Title

Synthetic Inflammation Imaging with PatchGAN Deep Learning Networks

Permalink

https://escholarship.org/uc/item/8xb9s0fj

Journal

Bioengineering, 10(5)

ISSN

0178-2029

Authors

Tolpadi, Aniket A
Luitjens, Johanna
Gassert, Felix G
et al.

Publication Date

2023

DOI

10.3390/bioengineering10050516

Peer reviewed

Article

Synthetic Inflammation Imaging with PatchGAN Deep Learning Networks

Aniket A. Tolpadi [1,2,*] , **Johanna Luitjens** [2,3] , **Felix G. Gassert** [2,4] , **Xiaojuan Li** [5] , **Thomas M. Link** [2] , **Sharmila Majumdar** [2] and **Valentina Pedoia** [2]

1   Department of Bioengineering, University of California, Berkeley, CA 94720, USA
2   Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA 94158, USA
3   Department of Radiology, Klinikum Großhadern, Ludwig-Maximilians-Universität, 81377 Munich, Germany
4   Department of Radiology, Klinikum Rechts der Isar, School of Medicine, Technical University of Munich, 81675 Munich, Germany
5   Department of Biomedical Imaging, Cleveland Clinic, Cleveland, OH 44106, USA
*   Correspondence: aniket.tolpadi@ucsf.edu

Abstract: Background: Gadolinium (Gd)-enhanced Magnetic Resonance Imaging (MRI) is crucial in several applications, including oncology, cardiac imaging, and musculoskeletal inflammatory imaging. One use case is rheumatoid arthritis (RA), a widespread autoimmune condition for which Gd MRI is crucial in imaging synovial joint inflammation, but Gd administration has well-documented safety concerns. As such, algorithms that could synthetically generate post-contrast peripheral joint MR images from non-contrast MR sequences would have immense clinical utility. Moreover, while such algorithms have been investigated for other anatomies, they are largely unexplored for musculoskeletal applications such as RA, and efforts to understand trained models and improve trust in their predictions have been limited in medical imaging. **Methods**: A dataset of 27 RA patients was used to train algorithms that synthetically generated post-Gd IDEAL wrist coronal T_1-weighted scans from pre-contrast scans. UNets and PatchGANs were trained, leveraging an anomaly-weighted L_1 loss and global generative adversarial network (GAN) loss for the PatchGAN. Occlusion and uncertainty maps were also generated to understand model performance. **Results**: UNet synthetic post-contrast images exhibited stronger normalized root mean square error (nRMSE) than Patch-GAN in full volumes and the wrist, but PatchGAN outperformed UNet in synovial joints (UNet nRMSEs: volume = 6.29 ± 0.88, wrist = 4.36 ± 0.60, synovial = 26.18 ± 7.45; PatchGAN nRMSEs: volume = 6.72 ± 0.81, wrist = 6.07 ± 1.22, synovial = 23.14 ± 7.37; n = 7). Occlusion maps showed that synovial joints made substantial contributions to PatchGAN and UNet predictions, while uncertainty maps showed that PatchGAN predictions were more confident within those joints. **Conclusions**: Both pipelines showed promising performance in synthesizing post-contrast images, but PatchGAN performance was stronger and more confident within synovial joints, where an algorithm like this would have maximal clinical utility. Image synthesis approaches are therefore promising for RA and synthetic inflammatory imaging.

Keywords: image synthesis; inflammatory imaging; deep learning; rheumatoid arthritis; magnetic resonance imaging

1. Introduction

Rheumatoid arthritis (RA) is a widespread autoimmune disorder observed in 0.5–1.0% of the American population, with incidence rates being two to three times higher in women than in men [1]. RA mainly affects the joints, typically the hands and feet, and is characterized by synovial joint inflammation. In the joints it can lead to bone tissue erosions and soft tissue breakdown, often inducing stiffness and debilitating pain, but may also show systemic effects in the skin, heart or lungs if left untreated [2]. It is typically diagnosed

through a holistic assessment that begins with a medical history examination, paying particular attention to pain, swelling, peripheral joint pain, and swelling/tenderness, all of which can be indicative of RA. Furthermore, laboratory tests for rheumatoid factor (RF), C-reactive protein (CRP), and erythrocyte sedimentation rate (ESR) are often performed to confirm other RA indications. Lastly, medical imaging plays a crucial role in distinguishing inflammatory phenotypes, providing additional evidence to confirm RA [3]. Once diagnosed, RA is usually treated with Disease-Modifying Anti-Rheumatic Drugs (DMARDs), which see 75–80% of patients attain intended treatment outcomes, but 90% when initiated in the early stages of RA [4]. Robust tools such as imaging are thus necessary for screening and diagnosing RA at early stages, maximizing the odds of successful treatment.

Radiographs have traditionally been the clinical standard imaging modality for RA diagnosis, as their acquisition is quick, inexpensive, and widely accessible, yielding two-dimensional images that are effective in visualizing late-stage bone erosions [5]. In recent years, however, Magnetic Resonance Imaging (MRI) has gained prominence despite its higher costs and longer acquisition time, producing three-dimensional anatomic images with excellent depiction of soft tissues and sharp details [6]. As a result, it has emerged as a superior option for visualizing early-stage bone erosions and bone marrow edema (BME) that can result from RA [7]. An added advantage of MR is the ability to administer contrast agents such as Gadolinium (Gd) prior to scans, altering the magnetic properties of underlying tissue to improve the visualization of numerous pathologies [8]. In RA imaging, a post-contrast Gd MRI can better distinguish active soft tissue RA sites in joints, such as synovitis, from general effusion [9], conveying critical information that conventional MRI cannot provide [10]. However, Gd administration has long-term concerns such as deposition in brain and bone [11,12], is contra-indicated in patient subgroups such as those with renal diseases and pregnant women [13], and, more generally, adds scan time, cost, and patient discomfort to the imaging protocol. As such, if post-contrast MR images could be synthetically generated without Gd administration, the implications for RA diagnosis and other musculoskeletal (MSK) inflammatory conditions or even sarcomas would be significant.

The problem posed by this clinical context is one of "image synthesis", or the designing of algorithms to generate images from some input. While these inputs can be multimodal, including text or patches of images, the focus here will be on synthesis algorithms that accept full image inputs [14,15]. For image synthesis tasks, deep learning (DL), and particularly convolutional neural networks (CNNs) [16], have taken on an outsized role in recent years. When trained with sufficiently large datasets, CNN filters can be optimized for a given task, with filters in early network layers typically being sensitive to generic features such as edges, while those in later layers are typically sensitive to far more complex, task-specific features [17]. The UNet is a commonly used image synthesis algorithm in which inputted images are encoded by convolutional filters into a low-resolution, high-dimensional representation that is decoded using deconvolutional filters, yielding an output image. Originally designed for segmentation, the UNet has seen substantial application in image synthesis for its ease of training and relatively low dataset size requirements compared to other DL approaches [18]. Another prominent approach is generative adversarial networks (GANs), where an image-to-image translation network such as a UNet ("generator") is paired with a discriminator network that is trained to distinguish between synthetic and real images [19]. By setting up training as a min-max game in which generator and discriminator networks continually try to fool one another, substantially sharper images can be obtained, although GANs are more difficult to train and are prone to hallucinating artifacts compared to conventional approaches [20]. Other approaches such as variational autoencoders (VAEs) and transformer networks have been investigated in this space [21,22].

These methods have seen considerable application for medical imaging tasks. In brain MRI, image synthesis has been studied for the reduction or elimination of the Gd dosage required for post-contrast tumor imaging. In several studies, standard UNet or encoder-

decoder style architectures accepted reduced-dose Gd post-contrast images and/or other MR sequences as inputs, were trained to predict full-dose post-contrast Gd images, and quantified model efficacy through radiologist assessment or the suitability of synthetic images for downstream tasks [23–25]. Another approach in eliminating Gd dosage for brain MRI used an innovative training scheme, training a network for tumor detection and passing convolutional feature maps from that network as inputs to a conventional image synthesis architecture. This allowed the image synthesis architecture to focus on pathologic regions when optimizing parameters to produce synthetic post-contrast images [26]. Some approaches beyond image synthesis have also been investigated to eliminate the need for Gd administration. For instance, Gd is administered in cardiac MRI to identify regions of myocardial infarction. Here, DL pipelines have been developed to accept exclusively non-contrast MR images as inputs, localize the left ventricle, extract motion-based features inherent to cardiac MRI, and integrate both to predict if a patient suffered from infarction [27,28]. On the other hand, features from non-contrast MR sequences such as synthetic MRI and diffusion weighted imaging (DWI) have proven effective in differentiating benign and metastatic retropharyngeal lymph nodes, a task that usually requires a post-contrast MRI [29]. Also worthy of mention are recent image synthesis applications in biomedical imaging outside of MRI: in histopathology, standard image synthesis generator networks have been paired with multiple discriminators to generate synthetic stained images, while in microscopy, GAN image synthesis pipelines have been applied for synthetic cell painting, identifying cellular components from brightfield microscopy images [30,31].

These works mark substantial progress, with well-validated frameworks yielding promising results on a wide variety of biomedical image synthesis tasks, including post-contrast MR image synthesis. That said, there are some clear gaps in the literature. For RA imaging, the authors are not aware of any previous work developing post-contrast MR image synthesis algorithms. Such algorithms would have immense clinical utility, synthesizing post-Gd images that could be used to identify synovitis and active inflammation sites in RA patients, while eliminating the risks associated with administering Gd. More generally, Gd is used in brain imaging to identify tumors and distinguish tumor types, while in cardiac imaging it helps identify myocardial infarction sites, among others; in MSK, however, it is administered to image inflammation. Synthetic inflammatory MSK imaging has seen little to no investigation in previous works. Particularly in comparison with brain applications, synthetic Gd dosage reduction in MSK applications, such as wrist imaging, brings about additional challenges such as severe motion artifacts, reduced signal-to-noise ratio (SNR), and considerably smaller datasets [32]. Lastly, despite all these image synthesis works in biomedical applications, efforts to understand the basis of model predictions have been limited; this work would be critical for radiologists to gain confidence in model predictions, a prerequisite for eventual clinical deployment. As such, post-contrast MSK MR image synthesis confers numerous unique challenges that must be managed methodologically, and has been largely unexplored, making it ripe for an initial proof-of-concept study.

This is precisely the niche this work seeks to fill: the purpose of this study was to develop DL pipelines that generate synthetic post-contrast wrist MR images from their pre-contrast counterparts [33], thereby marking the first known effort for synthetic MSK inflammatory imaging. We use image quality metrics to assess the diagnostic and perceptual quality of model-generated synthetic post-contrast images relative to true post-contrast images. We also generate occlusion and uncertainty maps to better understand model performance, making its predictions more trustworthy. More specifically, the contributions and novelty of our work are as follows:

1. To our knowledge, this proof-of-concept study is the first application of DL techniques for generating synthetic post-contrast images for MSK inflammatory imaging.
2. We show that our trained pipelines perform strongly with regards to predicting post-contrast image appearance, particularly in regions afflicted with synovitis, where these models would see the most clinical utility.

3. We investigate the deconvolution operator, checkerboarding artifacts that can be intrinsic to architectures that use it, and how they surface in conventional and adversarial network training schemes.

4. We conduct a rigorous analysis of model predictions, identifying regions in pre-contrast image inputs that were most important to predicted post-contrast images, and regions in which predictions were most uncertain. This provides a straightforward framework that can be used to understand predictions made by image synthesis architectures in biomedical imaging applications.

2. Materials and Methods

2.1. Study Group

All studies performed in this retrospective study were Health Insurance Portability and Accountability Act (HIPAA) compliant, approved by the UCSF Institutional Review Board (Human Research Protection Program, IRB# 12-10418) and registered under Clinical Trial NCT01773681. Informed consent was obtained from all study participants. Twenty-seven UCSF patients with RA were recruited that met the following criteria: at least 18 years old and fulfilled the 2010 ACR/EULAR criteria for the classification of RA. Patients were treated with either methotrexate or a combination of methotrexate and tumor necrosis factor alpha inhibitors (anti-TNFα) based on RA disease activity; intended sample sizes were thus as large as feasible given the exclusion criteria and the requirements of informed consent from study participants. Data was collected from patients as part of this cohort from 20 March 2014 to 8 February 2018. Patients were imaged at baseline, 3-months, and 1-year follow-up time points, conducting MR imaging, sampling serum to measure ESR, and recording clinical notes at each time point. As the dataset used in this study was from a UCSF clinical trial, data privacy and patient confidentiality concerns prevent its public release, but codes used in generating results can be obtained from the authors upon reasonable request.

2.2. MR Acquisition

All patients underwent a standardized protocol that included coronal T_1 IDEAL scans pre- and post-Gd administration on a 3.0-T wide bore scanner (MR Discovery 750w, GE Healthcare, Waukesha, WI, USA) using 8-channel HD wrist array coils (GE Healthcare, Waukesha, WI, USA). Scans were done with acquisition matrices of 384×256 (n = 58) or 256×224 (n = 6), a slice thickness of 2 mm, a TR of 457 to 793 ms, and a TE of 10.06–12.48 ms. Complete acquisition parameters for both sequences can be found in Table A1.

