and shape-based representations while being fully compatible with the state-of-the-art recognition backbones, similar to RGB imaging. Additionally, the proposed ORB modality can be fused with RGB or other imaging modalities to significantly improve their performance. This modality is modular, quick to compute, and highly resistant to distortions such as clothing changes, occlusion, pose variations, and varying imaging distances. We presented results from four public datasets of varying complexities, and compared our performance with the state-of-the-art models that utilize RGB and other multimodal information. We also reported results on the BRIAR task, covering both Phase 1 and Phase 2 evaluation protocols. By isolating

various distortions—such as clothing changes, environmental changes, and different imaging ranges—we demonstrate that ORB either outperforms other modalities or significantly enhances the performance of any modality it is combined with. Through quantitative analysis, we showed that our proposed modality consistently delivers robust performance under challenging conditions and significantly boosts baseline performance. This approach is invaluable for real-world human recognition tasks, particularly in scenarios involving clothing changes, long-range recognition in unconstrained environments where face, body, or gait features are less effective, and when faster recognition is required.

Towards the applicability of domain-inspired algorithms in drug discovery, we showed a specific case of full-atom synthesis in intrinsically disordered protein A β 42, associated with Alzheimer’s. We propose a novel sampling technique that can be used with any generative model, and successfully tackles the dynamic nature of the A β 42 monomer. We use the sampling technique in conjunction with a hybrid autoencoder-type network (coined ProFormX). We show from extensive analysis that ProFormX can generate stable, diverse, and unique conformations. Further analysis demonstrates the biological relevance of synthesized conformations, and that ProFormX can sample from energy regions originally absent in the raw MD simulation.

The insights gained through this work not only enhance the applicability of deep learning in challenging environments but also lay the groundwork for future research. By identifying persistent challenges and proposing viable solutions, this dissertation serves as a foundation for continued efforts to improve the integration of deep learning into diverse domains, ensuring its impact is both meaningful and enduring.

Bibliography

- [1] Mars 2020 perseverance rover.
- [2] National cancer institute genomic data commons data portal.
- [3] Vasari research project.
- [4] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 683–687. IEEE, 2019.
- [5] L Achenbach, J Bailey, R Barnes, J Baross, C Bertka, P Boston, E Boyd, M Cable, I Chen, F Ciesla, et al. NASA Astrobiology strategy 2015. *NASA Astrobiology*, pages 1–236, 2015.
- [6] Obinna Agbodike, Weijin Zhang, Jenhui Chen, and Lei Wang. Face and body-shape integration model for cloth-changing person re-identification. *Image and Vision Computing*, 140:104843, 2023.
- [7] Andrey Alekseev and Anatoly Bobe. Gabornet: Gabor filters with learnable parameters in deep convolutional neural network. In *2019 International Conference on Engineering and Telecommunication (EnT)*, pages 1–4. IEEE, 2019.
- [8] Hazrat Ali, Rafiul Biswas, Farida Ali, Uzair Shah, Asma Alamgir, Osama Mousa, and Zubair Shah. The role of generative adversarial networks in brain mri: a scoping review. *Insights into Imaging*, 13(1):1–15, 2022.
- [9] Sadiq Alinsaif, Jochen Lang, Alzheimer’s Disease Neuroimaging Initiative, et al. 3d shearlet-based descriptors combined with deep features for the classification of alzheimer’s disease based on mri data. *Computers in Biology and Medicine*, 138:104879, 2021.
- [10] Mohammed AlQuraishi. Machine learning in protein structure prediction. *Current opinion in chemical biology*, 65:1–8, 2021.
- [11] Ross P Anderson, Christina R Woltz, Nicholas J Tosca, Susannah M Porter, and Derek EG Briggs. Fossilisation processes and our reading of animal antiquity. *Trends in Ecology and Evolution*, 38(11):1060–1071, 2023.

- [12] Véronique Ansan, D Loizeau, N Mangold, S Le Mouélic, J Carter, F Poulet, Gilles Dromart, A Lucas, J-P Bibring, A Gendrin, et al. Stratigraphy, mineralogy, and origin of layered deposits inside terby crater, Mars. *Icarus*, 211(1):273–304, 2011.
- [13] Wajeeha Ansar, Muhammad Moazam Fraz, Muhammad Shahzad, Imad Gohar, Sajid Javed, and Soon Ki Jung. Two stream deep cnn-rnn attentive pooling architecture for video-based person re-identification. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 23rd Iberoamerican Congress, CIARP 2018, Madrid, Spain, November 19-22, 2018, Proc. 23*, pages 654–661. Springer, 2019.
- [14] Ali Ari and Davut Hanbay. Deep learning based brain tumor classification and detection system. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(5):2275–2286, 2018.
- [15] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, 2017.
- [16] Shehreen Azad and Yogesh Singh Rawat. Activity-biometrics: Person identification from daily activities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 287–296, 2024.
- [17] Slawomir Bak, Etienne Corvee, Francois Bremond, and Monique Thonnat. Person re-identification using spatial covariance regions of human body parts. In *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 435–440. IEEE, 2010.
- [18] Evgeny E Baraboshkin, Leyla S Ismailova, Denis M Orlov, Elena A Zhukovskaya, Georgy A Kalmykov, Oleg V Khotylev, Evgeny Yu Baraboshkin, and Dmitry A Koroteev. Deep convolutions for in-depth automated rock typing. *Computers & Geosciences*, 135:104330, 2020.
- [19] Shane Barratt and Rishi Sharma. A note on the inception score. *arXiv preprint arXiv:1801.01973*, 2018.
- [20] Lennart Barthel, Martin Hadamitzky, Philipp Dammann, Manfred Schedlowski, Ulrich Sure, Basant Kumar Thakur, and Susann Hetze. Glioma: molecular signature and crossroads with tumor microenvironment. *Cancer and Metastasis Reviews*, pages 1–23, 2022.
- [21] Stefan Bauer, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes. A survey of mri-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology*, 58(13):R97, 2013.
- [22] Elaine L Bearer, John S Lowengrub, Hermann B Frieboes, Yao-Li Chuang, Fang Jin, Steven M Wise, Mauro Ferrari, David B Agus, and Vittorio Cristini. Multiparameter computational modeling of tumor invasion. *Cancer research*, 69(10):4493–4501, 2009.

- [23] Mario Beauchemin, Keith PB Thomson, and G Edwards. On the Hausdorff distance used for the evaluation of segmentation results. *Canadian Journal of Remote Sensing*, 24(1):3–8, 1998.
- [24] Niha Beig, Jay Patel, Prateek Prasanna, Sasan Partovi, Vinay Varadan, Anant Madabhushi, and Pallavi Tiwari. Radiogenomic analysis of hypoxia pathway reveals computerized mri descriptors predictive of overall survival in glioblastoma. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 101341U. International Society for Optics and Photonics, 2017.
- [25] J F Bell III, J N Maki, G L Mehall, M A Ravine, M A Caplinger, Z J Bailey, S Brylow, J A Schaffner, K M Kinch, M B Madsen, et al. The Mars 2020 Perseverance Rover Mast Camera Zoom (Mastcam-Z) multispectral, stereoscopic imaging investigation. *Space Science Reviews*, 217(24), 2021.
- [26] Sean Benson and Regina Beets-Tan. Gan-based anomaly detection in multi-modal mri images. *bioRxiv*, 2020.
- [27] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984.
- [28] Xiaojun Bi and Hao Wang. Appearance-pose joint coordinates information collaboration model for clothes-changing person re-identification. *Expert Systems with Applications*, 241:122473, 2024.
- [29] Abhishek Bihani, Hugh Daigle, Javier E Santos, Christopher Landry, Maša Prodanović, and Kitty Milliken. MudrockNet: Semantic segmentation of mudrock sem images through deep learning. *Computers & Geosciences*, 158:104952, 2022.
- [30] Maryam Moradi Binabaj, Afsane Bahrami, Soodabeh ShahidSales, Marjan Joodi, Mona Joudi Mashhad, Seyed Mahdi Hassanian, Kazem Anvari, and Amir Avan. The prognostic value of mgmt promoter methylation in glioblastoma: a meta-analysis of clinical trials. *Journal of cellular physiology*, 233(1):378–386, 2018.
- [31] Alceu Bissoto, Eduardo Valle, and Sandra Avila. Gan-based data augmentation and anonymization for skin-lesion analysis: A critical review. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1847–1856, 2021.
- [32] Zuhir Bodalal, Stefano Trebeschi, Thi Dan Linh Nguyen-Kim, Winnie Schats, and Regina Beets-Tan. Radiogenomics: bridging imaging and genomics. *Abdominal radiology*, 44(6):1960–1984, 2019.
- [33] Rosalin Bonetta and Gianluca Valentino. Machine learning techniques for protein function prediction. *Proteins: Structure, Function, and Bioinformatics*, 88(3):397–413, 2020.
- [34] Ali Borji. Pros and cons of gan evaluation measures. *Computer vision and image understanding*, 179:41–65, 2019.

- [35] Katarzyna Borys, Yasmin Alyssa Schmitt, Meike Nauta, Christin Seifert, Nicole Krämer, Christoph M Friedrich, and Felix Nensa. Explainable ai in medical imaging: An overview for clinical practitioners—beyond saliency-based xai approaches. *European journal of radiology*, 162:110786, 2023.
- [36] Tanja Bosak, Kelsey R Moore, Jian Gong, and John P Grotzinger. Searching for biosignatures in sedimentary rocks from early Earth and Mars. *Nature Reviews Earth & Environment*, 2(7):490–506, 2021.
- [37] David J Bottjer, James W Hagadorn, Stephen Q Dornbos, et al. The cambrian substrate revolution. *GSA today*, 10(9):1–7, 2000.
- [38] Christopher Bowles, Liang Chen, Ricardo Guerrero, Paul Bentley, Roger Gunn, Alexander Hammers, David Alexander Dickie, Maria Valdés Hernández, Joanna Wardlaw, and Daniel Rueckert. Gan augmentation: Augmenting training data using generative adversarial networks. *arXiv preprint arXiv:1810.10863*, 2018.
- [39] Andrew Brock. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.
- [40] Adrian Brown, Malcolm Walter, and Thomas Cudahy. Short-wave infrared reflectance investigation of sites of paleobiological interest: applications for mars exploration. *Astrobiology*, 4(3):359–376, 2004.
- [41] Mateusz Buda, Ashirbani Saha, and Maciej A Mazurowski. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in biology and medicine*, 109:218–225, 2019.
- [42] Sherry L Cady, Jack D Farmer, John P Grotzinger, J William Schopf, and Andrew Steele. Morphological biosignatures and the search for life on Mars. *Astrobiology*, 3(2):351–368, 2003.
- [43] Mengqiu Cao, Shiteng Suo, Xiao Zhang, Xiaoqing Wang, Jianrong Xu, Wei Yang, and Yan Zhou. Qualitative and quantitative mri analysis in idh1 genotype prediction of lower-grade gliomas: A machine learning approach. *BioMed research international*, 2021, 2021.
- [44] Van Thang Cao, Tae-Young Jung, Shin Jung, Shu-Guang Jin, Kyung-Sub Moon, In-Young Kim, Sam-Suk Kang, Chang-Soo Park, Kyung-Hwa Lee, and Hong-Jae Chae. The correlation and prognostic significance of mgmt promoter methylation and mgmt protein in glioblastomas. *Neurosurgery*, 65(5):866–875, 2009.
- [45] Leonardo Carrer and Lorenzo Bruzzone. A novel approach to the detection and imaging of candidate Martian subglacial water bodies by radar sounder data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021.
- [46] JA1 Carrillo, A Lai, PL Nghiemphu, HJ Kim, HS Phillips, S Kharbanda, P Mof-takhar, S Lalaezari, W Yong, BM Ellingson, et al. Relationship between tumor enhancement, edema, idh1 mutational status, mgmt promoter methylation, and survival in glioblastoma. *American Journal of Neuroradiology*, 33(7):1349–1355, 2012.

- [47] Rita Casadio, Pier Luigi Martelli, and Castrense Savojardo. Machine learning solutions for predicting protein-protein interactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(6):e1618, 2022.
- [48] Ahmad Chaddad, Saima Rathore, Mingli Zhang, Christian Desrosiers, and Tamim Niazi. Deep radiomic features from mri scans predict survival outcome of recurrent glioblastoma. *arXiv preprint arXiv:1911.06687*, 2019.
- [49] Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuvver M Rao, et al. Interpretability of deep learning models: A survey of results. In *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smart-world/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)*, pages 1–6. IEEE, 2017.
- [50] Stanley H Chan. Tilt-then-blur or blur-then-tilt? clarifying the atmospheric turbulence model. *IEEE Signal Processing Letters*, 29:1833–1837, 2022.
- [51] Abel Chandra, Laura Tünnermann, Tommy Löfstedt, and Regina Gratz. Transformer-based deep learning for predicting protein properties in the life sciences. *Elife*, 12:e82819, 2023.
- [52] Chia-En Chang and Michael K Gilson. Tork: Conformational analysis method for molecules and complexes. *Journal of computational chemistry*, 24(16):1987–1998, 2003.
- [53] P Chang, J Grinband, BD Weinberg, M Bardis, M Khy, G Cadena, M-Y Su, S Cha, CG Filippi, D Bota, and P Baldi. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *American Journal of Neuroradiology*, 39(7):1201–1207, 2018.
- [54] Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 44(7):3467–3478, 2021.
- [55] Chunhui Chen, Joon Huang Chuah, Raza Ali, and Yizhou Wang. Retinal vessel segmentation using deep learning: a review. *IEEE Access*, 9:111985–112004, 2021.
- [56] Guangyi Chen, Yongming Rao, Jiwen Lu, and Jie Zhou. Temporal coherence or temporal motion: Which is more critical for video-based person re-identification? In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 660–676. Springer, 2020.
- [57] Jiaxing Chen, Xinyang Jiang, Fudong Wang, Jun Zhang, Feng Zheng, Xing Sun, and Wei-Shi Zheng. Learning 3d shape feature for texture-insensitive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8146–8155, 2021.