2.3. Anomaly Segmentations and Evaluations

In post-contrast images, synovitis was segmented in the following synovial joints: intercarpal joints, carpometacarpal joints, the radioulnar joint, and radiolunar joints. Regions with bone marrow edema (BME) were segmented in the following bones: the first to fifth metacarpals, capitate, hamate, lunate, pisiform, scaphoid, trapezium, trapezoid, triquetrum, ulna, and radius. Anomaly segmentations were performed by a radiologist with over 30 years of experience (T.L.) using the Image Processing Package (version 6.43.01) developed by the University of California, San Francisco Musculoskeletal Quantitative Imaging Research Group.

T.L. also quantified synovitis severity for each patient at each time point with the Rheumatoid Arthritis Magnetic Resonance Imaging Score (RAMRIS) for synovitis [34], a 0–9 scale in which a higher score is associated with more severe imaging findings of RA.

Lastly, bounding boxes delineating wrist tissue and background were drawn using the software MD.ai by a radiologist with two years of experience (J.L.), such that reconstruction metrics for synthetic post-Gd images could be evaluated solely in wrist tissue and not be sensitive to textures and noise in background pixels.

2.4. Image Preprocessing

Six of 64 acquired imaging volumes had slices that were 256×256 pixels, with the remainder being 512×512; the slices of these six volumes were upsampled to 512×512 using third-order b-spline interpolation. Pre-Gd volumes were then registered to post-Gd volumes with a three-step process: (1) translation, (2) affine, and (3) third order b-spline registration (maximum iterations = 256, 256, 512, respectively; Advanced Mattes Mutual Information [35] criterion for all). B-spline registration was only done for scans where the structural similarity index (SSIM) [36] between pre and post-Gd acquisitions was above 0.5; other scans had motion artifacts so severe that non-rigid registration was not possible. All registrations were performed using SimpleITK 2.0.0 in Python (version 3.7.11) [37–39]. Example slices before and after registration can be found in Figure A1. Pixel values in the slices of pre-Gd scans were scaled such that the middle 95% of pixel values were between 0 and 1. The unscaled pixel values in pre-Gd slices that corresponded to 0 and 1 in the scaled slices were also mapped to 0 and 1 in the post-Gd slices, thereby scaling post-Gd slices while preserving the relative enhancement across the volume.

2.5. Data Partitioning

The data were partitioned into training, validation, and test datasets, splitting such that all scans from a given patient were in only one of the three datasets. Furthermore, four patients without imaging findings of synovitis were in the dataset (RAMRIS synovitis of 0); splits ensured at least 1 of these patients were in each of training, validation and test. Splits were intended to maintain similar age, BMI, and ESR across the three datasets, but the relatively small overall dataset required some compromise. The full characteristics of the data splits can be found in Table 1.

Table 1. Full Dataset and Splits Information. Demographics and patient information for the entire dataset and splits into training, validation and test. All data are presented as mean $\pm$ 1 s.d. The dataset consisted of 27 patients diagnosed with RA, each of whom were scanned up to three times (baseline, 3-month, and 1-year follow-up after one of two treatments). Data splitting was done at a patient level while ensuring each of the training, validation and test datasets included at least one patient with a RAMRIS synovitis of 0. The small dataset size and splitting conditions caused slight imbalances in demographic and health variables across the splits.

	Train	**Validation**	**Test**	**Full**
Age	53.38 ± 13.50	45.94 ± 16.16	52.12 ± 18.60	52.41 ± 14.65
BMI	29.35 ± 8.90	25.32 ± 3.06	28.33 ± 1.26	28.79 ± 8.03
ESR [mm/h]	29.06 ± 26.07	32.00 ± 24.00	27.00 ± 20.12	29.05 ± 25.32
RAMRIS Synovitis	4.57 ± 2.13	2.33 ± 2.62	1.67 ± 1.25	4.00 ± 2.37
Slices	783	87	105	975
Volumes	51	6	7	64

2.6. Network Architecture

All network architectures were implemented in PyTorch (version 1.10.2). Two-dimensional UNet [18] architectures were used as image-to-image synthesizers in our approaches, accepting as input a pre-processed pre-Gd coronal T_1 IDEAL slice and outputting the corresponding synthetic post-Gd slice. A baseline UNet model was trained, and in a separate pipeline version, an identical UNet was treated as a PatchGAN generator and paired with a PatchGAN discriminator [40]. The PatchGAN discriminator accepted concatenated inputs of the pre-processed pre-Gd slice and either the corresponding synthetic post-Gd slice or the ground truth post-Gd slice, yielding a 16×16 output in which each output pixel had a corresponding receptive field "patch" in the concatenated inputs. The 16×16 outputs were trained to predict whether synthetic post-Gd generator outputs were real or synthetic. Multiple baseline UNet and PatchGAN generator versions were trained: one set in which all steps of the UNet/generator decoding path used a deconvo-

lution operator, and another in which the deconvolutions were replaced by either a 2 × 2 bilinear upsampling interpolation operator followed by a convolution [41], or just the 2 × 2 bilinear interpolation. The exact network architecture and layers can be seen in Figure 1. Weights for the UNets, UNet generators, and PatchGAN discriminators were initialized randomly to have a mean of 0 and a standard deviation of 0.02.
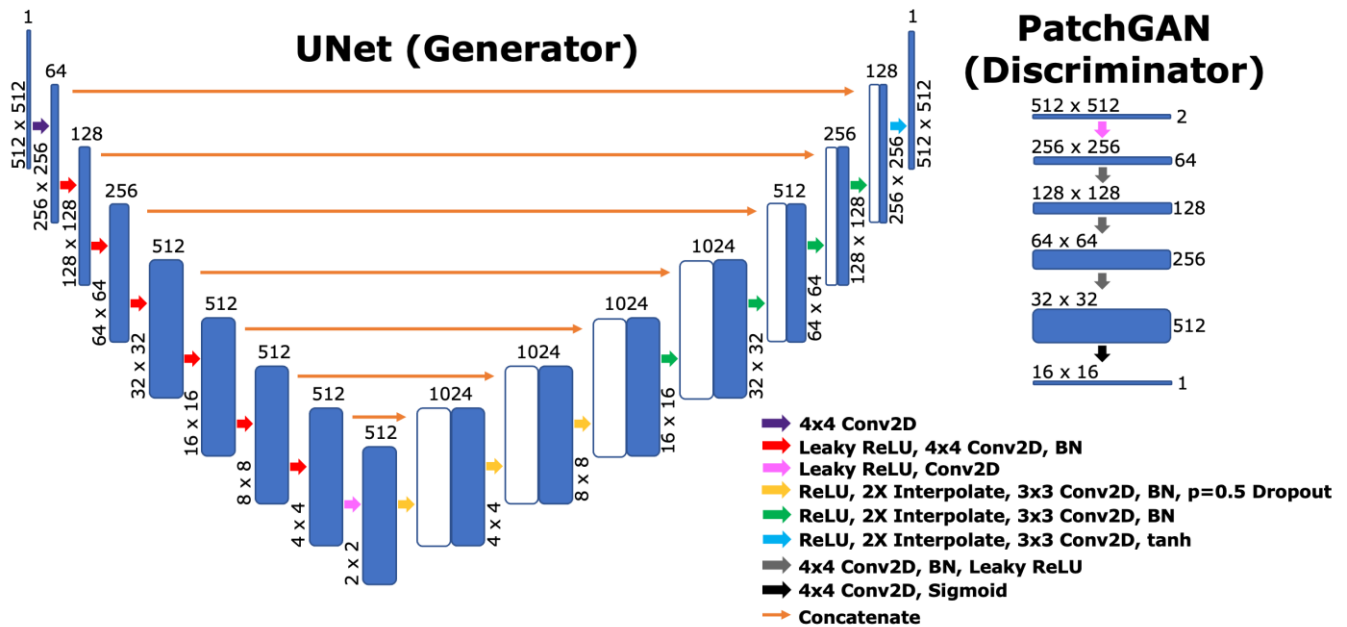

Figure 1. Network Architectures. The baseline UNet and PatchGAN generators used identical architectures, while the PatchGAN pipeline also trained a discriminator whose architecture is pictured. All generator encoding path convolutions had a stride of 2 and a padding 1, while all decoding path convolutions had a stride and padding of 1. The first three discriminator convolutions had a stride of 2 and a padding of 1, while the final two had a stride and a padding of 1. For PatchGAN and UNet pipelines with deconvolutions, all "2X interpolate, 4 × 4 Conv2D" steps would be replaced by 4 × 4 2D transposed convolutions with a stride of 2 and a padding of 1. All leaky ReLU layers had a negative slope of 0.2.

2.7. Training Details

The baseline UNets were trained with a weighted L_1 loss, as shown below in Equation (1), with loss function variables as follows: n = number of samples; S_i = anomaly segmentation mask for slice i; $\hat{y}_i$ = synthetic post-Gd image slice; y_i = ground truth post-Gd slice. The anomaly segmentation mask S_i used to weight the L_1 loss was calculated as follows: anomaly segmentations were turned into binary masks, any pixel more than 20 pixels from the nearest anomaly was set to a background value λ_B, pixels within anomalies were set to 1, and intermediate pixels were set to a range from λ_B to 1 based on their Euclidean distance from an anomaly segmentation. A sample distance map can be found in Figure A2.

$$L_{UNet} = \frac{1}{n}\sum_{i=0}^{n} S_i(\hat{y}_i - y_i);$$ (1)

On the other hand, PatchGAN generators were trained with the same weighted L_1 loss and a GAN loss, as shown in Equation (2), while PatchGAN discriminators were trained with the loss function shown in Equation (3). Additional variables for these loss functions are as follows: x_i = pre-Gd image slice; $D(a, b)$ = PatchGAN discriminator output for concatenated inputs a and b; λ_{L_1} = anomaly-weighted L_1 loss weighting for generator; λ_{GAN} = discriminator loss weighting for generator. With this loss function setup, the discriminator was trained to predict values of 1 when fed ground truth data and 0 when fed generator predictions, while the generator was trained to do the opposite. For any training

batch, the following scheme was followed: (1) synthetic post-Gd generator predictions were calculated; (2) pre-Gd, synthetic post-Gd, and ground truth post-Gd images were used to calculate L_{Dis} and update discriminator parameters; (3) synthetic post-Gd generator predictions and corresponding discriminator outputs were recalculated with new model parameters, L_{Gen} was calculated, and generator parameters were updated; (4) steps (1) and (2) were repeated again to update the discriminator parameters. This approach of two discriminator steps and one generator step per training batch was empirically useful in yielding similar generator and discriminator strength during training.

$$L_{Gen} = \frac{1}{n}\sum\nolimits_{i=0}^{n} \lambda_{L_1} S_i(\hat{y}_i - y_i) - \lambda_{GAN} \log D(x_i, \hat{y}_i); \tag{2}$$

$$L_{Dis} = \frac{1}{2n}\sum\nolimits_{i=0}^{n} \log D(x_i, \hat{y}_i) - \log D(x_i, y_i). \tag{3}$$

Baseline UNets, PatchGAN generators, and PatchGAN discriminators were all trained with a learning rate of 0.001, an Adam optimizer (β_1 = 0.5, β_2 = 0.999), and batch size of 1 to ensure that full batches fit on a single GPU [42]. All pipelines were trained on an NVIDIA Titan Xp 12 GB GPU. For baseline UNet and PatchGAN generator inputs, the following augmentations were done on the training set, each with a probability 0.5: [−2,2] degree random rotation, [−10,10] pixel random translation along both directions in a slice, [−5,5] percent random zoom, and Gaussian noise addition with a mean of 0 and standard deviation of 0.02. Training was done in two stages: initially for 10 epochs in a hyperparameter search to optimize λ_{GAN} and λ_B (more thoroughly described in the following subsection), and finally for 35 epochs with optimized parameters. With 783 pairs of pre and post-Gd slices seen in the training set, this means that 27,405 total slices were seen by all selected models during training (3045 additional slices for validation).

2.8. Hyperparameter Search and Model Selection

For each of the four pipelines trained (UNet and PatchGAN, both with and without deconvolutions), grid hyperparameter searches were carried out to optimize the background pixel weighting in segmentation distance maps (0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2) and λ_{GAN} (0.001–0.01, spaced by 0.001). λ_{L_1} was held constant at 1 for all searches. In hyperparameter searches, models were trained for 10 epochs and model performances were evaluated on the validation set. The most promising parameter set for each of the four pipelines was then trained from scratch for 35 epochs to yield the final models.

The selection of optimal parameter sets was done through a combination of standard reconstruction metrics and visual inspection. For each of the four pipelines, SSIM and normalized root mean square error (nRMSE) were used to screen for top candidate models, whose performance on the validation set was then assessed by visual inspection. The primary criteria for evaluating model performance were (1) the synthesis of new information not obvious from pre-Gd scans, (2) the preservation of sharp textures in synthetic post-Gd scans compared to ground truth post-Gd scans, and (3) the absence of obvious algorithm-generated artifacts that may cause a radiologist to lose confidence in the reconstructed image quality.

2.9. Model Performance Evaluation

The assessment of whether to use or omit the deconvolutions in the UNet decoding path was done visually for the UNet and PatchGAN approaches; the best performing models for both methods were then used for a more rigorous analysis. The quantitative assessment of synthetic post-Gd image quality was performed using three standard reconstruction metrics: SSIM, nRMSE, and peak signal-to-noise ratio (PSNR) [43]. Due to the slight misregistration of corresponding slices that may have been present even after previous preprocessing, metrics were presented both with and without slice-wise registration: ((1) 256-iteration translation, (2) 256-iteration affine, and then (3) 512-iteration

third order b-spline with a transformation bending penalty of 500, all with the Advanced Mattes Mutual Information criterion). The slice-wise registration was solely for the calculation of model performance metrics; only unregistered model outputs are presented in figures. The reconstruction metrics were evaluated per-volume in the following regions: full imaging volumes, wrist anatomy bounding boxes, and synovial joints. While these metrics do not correlate well with gold-standard radiologist annotations when evaluated on full image volumes or slices, they are widely used in the image reconstruction and image synthesis literature, and thus facilitate easy comparison of model performance with those performing similar tasks [44,45]. Furthermore, our dataset affords us wrist and anomaly bounding boxes; the calculation of these metrics specifically in these regions—one discarding background, and another focusing specifically on tissues of highest clinical interest when administering Gadolinium—can overcome the limitations of these metrics when used conventionally, affording them more clinical significance.

2.10. Enhancement Maps

For UNet, PatchGAN, and ground truth post-Gd images, pixels among the top 10% in predicted signal enhancement were identified. Enhancement maps were shown as follows: pre-Gd slice, post-Gd slice, and post-Gd slice with the degree of enhancement overlaid for the most enhancing pixels (top 10%), colored by the predicted extent of the enhancement. For visual consistency, colormap ranges for the enhancement map were calculated with respect to the enhancement observed in ground truth, with the same ranges being used for the maps regardless of algorithmic approach.

2.11. Occlusion Maps

For each slice, pre-contrast IDEAL T_1 images were pre-processed using previously described techniques, which were used as inputs for UNet and PatchGAN generator architectures, generating network outputs. The pixel values were then set to zero in a 32×32 occlusion, and the occluded image was fed through the same architecture, recording the absolute difference in predicted pixel magnitude as compared to the unoccluded image. This procedure was repeated for all 32×32 occlusions throughout the slice (with a stride length of 8), summing up the predicted changes in pixel magnitudes in an aggregate array and dividing each pixel by the number of occlusions in which it was contained. The aggregate array values were then min-max normalized, divided by pre-contrast IDEAL T_1 pixel values (to incorporate into resulting maps information for regions other than areas of high pixel intensity), and again min-max normalized, yielding occlusion maps. For display purposes, the maps are thresholded such that only the top 5% of the occlusion map magnitudes were visualized.

2.12. Uncertainty Maps

The uncertainty maps of the model predictions were generated by corrupting the latent representations of a given slice [46]. Namely, for 100 iterations, Gaussian noise with a mean of 0 and a standard deviation of 0.5 was added to the encoding path outputs at each of the eight levels (seven layers that were concatenated to the corresponding decoding path levels and the bottom of the encoder). The variance of the predicted pixel intensities from these 100 perturbed latent spaces was then calculated, min-max normalized, and thresholded for display purposes such that only the 15% most variant pixels would display, thereby generating uncertainty maps for each slice.

2.13. Statistical Analysis

To assess if synthetic post-Gd scans provided significant improvements over baseline pre-Gd images, 2-sample *t*-tests [47] were conducted. On a per-scanned-volume-basis, these tests compared the metrics of model outputs (nRMSE, SSIM, PSNR) to those of the pre-Gd scanned volumes; a Bonferroni correction [48] was applied when necessary to adjust for multiple comparisons.

3. Results

The hyperparameter search results are presented on the validation set, which was used to select optimal values for λ_{GAN} and λ_B in training loss functions. The results from finalized models are presented on the test set, on which finalized models were run just one time. Key demographic information on the test set is available in Table 1.

3.1. Model Parameter Selection

The reconstruction performance metrics evaluating the similarity of the synthetic post-Gd model outputs to ground truth were calculated for all 70 tested hyperparameter combinations for each of the four model type configurations (PatchGAN and baseline UNet, with and without decoding path deconvolutions). Sample results are shown for PatchGAN without generator deconvolutions for SSIMs in Table A2, and for nRMSEs in Table A3. Hyperparameter combinations with strong performances in either approach were carried onto a visual inspection of post-Gd synthesis performance, an example of which is shown for several hyperparameter combinations in Figure A3, also for the PatchGAN without deconvolutions. Hyperparameters associated with the selected best models through this process are listed below:

1. PatchGAN, no deconvolutions: $\lambda_B = 0.05$, $\lambda_{GAN} = 0.01$;
2. PatchGAN, with deconvolutions: $\lambda_B = 0.15$, $\lambda_{GAN} = 0.001$;
3. UNet, no deconvolutions: $\lambda_B = 0.05$;
4. UNet, with deconvolutions: $\lambda_B = 0.15$.

3.2. Utility of Deconvolution Operators in Baseline UNet and PatchGAN Generator Decoders

A comparison of sample synthetic post-Gd slices with and without deconvolutions in the UNet decoding path can be found in Figure 2, while a comparison of synthetic post-Gd slices with and without deconvolutions in the PatchGAN generator decoding path can be found in Figure 3. In baseline UNet pipelines, checkerboarding artifacts were apparent when deconvolutions were used, particularly in regions of relatively homogenous pixel values, such as the muscles around the radius and ulna. When those deconvolutions were replaced by 2×2 upsampling and standard convolutions, the checkerboarding artifacts were largely absent. These checkerboarding artifacts were less apparent in both PatchGAN pipelines, but in the version that used deconvolutions, they were evident at the extended boundaries of sharp changes in pixel intensities. Checkerboarding was thus best avoided by PatchGAN and UNet pipelines without deconvolutions, and these pipeline versions were selected as top-performing pipelines for both approaches in the remaining experiments.

3.3. Standard Reconstruction Metrics Performance

Standard reconstruction metrics across the test set are shown in Table 2 for full imaging volumes, wrist volumes, and synovial joints. Both synthetic post-Gd volumes had showed significant improvements over pre-Gd volumes in PSNR and nRMSE, with the baseline UNet pipeline also showing significantly higher SSIM. While the UNet baseline model showed stronger performance in all metrics within full volumes and the wrist, the PatchGAN showed stronger reconstruction performance in synovial joints when measured by nRMSE and PSNR.
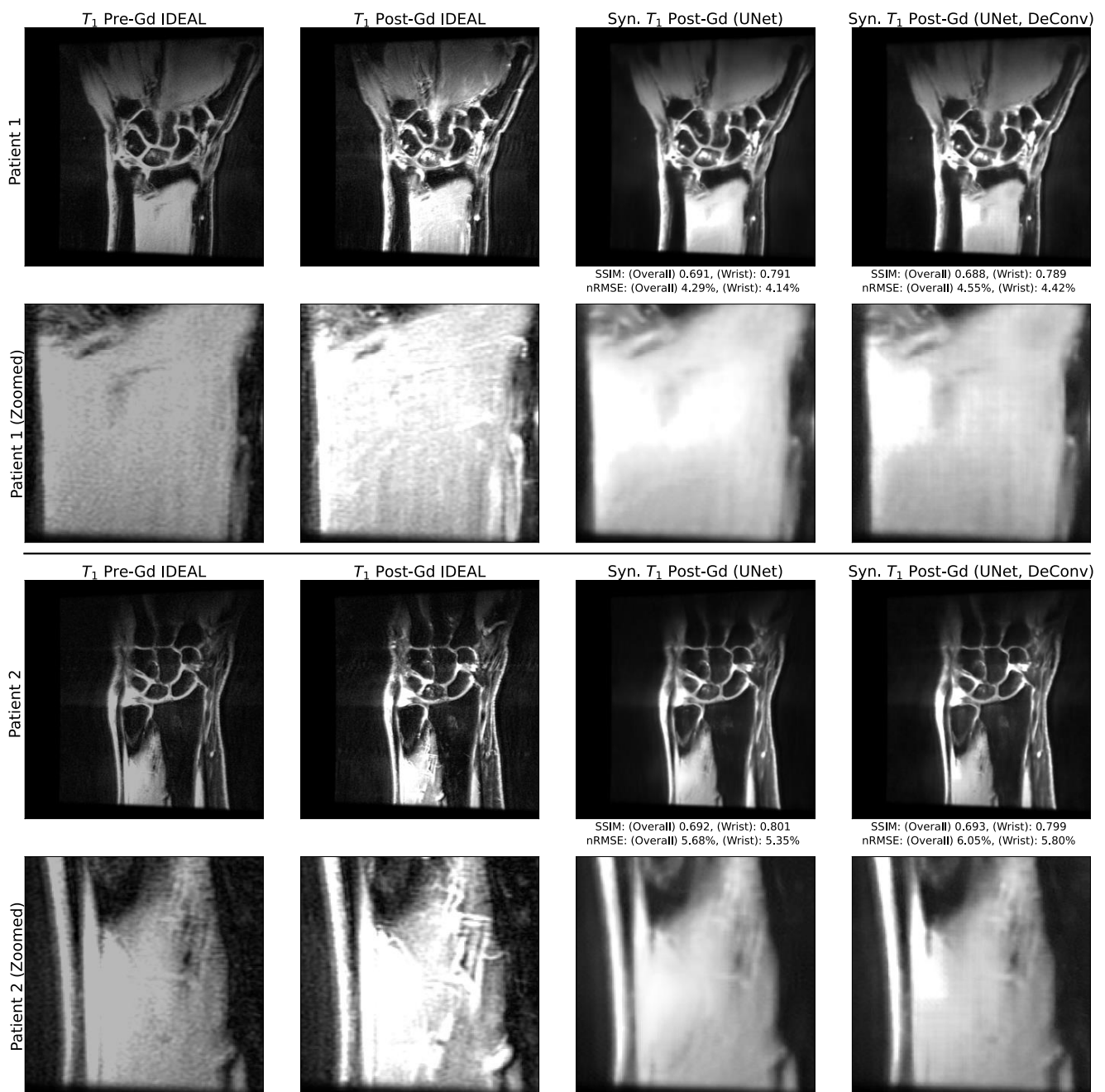
Figure 2. Network Performance with and without Deconvolutions in Decoding Path of Baseline UNet. The performance on example test set slices for baseline UNet, with and without decoding path deconvolutions, with zoomed insets. The use of decoding path deconvolutions in baseline UNets induces checkerboarding artifacts in larger regions of relatively homogenous pixel values, such as the forearm muscle insets (particularly evident in patient 1). When replaced with convolution and interpolation operators, these artifacts were substantially mitigated, making this the preferred architecture when training baseline UNets.
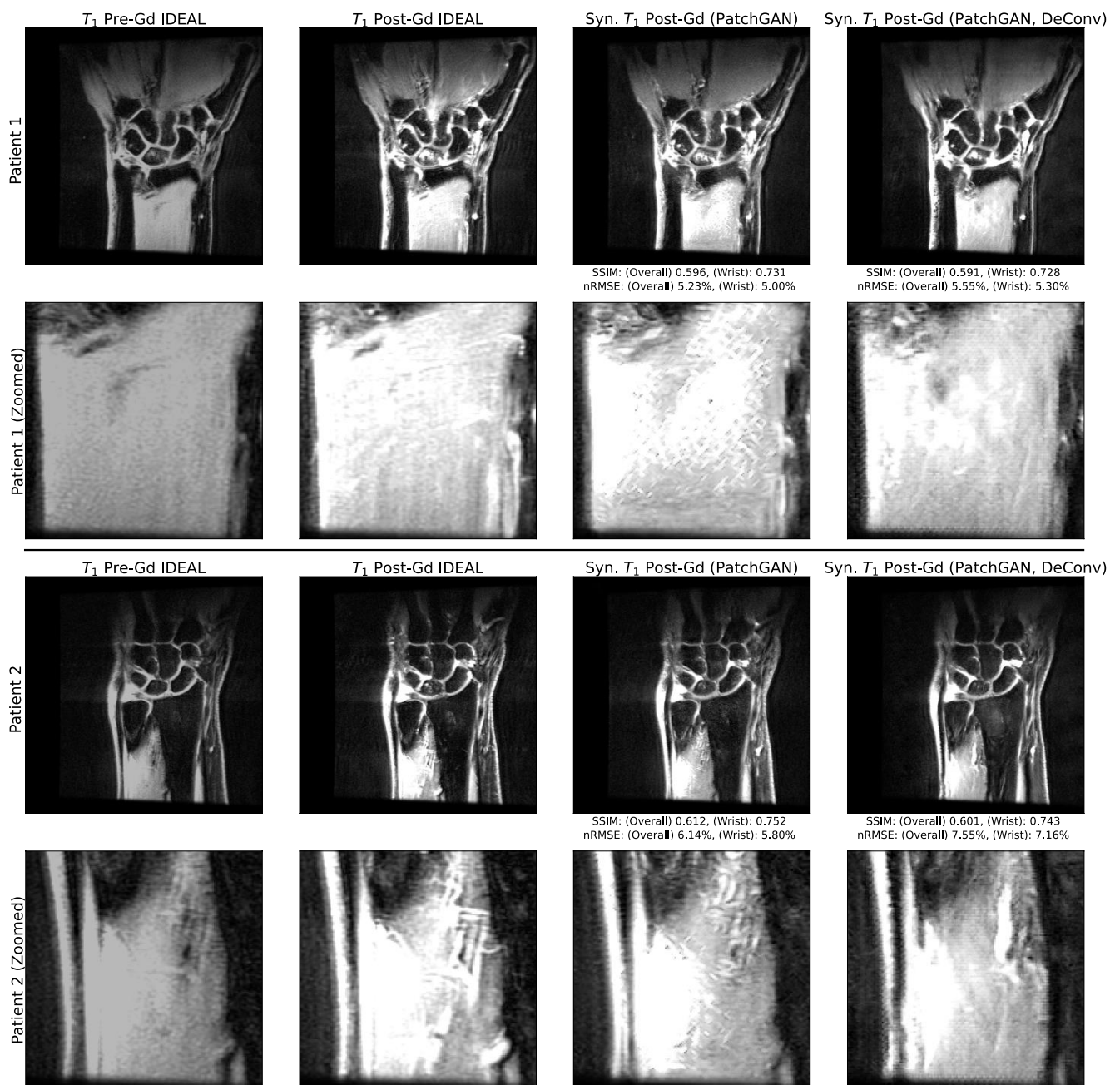
Figure 3. Network Performance with and without Deconvolutions in Decoding Path of PatchGAN Generator. The performance on example test set slices for PatchGAN pipelines, with and without generator decoding deconvolutions, with zoomed insets. At sharp transitions in pixel intensities, such as intersections of the radius and ulna with muscles displayed in insets, clear checkerboarding is observed when deconvolutions are used. This was substantially reduced when deconvolutions were replaced with convolutions and interpolation; PatchGAN generators with these decoding path operations were thus used when training PatchGAN pipelines in the remainder of this paper.