- [58] Junhua Chen, Inigo Bermejo, Andre Dekker, and Leonard Wee. Generative models improve radiomics performance in different tasks and different datasets: An experimental study. *arXiv preprint arXiv:2109.02252*, 2021.
- [59] Xianjie Chen, Roozbeh Mottaghi, Xiaobai Liu, Sanja Fidler, Raquel Urtasun, and Alan Yuille. Detect what you can: Detecting and representing objects using holistic models and body parts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1971–1978, 2014.
- [60] Xin Chen, Xizhao Luo, Jian Weng, Weiqi Luo, Huiting Li, and Qi Tian. Multi-view gait image generation for cross-view gait recognition. *IEEE Trans. on Image Processing*, 30:3041–3055, 2021.
- [61] Xu Chen, Bryan M Williams, Srinivasa R Vallabhaneni, Gabriela Czanner, Rachel Williams, and Yalin Zheng. Learning active contour models for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11632–11640, 2019.
- [62] Yehansen Chen, Lin Wan, Zhihang Li, Qianyan Jing, and Zongyuan Sun. Neural feature search for rgb-infrared person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 587–597, 2021.
- [63] Bowen Cheng, Anwesa Choudhuri, Ishan Misra, Alexander Kirillov, Rohit Girdhar, and Alexander G Schwing. Mask2former for video instance segmentation. *arXiv preprint arXiv:2112.10764*, 2021.
- [64] Hong-Bin Cheng, Wu Yue, Chen Xie, Ru-You Zhang, Shao-Shan Hu, and Zhi Wang. Idh1 mutation is associated with improved overall survival in patients with glioblastoma: a meta-analysis. *Tumor Biology*, 34(6):3555–3559, 2013.
- [65] Gloria C Chiang, Ilhami Kovanlikaya, Changho Choi, Rohan Ramakrishna, Rajiv Magge, and Dikoma C Shungu. Magnetic resonance spectroscopy, positron emission tomography and radiogenomics—relevance to glioma. *Frontiers in Neurology*, 9:33, 2018.
- [66] Nicholas Chimitt, Xingguang Zhang, Zhiyuan Mao, and Stanley H Chan. Real-time dense field phase-to-space simulation of imaging through atmospheric turbulence. *IEEE Transactions on Computational Imaging*, 8:1159–1169, 2022.
- [67] Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563, 2021.
- [68] Maria JM Chuquicusma, Sarfaraz Hussein, Jeremy Burt, and Ulas Bagci. How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 240–244. IEEE, 2018.

- [69] Marco Domenico Cirillo, David Abramian, and Anders Eklund. Vox2vox: 3d-gan for brain tumour segmentation. *arXiv preprint arXiv:2003.13653*, 2020.
- [70] Fusun Citak-Er, Zeynep Firat, Ilhami Kovanlikaya, Ugur Ture, and Esin Ozturk-Isik. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3t. *Computers in Biology and Medicine*, 99:154–160, 2018.
- [71] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, and L Tarbox. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.
- [72] Elizabeth K Cole, Frank Ong, Shreyas S Vasanaawala, and John M Pauly. Fast unsupervised mri reconstruction without fully-sampled ground truth data using generative adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3988–3997, 2021.
- [73] Dov Corenblit, Olivier Decaux, Sébastien Delmotte, Jean-Pierre Toumazet, Florent Arrignon, Marie-Françoise André, José Darrozes, Neil S Davies, Frédéric Julien, Thierry Otto, et al. Signatures of life detected in images of rocks using neural network analysis demonstrate new potential for searching for biosignatures on the surface of mars. *Astrobiology*, 2023.
- [74] David Cornett, Joel Brogan, Nell Barber, Deniz Aykac, Seth Baird, Nicholas Burchfield, Carl Dukes, Andrew Duncan, Regina Ferrell, Jim Goddard, et al. Expanding accurate person recognition to new altitudes and ranges: The briar dataset. In *Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023.
- [75] Gabriela Csurka, Diane Larlus, Florent Perronnin, and France Meylan. What is a good evaluation measure for semantic segmentation? In *Bmvc*, volume 27, pages 10–5244. Citeseer, 2013.
- [76] Shaoguo Cui, Mingjun Wei, Chang Liu, and Jingfeng Jiang. Gan-segnet: A deep generative adversarial segmentation network for brain tumor semantic segmentation. *Intl. Journal of Imaging Systems and Technology*, 2021.
- [77] Zhenyu Cui, Jiahuan Zhou, Yuxin Peng, Shiliang Zhang, and Yaowei Wang. Dcr-reid: Deep component reconstruction for cloth-changing person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(8):4415–4428, 2023.
- [78] Salman UH Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Cukur. Image synthesis in multi-contrast mri with conditional generative adversarial networks. *IEEE Transactions on Medical Imaging*, 38(10):2375–2388, 2019.
- [79] Neil S Davies, Alexander G Liu, Martin R Gibling, and Randall F Miller. Resolving miss conceptions and misconceptions: a geological approach to sedimentary surface textures generated by microbial and abiotic processes. *Earth-Science Reviews*, 154:210–246, 2016.

- [80] Ruggero De Maria Marchiano, Gabriele Di Sante, Geny Piro, Carmine Carbone, Giampaolo Tortora, Luca Boldrini, Antonella Pietragalla, Gennaro Daniele, Maria Tredicine, Alfredo Cesario, and V Valentini. Translational research in the era of precision medicine: Where we are and where we will go. *Journal of Personalized Medicine*, 11(3):216, 2021.
- [81] S Deepak and PM Ameer. MSG-GAN based synthesis of brain MRI with meningioma for data augmentation. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–6, 2020.
- [82] David J Des Marais, Josph A Nuth III, Louis J Allamandola, Alan P Boos, Jack D Farmer, Tori M Hoehler, Bruce M Jakosky, Victoria S Meadows, Andrew Pohorille, Bruce Runnegar, and Alfred M Spormann. The NASA Astrobiology roadmap. *Astrobiology*, 8(4), 2008.
- [83] Haibin Di, Leigh Truelove, Cen Li, and Aria Abubakar. Accelerating seismic fault and stratigraphy interpretation with deep CNNs: A case study of the taranaki basin, new zealand. *The Leading Edge*, 39(10):727–733, 2020.
- [84] Foivos I Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, 2020.
- [85] Yongkang Ding, Yinghao Wu, Anqi Wang, Tiantian Gong, and Liyan Zhang. Disentangled body features for clothing change person re-identification. *Multimedia Tools and Applications*, pages 1–22, 2024.
- [86] Tuan Quang Dinh, Yunyang Xiong, Zhichun Huang, Tien Vo, Akshay Mishra, Won Hwa Kim, Sathya Ravi, and Vikas Singh. Performing group difference testing on graph structured data from gans: Analysis and applications in neuroimaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [87] Tara Djokic, Martin J Van Kranendonk, Kathleen A Campbell, Malcom R Walter, and Colin R Ward. Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. *Nature Communications*, 8(15263), 2017.
- [88] Pia Domschke, Dumitru Trucu, Alf Gerisch, and Mark AJ Chaplain. Mathematical modelling of cancer invasion: implications of cell adhesion variability for tumour infiltrative growth patterns. *Journal of Theoretical Biology*, 361:41–60, 2014.
- [89] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.
- [90] Zolnamar Dorjsembe, Sodtavilan Odonchimed, and Furen Xiao. Three-dimensional medical image synthesis with denoising diffusion probabilistic models. In *Medical Imaging with Deep Learning*, 2022.

- [91] Zolnamar Dorjsembe, Hsing-Kuo Pao, Sodtavilan Odonchimed, and Furen Xiao. Conditional diffusion models for semantic 3d brain mri synthesis. *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [92] Sylvain Douté and Cheng Jiang. Small-scale topographical characterization of the Martian surface with in-orbit imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1):447–460, 2019.
- [93] Mary L Droser, Scott Evans, Lidya G Tarhan, Rachel L Surprenant, Ian V Hughes, Emmy B Hughes, and James G Gehling. What happens between depositional events, stays between depositional events: The significance of organic mat surfaces in the capture of Ediacara communities and the sedimentary rocks that preserve them. *Frontiers in Earth Science*, page 198, 2022.
- [94] Mary L Droser, James G Gehling, Lidya G Tarhan, Scott D Evans, Christine MS Hall, Ian V Hughes, Emily B Hughes, Mary E Dzaugis, Matthew P Dzaugis, Peter W Dzaugis, et al. Piecing together the puzzle of the ediacara biota: excavation and reconstruction at the ediacara national heritage site nilpena (south australia). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 513:132–145, 2019.
- [95] Mary L Droser, Lidya G Tarhan, and James G Gehling. The rise of animals in a changing environment: Global ecological innovation in the late ediacaran. *Annual Review of Earth and Planetary Sciences*, 45:593–617, 2017.
- [96] Yuning Du, Haizhou Ai, and Shihong Lao. Evaluation of color spaces for person re-identification. In *Proc. of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 1371–1374. IEEE, 2012.
- [97] Murat Dundar, Bethany L Ehlmann, and Ellen K Leask. Machine-learning-driven new geologic discoveries at mars rover landing sites: Jezero and ne syrtis. *arXiv preprint arXiv:1909.02387*, 2019.
- [98] El-Sayed A El-Dahshan, Heba M Mohsen, Kenneth Revett, and Abdel-Badeeh M Salem. Computer-aided diagnosis of human brain tumor through mri: A survey and a new algorithm. *Expert Systems with Applications*, 41(11):5526–5545, 2014.
- [99] Ahmed Elazab, Changmiao Wang, Syed Jamal Safdar Gardezi, Hongmin Bai, Qing-mao Hu, Tianfu Wang, Chunqi Chang, and Baiying Lei. Gp-gan: Brain tumor growth prediction using stacked 3d generative adversarial networks from longitudinal mr images. *Neural Networks*, 132:321–332, 2020.
- [100] Benjamin M Ellingson. Radiogenomics and imaging phenotypes in glioblastoma: novel observations and correlation with molecular characteristics. *Current neurology and neuroscience reports*, 15(1):506, 2015.
- [101] Anette Eltner, Andreas Kaiser, Carlos Castillo, Gilles Rock, Fabian Neugirg, and Antonio Abellán. Image-based surface reconstruction in geomorphometry—merits, limits and developments. *Earth Surface Dynamics*, 4(2):359–389, 2016.

- [102] Ebrahim Emami, Touqeer Ahmad, George Bebis, Ara Nefian, and Terry Fong. Crater detection using unsupervised algorithms and convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5373–5383, 2019.
- [103] Aleksandr Emchinov. A deep learning approach to glioblastoma radiogenomic classification using brain mri. In *International MICCAI Brainlesion Workshop*, pages 345–356. Springer, 2021.
- [104] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.
- [105] Oriol Falivene, Neal C Auchter, Rafael Pires de Lima, Luuk Kleipool, John G Solum, Pedram Zarian, Rachel W Clark, and Irene Espejo. Lithofacies identification in cores using deep learning segmentation and the role of geoscientists: Turbidite deposits (gulf of mexico and north sea). *AAPG Bulletin*, 106(7):1357–1372, 2022.
- [106] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. Gaitpart: Temporal part-based model for gait recognition. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [107] Rui Fan, Mohammad Junaid Bocus, Yilong Zhu, Jianhao Jiao, Li Wang, Fulong Ma, Shanshan Cheng, and Ming Liu. Road crack detection using deep convolutional neural network and adaptive thresholding. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 474–479. IEEE, 2019.
- [108] Yuheng Fan, Hanxi Liao, Shiqi Huang, Yimin Luo, Huazhu Fu, and Haikun Qi. A survey of emerging applications of diffusion probabilistic models in mri. *Meta-Radiology*, page 100082, 2024.
- [109] Giuseppe Nicolò Fanelli, Dario Grassini, Valerio Ortenzi, Francesco Pasqualetti, Nicola Montemurro, Paolo Perrini, Antonio Giuseppe Naccarato, and Cristian Scatena. Decipher the glioblastoma microenvironment: the first milestone for new groundbreaking therapeutic strategies. *Genes*, 12(3):445, 2021.
- [110] Jiangxiong Fang, Huaxiang Liu, Liting Zhang, Jun Liu, and Hesheng Liu. Region-edge-based active contours driven by hybrid and local fuzzy region-based energy for image segmentation. *Information Sciences*, 546:397–419, 2021.
- [111] Kenneth A Farley, Kenneth H Williford, Kathryn M Stack, Rohit Bhartia, Al Chen, Manuel de la Torre, Kevin Hand, Yulia Goreva, Christopher DK Herd, Ricardo Hueso, et al. Mars 2020 mission overview. *Space Science Reviews*, 216(8):1–41, 2020.
- [112] Anahita Fathi Kazerooni, Spyridon Bakas, Hamidreza Saligheh Rad, and Christos Davatzikos. Imaging signatures of glioblastoma molecular characteristics: A radiogenomics review. *Journal of Magnetic Resonance Imaging*, 2019.

- [113] Anahita Fathi Kazerooni, Spyridon Bakas, Hamidreza Saligheh Rad, and Christos Davatzikos. Imaging signatures of glioblastoma molecular characteristics: A radiogenomics review. *Journal of Magnetic Resonance Imaging*, 2019.
- [114] Matteo Ferro, Ottavio de Cobelli, Mihai Dorin Vartolomei, Giuseppe Lucarelli, Felice Crocetto, Biagio Barone, Alessandro Sciarra, Francesco Del Giudice, Matteo Muto, Martina Maggi, and G Carrieri. Prostate cancer radiogenomics—from imaging to molecular characterization. *International Journal of Molecular Sciences*, 22(18):9971, 2021.
- [115] Mehdi Foroozandeh and Anders Eklund. Synthesizing brain tumor images and annotations by combining progressive growing gan and spade. *arXiv preprint arXiv:2009.05946*, 2020.
- [116] Greg J Freeman, Alan C Bovik, and John W Holt. Automated detection of near surface Martian ice layers in orbital radar data. In *2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI)*, pages 117–120. IEEE, 2010.
- [117] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.
- [118] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.
- [119] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 289–293. IEEE, 2018.
- [120] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 289–293. IEEE, 2018.
- [121] Marc D Fries, Carina Lee, Rohit Bhartia, Joseph Razzell Hollis, Luther W Beegle, Kyle Uckert, Trevor G Graff, William Abbey, Zachary Bailey, Eve L Berger, et al. The SHERLOC calibration target on the Mars 2020 Perseverance rover: Design, operations, outreach, and future human exploration functions. *Space Science Reviews*, 218(6):1–33, 2022.
- [122] Huazhu Fu, Yanwu Xu, Stephen Lin, Damon Wing Kee Wong, and Jiang Liu. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 132–139. Springer, 2016.