Table 2. Coronal IDEAL Post-Gd T_1 Image Synthesis Performance for Select Pipelines. Standard reconstruction metrics of the PatchGAN and baseline UNet pipelines were evaluated on a per-patient basis within the test set (n = 7) for entire imaging volumes ("full"), wrist tissue in each volume ("wrist"), and synovial joints. Metrics were calculated with and without three-stage nonlinear registration of synthetic post-Gd volumes to ground truth. UNet pipelines reflect the stronger bulk reconstruction metrics in full volumes and within wrist tissue, whereas the PatchGAN pipeline shows stronger performance in the synovial joints in which an algorithm like this would have most clinical utility, and was therefore the stronger model. The 2-sample t-tests with Bonferroni correction showed that nearly all pipelines offered significantly better reconstruction metrics than Pre-Gd baselines (n = 7; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

		Full	Wrist Only	Synovial Joints
Pre-Gd	nRMSE	26.30 ± 9.16	17.82 ± 6.31	260.24 ± 158.56
	PSNR	17.77 ± 0.95	22.99 ± 0.91	8.94 ± 1.64
	SSIM	0.60 ± 0.03	0.94 ± 0.00	
PatchGAN Registered	nRMSE	6.72 ± 0.81 ***	6.07 ± 1.22 ***	23.14 ± 7.37 **
	PSNR	20.77 ± 0.65 ***	25.40 ± 1.24 **	12.10 ± 1.34 **
	SSIM	0.58 ± 0.02	0.94 ± 0.01	
PatchGAN Unregistered	nRMSE	8.46 ± 1.03 ***	7.68 ± 1.41 **	28.96 ± 10.57 **
	PSNR	19.85 ± 0.69 ***	24.38 ± 1.21 *	11.23 ± 1.52 *
	SSIM	0.56 ± 0.02 *	0.94 ± 0.01	
UNet Registered	nRMSE	6.29 ± 0.88 ***	4.36 ± 0.60 ***	26.18 ± 7.45 **
	PSNR	22.03 ± 0.60 ***	27.13 ± 0.69 ***	11.58 ± 0.93 **
	SSIM	0.69 ± 0.02 ***	0.95 ± 0.00 **	
UNet Unregistered	nRMSE	7.73 ± 1.03 ***	5.38 ± 0.73 ***	29.69 ± 7.60 **
	PSNR	21.20 ± 0.62 ***	26.20 ± 0.77 ***	10.98 ± 0.87 *
	SSIM	0.68 ± 0.02 ***	0.95 ± 0.01 *	

3.4. Comparison of Reconstruction Performance across Synovitis Severity

The image quality metrics for synthetic post-Gd volumes are shown in Table A4 for test set patients without imaging findings of RA synovitis (RAMRIS synovitis = 0, n = 2) and those with imaging findings of RA synovitis (RAMRIS synovitis > 0, n = 5). Though the sample size limits the power of these conclusions, the metrics were slightly stronger for RAMRIS > 0 than for RAMRIS = 0. Visual examples of the reconstructed post-Gd volumes for a RAMRIS = 0 and RAMRIS > 0 patient are shown in Figure 4. In the RAMRIS = 0 patient with no imaging findings of synovitis, the absence of synovial enhancement was captured by both pipelines, whereas in the RAMRIS > 0 patient, UNet and PatchGAN pipelines illuminated similar enhancement patterns in intercarpal regions, with the PatchGAN pipeline depicting sharper enhancement pattern contours, particularly in the muscles and bones.

3.5. Enhancement Maps Analysis

Enhancement maps are shown for an example slice for the PatchGAN and UNet models, as well as ground truth, in Figure 5. The enhancement maps show that for the PatchGAN model, general magnitudes of uptake were much more accurately preserved than for the UNet, most notably across intercarpal joints. The predicted enhancement locations were visually very similar for both pipelines.

3.6. Occlusion and Uncertainty Maps Analysis

Occlusion maps for the UNet and PatchGAN pipelines in sample test set slices are shown in Figure 6. Encouragingly, occlusion maps for both pipelines show a substantial focus on intercarpal joint regions in terms of their relative importance to the predicted pixel values. Peripherally to the intercarpal joint, the occlusion maps show some focus on muscles as well, perhaps slightly more so for the UNet than for the PatchGAN. On

the other hand, the uncertainty maps are shown in an example test set slice for UNet and PatchGAN pipelines in Figure 7. The UNet shows considerable uncertainty in intercarpal joint region predicted pixel values, whereas for the PatchGAN, uncertainty was highest in the background and within the muscles. PatchGAN also showed some uncertainty in predictions within bones such as the radius and ulna, as well as within bone marrow edema regions; notably, however, uncertainty was limited in the synovial joints.
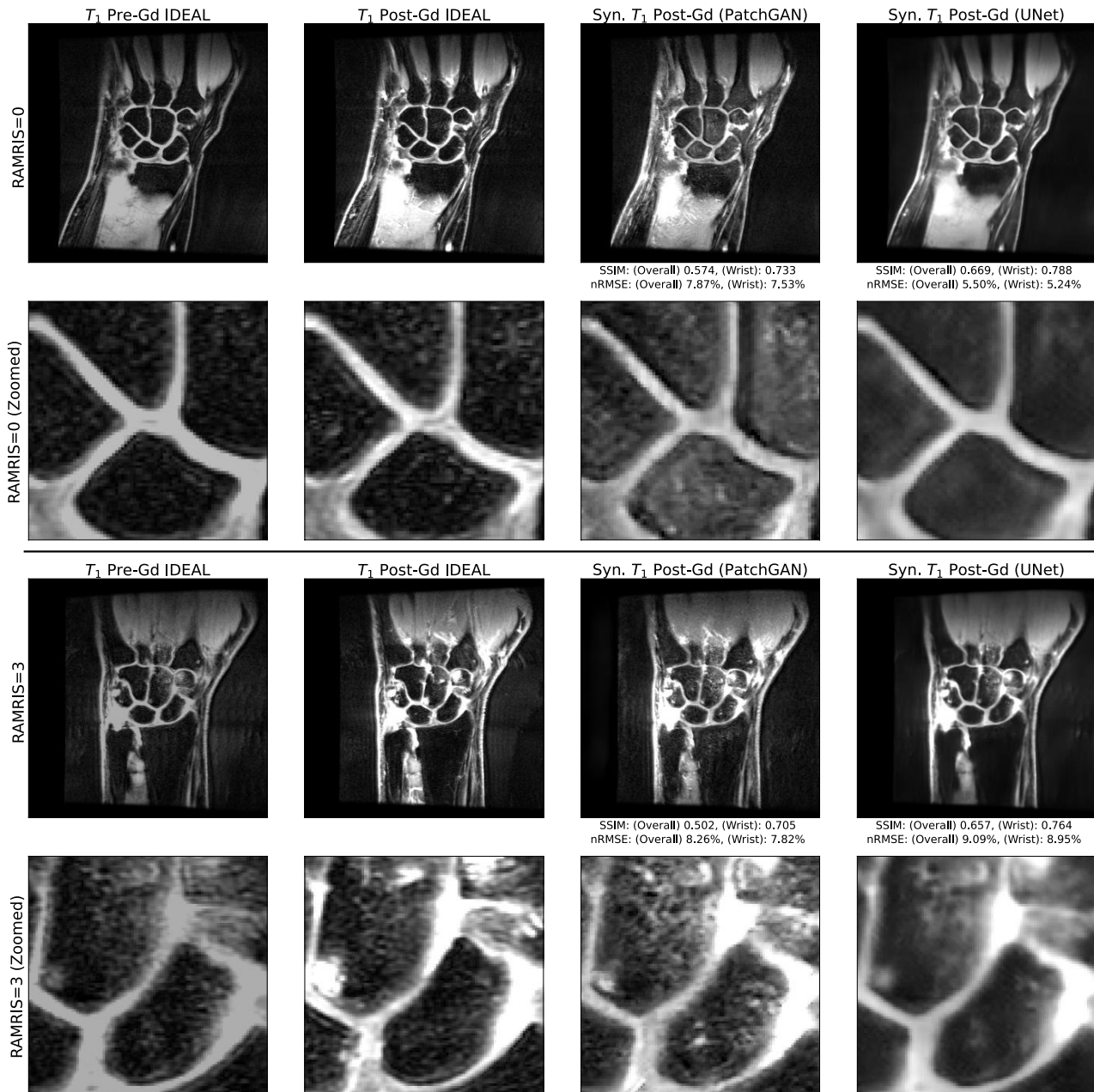

Figure 4. Visual Comparison of Reconstructed Post-Gadolinium Images with and without Imaging Findings of RA. Two example test set slices reconstructed by baseline UNet and PatchGAN pipelines for patients with and without imaging findings of RA (RAMRIS = 3, RAMRIS = 0, respectively). There was little to no enhancement in the synovial joints of the RAMRIS = 0 patient, which is captured by both pipelines, as seen in the zoomed insets. In the RAMRIS = 3 patient, the contours of enhancement in the zoomed inset were captured well within the intercarpal joint for both pipelines, with noise distribution patterns better reconstructed by the PatchGAN. The reconstruction performance thus shows promise for patients with and without imaging findings of RA.
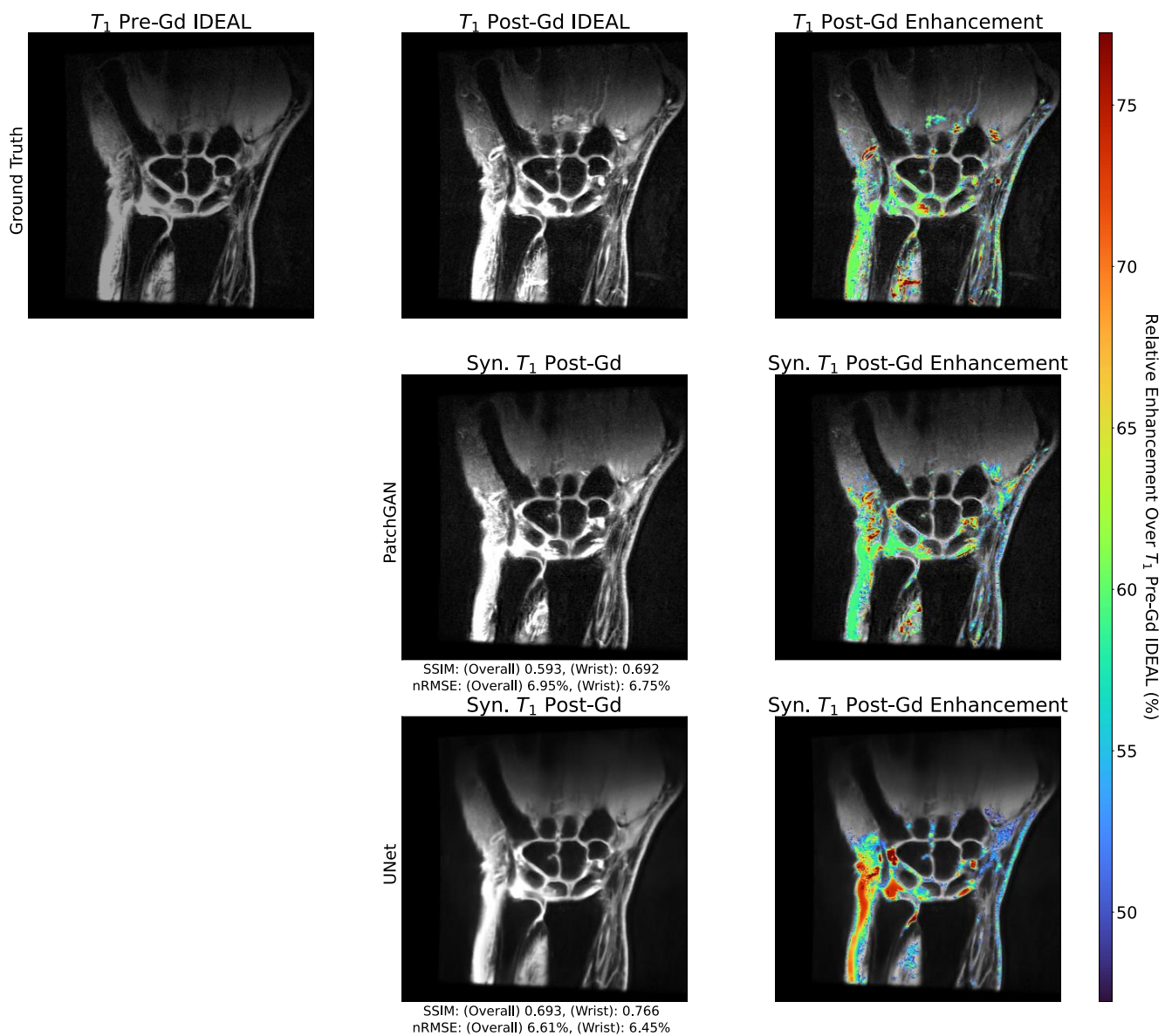
Figure 5. Predicted Gadolinium Enhancement Maps with PatchGAN, UNet, and Ground Truth Models. Enhancement maps were generated by identifying the magnitude of pixel intensity increase from synthetic or ground truth Post-Gd slices compared to corresponding pre-Gd slices, and by highlighting the top 10%. While the performance in preserving the location of these top 10% of enhancing pixels was similar for the baseline UNet and PatchGAN, the enhancement magnitudes were far better preserved globally by the PatchGAN, including intercarpal regions susceptible to synovitis. These maps reflect the long-term vision of a pipeline like this: given a pre-Gd scan, the algorithm can identify locations susceptible to synovitis and distinguish active inflammatory sites from general effusion with additional model development.
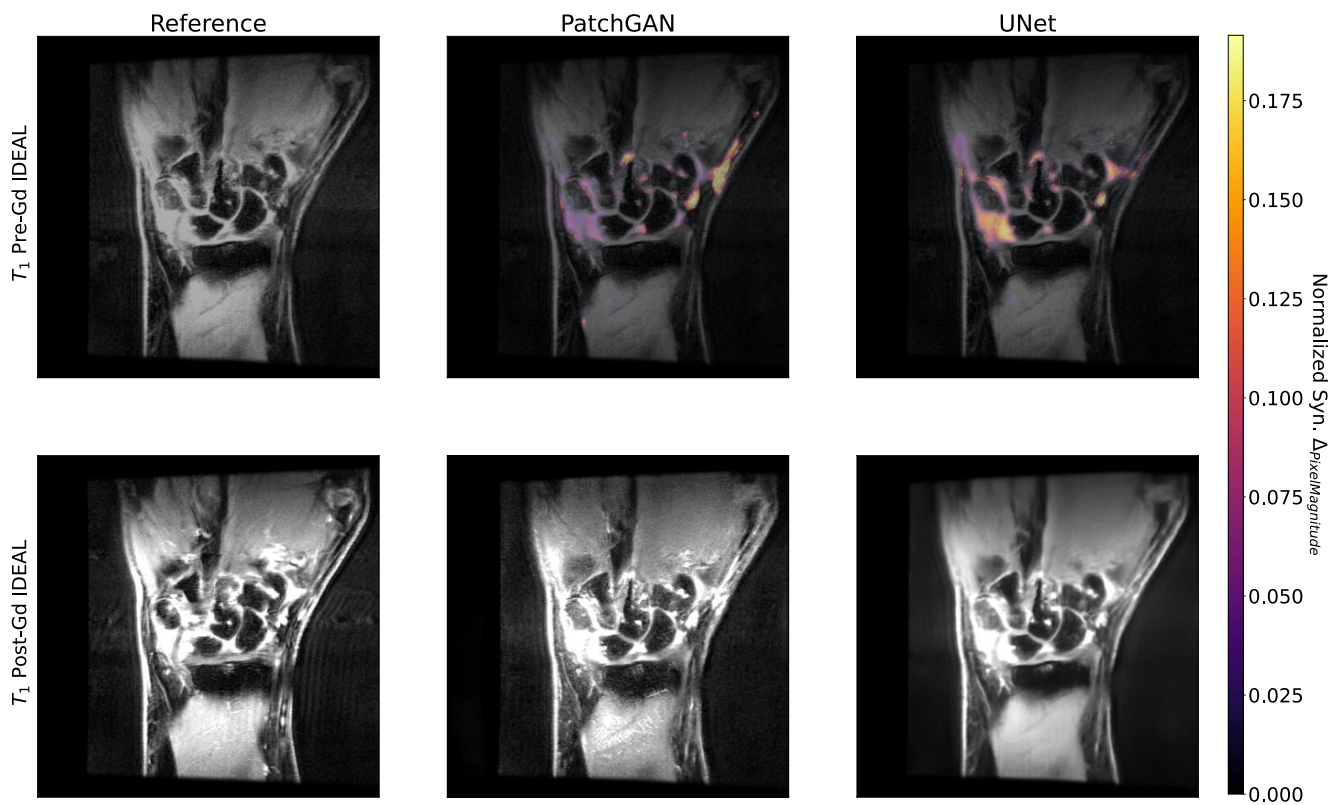
Figure 6. Occlusion Maps for PatchGAN and UNet Pipelines. Occlusion maps were generated for PatchGAN and UNet by occluding 32×32 patches of the input slices and assessing changes in predicted pixel values compared to unoccluded slices. Occlusion maps were then normalized by pre-Gd pixel intensities and thresholded to identify hotspots most impactful in model predictions. For UNet and PatchGAN, hotspots primarily included intercarpal joint regions. Particularly for the UNet, the maps also showed some emphasis on the forearm muscles. Given that the synovial joints are where an inflammatory imaging algorithm would see the most utility, the fact that both algorithms placed heavy emphasis on the intercarpal regions was promising, indicating that both focused on synovitis-relevant regions to make predictions.
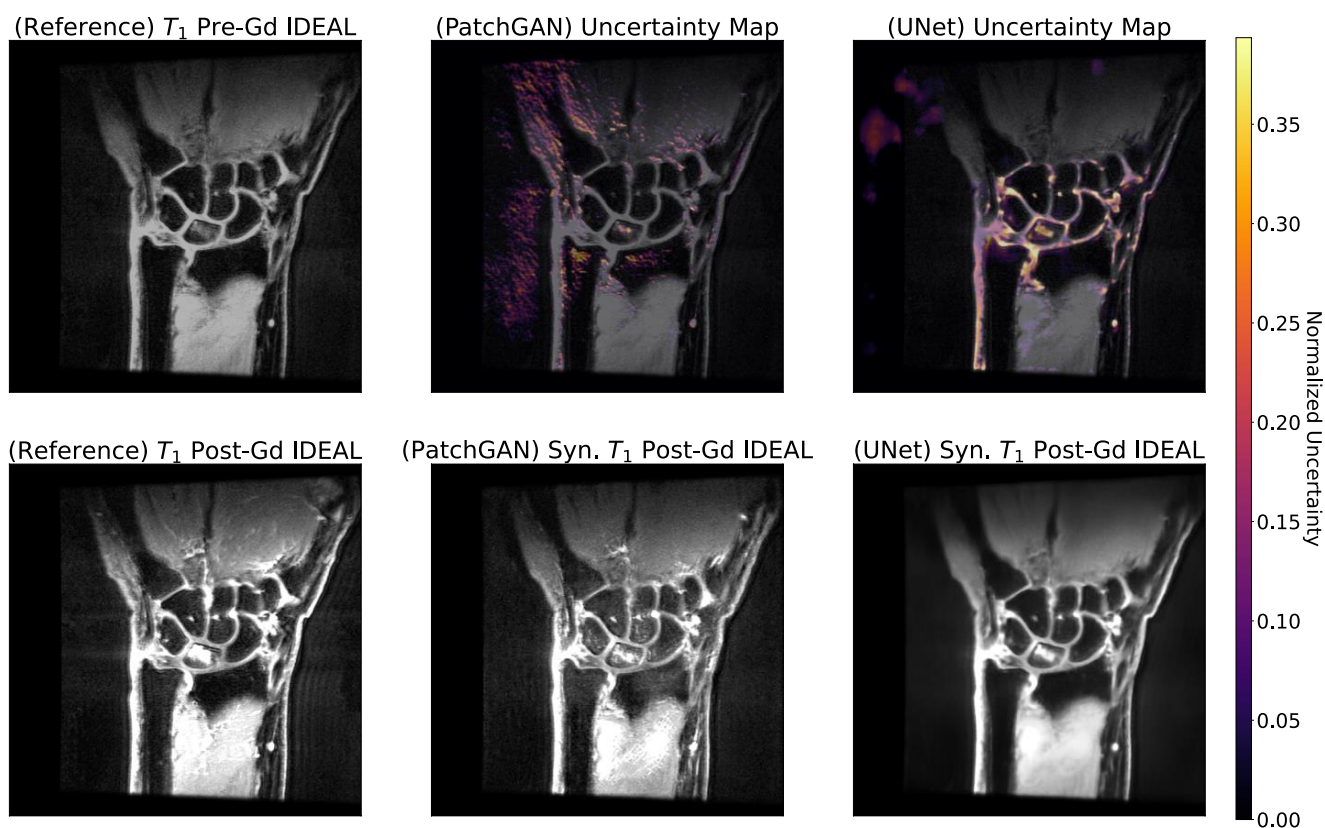
Figure 7. Uncertainty Maps for PatchGAN and UNet Pipelines. Uncertainty maps for the PatchGAN generator and baseline UNet were generated by corrupting the latent space of all encoding path outputs, adding Gaussian noise with a mean of 0 and a standard deviation of 0.5 for 100 iterations, and calculating the variance in the predicted pixel magnitudes for each output pixel across these iterations. The most variant pixels were designated as the most uncertain ones. For the PatchGAN, the uncertain regions were mainly in the background, muscles, and within bones. For baseline UNet, the uncertainty maps placed a heavy emphasis on the intercarpal joint, with some residual highlighting of background. In conjunction with occlusion maps, PatchGAN generator predictions were more confident and less uncertain within intercarpal joint regions compared to the baseline UNet. Considering that the intercarpal joint is crucial for synovitis diagnosis and is where both algorithms would be the most useful, the PatchGAN's confident predictions within it were promising.

4. Discussion

In this work, we developed multiple strong-performing DL pipelines that synthetically generate post-contrast coronal IDEAL T_1 wrist MR images from pre-contrast coronal IDEAL T_1 wrist images, marking steps toward synthetic inflammatory imaging of MSK tissues for conditions such as RA. Reconstruction metrics show reasonably strong performances for UNet and PatchGAN pipelines without generator decoding path deconvolutions—PatchGAN nRMSEs in the wrist were 7.68 ± 1.41 (6.07 ± 1.22 after registration, mean $\pm$ standard deviation (s.d.)) and for the UNet they were 5.38 ± 0.73 (4.36 ± 0.60 after registration, mean $\pm$ s.d.). Standard reconstruction metrics—nRMSE, PSNR, and SSIM—showed the UNet to have superior performance across full volumes and within the wrist, but purely in the synovial joints, where a pipeline like this would see the most utility, the PatchGAN outperformed the UNet. These findings provide yet additional evidence to a growing body of literature which suggests that standard reconstruction metrics do not provide great correlation with clinically useful metrics when evaluated in a classical fashion (across an entire tissue) [44,45,49]. This, in addition to a perceptually stronger performance replicating sharper textures (particularly within muscles and bones, but at times in the synovial joints as well), shows the PatchGAN pipeline without deconvolutions

to be the strongest tested version and with the most potential for eventual clinical use with further development. Additionally, enhancement maps showed that while both pipelines exhibited similar performance in identifying the location of the top 10% of enhancing pixels, the PatchGAN did a substantially better job in preserving the enhancement magnitudes. These trends particularly held in the muscles and vessels, but also in many synovial joints.

To build clinicians' trust in medical image processing algorithms, experiments such as the proposed occlusion map and uncertainty analyses are vital to address the criticism of deep learning algorithms being "black boxes". These techniques yielded notable insights in the PatchGAN and UNet pipelines: occlusion maps showed that both pipelines focused heavily on intercarpal regions and synovial joints as a basis for generating model predictions. At the same time, uncertainty maps yielded diverging conclusions: whereas the PatchGAN was most uncertain in background, muscles, and within bones, the UNet pipeline was the most uncertain within the intercarpal joints themselves. Given that intercarpal joints—and more generally synovial joints—are where a synthetic inflammatory imaging algorithm would see maximal utility in RA imaging, it is extremely encouraging that the PatchGAN based much of its predictions on the intercarpal joints and was relatively confident in its predictions. This, combined with the superior reconstruction metrics obtained in synovial joints by the PatchGAN as compared to the UNet, confirms it to be the pipeline with the most potential for clinical utility, and indicates that the combination of a GAN and a focused, ROI-based loss can yield promising results for optimizing image synthesis algorithms. Uncertainty and occlusion map approaches such as those applied in this work are straightforward to implement and can be extended to other deep learning applications such as image synthesis, image segmentation, and image reconstruction. In doing so, they can make the findings of such algorithms easier to interpret while providing valuable insights into how they work. From a clinical perspective, they can not only build trust in algorithm outputs, but also direct a radiologist's attention to uncertain regions in an image that require closer examination.