- [123] Shohei Fujita, Akifumi Hagiwara, Yujiro Otsuka, Masaaki Hori, Naoyuki Takei, Ken-Pin Hwang, Ryusuke Irie, Christina Andica, Koji Kamagata, Toshiaki Akashi, et al. Deep learning approach for generating mra images from 3d quantitative synthetic mri without additional scans. *Investigative radiology*, 55(4):249–256, 2020.
- [124] Wenhao Gao, Sai Pooja Mahajan, Jeremias Sulam, and Jeffrey J Gray. Deep learning in protein structural modeling and design. *Patterns*, 1(9), 2020.
- [125] Yuan Gao, Yingchao Liu, Yuanyuan Wang, Zhifeng Shi, and Jinhua Yu. A universal intensity standardization method based on a many-to-one weak-paired cycle generative adversarial network for magnetic resonance images. *IEEE Transactions on Medical Imaging*, 38(9):2059–2069, 2019.
- [126] Roberto García-Figueiras, Sandra Baleato-González, Anwar R Padhani, Antonio Luna-Alcalá, Ana Marhuenda, Joan C Vilanova, Iria Osorio-Vázquez, Anxo Martínez-de Alegría, and Antonio Gómez-Caamaño. Advanced imaging techniques in evaluation of colorectal cancer. *Radiographics*, 38(3):740–765, 2018.
- [127] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.
- [128] Nathan Gaw, Andrea Hawkins-Daarud, Leland S Hu, Hyunsoo Yoon, Lujia Wang, Yanzhe Xu, Pamela R Jackson, Kyle W Singleton, Leslie C Baxter, Jennifer Eschbacher, et al. Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric mri. *Scientific reports*, 9(1):10063, 2019.
- [129] J G Gehling. Environmental interpretation and a sequence stratigraphic framework for the terminal Proterozoic Ediacara Member within the Rawnsley Quartzite, South Australia. *Precambrian Research*, 100(1-3):65–95, 2000.
- [130] James G Gehling. Microbial mats in terminal Proterozoic siliciclastics: Ediacaran death masks. *PALAIOS*, 14(1):40–57, 1999.
- [131] James G Gehling and Mary L Droser. Textured organic surfaces associated with the Ediacara biota in South Australia. *Earth-Science Reviews*, 96:196–206, 2009.
- [132] James G Gehling and Mary L Droser. Ediacaran stratigraphy and the biota of the Adelaide Geosyncline, South Australia. *Episodes Journal of International Geoscience*, 35(1):236–246, 2012.
- [133] Christoph Geisenberger, Andreas Mock, Rolf Warta, Carmen Rapp, Christian Schwager, Andrey Korshunov, Ann-Katrin Nied, David Capper, Benedikt Brors, Christine Jungk, and D Jones. Molecular profiling of long-term survivors identifies a subgroup of glioblastoma characterized by chromosome 19/20 co-gain. *Acta Neuropathologica*, 130(3):419–434, 2015.

- [134] Olivier Gevaert, Lex A Mitchell, Achal S Achrol, Jiajing Xu, Sebastian Echegaray, Gary K Steinberg, Samuel H Cheshier, Sandy Napel, Greg Zaharchuk, and Sylvia K Plevritis. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology*, 273(1):168–174, 2014.
- [135] Stavroula G Giatili and Georgios S Stamatakis. A detailed numerical treatment of the boundary conditions imposed by the skull on a diffusion–reaction model of glioma tumor growth. clinical validation aspects. *Applied Mathematics and Computation*, 218(17):8779–8799, 2012.
- [136] Murray Gingras, James W Hagadorn, Adolf Seilacher, Stefan V Lalonde, Ernesto Pecoits, Daniel Petrash, and Kurt O Konhauser. Possible evolution of mobile animals in association with microbial mats. *Nature Geoscience*, 4(6):372–375, 2011.
- [137] S Rosas González, I Zemmoura, and C Tauber. Deep convolutional neural network to predict 1p19q co-deletion and idh1 mutation status from mri in low grade gliomas. In *10th International Conference on Pattern Recognition Systems (ICPRS-2019)*, pages 7–11. IET, 2019.
- [138] Sonal Gore and Jayant Jagtap. Radiogenomic analysis: 1p/19q codeletion based subtyping of low-grade glioma by analysing advanced biomedical texture descriptors. *Journal of King Saud University-Computer and Information Sciences*, 34(10):8449–8458, 2022.
- [139] Lars J Grimm and Maciej A Mazurowski. Breast cancer radiogenomics: current status and future directions. *Academic Radiology*, 27(1):39–46, 2020.
- [140] Shresth Grover, Vibhav Vineet, and Yogesh Rawat. Revealing the unseen: Benchmarking video action recognition under occlusion. *Advances in Neural Information Processing Systems*, 36, 2024.
- [141] Xinqian Gu, Hong Chang, Bingpeng Ma, Shutao Bai, Shiguang Shan, and Xilin Chen. Clothes-changing person re-identification with rgb modality only. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1060–1069, 2022.
- [142] Roberto Lo Gullo, Isaac Daimiel, Elizabeth A Morris, and Katja Pinker. Combining molecular and imaging metrics in cancer: radiogenomics. *Insights into imaging*, 11(1):1–17, 2020.
- [143] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein GANs. In *Advances in Neural Information Processing Systems*, pages 5767–5777, 2017.
- [144] Peini Guo, Hong Liu, Jianbing Wu, Guoquan Wang, and Tao Wang. Semantic-aware consistency network for cloth-changing person re-identification. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 8730–8739, 2023.

- [145] Yuxiang Guo, Cheng Peng, Chun Pong Lau, and Rama Chellappa. Multi-modal human authentication using silhouettes, gait, and rgb. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–7. IEEE, 2023.
- [146] David A Gutman, Lee AD Cooper, Scott N Hwang, Chad A Holder, JingJing Gao, Tarun D Aurora, William D Dunn Jr, Lisa Scarpace, Tom Mikkelsen, Rajan Jain, et al. Mr imaging predictors of molecular profile and survival: multi-institutional study of the tcga glioblastoma data set. *Radiology*, 267(2):560–569, 2013.
- [147] Adel Hafiane, Pierre Vieyres, and Alain Delbos. Deep learning with spatiotemporal consistency for nerve segmentation in ultrasound images. *arXiv preprint arXiv:1706.05870*, 2017.
- [148] Changhee Han, Hideaki Hayashi, Leonardo Rundo, Ryosuke Araki, Wataru Shimoda, Shinichi Muramatsu, Yujiro Furukawa, Giancarlo Mauri, and Hideki Nakayama. Gan-based synthetic brain mr image generation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 734–738. IEEE, 2018.
- [149] Changhee Han, Kohei Murao, Tomoyuki Noguchi, Yusuke Kawata, Fumiya Uchiyama, Leonardo Rundo, Hideki Nakayama, and Shin’ichi Satoh. Learning more with less: Conditional pggan-based data augmentation for brain metastases detection using highly-rough annotation on mr images. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 119–127, 2019.
- [150] Changhee Han, Leonardo Rundo, Ryosuke Araki, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama, and Hideaki Hayashi. Infinite brain tumor images: Can gan-based data augmentation improve tumor detection on mr images? In *Proc. Meeting on Image Recognition and Understanding (MIRU 2018)*, Sapporo, Japan, 2018.
- [151] Changhee Han, Leonardo Rundo, Ryosuke Araki, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama, and Hideaki Hayashi. Infinite brain mr images: Pggan-based data augmentation for tumor detection. In *Neural approaches to dynamics of signal exchanges*, pages 291–303. Springer, 2020.
- [152] Changhee Han, Leonardo Rundo, Ryosuke Araki, Yudai Nagano, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama, and Hideaki Hayashi. Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection. *IEEE Access*, 7:156966–156977, 2019.
- [153] Changhee Han, Leonardo Rundo, Kohei Murao, Tomoyuki Noguchi, Yuki Shimahara, Zoltán Ádám Milacski, Saori Koshino, Evis Sala, Hideki Nakayama, and Shin’ichi Satoh. Madgan: unsupervised medical anomaly detection gan using multiple adjacent brain mri slice reconstruction. *BMC bioinformatics*, 22(2):1–20, 2021.
- [154] Ke Han, Shaogang Gong, Yan Huang, Liang Wang, and Tieniu Tan. Clothing-change feature augmentation for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22066–22075, 2023.

- [155] Lichy Han and Maulik R Kamdar. Mri to mgmt: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium*, pages 331–342. World Scientific, 2018.
- [156] Zefang Han, Hong Shangguan, Xiong Zhang, Pengcheng Zhang, Xueying Cui, and Huiying Ren. A dual-encoder-single-decoder based low-dose ct denoising network. *IEEE Journal of Biomedical and Health Informatics*, 26(7):3251–3260, 2022.
- [157] Hana LP Harpold, Ellsworth C Alvord Jr, and Kristin R Swanson. The evolution of mathematical modeling of glioma proliferation and invasion. *Journal of Neuropathology & Experimental Neurology*, 66(1):1–9, 2007.
- [158] Christian Hartmann, Bettina Hentschel, Matthias Simon, Manfred Westphal, Gabriele Schackert, Jörg C Tonn, Markus Loeffler, Guido Reifenberger, Torsten Pietsch, Andreas von Deimling, et al. Long-term survival in primary glioblastoma with versus without isocitrate dehydrogenase mutations. *Clinical Cancer Research*, 19(18):5146–5157, 2013.
- [159] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.
- [160] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.
- [161] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1904–1916, 2015.
- [162] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [163] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15013–15022, 2021.
- [164] Monika E Hegi, Annie-Claire Diserens, Thierry Gorlia, Marie-France Hamou, Nicolas De Tribolet, Michael Weller, Johan M Kros, Johannes A Hainfellner, Warren Mason, Luigi Mariani, et al. Mgmt gene silencing and benefit from temozolomide in glioblastoma. *New England Journal of Medicine*, 352(10):997–1003, 2005.
- [165] Dieter Henrik Heiland, Carl Philipp Simon-Gabriel, Theo Demerath, Gerrit Haaker, Dietmar Pfeifer, Elias Kellner, Valerij G Kiselev, Ori Staszewski, Horst Urbach, Astrid Weyerbrock, et al. Integrative diffusion-weighted imaging and radiogenomic network analysis of glioblastoma multiforme. *Scientific Reports*, 7(1):1–9, 2017.

- [166] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.
- [167] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.
- [168] S Heymann, K Müller, A Smolic, B Froehlich, and T Wiegand. Sift implementation and optimization for general-purpose gpu. 2007.
- [169] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. *Iclr*, 2(5):6, 2017.
- [170] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- [171] T Hoang Ngan Le, Khoa Luu, Chi Nhan Duong, Kha Gia Quach, Thanh Dat Truong, Kyle Sadler, and Marios Savvides. Active contour model in deep learning era: a revise and review. In *Applications of Hybrid Metaheuristic Algorithms for Image Processing*, pages 231–260. Springer, 2020.
- [172] Eun Kyoung Hong, Seung Hong Choi, Dong Jae Shin, Sang Won Jo, Roh-Eul Yoo, Koung Mi Kang, Tae Jin Yun, Ji-Hoon Kim, Chul-Ho Sohn, Sung-Hye Park, et al. Radiogenomics correlation between mr imaging features and major genetic profiles in glioblastoma. *European radiology*, 28(10):4350–4361, 2018.
- [173] Peixian Hong, Tao Wu, Ancong Wu, Xintong Han, and Wei-Shi Zheng. Fine-grained shape-appearance mutual learning for cloth-changing person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10513–10522, 2021.
- [174] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010.
- [175] Guo X Hu, Bao L Hu, Zhong Yang, Li Huang, and Ping Li. Pavement crack detection method based on deep learning models. *Wireless Communications and Mobile Computing*, 2021, 2021.
- [176] Leland S Hu, Shuluo Ning, Jennifer M Eschbacher, Leslie C Baxter, Nathan Gaw, Sara Ranjbar, Jonathan Plasencia, Amylou C Dueck, Sen Peng, Kris A Smith, and P Nakaji. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. *Neuro-oncology*, 19(1):128–137, 2017.
- [177] Siyuan Huang, Ram Prabhakar Kathirvel, Chun Pong Lau, and Rama Chellappa. Whole-body detection, recognition and identification at altitude and range. *arXiv preprint arXiv:2311.05725*, 2023.

- [178] Yan Huang, Qiang Wu, Jingsong Xu, Yi Zhong, and ZhaoXiang Zhang. Clothing status awareness for long-term person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11895–11904, 2021.
- [179] Yan Huang, Qiang Wu, Zhang Zhang, Caifeng Shan, Yi Zhong, and Liang Wang. Meta clothing status calibration for long-term person re-identification. *IEEE Transactions on Image Processing*, 2024.
- [180] Yan Huang, Jingsong Xu, Qiang Wu, Yi Zhong, Peng Zhang, and Zhaoxiang Zhang. Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(10):3459–3471, 2019.
- [181] Yukun Huang, Xueyang Fu, Liang Li, and Zheng-Jun Zha. Learning degradation-invariant representation for robust real-world person re-identification. *International Journal of Computer Vision*, 130(11):2770–2796, 2022.
- [182] Fantine Huot, Robert Clapp, Biondo Biondi, Bruce Power, and Joe Stefani. Stratigraphy estimation from seismic data using deep learning. In *SEG Technical Program Expanded Abstracts 2019*, pages 2378–2382. Society of Exploration Geophysicists, 2019.
- [183] John Ingraham, Vikas Garg, Regina Barzilay, and Tommi Jaakkola. Generative models for graph-based protein design. *Advances in neural information processing systems*, 32, 2019.
- [184] Ashif Sikandar Iquebal and Satish TS Bukkapatnam. Consistent estimation of the max-flow problem: Towards unsupervised image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [185] Jyoti Islam and Yanqing Zhang. Gan-based synthetic brain pet image generation. *Brain informatics*, 7(1):1–12, 2020.
- [186] Mobarakol Islam, Navodini Wijethilake, and Hongliang Ren. Glioblastoma multiforme prognosis: Mri missing modality generation, segmentation and radiogenomic survival prediction. *Computerized Medical Imaging and Graphics*, page 101906, 2021.
- [187] Haruka Itakura, Achal S Achrol, Lex A Mitchell, Joshua J Loya, Tiffany Liu, Erick M Westbroek, Abdullah H Feroze, Scott Rodriguez, Sebastian Echegaray, Tej D Azad, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. *Science translational medicine*, 7(303):303ra138–303ra138, 2015.
- [188] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7. IEEE, 2020.
- [189] Emad Sami Jaha and Mark S Nixon. From clothing to identity: Manual and automatic soft biometrics. *IEEE Transactions on Information Forensics and Security*, 11(10):2377–2390, 2016.