The exploration of architectural designs also yielded interesting insights. Checkerboarding artifacts have long been reported as a shortcoming of CNNs, and more specifically UNets, with many strategies being proposed to mitigate them [50–52]. Our investigation of UNet pipelines with and without one such mitigating strategy—replacing deconvolutions with interpolation and standard convolutions—showed checkerboarding artifacts to be widespread in larger areas of relatively homogenous pixel intensity with the standard deconvolutions, but absent with the mitigating strategy implemented. When paired with a PatchGAN discriminator, even a UNet generator with deconvolutions resolved the checkerboarding artifacts in larger homogenous pixel intensity areas, but saw minor checkerboarding emerge at the boundaries between pixel intensities. Checkerboarding artifacts are thus intrinsic to the standard UNet architecture, and among the tasks a discriminator must learn in adversarial training is their removal. When deconvolutions are replaced with interpolation and standard convolutions, the artifact removal responsibility is simplified for a GAN discriminator, in theory allowing the discriminator to focus on more minute differences between real and synthetic images and, thus, possibly producing stronger synthetic images. These lessons can be translated to GAN training strategies in other settings—training schemes may yield stronger results after the thorough inspection of generator architectures to ensure that obvious artifacts are not intrinsic to the network design.

It is clear from our work that larger sample sizes are needed to derive statistical conclusions with more power and to assess algorithm efficacy stratifying by race, RA status, and others. However, this study nonetheless serves as a strong proof-of-concept indicating the potential for DL algorithms to synthesize post-contrast images for inflammatory imaging in MSK applications. Importantly, these algorithms can synthesize images in a negligible amount of time, essentially providing free information for radiologists examining inflammation, even for the many patients for whom contrast MR sequences would otherwise not be prescribed. With additional validation, and through building clinicians' trust in

these algorithms, they can allow for safer, more comfortable, and less time-consuming RA diagnosis and treatment through synthetic imaging. Beyond the proof-of-concept wrist RA post-contrast synthesis, this work can seed new efforts in other MSK applications such as synthetic RA imaging in other joints [53], synthetic screening for sarcoma [54], more thorough investigations associating contrast and non-contrast MRI of Hoffa's fat pad with pain [55], larger cohort studies assessing bone perfusion [56], and safer imaging techniques to diagnose spondylodiscitis [57]. In all these applications, Gd is administered in standard imaging protocols, so similar datasets can be curated and used to train synthetic post-contrast imaging algorithms to reduce and hopefully eliminate the need for Gd administration. Furthermore, validated algorithms could synthesize post-contrast images from existing large datasets such as the Osteoarthritis Initiative (OAI), K2S, and fastMRI+ to allow for large cohort studies to facilitate a better understanding of inflammation [58–60].

This study had several limitations. Ideally, there would be a true comparison of algorithm performance in patients with and without RA to ensure strong performance in both, but ethical considerations prevented us from administering Gd to healthy controls. In the absence of this, we used RAMRIS scores to stratify RA patients into subgroups of those with and without imaging findings of RA for a pseudo-control study, but this is not a true control study. Furthermore, the desire to compare algorithm performance in patients with and without imaging findings of RA in a pseudo-control study, combined with the small dataset size, led to some imbalance in demographic characteristics across training, validation, and test datasets. Namely, test set patients had the least severe RA. Additionally, pre-Gd coronal IDEAL images were registered to corresponding post-Gd images in data preprocessing. Radiologist anomaly segmentations were performed only on post-Gd images, so doing so allowed segmentations to be used in weighting loss functions and assessing model performance in anomalous regions, but this registration step would not be possible at the inference time. There was thus a tradeoff between optimizing trained algorithms for strong performance in synovial joints and using a realistic workflow for eventual clinical utility; the authors viewed the former as more important in a proof-of-concept approach. Lastly, standard imaging protocols would typically use T_1 pre-contrast scans and fat-saturated post-contrast T_1 scans for RA imaging. Our approach used IDEAL scans before and after contrast administration, as these sequences were available in our dataset, but for true clinical translation an algorithm should be trained on these other sequences. The structure of our dataset thus conferred many limitations on our work, but nonetheless, it represents a meaningful first step towards making synthetic inflammatory imaging a larger research focus for the MSK community.

5. Conclusions

To the best of the authors' knowledge, our work marks the first concerted effort at leveraging DL for synthetic inflammation imaging for an MSK application. We developed PatchGAN and baseline UNet pipelines that showed strong performance synthesizing post-contrast IDEAL T_1 images from corresponding pre-contrast IDEAL T_1 images, with the PatchGAN pipeline outperforming the UNet in synovial joints, generating more accurate and confident predictions where a model would have the most utility. The PatchGAN also showed magnitudes of signal enhancement that more closely match that of ground truth images and retained sharp textures in synthetic images. As such, the PatchGAN model was particularly promising in synthesizing post-contrast inflammatory images, and with further development, it could reduce or eliminate the need for Gadolinium administration in treating patients with RA. There are numerous future directions for research: (1) more sophisticated GANs such as CycleGAN can be implemented to improve the sharpness in reconstructed images; (2) generator architectures that learn registration transforms and predict images can also be investigated, eliminating the need for the registering of pre-Gd images to post-Gd images, which would not be possible at inference time in the clinic; (3) investigating other loss functions, such as other types of GAN distances; and (4) assessing model robustness by inferring from conventional wrist coronal T_1 scans

to evaluate predicted post-contrast scans on conventionally used clinical sequences in inflammatory imaging. For substantial progress, however, the MSK field will require concerted efforts to curate larger datasets for inflammatory RA conditions that will allow for more statistically powerful conclusions, more complicated models, and comparisons across population subgroups. Our hope is that the promise of our results can motivate efforts to do so.

Appendix A

Figure A1. Example Registrations. Coronal IDEAL pre-Gd slices were registered to coronal IDEAL post-Gd slices to account for motion and slight alterations in patient position that occurred between the sequences. Registration was done on a per-volume basis in a three-stage algorithm: (1) translation, (2) affine, and (3) third order b-spline registration (maximum iterations = 256, 256, 512, respectively; Advanced Mattes Mutual Information criteria). B-spline registration was done for patients where SSIM between unregistered Pre-Gd and Post-Gd scans was above 0.5, which was used as a proxy for detecting motion artifacts so severe that any non-linear registration would lead to overfitting.
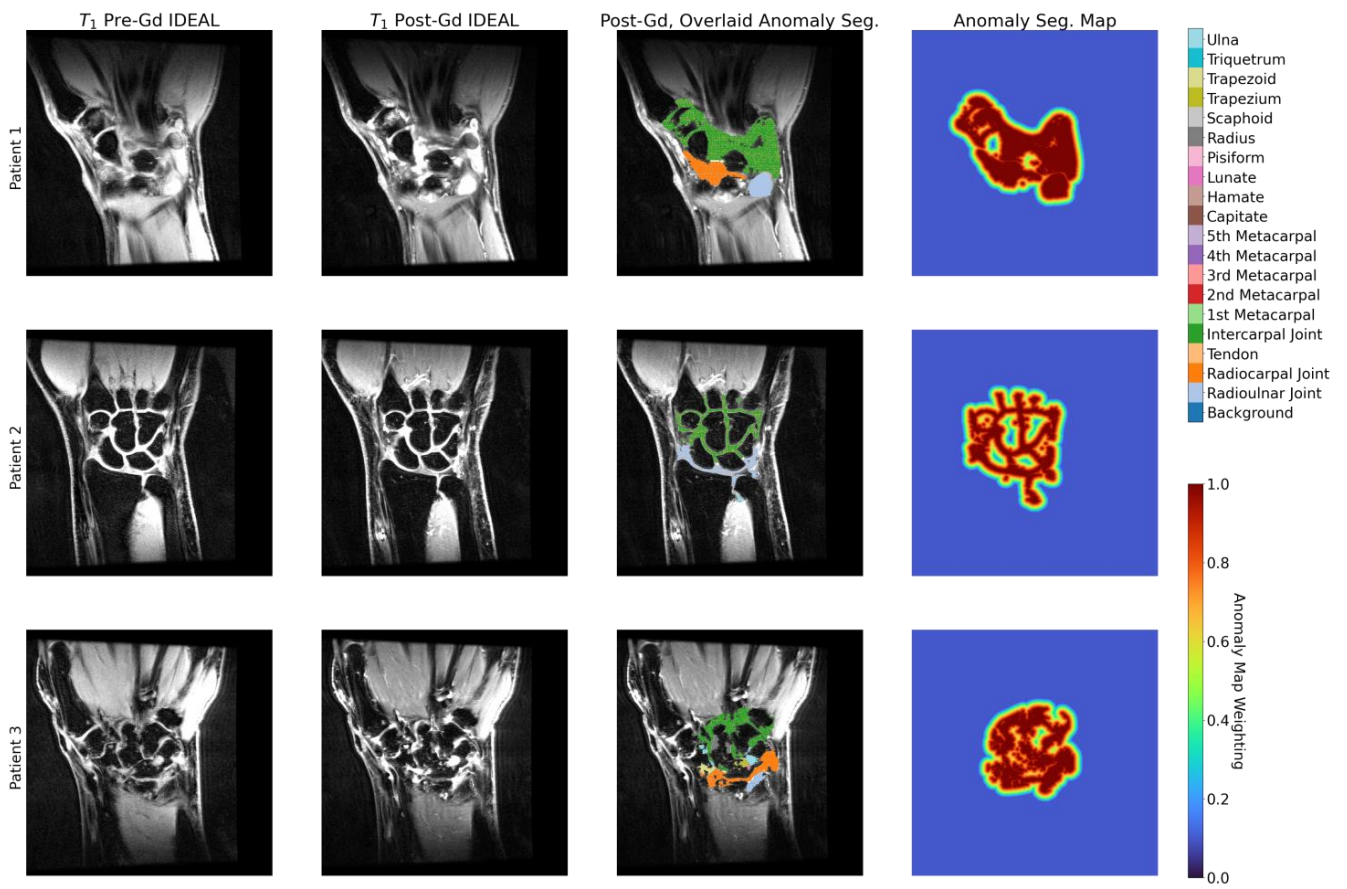
Figure A2. Example Anomaly Distance Map. Example anomaly segmentations and corresponding anomaly segmentation maps that would result and be used to weight pixel-based L_1 loss functions during training.
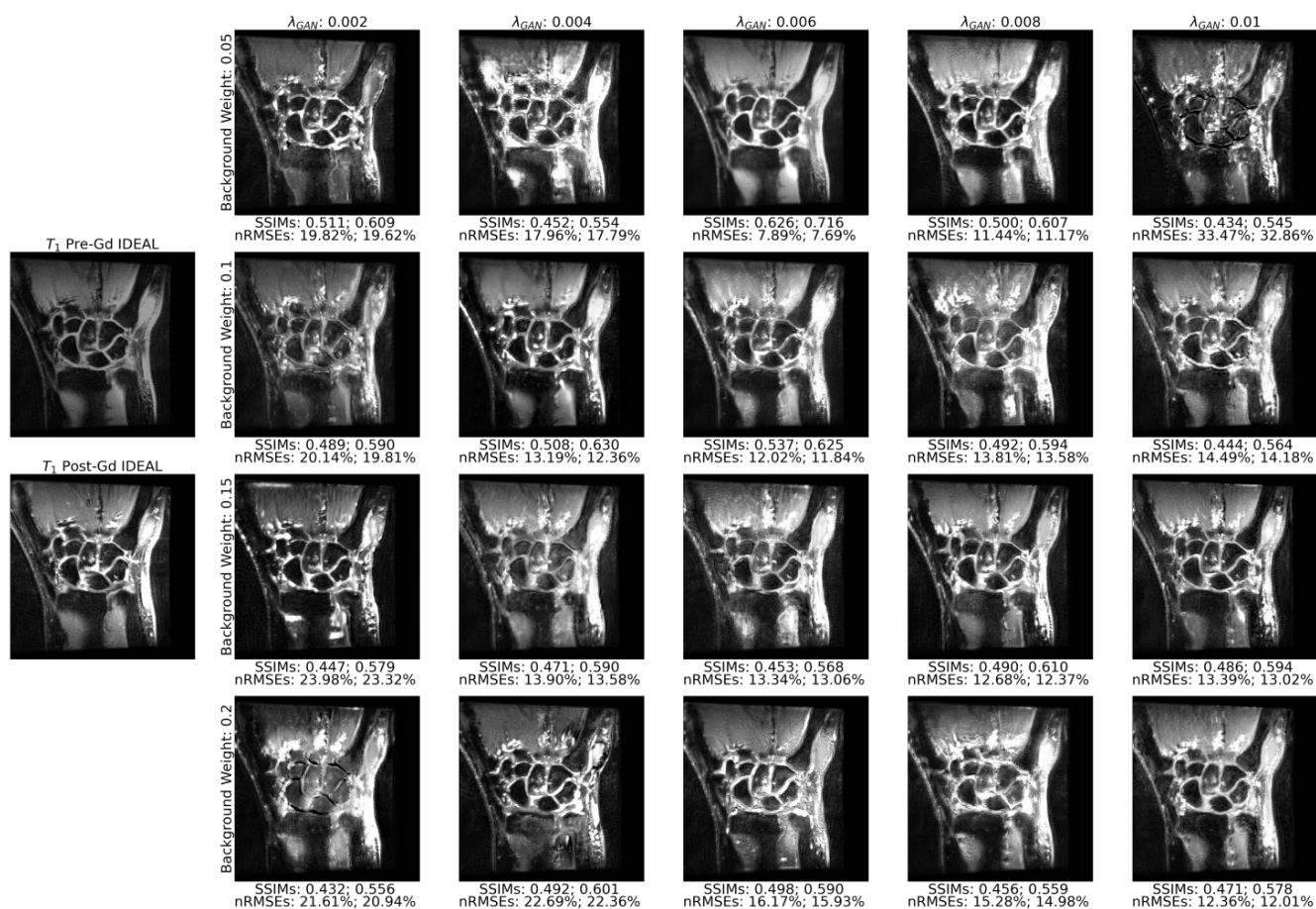
Figure A3. Sample Hyperparameter Search Slices. Examples of 10-epoch training results across one slice for the PatchGAN pipeline. The metrics are listed below the images, the first for the entire imaging slice and the second for only wrist tissue. After initially screening hyperparameter configurations using the SSIM and nRMSE validation sets, optimal hyperparameter combinations were selected based on visual inspection, with the primary criteria being the synthesis of new information, the fidelity of reconstructed volumes to ground truth, and the absence of obvious algorithm-generated artifacts. Models with optimal hyperparameter sets were then trained from scratch for 35 epochs.

Table A1. MR Acquisition Parameters. Acquisition parameters for the 64 scanned volume pairs in the 27-patient dataset of RA patients. The parameters for any given patient at any given time point were the same for both pre- and post-Gadolinium coronal T_1 IDEAL wrist scans.

Scanner	GE Signa Discovery MR750w
Coil	8-channel HD Wrist Array
Field Strength	3T
Slice Thickness	2 mm
Spacing between Slices	2 mm
TR	457–793 ms
TE	10.06–12.48 ms
Frequency	127.8 Hz
Bandwidth	195.3 Hz (384 × 256, n = 58), 390.6 Hz (256 × 224, n = 6)
Acquisition Matrix	384 × 256 (n = 58), 256 × 224 (n = 6)
Flip Angle	90 (n = 2) or 111 (n = 62)
SAR	1.578–3.259
Pixel Spacing	0.234 × 0.234 mm (n = 58), 0.469 × 0.469 mm (n = 6)

Table A2. SSIM for Full Volumes and Wrist Tissue from Hyperparameter Search. SSIMs obtained for 10-epoch trains of all 70 hyperparameter combinations for the PatchGAN pipeline. The top 12 performing parameter sets by SSIMs are highlighted in bold, and were visually examined to find the best-performing parameter set.

λ_{GAN}		λ_B						
		0	0.025	0.050	0.075	0.100	0.150	0.200
0.001	Full	0.596	0.590	0.593	0.560	0.584	0.573	0.572
	Wrist	0.713	0.713	**0.718**	0.708	0.711	0.712	0.705
0.002	Full	0.555	0.553	0.548	0.537	0.567	0.557	0.556
	Wrist	0.697	0.709	0.673	0.662	0.715	0.694	0.690
0.003	Full	0.538	0.577	0.541	0.590	0.538	0.634	0.592
	Wrist	0.681	0.703	0.689	0.712	0.665	**0.747**	**0.725**
0.004	Full	0.496	0.550	0.582	0.594	0.548	0.555	0.581
	Wrist	0.713	0.680	0.705	0.711	0.709	0.707	0.709
0.005	Full	0.591	0.554	0.585	0.575	0.584	0.622	0.578
	Wrist	**0.721**	0.698	0.708	0.692	**0.721**	**0.738**	0.716
0.006	Full	0.519	0.527	0.596	0.583	0.565	0.505	0.577
	Wrist	0.695	0.687	0.705	0.703	0.698	0.687	0.703
0.007	Full	0.557	0.587	0.607	0.491	0.579	0.555	0.56
	Wrist	0.712	0.714	**0.722**	0.682	0.703	0.695	0.685
0.008	Full	0.561	0.567	0.585	0.553	0.565	0.526	0.567
	Wrist	0.694	**0.732**	0.715	0.688	0.717	0.675	0.681
0.009	Full	0.560	0.510	0.624	0.550	0.571	0.580	0.507
	Wrist	0.684	0.674	**0.748**	0.705	0.693	0.704	0.703
0.010	Full	0.518	0.496	0.574	0.582	0.502	0.567	0.632
	Wrist	0.689	**0.720**	**0.733**	0.701	0.646	0.694	**0.737**

Table A3. nRMSEs for Full Volumes and Wrist Tissue from Hyperparameter search. nRMSEs obtained for 10-epoch trains of all 70 hyperparameter combinations for the PatchGAN pipeline. The top 12 performing parameter sets by nRMSEs are highlighted in bold and were visually examined to find the best-performing parameter set.

λ_{GAN}		λ_B						
		0	0.025	0.050	0.075	0.100	0.150	0.200
0.001	Full	11.5	6.1	11.4	7.8	14.8	8.7	24.1
	Wrist	11.1	**5.8**	10.8	**7.4**	14.5	**8.2**	23.1
0.002	Full	12.6	9.7	6.2	10	23.9	20.3	20.3
	Wrist	12	9.1	**5.9**	9.5	23.7	20	20.1
0.003	Full	10.6	12	9.9	10.6	38.4	10.1	12.6
	Wrist	10.2	11.2	9.5	10	38.1	9.6	11.7
0.004	Full	9.2	18.5	22.5	11.6	10	33.8	10
	Wrist	8.8	18	21.8	11.1	9.7	33.4	9.5
0.005	Full	12.8	16.7	11.8	11.3	15.3	24.5	18.9
	Wrist	12.4	15.9	11.3	10.6	14.9	24.2	18.3
0.006	Full	21.2	6.2	17.9	13.1	5.8	14.5	12.3
	Wrist	20.9	**5.9**	17	12.6	**5.5**	13.9	11.1
0.007	Full	10.1	6.4	22	11.4	10.4	8.9	19.1
	Wrist	9.5	**6.1**	21.4	11	9.9	**8.4**	18.1
0.008	Full	9.9	11.4	9.4	8.9	9.6	9.8	7.9
	Wrist	9.3	10.9	9	8.4	9.2	8.9	**7.5**
0.009	Full	15.3	47	10.5	12.7	9.9	10.7	13.2
	Wrist	14.7	**7.4**	10	12.4	9.5	10.3	12.6
0.010	Full	10.9	12.9	11	10.6	15.4	6.8	8.7
	Wrist	10.5	12.6	10.4	10.1	14.7	**6.4**	**8.4**

Table A4. Reconstruction Metrics for Patients with and without Imaging Findings of RA. Bulk reconstruction metrics in full imaging volumes, wrist tissue, and synovial joints in patients without imaging findings of synovitis (RAMRIS = 0, n = 2) and patients with imaging findings of synovitis (RAMRIS > 0, n = 5) within the test set. All metrics were evaluated on a per-patient basis. Small sample sizes prevent proper statistical comparisons, but reconstruction metrics were generally stronger for RAMRIS > 0 patients, as expected given the strong bias of this dataset towards RAMRIS > 0 patients (23 of 27 patients).

		RAMRIS = 0			RAMRIS > 0		
		Full	**Wrist Only**	**Synovial Joints**	**Full**	**Wrist Only**	**Synovial Joints**
Pre-Gd	nRMSE	23.95 ± 4.71	23.72 ± 4.83	133.52 ± 43.31	27.24 ± 10.27	26.94 ± 10.43	310.93 ± 159.54
	PSNR	17.76 ± 0.38	17.89 ± 0.38	9.38 ± 0.40	17.77 ± 1.10	17.95 ± 1.23	8.77 ± 1.90
	SSIM	0.62 ± 0.03	0.75 ± 0.01		0.59 ± 0.03	0.73 ± 0.01	
PatchGAN Reg.	nRMSE	7.13 ± 0.80	6.85 ± 0.77	28.20 ± 11.85	6.55 ± 0.75	6.26 ± 0.73	21.12 ± 2.37
	PSNR	20.44 ± 0.52	20.68 ± 0.51	11.53 ± 2.19	20.90 ± 0.64	21.16 ± 0.65	12.33 ± 0.64
	SSIM	0.58 ± 0.02	0.72 ± 0.00		0.58 ± 0.02	0.73 ± 0.01	
PatchGAN Unreg.	nRMSE	8.93 ± 0.69	8.65 ± 0.68	36.44 ± 17.23	8.28 ± 1.09	7.96 ± 1.06	25.98 ± 2.54
	PSNR	19.54 ± 0.34	19.73 ± 0.33	10.60 ± 2.45	19.98 ± 0.75	20.21 ± 0.75	11.49 ± 0.76
	SSIM	0.56 ± 0.02	0.70 ± 0.00		0.56 ± 0.02	0.71 ± 0.01	
UNet Reg.	nRMSE	7.26 ± 0.20	7.00 ± 0.17	28.74 ± 10.61	5.90 ± 0.73	5.67 ± 0.76	25.15 ± 5.39
	PSNR	21.36 ± 0.07	21.57 ± 0.04	11.26 ± 1.58	22.30 ± 0.49	22.54 ± 0.51	11.71 ± 0.40
	SSIM	0.68 ± 0.02	0.78 ± 0.01		0.69 ± 0.02	0.79 ± 0.01	
UNet Unreg.	nRMSE	8.35 ± 0.45	8.11 ± 0.45	31.51 ± 9.96	7.48 ± 1.09	7.24 ± 1.09	28.96 ± 6.28
	PSNR	20.82 ± 0.15	21.00 ± 0.14	10.67 ± 1.36	21.36 ± 0.67	21.57 ± 0.67	11.10 ± 0.51
	SSIM	0.67 ± 0.02	0.77 ± 0.01		0.68 ± 0.02	0.78 ± 0.01	

References

1. Silman, A.J.; Pearson, J.E. Epidemiology and Genetics of Rheumatoid Arthritis. *Arthritis Res.* **2002**, *4* (Suppl. S3), S265–S272. [CrossRef]
2. Aletaha, D.; Smolen, J.S. Diagnosis and Management of Rheumatoid Arthritis: A Review. *JAMA* **2018**, *320*, 1360–1372. [CrossRef]
3. Taylor, P.C. Update on the Diagnosis and Management of Early Rheumatoid Arthritis. *Clin. Med.* **2020**, *20*, 561. [CrossRef]
4. Goekoop-Ruiterman, Y.P.M.; De Vries-Bouwstra, J.K.; Allaart, C.F.; Van Zeben, D.; Kerstens, P.J.S.M.; Hazes, J.M.W.; Zwinderman, A.H.; Ronday, H.K.; Han, K.H.; Westedt, M.L.; et al. Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis (the BeSt Study): A Randomized, Controlled Trial. *Arthritis Rheumatol.* **2005**, *52*, 3381–3390. [CrossRef]
5. Kgoebane, K.; Ally, M.M.T.M.; Duim-Beytell, M.C.; Suleman, F.E. The Role of Imaging in Rheumatoid Arthritis. *S. Afr. J. Radiol.* **2018**, *22*, 6. [CrossRef]
6. De Schepper, A.M.; De Beuckeleer, L.; Vandevenne, J.; Somville, J. Magnetic Resonance Imaging of Soft Tissue Tumors. *Eur. Radiol.* **2000**, *10*, 213–222. [CrossRef]
7. Rubin, D.A. MRI and Ultrasound of the Hands and Wrists in Rheumatoid Arthritis. I. Imaging Findings. *Skelet. Radiol.* **2019**, *48*, 677–695. [CrossRef]
8. Zhou, Z.; Lu, Z.R. Gadolinium-Based Contrast Agents for MR Cancer Imaging. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2012**, *5*, 1–18. [CrossRef]
9. Eshed, I.; Feist, E.; Althoff, C.E.; Hamm, B.; Konen, E.; Burmester, G.R.; Backhaus, M.; Hermann, K.G.A. Tenosynovitis of the Flexor Tendons of the Hand Detected by MRI: An Early Indicator of Rheumatoid Arthritis. *Rheumatology (Oxford)* **2009**, *48*, 887–891. [CrossRef]
10. Tamai, M.; Kawakami, A.; Uetani, M.; Fukushima, A.; Arima, K.; Fujikawa, K.; Iwamoto, N.; Aramaki, T.; Kamachi, M.; Nakamura, H.; et al. Magnetic Resonance Imaging (MRI) Detection of Synovitis and Bone Lesions of the Wrists and Finger Joints in Early-Stage Rheumatoid Arthritis: Comparison of the Accuracy of Plain MRI-Based Findings and Gadolinium-Diethylenetriamine Pentaacetic Acid-Enhanced MRI-Based Findings. *Mod. Rheumatol.* **2012**, *22*, 654–658. [CrossRef]
11. Boyd, A.S.; Zic, J.A.; Abraham, J.L. Gadolinium Deposition in Nephrogenic Fibrosing Dermopathy. *J. Am. Acad. Dermatol.* **2007**, *56*, 27–30. [CrossRef] [PubMed]
12. Murata, N.; Gonzalez-Cuyar, L.F.; Murata, K.; Fligner, C.; Dills, R.; Hippe, D.; Maravilla, K.R. Macrocyclic and Other Non-Group 1 Gadolinium Contrast Agents Deposit Low Levels of Gadolinium in Brain and Bone Tissue: Preliminary Results from 9 Patients with Normal Renal Function. *Investig. Radiol.* **2016**, *51*, 447–453. [CrossRef] [PubMed]
13. Gulani, V.; Calamante, F.; Shellock, F.G.; Kanal, E.; Reeder, S.B. Gadolinium Deposition in the Brain: Summary of Evidence and Recommendations. *Lancet Neurol.* **2017**, *16*, 564–570. [CrossRef] [PubMed]

14. Tseng, H.Y.; Lee, H.Y.; Jiang, L.; Yang, M.H.; Yang, W. RetrieveGAN: Image Synthesis via Differentiable Patch Retrieval. *Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.)* **2020**, *12353*, 242–257. [CrossRef]

15. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-Shot Text-to-Image Generation. *Int. Conf. Mach. Learn.* **2021**, *139*, 8821–8831.