- [190] Neema Jamshidi, Maximilian Diehn, Markus Bredel, and Michael D Kuo. Illuminating radiogenomic characteristics of glioblastoma multiforme through integration of mr imaging, messenger rna expression, and dna copy number variation. *Radiology*, 270(1):1–2, 2014.
- [191] Giacomo Janson, Gilberto Valdes-Garcia, Lim Heo, and Michael Feig. Direct generation of protein conformational ensembles via machine learning. *Nature Communications*, 14(1):774, 2023.
- [192] Emmanuelle J Javaux. *Eukaryotic Membranes and Cytoskeleton Origins and Evolution*, volume 607. Springer, 2007.
- [193] Emmanuelle J Javaux. Challenges in evidencing the earliest traces of life. *Nature*, 572:451–460, 2019.
- [194] Biswajit Jena, Sanjay Saxena, Gopal Krishna Nayak, Antonella Balestrieri, Neha Gupta, Narinder N Khanna, John R Laird, Manudeep K Kalra, Mostafa M Fouda, Luca Saba, et al. Brain tumor characterization using radiogenomics in artificial intelligence framework. *Cancers*, 14(16):4052, 2022.
- [195] Baode Jiang, Xiaoya An, Shaofen Xu, and Zhanlong Chen. Intelligent image semantic segmentation: A review through deep learning techniques for remote sensing image analysis. *Journal of the Indian Society of Remote Sensing*, pages 1–14, 2022.
- [196] Grisel Jimenez Soto, Mirza Arshad Beg, Michael C Poppelreiter, and Khaidi Rahmatsyah. Automatic extraction of modern reefs satellite images geometries using computer vision. In *EGU General Assembly Conference Abstracts*, page 13893, 2020.
- [197] Xin Jin, Tianyu He, Kecheng Zheng, Zhiheng Yin, Xu Shen, Zhen Huang, Ruoyu Feng, Jianqiang Huang, Zhibo Chen, and Xian-Sheng Hua. Cloth-changing person re-identification from a single image with gait prediction and regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14278–14287, 2022.
- [198] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE Transactions on Visualization and Computer Graphics*, 2019.
- [199] P Jonnalagedda, RL Surprenant, ML Droser, and B Bhanu. Automated remote biosignature detection using modified spaceseg model. *LPI Contributions*, 2678:2617, 2022.
- [200] Padmaja Jonnalagedda and Bir Bhanu. Novel body biometric for long-range recognition under extreme conditions. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9. IEEE, 2023.
- [201] Padmaja Jonnalagedda, Daniel Schmolze, and Bir Bhanu. MVPNets: Multi-viewing path deep learning neural networks for magnification invariant diagnosis in breast

- cancer. In *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 189–194. IEEE, 2018.
- [202] Padmaja Jonnalagedda, Rachel Surprenant, Mary Droser, and Bir Bhanu. SPACESeg: Automated detection of bed junction morphologies indicating signs of life in Ediacaran period. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 2095–2104, 2021.
- [203] Padmaja Jonnalagedda, Rachel L Surprenant, Mary L Droser, and Bir Bhanu. Searching for life: End-to-end automated detection and characterization of ediacaran biosignatures. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.
- [204] Padmaja Jonnalagedda, Brent Weinberg, Jason Allen, and Bir Bhanu. Feature disentanglement to aid imaging biomarker characterization for genetic mutations. In *Medical Imaging with Deep Learning*, pages 349–364. PMLR, 2020.
- [205] Padmaja Jonnalagedda, Brent Weinberg, Jason Allen, Taejin L Min, Shiv Bhanu, and Bir Bhanu. SAGE: Sequential attribute generator for analyzing glioblastomas using limited dataset. In *25th International Conference on Pattern Recognition (ICPR)*, pages 4941–4948. IEEE, 2021.
- [206] Padmaja Jonnalagedda, Brent Weinberg, Taejin L Min, Shiv Bhanu, and Bir Bhanu. A comprehensive radiogenomic feature characterization of 19/20 co-gain in glioblastoma. *IEEE Transactions on Artificial Intelligence*, 2024.
- [207] William L Jorgensen, Jayaraman Chandrasekhar, Jeffery D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [208] R Joseph, L Graham, B Budel, et al. Mars: Algae, Lichens, Fossils, Minerals, Microbial Mats, and Stromatolites in Gale Crater. *Journal of Astrobiology and Space Science Reviews*, 3(1):40–111, 2020.
- [209] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.
- [210] Vasileios G Kanas, Evangelia I Zacharaki, Ginu A Thomas, Pascal O Zinn, Vasileios Megalooikonomou, and Rivka R Colen. Learning mri-based classification models for mgmt methylation status prediction in glioblastoma. *Computer Methods and Programs in Biomedicine*, 140:249–257, 2017.
- [211] Nikolaos Karianakis, Zicheng Liu, Yinpeng Chen, and Stefano Soatto. Person depth reid: Robust person re-identification with commodity depth sensors. *arXiv preprint arXiv:1705.09882*, 2017.

- [212] Animesh Karnewar and Oliver Wang. MSG-GAN: Multi-scale gradients for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7799–7808, 2020.
- [213] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
- [214] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.
- [215] Suniti Karunatilake, Scott M McLennan, Kenneth E Herkenhoff, Jonathan M Husch, Craig Hardgrove, and JR Skok. A Martian case study of segmenting images automatically for granulometry and sedimentology, part 2: Assessment. *Icarus*, 229:408–417, 2014.
- [216] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [217] Sukhpal Kaur, Himanshu Aggarwal, and Rinkle Rani. Mr image synthesis using generative adversarial networks for parkinson’s disease classification. In *Proceedings of International Conference on Artificial Intelligence and Applications*, pages 317–327. Springer, 2021.
- [218] Mahmut Kaya and Hasan Şakir Bilge. Deep metric learning: A survey. *Symmetry*, 11(9):1066, 2019.
- [219] Amirhossein Kazerouni, Ehsan Khodapanah Aghdam, Moein Heidari, Reza Azad, Mohsen Fayyaz, Ilker Hacihaliloglu, and Dorit Merhof. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis*, 88:102846, 2023.
- [220] Koshino Kazuhiro, Rudolf A Werner, Fujio Toriumi, Mehrbod S Javadi, Martin G Pomper, Lilja B Solnes, Franco Verde, Takahiro Higuchi, and Steven P Rowe. Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images. *Tomography*, 4(4):159–163, 2018.
- [221] Onur Keleş, M Akın Yılmaz, A Murat Tekalp, Cansu Korkmaz, and Zafer Doğan. On the computation of psnr for a set of images or video. In *2021 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2021.
- [222] Ronald Kemker, Carl Salvaggio, and Christopher Kanan. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:60–77, 2018.
- [223] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN’95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE, 1995.

- [224] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International Conference on Medical Imaging with Deep Learning*, pages 285–296. PMLR, 2019.
- [225] Mena Nagy A Khalaf, Taysir Hassan A Soliman, and Sara Salah Mohamed. Generating nature-resembling tertiary protein structures with advanced generative adversarial networks (gans). *International Journal of Advanced Computer Science and Applications*, 14(8), 2023.
- [226] Mahyar Khayatkhoei, Maneesh K Singh, and Ahmed Elgammal. Disconnected manifold learning for generative adversarial networks. In *Advances in Neural Information Processing Systems*, pages 7343–7353, 2018.
- [227] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33:18661–18673, 2020.
- [228] Philipp Kickingeder, David Bonekamp, Martha Nowosielski, Annkathrin Kratz, Martin Sill, Sina Burth, Antje Wick, Oliver Eidel, Heinz-Peter Schlemmer, Alexander Radbruch, et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional mr imaging features. *Radiology*, 281(3):907–918, 2016.
- [229] Jonghun Kim and Hyunjin Park. Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7604–7613, 2024.
- [230] Yangjin Kim, Hyejin Jeon, and Hans Othmer. The role of the tumor microenvironment in glioblastoma: A mathematical model. *IEEE Transactions on Biomedical Engg.*, 64(3):519–527, 2016.
- [231] Madhu Kiran, Amran Bhuiyan, Louis-Antoine Blais-Morin, Ismail Ben Ayed, Eric Granger, et al. Flow guided mutual attention for person re-identification. *Image and Vision Computing*, 2021.
- [232] Iasonas Kokkinos. Pushing the boundaries of boundary detection using deep learning. *arXiv preprint arXiv:1511.07386*, 2015.
- [233] Filippos Konidaris, Thanos Tagaris, Maria Sdraka, and Andreas Stafylopatis. Generative adversarial networks as an advanced data augmentation technique for MRI data. In *VISIGRAPP (5: VISAPP)*, pages 48–59, 2019.
- [234] Branko Kordić, Borna Lužar-Oberiter, Kristina Pikelj, Bojan Matoš, and Goran Vlastelica. Integration of terrestrial laser scanning and uas photogrammetry in geological studies: examples from croatia. *Periodica Polytechnica. Civil Engineering*, 63(4):989, 2019.

- [235] P Korfiatis and B Erickson. Deep learning can see the unseeable: predicting molecular markers from mri of brain gliomas. *Clinical Radiology*, 74(5):367–373, 2019.
- [236] Panagiotis Korfiatis, Timothy L Kline, Daniel H Lachance, Ian F Parney, Jan C Buckner, and Bradley J Erickson. Residual deep convolutional neural network predicts mgmt methylation status. *Journal of digital imaging*, 30(5):622–628, 2017.
- [237] Kazuhiro Koshino, Rudolf A Werner, Martin G Pomper, Ralph A Bundschuh, Fujio Toriumi, Takahiro Higuchi, and Steven P Rowe. Narrative review of generative adversarial networks in medical and molecular imaging. *Annals of Translational Medicine*, 9(9), 2021.
- [238] Tabea Kossen, Pooja Subramaniam, Vince I Madai, Anja Hennemuth, Kristian Hildebrand, Adam Hilbert, Jan Sobesky, Michelle Livne, Ivana Galinovic, Ahmed A Khalil, et al. Synthesizing anonymized and labeled TOF-MRA patches for brain vessel segmentation using generative adversarial networks. *Computers in Biology and Medicine*, 131:104254, 2021.
- [239] Sethu Arun Kumar, Thirumoorthy Durai Ananda Kumar, Narasimha M Beeraka, Gurubasavaraj Veeranna Pujar, Manisha Singh, Handattu Sankara Narayana Akshatha, and Meduri Bhagyalalitha. Machine learning and deep learning in data-driven decision making of drug discovery and challenges in high-quality data acquisition in the pharmaceutical industry. *Future Medicinal Chemistry*, 14(4):245–270, 2022.
- [240] Maria Elena Laino, Pierandrea Cancian, Letterio Salvatore Politi, Matteo Giovanni Della Porta, Luca Saba, and Victor Savevski. Generative adversarial networks in brain imaging: A narrative review. *Journal of Imaging*, 8(4):83, 2022.
- [241] Antonio Lavecchia. Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug discovery today*, 24(10):2017–2032, 2019.
- [242] Kyung Won Lee. *Learning Appearance Features Using Soft-Biometrics for Long-Term Person Re-identification*. PhD thesis, State University of New York at Buffalo, 2023.
- [243] Kevin W Lewis, Oded Aharonson, John P Grotzinger, Steven W Squyres, James F Bell III, Larry S Crumpler, and Mariek E Schmidt. Structure and stratigraphy of home plate from the spirit Mars exploration rover. *Journal of Geophysical Research: Planets*, 113(E12), 2008.
- [244] Daiqing Li, Junlin Yang, Karsten Kreis, Antonio Torralba, and Sanja Fidler. Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8300–8311, 2021.
- [245] Hanjun Li, Gaojie Wu, and Wei-Shi Zheng. Combined depth space based architecture search for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6729–6738, 2021.

- [246] Jianing Li, Jingdong Wang, Qi Tian, Wen Gao, and Shiliang Zhang. Global-local temporal representations for video person re-identification. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pages 3958–3967, 2019.
- [247] Jiaojiao Li, Shunyao Zi, Rui Song, Yunsong Li, Yinlin Hu, and Qian Du. A stepwise domain adaptive segmentation network with covariate shift alleviation for remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.
- [248] Lu Li, Ran Wang, Jingyu Lin, Zhouxuan Xiao, and Yuanxiu Hui. A novel approach for extraction of ripple mark parameters based on SfM. *Sedimentary Geology*, 392:105523, 2019.
- [249] Peike Li, Yunqiu Xu, Yunchao Wei, and Yi Yang. Self-correction for human parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3260–3271, 2020.
- [250] Shuai Li, Hongze Han, Dong Sui, Aimin Hao, and Hong Qin. A novel radiogenomics framework for genomic and image feature correlation using deep learning. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 899–906. IEEE, 2018.
- [251] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2285–2294, 2018.
- [252] Weijia Li, Saihui Hou, Chunjie Zhang, Chunshui Cao, Xu Liu, Yongzhen Huang, and Yao Zhao. An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13824–13833, 2023.
- [253] Yu-Jhe Li, Zhengyi Luo, Xinshuo Weng, and Kris M Kitani. Learning shape representations for clothing variations in person re-identification. *arXiv preprint arXiv:2003.07340*, 2020.
- [254] Yu-Jhe Li, Xinshuo Weng, and Kris M Kitani. Learning shape representations for person re-identification under clothing change. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2432–2441, 2021.
- [255] Zeju Li, Yuanyuan Wang, Jinhua Yu, Yi Guo, and Wei Cao. Deep learning based radiomics (dlr) and its usage in noninvasive idh1 prediction for low grade glioma. *Scientific Reports*, 7(1):1–11, 2017.
- [256] Zhongbin Li, Wenzhong Shi, Qunming Wang, and Zelang Miao. Extracting man-made objects from high spatial resolution remote sensing images via fast level set evolutions. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2):883–899, 2014.