16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

17. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. *Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.)* **2013**, *8689 LNCS*, 818–833. [CrossRef]

18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *LNCS* **2015**, *9351*, 234–241.

19. Huang, H.; Yu, P.S.; Wang, C. An Introduction to Image Synthesis with Generative Adversarial Nets. *arXiv* **2018**, arXiv:1803.04469. [CrossRef]

20. Fulgeri, F.; Fabbri, M.; Alletto, S.; Calderara, S.; Cucchiara, R. Can Adversarial Networks Hallucinate Occluded People with a Plausible Aspect? *Comput. Vis. Image Underst.* **2019**, *182*, 71–80. [CrossRef]

21. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. In Proceedings of the 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, 14–16 April 2014.

22. Esser, P.; Rombach, R.; Ommer, B. Taming Transformers for High-Resolution Image Synthesis. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12868–12878. [CrossRef]

23. Calabrese, E.; Rudie, J.D.; Rauschecker, A.M.; Villanueva-Meyer, J.E.; Cha, S. Feasibility of Simulated Postcontrast Mri of Glioblastomas and Lower-Grade Gliomas by Using Three-Dimensional Fully Convolutional Neural Networks. *Radiol. Artif. Intell.* **2021**, *3*, e200276. [CrossRef]

24. Gong, E.; Pauly, J.M.; Wintermark, M.; Zaharchuk, G. Deep Learning Enables Reduced Gadolinium Dose for Contrast-Enhanced Brain MRI. *J. Magn. Reson. Imaging* **2018**, *48*, 330–340. [CrossRef]

25. Pasumarthi, S.; Tamir, J.I.; Christensen, S.; Zaharchuk, G.; Zhang, T.; Gong, E. A Generic Deep Learning Model for Reduced Gadolinium Dose in Contrast-Enhanced Brain MRI. *Magn. Reason. Med.* **2021**, *86*, 1687–1700. [CrossRef]

26. Xie, H.; Lei, Y.; Wang, T.; Roper, J.; Axente, M.; Bradley, J.D.; Liu, T.; Yang, X. Magnetic Resonance Imaging Contrast Enhancement Synthesis Using Cascade Networks with Local Supervision. *Med. Phys.* **2022**, *49*, 3278–3287. [CrossRef]

27. Xu, C.; Xu, L.; Gao, Z.; Zhao, S.; Zhang, H.; Zhang, Y.; Du, X.; Zhao, S.; Ghista, D.; Liu, H.; et al. Direct Delineation of Myocardial Infarction without Contrast Agents Using a Joint Motion Feature Learning Architecture. *Med. Image Anal.* **2018**, *50*, 82–94. [CrossRef]

28. Zhang, N.; Yang, G.; Gao, Z.; Xu, C.; Zhang, Y.; Shi, R.; Keegan, J.; Xu, L.; Zhang, H.; Fan, Z.; et al. Deep Learning for Diagnosis of Chronic Myocardial Infarction on Nonenhanced Cardiac Cine MRI. *Radiology* **2019**, *291*, 606–607. [CrossRef]

29. Wang, P.; Hu, S.; Wang, X.; Ge, Y.; Zhao, J.; Qiao, H.; Chang, J.; Dou, W.; Zhang, H. Synthetic MRI in Differentiating Benign from Metastatic Retropharyngeal Lymph Node: Combination with Diffusion-Weighted Imaging. *Eur. Radiol.* **2023**, *33*, 152–161. [CrossRef]

30. Kausar, T.; Kausar, A.; Ashraf, M.A.; Siddique, M.F.; Wang, M.; Sajid, M.; Siddique, M.Z.; Haq, A.U.; Riaz, I. SA-GAN: Stain Acclimation Generative Adversarial Network for Histopathology Image Analysis. *Appl. Sci.* **2021**, *12*, 288. [CrossRef]

31. Cross-Zamirski, J.O.; Mouchet, E.; Williams, G.; Schönlieb, C.B.; Turkki, R.; Wang, Y. Label-Free Prediction of Cell Painting from Brightfield Images. *Sci. Rep.* **2022**, *12*, 10001. [CrossRef]

32. Vassa, R.; Garg, A.; Omar, I.M. Magnetic Resonance Imaging of the Wrist and Hand. *Pol. J. Radiol.* **2020**, *85*, e461. [CrossRef]

33. Reeder, S.B.; McKenzie, C.A.; Pineda, A.R.; Yu, H.; Shimakawa, A.; Brau, A.C.; Hargreaves, B.A.; Gold, G.E.; Brittain, J.H. Water–Fat Separation with IDEAL Gradient-Echo Imaging. *J. Magn. Reson. Imaging* **2007**, *25*, 644–652. [CrossRef]

34. Østergaard, M.; Peterfy, C.G.; Bird, P.; Gandjbakhch, F.; Glinatsi, D.; Eshed, I.; Haavardsholm, E.A.; Lillegraven, S.; Bøyesen, P.; Ejbjerg, B.; et al. The OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging (MRI) Scoring System: Updated Recommendations by the OMERACT MRI in Arthritis Working Group. *J. Rheumatol.* **2017**, *44*, 1706–1712. [CrossRef]

35. Mattes, D.; Haynor, D.; Vesselle, H.; Lewellyn, T.; Eubank David Mattes, W.; Haynor, D.R.; Lewellyn, T.K.; Eubank, W. Nonrigid Multimodality Image Registration. *Med. Imaging 2001 Image Process.* **2001**, *4322*, 1609–1620. [CrossRef]

36. Dosselmann, R.; Yang, X.D. A Comprehensive Assessment of the Structural Similarity Index. *Signal Image Video Process* **2009**, *5*, 81–91. [CrossRef]

37. Beare, R.; Lowekamp, B.; Yaniv, Z. Image Segmentation, Registration and Characterization in R with SimpleITK. *J. Stat. Softw.* **2018**, *86*, 8. [CrossRef]

38. Yaniv, Z.; Lowekamp, B.C.; Johnson, H.J.; Beare, R. SimpleITK Image-Analysis Notebooks: A Collaborative Environment for Education and Reproducible Research. *J. Digit. Imaging* **2018**, *31*, 290. [CrossRef]

39. Lowekamp, B.C.; Chen, D.T.; Ibáñez, L.; Blezek, D. The Design of SimpleITK. *Front. Neuroinform.* **2013**, *7*, 45. [CrossRef]

40. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2016; pp. 5967–5976. [CrossRef]

41. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]

42. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015. [CrossRef]

43. Horé, A.; Ziou, D. Is There a Relationship between Peak-Signal-to-Noise Ratio and Structural Similarity Index Measure? *IET Image Process.* **2013**, *7*, 12–24. [CrossRef]

44. Knoll, F.; Murrell, T.; Sriram, A.; Yakubova, N.; Zbontar, J.; Rabbat, M.; Defazio, A.; Muckley, M.J.; Sodickson, D.K.; Zitnick, C.L.; et al. Advancing Machine Learning for MR Image Reconstruction with an Open Competition: Overview of the 2019 FastMRI Challenge. *Magn. Reason. Med.* **2020**, *84*, 3054–3070. [CrossRef]

45. Mason, A.; Rioux, J.; Clarke, S.E.; Costa, A.; Schmidt, M.; Keough, V.; Huynh, T.; Beyea, S. Comparison of Objective Image Quality Metrics to Expert Radiologists' Scoring of Diagnostic Quality of MR Images. *IEEE Trans. Med. Imaging* **2020**, *39*, 1064–1072. [CrossRef]

46. Tomczak, A.; Gupta, A.; Ilic, S.; Navab, N.; Albarqouni, S. What Can We Learn About a Generated Image Corrupting Its Latent Representation? *Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.)* **2022**, *13436 LNCS*, 505–515.

47. De Winter, J.C.F. Using the Student's t-Test with Extremely Small Sample Sizes. *Pract. Assess. Res. Eval. Pract. Assess.* **2013**, *18*, 10. [CrossRef]

48. Nahler, G. Bonferroni Correction. *Dict. Pharm. Med.* **2009**, 18. [CrossRef]

49. Adamson, P.M.; Gunel, B.; Dominic, J.; Desai, A.D.; Spielman, D.; Vasanawala, S.; Pauly, J.M.; Chaudhari, A. SSFD: Self-Supervised Feature Distance as an MR Image Reconstruction Quality Metric. In Proceedings of the NeurIPS 2021 Workshop on Deep Learning and Inverse Problems, Virtual, 6–14 December 2021.

50. Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883. [CrossRef]

51. Sugawara, Y.; Shiota, S.; Kiya, H. Checkerboard Artifacts Free Convolutional Neural Networks. *APSIPA Trans. Signal Inf. Process.* **2019**, *8*, e9. [CrossRef]

52. Kamrul Hasan, S.M.; Linte, C.A. U-NetPlus: A Modified Encoder-Decoder U-Net Architecture for Semantic and Instance Segmentation of Surgical Instruments from Laparoscopic Images. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2019, Berlin, Germany, 23–27 July 2019; pp. 7205–7211. [CrossRef]

53. Guermazi, A.; Roemer, F.W.; Hayashi, D.; Crema, M.D.; Niu, J.; Zhang, Y.; Marra, M.D.; Katur, A.; Lynch, J.A.; El-Khoury, G.Y.; et al. Assessment of Synovitis with Contrast-Enhanced MRI Using a Whole-Joint Semiquantitative Scoring System in People with, or at High Risk of, Knee Osteoarthritis: The MOST Study. *Ann. Rheum. Dis.* **2011**, *70*, 805–811. [CrossRef]

54. Amini, B.; Murphy, W.A.; Haygood, T.M.; Kumar, R.; McEnery, K.W.; Madewell, J.E.; Mujtaba, B.M.; Wei, W.; Costelloe, C.M. Gadolinium-Based Contrast Agents Improve Detection of Recurrent Soft-Tissue Sarcoma at Mri. *Radiol. Imaging Cancer* **2020**, *2*, e190046. [CrossRef]

55. Crema, M.D.; Felson, D.T.; Roemer, F.W.; Niu, J.; Marra, M.D.; Zhang, Y.; Lynch, J.A.; El-Khoury, G.Y.; Lewis, C.E.; Guermazi, A. Peripatellar Synovitis: Comparison between Non-Contrast-Enhanced and Contrast-Enhanced MRI and Association with Pain. The MOST Study. *Osteoarthr. Cartil.* **2013**, *21*, 413–418. [CrossRef]

56. Lee, J.H.; Dyke, J.P.; Ballon, D.; Ciombor, D.M.K.; Tung, G.; Aaron, R.K. Assessment of Bone Perfusion with Contrast-Enhanced Magnetic Resonance Imaging: Imaging of Bone Marrow Edema Associated with Osteoarthritis and Avascular Necrosis. *Orthop. Clin. N. Am.* **2009**, *40*, 249. [CrossRef]

57. Salaffi, F.; Ceccarelli, L.; Carotti, M.; Di Carlo, M.; Polonara, G.; Facchini, G.; Golfieri, R.; Giovagnoni, A. Differentiation between Infectious Spondylodiscitis versus Inflammatory or Degenerative Spinal Changes: How Can Magnetic Resonance Imaging Help the Clinician? *Radiol. Med.* **2021**, *126*, 843. [CrossRef]

58. Tolpadi, A.A.; Bharadwaj, U.; Gao, K.T.; Bhattacharjee, R.; Gassert, F.G.; Luitjens, J.; Giesler, P.; Morshuis, J.N.; Fischer, P.; Hein, M.; et al. K2S Challenge: From Undersampled K-Space to Automatic Segmentation. *Bioengineering* **2023**, *10*, 267. [CrossRef]

59. Zhao, R.; Yaman, B.; Zhang, Y.; Stewart, R.; Dixon, A.; Knoll, F.; Huang, Z.; Lui, Y.W.; Hansen, M.S.; Lungren, M.P. FastMRI+: Clinical Pathology Annotations for Knee and Brain Fully Sampled Multi-Coil MRI Data. *Sci. Data.* **2022**, *9*, 152. [CrossRef]

60. Peterfy, C.G.; Schneider, E.; Nevitt, M. The Osteoarthritis Initiative: Report on the Design Rationale for the Magnetic Resonance Imaging Protocol for the Knee. *Osteoarthr. Cartil.* **2008**, *16*, 1433–1441. [CrossRef]