- [257] Xiaodan Liang, Ke Gong, Xiaohui Shen, and Liang Lin. Look into person: Joint body parsing & pose estimation network and a new benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(4):871–885, 2018.
- [258] Xiaodan Liang, Chunyan Xu, Xiaohui Shen, Jianchao Yang, Si Liu, Jinhui Tang, Liang Lin, and Shuicheng Yan. Human parsing with contextualized convolutional neural network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1386–1394, 2015.
- [259] Qinghua Liao, Michael C Owen, Olujide O Olubiyi, Bogdan Barz, and Birgit Strodel. Conformational transitions of the amyloid- β peptide upon copper (ii) binding and ph changes. *Israel Journal of Chemistry*, 57(7-8):771–784, 2017.
- [260] Beibei Lin, Shunli Zhang, Ming Wang, Lincheng Li, and Xin Yu. Gaitgl: Learning discriminative global-local feature representations for gait recognition. *arXiv preprint arXiv:2208.01380*, 2022.
- [261] Eugene Lin, Chieh-Hsin Lin, and Hsien-Yuan Lane. De novo peptide and protein design using generative adversarial networks: an update. *Journal of Chemical Information and Modeling*, 62(4):761–774, 2022.
- [262] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermesen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, 6:26286, 2016.
- [263] Chunxiao Liu, Shaogang Gong, Chen Change Loy, and Xinggang Lin. Person re-identification: What features are important? In *Computer Vision–ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7-13, 2012, Proceedings, Part I 12*, pages 391–401. Springer, 2012.
- [264] Feng Liu, Ryan Ashbaugh, Nicholas Chimitt, Najmul Hassan, Ali Hassani, Ajay Jaiswal, Minchul Kim, Zhiyuan Mao, Christopher Perry, Zhiyuan Ren, et al. Far-sight: A physics-driven whole-body biometric system at large distance and altitude. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6227–6236, 2024.
- [265] Feng Liu, Minchul Kim, ZiAng Gu, Anil Jain, and Xiaoming Liu. Learning clothing and pose invariant 3d shape representation for long-term person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19617–19626, 2023.
- [266] Feng Liu, Minchul Kim, Zhiyuan Ren, and Xiaoming Liu. Distilling clip with dual guidance for learning discriminative human body shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 256–266, 2024.

- [267] Haiqiang Liu, Meibao Yao, Xueming Xiao, and Yonggang Xiong. Rockformer: a u-shaped transformer network for Martian rock segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [268] Qian Liu and Pingzhao Hu. Extendable and explainable deep learning for pan-cancer radiogenomics research. *Current Opinion in Chemical Biology*, 66:102111, 2022.
- [269] Xiaokang Liu and Haijun Song. Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional neural networks. *Sedimentary Geology*, 410:105790, 2020.
- [270] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.
- [271] Zhihua Liu, Long Chen, Lei Tong, Feixiang Zhou, Zheheng Jiang, Qianni Zhang, Caifeng Shan, Yinhai Wang, Xiangrong Zhang, Ling Li, et al. Deep learning based brain tumor segmentation: a survey. *arXiv preprint arXiv:2007.09479*, 2020.
- [272] Zhihua Liu, Lei Tong, Long Chen, Feixiang Zhou, Zheheng Jiang, Qianni Zhang, Yinhai Wang, Caifeng Shan, Ling Li, and Huiyu Zhou. Canet: Context aware network for brain glioma segmentation. *IEEE Transactions on Medical Imaging*, 40(7):1763–1777, 2021.
- [273] Jarred C Lloyd, Morgan L Blades, John W Counts, Allan S Collins, Kathryn J Amos, Benjamin P Wade, James W Hall, Stephen Hore, Ashleigh L Ball, Sameh Shahin, and Matthew Drabsch. Neoproterozoic geochronology and provenance of the Adelaide Superbasin. *Precambrian Research*, 350, 2020.
- [274] Ilya Loshchilov and Frank Hutter. Sgdr:stochastic gradient descent with warm restarts. *arXiv preprint:1608.03983*, 2016.
- [275] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157. IEEE, 1999.
- [276] Chia-Feng Lu, Fei-Ting Hsu, Kevin Li-Chun Hsieh, Yu-Chieh Jill Kao, Sho-Jen Cheng, Justin Bo-Kai Hsu, Ping-Huei Tsai, Ray-Jade Chen, Chao-Ching Huang, Yun Yen, et al. Machine learning-based radiomics for molecular subtyping of gliomas. *Clinical Cancer Research*, 24(18):4429–4436, 2018.
- [277] Shangzhen Luan, Chen Chen, Baochang Zhang, Jungong Han, and Jianzhuang Liu. Gabor convolutional networks. *IEEE Transactions on Image Processing*, 27(9):4357–4366, 2018.
- [278] Kirsten Ludwig and Harley I Kornblum. Molecular markers in glioma. *Journal of Neuro-Oncology*, 134(3):505–512, 2017.

- [279] Thorsten Lühns, Christiane Ritter, Marc Adrian, Dominique Riek-Loher, Bernd Bohrmann, Heinz Döbeli, David Schubert, and Roland Riek. 3d structure of alzheimer’s amyloid- β (1–42) fibrils. *Proceedings of the National Academy of Sciences*, 102(48):17342–17347, 2005.
- [280] Qing Lyu and Ge Wang. Conversion between ct and mri images using diffusion and score-matching models. *arXiv preprint arXiv:2209.12104*, 2022.
- [281] Suyue Lyu, Shahin Sowlati-Hashjin, and Michael Garton. Proteinvae: variational autoencoder for translational protein design. *bioRxiv*, pages 2023–03, 2023.
- [282] Zhongxing Ma, Yifan Zhao, and Jia Li. Pose-guided inter-and intra-part relational transformer for occluded person re-identification. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1487–1496, 2021.
- [283] A Mahajan, A Sahu, R Ashtekar, T Kulkarni, S Shukla, U Agarwal, and K Bhattacharya. Glioma radiogenomics and artificial intelligence: road to precision cancer medicine. *Clinical Radiology*, 78(2):137–149, 2023.
- [284] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.
- [285] Owais A Malik, Idrus Puasa, and Daphne Teck Ching Lai. Segmentation for multi-rock types on digital outcrop photographs using deep learning techniques. *Sensors*, 22(21):8086, 2022.
- [286] Vishal Mandal, Lan Uong, and Yaw Adu-Gyamfi. Automated road crack detection using deep convolutional neural networks. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5212–5215. IEEE, 2018.
- [287] Shunan Mao, Shiliang Zhang, and Ming Yang. Resolution-invariant person re-identification. *arXiv preprint arXiv:1906.09748*, 2019.
- [288] Lynn Margulis, Michael F Dolan, and Michael Dolan. *Early life: Evolution on the Precambrian Earth*. Jones & Bartlett Learning, 2002.
- [289] Jack J Matthews, Alexander G Liu, Chuan Yang, Duncan McIlroy, Bruce Levell, and Daniel J Condon. A chronostratigraphic framework for the rise of the ediacaran macrobiota: new constraints from mistaken point ecological reserve, newfoundland. *Bulletin*, 133(3-4):612–624, 2021.
- [290] John Maynard, Sachi Okuchi, Stephen Wastling, Ayisha Al Busaidi, Ofra Almosawi, Wonderboy Mbatha, Sebastian Brandner, Zane Jaunmuktane, Ali Murat Koc, Mancini Laura, R Jäger, and S Thust. World health organization grade ii/iii glioma molecular status: prediction by mri morphologic features and apparent diffusion coefficient. *Radiology*, 296(1):111–121, 2020.

- [291] Maciej A Mazurowski. Radiogenomics: what it is and why it is important. *Journal of the American College of Radiology*, 12(8):862–866, 2015.
- [292] Maciej A Mazurowski, Jing Zhang, Katherine B Peters, and Hasan Hobbs. Computer-extracted mr imaging features are associated with survival in glioblastoma patients. *Journal of neuro-oncology*, 120:483–488, 2014.
- [293] S McMahon, T Bosak, JP Grotzinger, RE Milliken, RE Summons, M Daye, SA Newman, A Fraeman, KH Williford, and DEG Briggs. A field guide to finding fossils on Mars. *Journal of Geophysical Research: Planets*, 123:1012–1040, 2018.
- [294] Sean McMahon and Julie Cosmidis. False biosignatures on Mars: anticipating ambiguity. *Journal of the Geological Society*, 179(2), 2022.
- [295] William J McMahon, Alexander G Liu, Benjamin H Tindal, and Maarten G Kleinhans. Ediacaran life close to land: Coastal and shoreface habitats of the ediacaran macrobiota, the central flinders ranges, south australia. *Journal of Sedimentary Research*, 90(11):1463–1499, 2020.
- [296] Sachin Mehta, Mohammad Rastegari, Anat Caspi, Linda Shapiro, and Hannaneh Hajishirzi. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 552–568, 2018.
- [297] Taejin L Min, Jason W Allen, Jose E Velazquez Vega, Stewart G Neill, and Brent D Weinberg. Mri imaging characteristics of glioblastoma with concurrent gain of chromosomes 19 and 20. *Tomography*, 7(2):228–237, 2021.
- [298] Ajoy Mondal, Susmita Ghosh, and Ashish Ghosh. Robust global and local fuzzy energy based active contour for image segmentation. *Applied Soft Computing*, 47:191–215, 2016.
- [299] Elisa Moya-Sáez, Óscar Peña-Nogales, Rodrigo de Luis-García, and Carlos Alberola-López. A deep learning approach for synthetic mri based on two routine sequences and training with synthetic data. *Computer Methods and Programs in Biomedicine*, 210:106371, 2021.
- [300] Junxi Mu, Hao Liu, Jian Zhang, Ray Luo, and Hai-Feng Chen. Recent force field strategies for intrinsically disordered proteins. *Journal of chemical information and modeling*, 61(3):1037–1047, 2021.
- [301] Blake A Myers, Lucas Jaggernauth, Thomas M Metz, Matthew Q Hill, Veda Nandan Gandi, Carlos D Castillo, and Alice J O’Toole. Recognizing people by body shape using deep networks of images and words. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2023.
- [302] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI Brain Lesion Workshop*, pages 311–320. Springer, 2018.

- [303] Timothy Paul Nagle-McNaughton, Louis Anthony Scuderi, and Nicholas Erickson. Squeezing data from a rock: Machine learning for Martian science. *Geosciences*, 12(6):248, 2022.
- [304] Nithin Gopalakrishnan Nair, Kangfu Mei, and Vishal M Patel. A comparison of different atmospheric turbulence simulation methods for image restoration. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3386–3390. IEEE, 2022.
- [305] Sahil S Nalawade, F Yu Fang, Chandan Ganesh Bangalore Yogananda, Gowtham K Murugesan, Bhavya R Shah, Marco C Pinho, Benjamin C Wagner, Bruce Mickey, Toral R Patel, Baowei Fei, et al. Brain tumor idh, 1p/19q, and mgmt molecular classification using mri-based deep learning: Effect of motion and motion correction. *bioRxiv*, 2020.
- [306] Sandy Napel, Wei Mu, Bruna V Jardim-Perassi, Hugo JWL Aerts, and Robert J Gillies. Quantitative imaging of cancer in the postgenomic era: Radio (geno) mics, deep learning, and habitats. *Cancer*, 124(24):4633–4649, 2018.
- [307] NASEM. An astrobiology strategy for the search for life in the universe. *The National Academies of Sciences, Engineering, and Medicine*, 2019.
- [308] Mohamed A Naser and M Jamal Deen. Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images. *Computers in biology and medicine*, 121:103758, 2020.
- [309] National Academies of Sciences, Engineering, and Medicine, and others. Origins, worlds, and life: A decadal strategy for Planetary Science and Astrobiology 2023–2032. *The National Academies Press*, 2023.
- [310] Shubhangi Nema, Akshay Dudhane, Subrahmanyam Murala, and Srivatsava Naidu. Rescuenet: An unpaired gan for brain tumor segmentation. *Biomedical Signal Processing and Control*, 55:101641, 2020.
- [311] Kee Yuan Ngiam and Wei Khor. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5):e262–e273, 2019.
- [312] Huy Hoang Nguyen, Thi Nhung Ta, Ngoc Cuong Nguyen, Hung Manh Pham, Duc Minh Nguyen, et al. Yolo based real-time human detection for smart video surveillance at the edge. In *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, pages 439–444. IEEE, 2021.
- [313] Vuong D Nguyen, Khadija Khaldi, Dung Nguyen, Pranav Mantini, and Shishir Shah. Contrastive viewpoint-aware shape learning for long-term person re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1041–1049, 2024.

- [314] Enhao Ning, Changshuo Wang, Huang Zhang, Xin Ning, and Prayag Tiwari. Occluded person re-identification with deep learning: a survey and perspectives. *Expert Systems with Applications*, page 122419, 2023.
- [315] Nora Noffke. The criteria for the biogenicity of microbially induced sedimentary structures (MISS) in Archean and younger, sandy deposits. *Earth-Science Reviews*, 96(3):173–180, 2009.
- [316] Nora Noffke. *Geobiology: Microbial Mats in Sandy Deposits from the Archean Era to Today*. Springer-Verlag, 2010.
- [317] Nora Noffke. Ancient sedimentary structures in the <3.7 Ga Gillespie Lake Member, Mars, that resemble macroscopic morphology, spatial associations, and temporal succession in terrestrial microbialites. *Astrobiology*, 15(2):169–192, 2015.
- [318] Nora Noffke, Gisela Gerdes, Thomas Klenke, and Wolfgang E Krumbein. Microbially induced sedimentary structures: A new category within the classification of primary sedimentary structures. *Journal of Sedimentary Research*, 71(5):649–656, 2001.
- [319] Nicholas Nuechterlein, Beibin Li, Abdullah Feroze, Eric C Holland, Linda Shapiro, David Haynor, James Fink, and Patrick J Cimino. Radiogenomic modeling predicts survival-associated prognostic groups in glioblastoma. *Neuro-oncology advances*, 3(1):vdab004, 2021.
- [320] Allison C Nugent, Adam G Thomas, Margaret Mahoney, Alison Gibbons, Jarrod T Smith, Antoinette J Charles, Jacob S Shaw, Jeffrey D Stout, Anna M Namyst, Arshitha Basavaraj, et al. The NIMH intramural healthy volunteer dataset: A comprehensive MEG, MRI, and behavioral resource. *Scientific Data*, 9(1):1–10, 2022.
- [321] Robert J O’Callaghan and David R Bull. Combined morphological-spectral unsupervised image segmentation. *IEEE Transactions on Image Processing*, 14(1):49–62, 2004.
- [322] Kazunori Ogohara and Ryusei Gichu. Automated segmentation of textured dust storms on Mars remote sensing images using an encoder-decoder type convolutional neural network. *Computers & Geosciences*, 160:105043, 2022.
- [323] Gretchen R O’Neil, Lydia S Tackett, and Mike Meyer. Petrographic evidence for ediacaran microbial mat-targeted behaviors from the great basin, united states. *Pre-cambrian Research*, 345:105768, 2020.
- [324] Richard Osuala, Kaisar Kushibar, Lidia Garrucho, Akis Linardos, Zuzanna Szafranska, Stefan Klein, Ben Glocker, Oliver Diaz, and Karim Lekadir. Data synthesis and adversarial networks: A review and meta-analysis in cancer imaging. *Medical Image Analysis*, page 102704, 2022.
- [325] Nobuyuki Otsu et al. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

- [326] Deqiang Ouyang, Yonghui Zhang, and Jie Shao. Video-based person re-identification via spatio-temporal attentional and two-stream fusion convolutional networks. *Pattern Recognition Letters*, 117:153–160, 2019.
- [327] Leon F Palafox, Christopher W Hamilton, Stephen P Scheidt, and Alexander M Alvarez. Automated detection of geological landforms on Mars using convolutional neural networks. *Computers & Geosciences*, 101:48–56, 2017.
- [328] Shaoyan Pan, Tonghe Wang, Richard LJ Qiu, Marian Axente, Chih-Wei Chang, Junbo Peng, Ashish B Patel, Joseph Shelton, Sagar A Patel, Justin Roper, et al. 2d medical image synthesis using transformer-based denoising diffusion probabilistic model. *Physics in Medicine & Biology*, 68(10):105004, 2023.
- [329] Yongsheng Pan, Mingxia Liu, Yong Xia, and Dinggang Shen. Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [330] Athanasios N Papanicolaou, Panayiotis Diplas, Mahesh Balakrishnan, and Clinton L Dancey. Computer vision technique for tracking bed load movement. *Journal of Computing in Civil Engineering*, 13(2):71–79, 1999.
- [331] Ji Eun Park, Dain Eun, Ho Sung Kim, Ryoung Woo Jang, and Namkug Kim. Generative adversarial network for glioblastoma ensures morphologic variations and improves diagnostic model for isocitrate dehydrogenase mutant type. *Scientific Reports*, 11(1):1–11, 2021.
- [332] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, 2016.
- [333] Juan C Pérez, Motasem Alfarra, Guillaume Jeanneret, Adel Bibi, Ali Thabet, Bernard Ghanem, and Pablo Arbeláez. Gabor layers enhance network robustness. In *European Conference on Computer Vision*, pages 450–466. Springer, 2020.
- [334] Friedrich Pfluger. Matground structures and redox facies. *PALAIOS*, 14(1), 1999.
- [335] C-H Pham, Carlos Tor-Díez, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Passat, and François Rousseau. Simultaneous super-resolution and segmentation using a generative adversarial network: Application to neonatal brain mri. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 991–994. IEEE, 2019.
- [336] Katja Pinker, Fuki Shitano, Evis Sala, Richard K Do, Robert J Young, Andreas G Wibmer, Hedvig Hricak, Elizabeth J Sutton, and Elizabeth A Morris. Background, current role, and potential applications of radiogenomics. *Journal of Magnetic Resonance Imaging*, 47(3):604–620, 2018.

- [337] Letterio S Politi and Riccardo Levi. Editorial comment: Radiogenomics of glioblastoma: shifting the focus from tumor cells to immune microenvironment. *European Radiology*, 33(1):207–208, 2023.
- [338] S Polushkin, Y Volokitin, I Edelman, E Sadykhov, O Lokhanova, Y Murzaev, and S Pastushkov. Processing thin section photos with neural networks and computer vision. In *First EAGE Digitalization Conference and Exhibition*, volume 2020, pages 1–3. European Association of Geoscientists & Engineers, 2020.
- [339] Agus Pratondo, Chee-Kong Chui, and Sim-Heng Ong. Integrating machine learning with region-based active contour models in medical image segmentation. *Journal of Visual Communication and Image Representation*, 43:1–9, 2017.
- [340] Lajos Pusztai and W Fraser Symmans. Histopathologic and molecular markers of prognosis and response to therapy. In *Breast Cancer 2nd edition*, pages 323–343. Springer, 2008.
- [341] Lu Qi, Jason Kuen, Weidong Guo, Tiancheng Shen, Jiuxiang Gu, Wenbo Li, Jiaya Jia, Zhe Lin, and Ming-Hsuan Yang. Fine-grained entity segmentation. *arXiv preprint arXiv:2211.05776*, 2022.
- [342] Xuelin Qian, Wenxuan Wang, Li Zhang, Fangrui Zhu, Yanwei Fu, Tao Xiang, Yuguang Jiang, and Xiangyang Xue. Long-term cloth-changing person re-identification. *arXiv preprint arXiv:2005.12633*, 2020.
- [343] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. BasNet: Boundary-aware salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7479–7489, 2019.
- [344] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [345] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [346] Alessio Radice, Sankar Sarkar, and Francesco Ballio. Image-based lagrangian particle tracking in bed-load experiments. *Journal of Visualized Experiments: JoVE*, 2017.
- [347] Aayush J Rana and Yogesh S Rawat. Hybrid active learning via deep clustering for video action detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18867–18877, 2023.
- [348] Saima Rathore, Hamed Akbari, Martin Rozycki, Kalil G Abdullah, MacLean P Nasrallah, Zev A Binder, Ramana V Davuluri, Robert A Lustig, Nadia Dahmane, Michel Bilello, and D.M. O’Rourke. Radiomic mri signature reveals three distinct subtypes of

- glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond idh1. *Scientific Reports*, 8(1):1–12, 2018.
- [349] Dripta S Raychaudhuri, Calvin-Khang Ta, Arindam Dutta, Rohit Lal, and Amit K Roy-Chowdhury. Prior-guided source-free domain adaptation for human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14996–15006, 2023.
 - [350] Guido Reifenberger, Hans-Georg Wirsching, Christiane B Knobbe-Thomsen, and Michael Weller. Advances in the molecular genetics of gliomas—implications for classification and therapy. *Nature reviews Clinical oncology*, 14(7):434–452, 2017.
 - [351] TR Rejusha and Vipin Kumar KS. Artificial mri image generation using deep convolutional gan and its comparison with other augmentation methods. In *2021 international conference on communication, Control and Information Sciences (ICCISc)*, volume 1, pages 1–6. IEEE, 2021.
 - [352] Ruoxu Ren, Terence Hung, and Kay Chen Tan. A generic deep-learning-based approach for automated surface inspection. *IEEE Transactions on Cybernetics*, 48(3):929–940, 2017.
 - [353] Mina Rezaei, Haojin Yang, and Christoph Meinel. voxel-gan: adversarial framework for learning imbalanced brain tumor segmentation. In *International MICCAI Brain-lesion Workshop*, pages 321–333. Springer, 2018.
 - [354] Federico Ricciuti, Christian Onzaca, Roberto Berto, and Maria Teresa Galli. Unsupervised recognition of textural facies in borehole images: a novel and efficient two-stage segmentation method. In *OMC Med Energy Conference and Exhibition*. OnePetro, 2021.
 - [355] Vincenzo Rizzo. Why should geological criteria used on earth not be valid also for Mars? evidence of possible microbialites and algae in extinct Martian lakes. *International Journal of Astrobiology*, 19(3):283–294, 2020.
 - [356] Wes Robbins, Gabriel Bertocco, and Terrance E Boulton. Daliid: Distortion-adaptive learned invariance for identification models. *arXiv preprint arXiv:2302.05753*, 2023.
 - [357] Graham Roberts, Simon Y Haile, Rajat Sainju, Danny J Edwards, Brian Hutchinson, and Yuanyuan Zhu. Deep learning for semantic segmentation of defects in advanced stem images of steels. *Scientific Reports*, 9(1):1–12, 2019.
 - [358] Dan Rohle, Janeta Popovici-Muller, Nicolaos Palaskas, Sevin Turcan, Christian Grommes, Carl Campos, Jennifer Tsoi, Owen Clark, Barbara Oldrini, Evangelia Komisopoulou, et al. An inhibitor of mutant idh1 delays growth and promotes differentiation of glioma cells. *Science*, 340(6132):626–630, 2013.
 - [359] Remi Ronfard. Region-based strategies for active contour models. *International Journal of Computer Vision*, 13(2):229–251, 1994.

- [360] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, pages 234–241. Springer, 2015.
- [361] Lior Rubanenko, Sebastian Pérez-López, Joseph Schull, and Mathieu GA Lapôtre. Automatic detection and segmentation of Barchan Dunes on Mars and Earth using a convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:9364–9371, 2021.
- [362] Iresha Rubasinghe and Dulani Meedeniya. Ultrasound nerve segmentation using deep probabilistic programming. *Journal of ICT Research and Applications*, 13(3):241–256, 2019.
- [363] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.
- [364] Marcin Rudzki. Vessel detection method based on Eigenvalues of the Hessian matrix and its applicability to airway tree segmentation. In *Proceedings of the 11th International PhD Workshop OWD*, pages 100–105, 2009.
- [365] Aaron M Rutman and Michael D Kuo. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. *European journal of radiology*, 70(2):232–241, 2009.
- [366] Tanzila Saba, Ahmed Sameh Mohamed, Mohammad El-Affendi, Javeria Amin, and Muhammad Sharif. Brain tumor detection using fusion of hand crafted and deep learning features. *Cognitive Systems Research*, 59:221–230, 2020.
- [367] Tariq Sadad, Amjad Rehman, Asim Munir, Tanzila Saba, Usman Tariq, Noor Ayesha, and Rashid Abbasi. Brain tumor detection and multi-classification using advanced deep learning techniques. *Microscopy Research and Technique*, 84(6):1296–1308, 2021.
- [368] Lisa Millgård Sagberg, Daniel Høyer Iversen, Even Hovig Fyllingen, Asgeir Store Jakola, Ingerid Reinertsen, and Ole Solheim. Brain atlas for assessing the impact of tumor location on perioperative quality of life in patients with high-grade glioma: A prospective population-based cohort study. *NeuroImage: Clinical*, 21:101658, 2019.
- [369] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016.
- [370] M Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 420–429, 2018.

- [371] Prodip Kumar Sarker, Qingjie Zhao, and Md Kamal Uddin. Transformer-based person re-identification: A comprehensive review. *IEEE Transactions on Intelligent Vehicles*, 2024.
- [372] Nishank Saxena, Ruarri J Day-Stirrat, Amie Hows, and Ronny Hofmann. Application of deep learning for semantic segmentation of sandstone thin sections. *Computers & Geosciences*, 152:104778, 2021.
- [373] Bettina E Schirrmeister, Patricia Sanchez-Baracalldo, and David Wacey. Cyanobacterial evolution during the Precambrian. *International Journal of Astrobiology*, 15(3):187–204, 2016.
- [374] Frdric Schmidt, Sylvain Douté, and Bernard Schmitt. Wavanglet: An efficient supervised classifier for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1374–1385, 2007.
- [375] Samuel C Schon, James W Head, and Caleb I Fassett. An overfilled laucstrine system and progradational delta in Jezero crater, Mars: Implications for Nochian climate. *Planetary and Space Science*, 67(1):28–45, 2012.
- [376] Stefan Schroedl. Current methods and challenges for deep learning in drug discovery. *Drug Discovery Today: Technologies*, 32:9–17, 2019.
- [377] Mauricio Pamplona Segundo, Cole Hill, and Sudeep Sarkar. Long range gait matching using 3d body fitting with gait-specific motion constraints. In *Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 603–612, 2023.
- [378] Adolf Seilacher. Biomat-related lifestyles in the Precambrian. *PALAIOS*, 14(1):86–93, 1999.
- [379] Maximilian Seitzer. pytorch-fid: FID Score for PyTorch. <https://github.com/mseitzer/pytorch-fid>, August 2020. Version 0.3.0.
- [380] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [381] Ahmed Serag, Adrian Ion-Margineanu, Hammad Qureshi, Ryan McMillan, Marie-Judith Saint Martin, Jim Diamond, Paul O’Reilly, and Peter Hamilton. Translational ai and deep learning in diagnostic pathology. *Frontiers in Medicine*, 6:185, 2019.
- [382] Emre Sevgen, Joshua Moller, Adrian Lange, John Parker, Sean Quigley, Jeff Mayer, Poonam Srivastava, Sitaram Gayatri, David Hosfield, Maria Korshunova, et al. Prot-vae: protein transformer variational autoencoder for functional protein design. *bioRxiv*, pages 2023–01, 2023.

- [383] Jamal Hussain Shah, Zonghai Chen, Muhammad Sharif, Mussarat Yasmin, and Steven Lawrence Fernandes. A novel biomechanics-based approach for person re-identification by generating dense color sift salience features. *Journal of Mechanics in Medicine and Biology*, 17(07):1740011, 2017.
- [384] Pratik Shah, Francis Kendall, Sean Khozin, Ryan Goosen, Jianying Hu, Jason Laramie, Michael Ringel, and Nicholas Schork. Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digital Medicine*, 2(1):1–5, 2019.
- [385] Pratibha Sharma, Ashley Aaroe, Jiyong Liang, and Vinay K Puduvalli. Tumor microenvironment in glioblastoma: Current and emerging concepts. *Neuro-Oncology Advances*, 5(1):vdad009, 2023.
- [386] J Shen, Yan Chen, Yu Sun, Li Liu, YongXin Pan, and Wei Lin. Detection of biosignatures in terrestrial analogs of Martian regions: Strategical and technical assessments. *Earth and Planetary Physics*, 6(5):431–450, 2022.
- [387] Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine P Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 1–11. Springer, 2018.
- [388] Konstantin Shmelkov, Cordelia Schmid, and Karteek Alahari. How good is my gan? In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 213–229, 2018.
- [389] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.
- [390] Alcides Nóbrega Sial, Claudio Gaucher, Muthuvairavasamy Ramkumar, and Valderez Pinto Ferreira. *Chemostratigraphy Across Major Chronological Boundaries*. Wiley Online Library, 2019.
- [391] Md Mahfuzur Rahman Siddiquee, Zongwei Zhou, Nima Tajbakhsh, Ruibin Feng, Michael B Gotway, Yoshua Bengio, and Jianming Liang. Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 191–200, 2019.
- [392] Nyle Siddiqui, Praveen Tirupattur, and Mubarak Shah. Dvanet: Disentangling view and action features for multi-view action recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4873–4881, 2024.
- [393] A Siminovich, T Elperin, I Katra, JF Kok, R Sullivan, Simone Silvestro, and H Yizhaq. Numerical study of shear stress distribution over sand ripples under terrestrial and Martian conditions. *Journal of Geophysical Research: Planets*, 124(1):175–185, 2019.

- [394] Kihyuk Sohn, Honglak Lee, and Xinchun Yan. Learning structured output representation using deep conditional generative models. *Advances in Neural Information Processing Systems*, 28:3483–3491, 2015.
- [395] David Solarna, Alberto Gotelli, Jacqueline Le Moigne, Gabriele Moser, and Sebastiano B Serpico. Crater detection and registration of planetary images through marked point processes, multiscale decomposition, and region-based analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6039–6058, 2020.
- [396] Farzan Soleymani, Eric Paquet, Herna Viktor, Wojtek Michalowski, and Davide Spinello. Protein-protein interaction prediction with deep learning: A comprehensive review. *Computational and Structural Biotechnology Journal*, 20:5316–5341, 2022.
- [397] Farzan Soleymani, Eric Paquet, Herna Lydia Viktor, Wojtek Michalowski, and Davide Spinello. Protinteract: A deep learning framework for predicting protein-protein interactions. *Computational and Structural Biotechnology Journal*, 21:1324–1348, 2023.
- [398] Antoine Soloy, Imen Turki, Matthieu Fournier, Stéphane Costa, Bastien Peuziat, and Nicolas Lecoq. A deep learning-based method for quantifying and mapping the grain size on pebble beaches. *Remote Sensing*, 12(21):3659, 2020.
- [399] Chunfeng Song, Caifeng Shan, Yan Huang, and Liang Wang. Mask-guided contrastive attention and two-stream metric co-learning for person re-identification. *Neurocomputing*, 465:561–573, 2021.
- [400] Dong Song, Ray Luo, and Hai-Feng Chen. The idp-specific force field ff14idpsff improves the conformer sampling of intrinsically disordered proteins. *Journal of chemical information and modeling*, 57(5):1166–1178, 2017.
- [401] Wanru Song, Yahong Wu, Jieying Zheng, Changhong Chen, and Feng Liu. Extended global-local representation learning for video person re-identification. *IEEE Access*, 7:122684–122696, 2019.
- [402] Andrew N Stein, Thomas S Stepleton, and Martial Hebert. Towards unsupervised whole-object segmentation: Combining automated matting with boundary detection. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.
- [403] Michael D Story and Marco Durante. Radiogenomics. *Medical physics*, 45(11):e1111–e1122, 2018.
- [404] Alexey Strokach, David Becerra, Carles Corbi-Verge, Albert Perez-Riba, and Philip M Kim. Fast and flexible protein design using deep graph neural networks. *Cell systems*, 11(4):402–411, 2020.
- [405] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Pose-driven deep convolutional model for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3960–3969, 2017.

- [406] Cecilia Suarez, Felipe Maglietti, Mario Colonna, Karina Breitburd, and Guillermo Marshall. Mathematical modeling of human glioma growth based on brain topological structures: study of two clinical cases. *PloS one*, 7(6):e39616, 2012.
- [407] Dong Sui, Maozu Guo, Xiaoxuan Ma, Julian Baptiste, and Lei Zhang. Imaging biomarkers and gene expression data correlation framework for lung cancer radio-genomics analysis based on deep learning. *IEEE Access*, 9:125247–125257, 2021.
- [408] V Z Sun, K P Hand, K M Stack, K A Farley, S Milkovick, R Kronyak, J I Simon, K Hickman-Lewis, D Shuster, J F Bell III, et al. Exploring the Jezero Crater floor: Overview of results from the Mars 2020 Perseverance Rover’s first science campaign. Technical report, 53rd Lunar and Planetary Science Conference, 2022.
- [409] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proc. of the European Conference on Computer Vision*, pages 480–496, 2018.
- [410] Aravind Sundaresan, Brian Burns, Indranil Sur, Yi Yao, Xiao Lin, and Sujeong Kim. Human body model based id using shape and pose parameters. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9. IEEE, 2023.
- [411] Rachel L Surprenant, James G Gehling, and Mary L Droser. Biological and ecological insights from the preservational variability of funisia dorothea, ediacara member, south australia. *Palaaios*, 35(9):359–376, 2020.
- [412] Kristin R Swanson, Carly Bridge, JD Murray, and Ellsworth C Alvord Jr. Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. *Journal of the neurological sciences*, 216(1):1–10, 2003.
- [413] Wojciech Szopa, Thomas A Burley, Gabriela Kramer-Marek, Wojciech Kaspera, et al. Diagnostic and therapeutic biomarkers in glioblastoma: current status and future perspectives. *BioMed research international*, 2017, 2017.
- [414] Birra Taha, Taihui Li, Daniel Boley, Clark C Chen, and Ju Sun. Detection of isocitrate dehydrogenase mutated glioblastomas through anomaly detection analytics. *Neuro-surgery*, 2021.
- [415] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Trans. on Computer Vision and Applications*, 2018.
- [416] Ahmad Faleh Tamimi. Epidemiology and outcome of glioblastoma, Sep 2017.
- [417] Ahmad Faleh Tamimi and Malik Juweid. Epidemiology and outcome of glioblastoma. In *Glioblastoma*. Codon Publications, 2017.
- [418] Martin L Tanaka, W Debinski, and IK Puri. Hybrid mathematical model of glioma progression. *Cell proliferation*, 42(5):637–646, 2009.

- [419] Guoliang Tang, Zhijing Liu, and Jing Xiong. Distinctive image features from illumination and scale invariant keypoints. *Multimedia Tools and Applications*, 78:23415–23442, 2019.
- [420] Lidya G Tarhan, Mary L Droser, James G Gehling, and Matthew P Dzaugis. Microbial mat sandwiches and other anactualistic sedimentary features of the Ediacara Member (Rawnsley Quartzite, South Australia. *PALAIOS*, 32(3), 2017.
- [421] Sampada Tavse, Vijayakumar Varadarajan, Mrinal Bachute, Shilpa Gite, and Ketan Kotecha. A systematic literature review on applications of gan-synthesized images for brain mri. *Future Internet*, 14(12):351, 2022.
- [422] Atsushi Teramoto, Tetsuya Tsukamoto, Ayumi Yamada, Yuka Kiriya, Kazuyoshi Imaizumi, Kuniaki Saito, and Hiroshi Fujita. Deep learning approach to classification of lung cytological images: Two-step training using actual and synthesized images by progressive growing of generative adversarial networks. *PloS one*, 15(3):e0229951, 2020.
- [423] Hoang Thanh-Tung and Truyen Tran. Catastrophic forgetting and mode collapse in gans. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10. IEEE, 2020.
- [424] David R Thompson and Rebecca Castano. Performance comparison of rock detection algorithms for autonomous planetary geology. In *2007 IEEE Aerospace Conference*, pages 1–9. IEEE, 2007.
- [425] Niklas Thon, Simone Kreth, and Friedrich-Wilhelm Kreth. Personalized treatment strategies in glioblastoma: Mgmt promoter methylation status. *OncoTargets and therapy*, 6:1363, 2013.
- [426] J Tian and D-M Chen. Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. *International Journal of Remote Sensing*, 28(20):4625–4644, 2007.
- [427] Yaolin Tian, Xue Wan, Aodi Wu, and Guangyuan Zhao. Scene aware semantic crack segmentation from oblique drone imagery. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 585–592. IEEE, 2022.
- [428] Eric Timmer, Calla Knudson, and Murray Gingras. Applying deep learning for identifying bioturbation from core photographs. *AAPG Bulletin*, 105(4):631–638, 2021.
- [429] Simona Tomaselli, Veronica Esposito, Paolo Vangone, Nico AJ van Nuland, Alexandre MJJ Bonvin, Remo Guerrini, Teodorico Tancredi, Piero A Temussi, and Delia Picone. The α -to- β conformational transition of alzheimer’s $\alpha\beta$ -(1–42) peptide in aqueous media is reversible: a step by step conformational analysis suggests the location of β conformation seeding. *ChemBioChem*, 7(2):257–267, 2006.

- [430] Katarzyna Tomczak, Patrycja Czerwińska, and Maciej Wiznerowicz. The cancer genome atlas (tcga): an immeasurable source of knowledge. *Contemporary Oncology*, 19(1A):A68, 2015.
- [431] Betty C Tong and David H Harpole. Molecular markers for incidence, prognosis, and response to therapy. *Surgical Oncology Clinics*, 21(1):161–175, 2012.
- [432] Mirko Torrisi, Gianluca Pollastri, and Quan Le. Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal*, 18:1301–1310, 2020.
- [433] Eleftherios Trivizakis, Georgios Z Papadakis, Ioannis Souglakos, Nikolaos Papanikolaou, Lefteris Koumakis, Demetrios A Spandidos, Aristidis Tsatsakis, Apostolos H Karantanis, and Kostas Marias. Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care. *International Journal of Oncology*, 57(1):43–53, 2020.
- [434] Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and content for video generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1526–1535, 2018.
- [435] Robert Tycko. Solid-state nmr studies of amyloid fibril structure. *Annual review of physical chemistry*, 62(1):279–299, 2011.
- [436] Victor Uc-Cetina, Laura Alvarez-Gonzalez, and Anabel Martin-Gonzalez. A review on generative adversarial networks for data augmentation in person re-identification systems. *arXiv preprint arXiv:2302.09119*, 2023.
- [437] Uddeshya Upadhyay, Viswanath P Sudarshan, and Suyash P Awate. Uncertainty-aware gan with adaptive loss for robust mri image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3255–3264, 2021.
- [438] Kaja Urbańska, Justyna Sokołowska, Maciej Szmidt, and Paweł Sysa. Glioblastoma multiforme—an overview. *Contemporary Oncology*, 18(5):307, 2014.
- [439] Silvana Valdebenito, Daniela D’Amico, and Eliseo Eugenin. Novel approaches for glioblastoma treatment: Focus on tumor heterogeneity, treatment resistance, and computational tools. *Cancer Reports*, 2(6):e1220, 2019.
- [440] Bas HM Van der Velden, Hugo J Kuijf, Kenneth GA Gilhuijs, and Max A Viergever. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical Image Analysis*, 79:102470, 2022.
- [441] Ruben Villegas, Jimei Yang, Seunghoon Hong, Xunyu Lin, and Honglak Lee. Decomposing motion and content for natural video sequence prediction. *arXiv preprint arXiv:1706.08033*, 2017.

- [442] Dimitris Visvikis, Catherine Cheze Le Rest, Vincent Jaouen, and Mathieu Hatt. Artificial intelligence, machine (deep) learning and radio (geno) mics: definitions and nuclear medicine imaging applications. *European journal of nuclear medicine and molecular imaging*, 46(13):2630–2637, 2019.
- [443] W Patrick Walters and Regina Barzilay. Applications of deep learning in molecule generation and molecular property prediction. *Accounts of chemical research*, 54(2):263–270, 2020.
- [444] Marielle Aulikki Wälti, Francesco Ravotti, Hiromi Arai, Charles G Glabe, Joseph S Wall, Anja Böckmann, Peter Güntert, Beat H Meier, and Roland Riek. Atomic-resolution structure of a disease-relevant $\alpha\beta$ (1–42) amyloid fibril. *Proceedings of the National Academy of Sciences*, 113(34):E4976–E4984, 2016.
- [445] Fangbin Wan, Yang Wu, Xuelin Qian, and Yanwei Fu. When person re-identification meets changing clothes. *arXiv preprint arXiv:2003.04070*, 2020.
- [446] Fei Wang, Rainu Kaushal, and Dhruv Khullar. Should health care demand interpretable artificial intelligence or accept “black box” medicine?, 2020.
- [447] Guan-An Wang, Tianzhu Zhang, Yang Yang, Jian Cheng, Jianlong Chang, Xu Liang, and Zeng-Guang Hou. Cross-modality paired-images generation for rgb-infrared person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12144–12151, 2020.
- [448] Guanshuo Wang, Xiong Chen, Jialin Gao, Xi Zhou, and Shiming Ge. Self-guided body part alignment with relation transformers for occluded person re-identification. *IEEE Signal Processing Letters*, 28:1155–1159, 2021.
- [449] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple granularities for person re-identification. In *Proceedings of the 26th ACM International Conference on Multimedia*, pages 274–282, 2018.
- [450] Jia-Xin Wang, Si-Bao Chen, Chris HQ Ding, Jin Tang, and Bin Luo. RanPaste: Paste consistency and pseudo label for semisupervised remote sensing image semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2021.
- [451] Jian Wang, Feng Zhou, Shilei Wen, Xiao Liu, and Yuanqing Lin. Deep metric learning with angular loss. In *Proc. of the IEEE International Conference on Computer Vision*, pages 2593–2601, 2017.
- [452] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 43(10):3349–3364, 2020.
- [453] Qingfeng Wang, Xuehai Zhou, Chao Wang, Zhiqin Liu, Jun Huang, Ying Zhou, Changlong Li, Hang Zhuang, and Jie-Zhi Cheng. Wgan-based synthetic minority

- over-sampling technique: Improving semantic fine-grained classification for lung nodules in ct images. *IEEE Access*, 7:18450–18463, 2019.
- [454] Rongguang Wang, Vishnu Bashyam, Zhijian Yang, Fanyang Yu, Vasiliki Tassopoulou, Lasya P Sreepada, Sai Spandana Chintapalli, Dushyant Sahoo, Ioanna Skampardon, Konstantina Nikita, et al. Applications of generative adversarial networks in neuroimaging and clinical neuroscience. *arXiv preprint arXiv:2206.07081*, 2022.
 - [455] Ruxin Wang, Shuyuan Chen, Chaojie Ji, Jianping Fan, and Ye Li. Boundary-aware context neural network for medical image segmentation. *Medical Image Analysis*, 78:102395, 2022.
 - [456] Xuekuan Wang, Cairong Zhao, Duoqian Miao, Zhihua Wei, Renxian Zhang, and Tingfei Ye. Fusion of multiple channel features for person re-identification. *Neurocomputing*, 213:125–136, 2016.
 - [457] Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. Understanding how dimension reduction tools work: an empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization. *The Journal of Machine Learning Research*, 22(1):9129–9201, 2021.
 - [458] Zhikang Wang, Feng Zhu, Shixiang Tang, Rui Zhao, Lihuo He, and Jiangning Song. Feature erasing and diffusion network for occluded person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4754–4763, 2022.
 - [459] Shanshan Wei, Faqiang Shi, Yuexin Wang, Yilin Chou, and Xuemin Li. A deep learning model for automated sub-basal corneal nerve segmentation and evaluation using in vivo confocal microscopy. *Translational Vision Science & Technology*, 9(2):32–32, 2020.
 - [460] Michael Weller, Jorg Felsberg, Christian Hartmann, Hilmar Berger, Joachim P Steinbach, Johannes Schramm, Manfred Westphal, Gabriele Schackert, Matthias Simon, Jorg C Tonn, et al. Molecular predictors of progression-free and overall survival in patients with newly diagnosed glioblastoma: a prospective translational study of the german glioma network. *Journal of Clinical Oncology*, 27(34):5743–5750, 2009.
 - [461] Michael Weller, Roger Stupp, Guido Reifenberger, Alba A Brandes, Martin J Van Den Bent, Wolfgang Wick, and Monika E Hegi. Mgmt promoter methylation in malignant gliomas: ready for personalized medicine? *Nature Reviews Neurology*, 6(1):39–51, 2010.
 - [462] Frances Westall. Morphological biosignatures in early terrestrial and extraterrestrial materials. *Space Science Reviews*, 135:95–114, 2008.
 - [463] Frances Westall, Frederic Foucher, Nicolas Bost, Marylene Bertrand, Damien Loizeau, Jorge L Vago, Gerhard Kminek, Frederic Gaboyer, Kathleen A Campbell, Jean-Gabriel Breheret, Pascale Gautret, and Charles S Cockell. Biosignatures on Mars:

- what, where, and how? Implications for the search for Martian life. *Astrobiology*, 15(11), 2015.
- [464] Navodini Wijethilake, Mobarakol Islam, and Hongliang Ren. Radiogenomics model for overall survival prediction of glioblastoma. *Medical & Biological Engineering & Computing*, 58:1767–1777, 2020.
 - [465] Kenneth H Williford, Kenneth A Farley, Kathryn M Stack, Abigail C Allwood, David Beaty, Luther W Beegle, Rohit Bhartia, Adrian J Brown, Manuel de la Torre Juarez, Svein-Erik Hamran, Michale H Hecht, Joel A Hurowitz, Jose A Rodriguez-Manfredi, Sylvestre Maurice, Sarah Milkovich, and Roger C Wiens. *From Habitability to Life on Mars*. Elsevier, 2018.
 - [466] Chi Wah Wong and Ammar Chaudhry. Radiogenomics of lung cancer. *Journal of Thoracic Disease*, 12(9):5104, 2020.
 - [467] Ancong Wu, Wei-Shi Zheng, Hong-Xing Yu, Shaogang Gong, and Jianhuang Lai. Rgb-infrared cross-modality person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5380–5389, 2017.
 - [468] Junde Wu, Rao Fu, Huihui Fang, Yu Zhang, Yehui Yang, Haoyi Xiong, Huiying Liu, and Yanwu Xu. Medsegdiff: Medical image segmentation with diffusion probabilistic model. In *Medical Imaging with Deep Learning*, pages 1623–1639. PMLR, 2024.
 - [469] Shangxuan Wu, Ying-Cong Chen, Xiang Li, An-Cong Wu, Jin-Jie You, and Wei-Shi Zheng. An enhanced deep feature representation for person re-identification. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–8. IEEE, 2016.
 - [470] Yu Wu, Yao Lu, Weiqiang Chen, Jianping Fu, and Rong Fan. In silico experimentation of glioma microenvironment development and anti-tumor therapy. *PLoS computational biology*, 8(2):e1002355, 2012.
 - [471] Chang Xiao, Peilin Zhong, and Changxi Zheng. Bourgan: Generative networks with metric embeddings. In *Advances in Neural Information Processing Systems*, pages 2269–2280, 2018.
 - [472] Shuhai Xiao, Zhe Chen, Chuanming Zhou, and Xunlai Yuan. Surfing in and on microbial mats: Oxygen-related behavior of a terminal ediacaran bilaterian animal. *Geology*, 47(11):1054–1058, 2019.
 - [473] Shuhai Xiao and Marc Laflamme. On the eve of animal radiation: phylogeny, ecology and evolution of the ediacara biota. *Trends in Ecology & Evolution*, 24(1):31–40, 2009.
 - [474] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403, 2015.

- [475] Chi Xu, Yasushi Makihara, Xiang Li, Yasushi Yagi, and Jianfeng Lu. Gait recognition from a single image using a phase-aware gait cycle reconstruction network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, Proc., Part XIX 16*, pages 386–403. Springer, 2020.
- [476] Jiachen Xu, Jingyu Gong, Jie Zhou, Xin Tan, Yuan Xie, and Lizhuang Ma. SceneEncoder: Scene-aware semantic segmentation of point clouds with a learnable scene descriptor. *arXiv preprint arXiv:2001.09087*, 2020.
- [477] Jingjing Xu, Xuancheng Ren, Junyang Lin, and Xu Sun. Dp-gan: diversity-promoting generative adversarial network for generating informative and diversified text. *arXiv preprint arXiv:1802.01345*, 2018.
- [478] Ziyue Xu, Xiaosong Wang, Hoo-Chang Shin, Dong Yang, Holger Roth, Fausto Milletari, Ling Zhang, and Daguang Xu. Correlation via synthesis: End-to-end image generation and radiogenomic learning based on generative adversarial network. In *Medical Imaging with Deep Learning*, pages 857–866. PMLR, 2020.
- [479] Ankit Yadav and Dinesh Kumar Vishwakarma. Deep learning algorithms for person re-identification: state-of-the-art and research challenges. *Multimedia Tools and Applications*, 83(8):22005–22054, 2024.
- [480] Hai Yan, D Williams Parsons, Genglin Jin, Roger McLendon, B Ahmed Rasheed, Weishi Yuan, Ivan Kos, Ines Batinic-Haberle, Siân Jones, Gregory J Riggins, et al. Idh1 and idh2 mutations in gliomas. *New England journal of medicine*, 360(8):765–773, 2009.
- [481] Zengqiang Yan, Xin Yang, and Kwang-Ting Cheng. A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 23(4):1427–1436, 2018.
- [482] Jianwei Yang, Anitha Kannan, Dhruv Batra, and Devi Parikh. Lr-gan: Layered recursive generative adversarial networks for image generation. *arXiv preprint arXiv:1703.01560*, 2017.
- [483] Jinzhu Yang, Shuang Ma, Qi Sun, Wenjun Tan, Mengjia Xu, Nan Chen, and Dazhe Zhao. Improved Hessian multiscale enhancement filter. *Bio-medical Materials and Engineering*, 24(6):3267–3275, 2014.
- [484] Juntao Yang, Zhizhong Kang, Ze Yang, Juan Xie, Bin Xue, Jianfeng Yang, and Jinyou Tao. A laboratory open-set Martian rock classification method based on spectral signatures. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.
- [485] Qize Yang, Ancong Wu, and Wei-Shi Zheng. Person re-identification by contour sketch under moderate clothing change. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2029–2046, 2019.

- [486] Yang Yang, Diana Arseni, Wenjuan Zhang, Melissa Huang, Sofia Lövestam, Manuel Schweighauser, Abhay Kotecha, Alexey G Murzin, Sew Y Peak-Chew, Jennifer Macdonald, et al. Cryo-em structures of amyloid- β 42 filaments from human brains. *Science*, 375(6577):167–172, 2022.
- [487] Zhengwei Yang, Meng Lin, Xian Zhong, Yu Wu, and Zheng Wang. Good is bad: Causality inspired cloth-debiasing for cloth-changing person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1472–1481, 2023.
- [488] Zhengwei Yang, Xian Zhong, Zhun Zhong, Hong Liu, Zheng Wang, and Shin’Ichi Satoh. Win-win by competition: Auxiliary-free cloth-changing person re-identification. *IEEE Transactions on Image Processing*, 32:2985–2999, 2023.
- [489] Zhenzhen Yang, Jing Shao, and Yongpeng Yang. An improved cyclegan for data augmentation in person re-identification. *Big Data Research*, 34:100409, 2023.
- [490] Lingxiang Yao, Worapan Kusakunniran, Qiang Wu, and Jian Zhang. Gait recognition using a few gait frames. *PeerJ Computer Science*, 7:e382, 2021.
- [491] Paul H Yi and Jan Fritz. Radiology alchemy: Gan we do it? *Radiology: Artificial Intelligence*, 3(5):e210125, 2021.
- [492] Zhike Yi, Tao Chang, Shuai Li, Ruijun Liu, Jing Zhang, and Aimin Hao. Scene-aware deep networks for semantic segmentation of images. *IEEE Access*, 7:69184–69193, 2019.
- [493] CGB Yogananda, Bhavya Ramesh Shah, SS Nalawade, GK Murugesan, FF Yu, MC Pinho, BC Wagner, B Mickey, Toral R Patel, B Fei, et al. Mri-based deep-learning method for determining glioma mgmt promoter methylation status. *American Journal of Neuroradiology*, 42(5):845–852, 2021.
- [494] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5505–5514, 2018.
- [495] Kai Yu, Yuqiong Hu, Fan Wu, Qiufang Guo, Zenghui Qian, Waner Hu, Jing Chen, Kuanyu Wang, Xiaoying Fan, Xinglong Wu, et al. Surveying brain tumor heterogeneity by single-cell rna-sequencing of multi-sector biopsies. *National Science Review*, 7(8):1306–1318, 2020.
- [496] Xiaohui Yuan, Jianfang Shi, and Lichuan Gu. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, 169:114417, 2021.
- [497] Yuan Yuan, Jian’an Zhang, and Qi Wang. Deep gabor convolution network for person re-identification. *Neurocomputing*, 378:387–398, 2020.

- [498] Chaoyi Zhang, Yang Song, Sidong Liu, Scott Lill, Chenyu Wang, Zihao Tang, Yuyi You, Yang Gao, Alexander Klistorner, Michael Barnett, et al. Ms-gan: Gan-based semantic segmentation of multiple sclerosis lesions in brain magnetic resonance imaging. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8. IEEE, 2018.
- [499] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318*, 2018.
- [500] Lei Zhang, Fan Yang, Yimin Daniel Zhang, and Ying Julie Zhu. Road crack detection using deep convolutional neural network. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3708–3712. IEEE, 2016.
- [501] Quanshi Zhang and Song-Chun Zhu. Visual interpretability for deep learning: a survey. *arXiv preprint arXiv:1802.00614*, 2018.
- [502] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.
- [503] Xiong Zhang, Zefang Han, Hong Shangguan, Xinglong Han, Xueying Cui, and Anhong Wang. Artifact and detail attention generative adversarial networks for low-dose ct denoising. *IEEE Transactions on Medical Imaging*, 40(12):3901–3918, 2021.
- [504] Yi Zhang, Pengliang Ji, Angtian Wang, Jieru Mei, Adam Kortylewski, and Alan Yuille. 3d-aware neural body fitting for occlusion robust 3d human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9399–9410, 2023.
- [505] Ying Zhang and Shutao Li. Gabor-lbp based region covariance descriptor for person re-identification. In *2011 Sixth International Conference on Image and Graphics*, pages 368–371. IEEE, 2011.
- [506] Yu Zhang, Peter Tiño, Aleš Leonardis, and Ke Tang. A survey on neural network interpretability. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2021.
- [507] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.
- [508] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Relation-aware global attention for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3186–3195, 2020.
- [509] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2881–2890, 2017.

- [510] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3586–3593, 2013.
- [511] Yan Zhao, Baoqiang Ma, Pengbo Jiang, Debin Zeng, Xuotong Wang, and Shuyu Li. Prediction of alzheimer’s disease progression with multi-information generative adversarial network. *IEEE Journal of Biomedical and Health Informatics*, 25(3):711–719, 2020.
- [512] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. Pluralistic image completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1438–1447, 2019.
- [513] Chuanxia Zheng, TatJen Cham, and Jianfei Cai. Pluralistic free-form image completion. *International Journal of Computer Vision*, 129(10):2786–2805, 2021.
- [514] Jinkai Zheng, Xinchun Liu, Wu Liu, Lingxiao He, Chenggang Yan, and Tao Mei. Gait recognition in the wild with dense 3d representations and a benchmark. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
- [515] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proc. of the IEEE International Conference on Computer Vision*, pages 1116–1124, 2015.
- [516] Yuanjie Zheng, Xiaodan Sui, Yanyun Jiang, Tontong Che, Shaoting Zhang, Jie Yang, and Hongsheng Li. Symreg-gan: symmetric image registration with generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [517] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
- [518] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Object detectors emerge in deep scene CNNs. *arXiv preprint arXiv:1412.6856*, 2014.
- [519] Wujie Zhou, Xinyang Lin, Jingsheng Lei, Lu Yu, and Jenq-Neng Hwang. MFFENet: Multiscale feature fusion and enhancement network for rgb–thermal urban road scene parsing. *IEEE Transactions on Multimedia*, 24:2526–2538, 2021.
- [520] Xiaoli Zhou and Bir Bhanu. Integrating face and gait for human recognition at a distance in video. *IEEE Trans. on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1119–1137, 2007.
- [521] Xiaoli Zhou and Bir Bhanu. Feature fusion of side face and gait for video-based human identification. *Pattern Recognition*, 41(3):778–795, 2008.

- [522] Haidong Zhu, Wanrong Zheng, Zhaoheng Zheng, and Ram Nevatia. Sharc: Shape and appearance recognition for person identification in-the-wild. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6290–6300, 2024.
- [523] Haidong Zhu, Zhaoheng Zheng, and Ram Nevatia. Gait recognition using 3-d human body shape inference. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 909–918, 2023.
- [524] Junjie Zhu, Zhengxin Li, Haowei Tong, Zhouyu Lu, Ningjie Zhang, Ting Wei, and Hai-Feng Chen. Phanto-idp: compact model for precise intrinsically disordered protein backbone generation and enhanced sampling. *Briefings in Bioinformatics*, 25(1):bbad429, 2024.
- [525] Zhe Zhu, Ehab Albadawy, Ashirbani Saha, Jun Zhang, Michael R Harowicz, and Maciej A Mazurowski. Deep learning for identifying radiogenomic associations in breast cancer. *Computers in biology and medicine*, 109:85–90, 2019.
- [526] Pascal O Zinn, Bhanu Mahajan, Pratheesh Sathyan, Sanjay K Singh, Sadhan Majumder, Ferenc A Jolesz, and Rivka R Colen. Correction: Radiogenomic mapping of edema/cellular invasion mri-phenotypes in glioblastoma multiforme. *PLoS One*, 7(2), 2012.
- [527] Pascal O Zinn, Bhanu Majadan, Pratheesh Sathyan, Sanjay K Singh, Sadhan Majumder, Ferenc A Jolesz, and Rivka R Colen. Radiogenomic mapping of edema/cellular invasion mri-phenotypes in glioblastoma multiforme. *PloS one*, 6(10):e25451, 2011.
- [528] Nan Zou, Zhiyu Xiang, Yiman Chen, Shuya Chen, and Chengyu Qiao. Boundary-aware CNN for semantic segmentation. *IEEE Access*, 7:114520–114528, 2019.
- [529] Qin Zou, Zheng Zhang, Qingquan Li, Xianbiao Qi, Qian Wang, and Song Wang. Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE Transactions on Image Processing*, 28(3):1498–1512, 2018.