Title
Multi-modal Medical Imaging Registration

Author
Ramalingam, Deeban

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Multi-modal Medical Imaging Registration

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Computer Science

by

Deeban Ramalingam

2021

ABSTRACT OF THE THESIS

Multi-modal Medical Imaging Registration

by

Deeban Ramalingam

Master of Science in Computer Science

University of California, Los Angeles, 2021

Advisor: Professor Fabien Scalzo

Medical image registration automatically brings two images into maximal spatial and anatomical correspondence. The registration of brain images taken at different modalities and time points helps in designing a treatment plan in neurology besides assisting in surgery. The registration of stained and unstained skin samples in dermatology used in virtual histological staining helps the dermatologists by removing the labor and cost that is involved in invasive techniques such as skin biopsies. The registration of histological skin images can be more challenging than the registration of brain images due to repetitious image textures, and the various expansions and morphological changes of tissue samples caused by the contrast agent used to stain the skin. For these reasons, the standard registration approaches may not adequately address this area of medical image registration. Advanced registration techniques have been developed to overcome such challenges. Recently, machine learning methods have gained popularity in multi-modal medical image registration. The goal of the thesis is to apply the

standard, advanced, and machine learning-based multi-modal registration methods to the 3D microscopic histological dermatological data and to compare their performance.

The thesis of Deeban Ramalingam is approved.

Quanquan Gu

Yair Rivenson

Fabien Scalzo

Demetri Terzopoulos, Chair

University of California, Los Angeles

2021

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my Master Thesis Advisor, Dr. Professor Fabien Scalzo, for his support in my Thesis. He kept up with my progress and was open to clear any of my doubts. The Thesis would not have been possible without the extraordinary guidance of my Advisor.

My sincere gratitude to Dr. Yair Rivenson for introducing me to Dr. Professor Fabien Scalzo. I would like to thank Dr. Professor Aydogan Ozcan, Dr. Yair Rivenson, and Jingxi Li for providing the 3D microscopic dermatological data necessary to conduct my experiments required for the Thesis. I would also like to thank my family for their unconditional support throughout my professional endeavors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Medical image registration is a useful technique that automatically brings two images into anatomical correspondence. In neurology, for example, registering images of a patient's brain taken at different modalities and time points helps physicians and neurosurgeons to identify a disease, plan a treatment, and perform surgery [1]. In virtual histological staining, registering unstained and stained skin samples acquired from a non-invasive microscopy technique assists dermatologists by removing the labor and cost involved in taking a skin biopsy from a patient for the diagnosis of diseases [2].

While deformations in the brain caused by injuries like strokes pose a challenge for neurological image registration, the registration of histological images can be more challenging due to the following reasons: (i) large image sizes; (ii) repetitive textures in the images that make it difficult to find distinguishing landmarks; (iii) various expansions and morphological changes to tissue samples caused by the contrast agent that is used to stain tissues; (iv) differences in the images caused by various acquisition methods; and (v) local structural differences within a 3D image stack [3]. For these reasons, the registration of skin samples is often beyond standard registration approaches. In fact, such approaches are outperformed by complex transformations that involve an initial rigid alignment followed by a non-rigid registration method [3]. Recently, machine learning methods have also shown promise in multi-modal medical image registration.

2. RELATED WORK

2.1. STANDARD AND ADVANCED REGISTRATION APPROACHES

The software MATLAB [5] provides the Image Processing Toolbox, which contains the interactive Registration Estimator App, intensity-based automatic image registration, and control point registration. The Registration Estimator App enables the user to interactively register a source image to a target image. Different registration techniques and settings can be selected. The app offers a quantitative registration quality measure and also provides the transformation matrix associated with the registration. The intensity-based automatic image registration aligns the corresponding pixels in each image based on their relative intensity patterns. The registration technique can be applied to both 2D and 3D monomodal and multimodal image pairs, and it is also useful for the registration of a large number of images. The user can specify the types of geometric transformation applied to the source image. The registration technique iteratively adjusts the parameters of the geometric transformation to optimize the spatial correspondence between a pair of images.

A localized displacement field can also be estimated using this optimization method and is applied as a nonrigid transformation to the source image. The control point registration enables the user to select features that are common to the image pairs. The registration technique is useful when specific anatomical features that need to be aligned are preferred to the less informative, repetitive features that are detected using an automated feature detection strategy. The control point registration supports a variety of global and nonrigid transformations. MATLAB also provides the Computer Vision Toolbox that offers automated feature detection

and matching. Automated feature detection and extraction can be used to detect corresponding features like corners and blobs that are from the source and target images and to estimate a geometric transformation that aligns the matched features.

An intensity-based medical image registration software is presented in [4]. The software formulates the image registration problem as an optimization problem. A cost function is minimized using an iterative optimization method with respect to the parameters of a registration transformation embedded in a multiresolution scheme. Some examples of cost functions include mean squared difference (MSD), normalized correlation (NC), and variants of mutual information (MI). The transformation models that are supported by the software include translation, rigid, similarity, affine, and nonrigid, such as B-spline, thin-plate spline, and elastic body spline. When a nonrigid transformation is used, a regularization term that penalizes undesired deformations can be added to the cost function. The regularization term is often expressed in terms of first- and second-order spatial derivatives of the transformation. The cost functions that are available to this software include gradient descent, quasi-Newton, nonlinear conjugate gradient, evolution strategy, and various stochastic gradient descent methods. Some examples of multi-resolution strategies include image pyramid schemes, varying B-spline grid resolutions, and adjusting the number of histogram bins used in computing the MI.

Three experiments were used to validate the effectiveness of the software. In the first experiment, the software was tested on 50 MR prostate scans from different patients. The Dice Score Coefficient (DSC) was used to compare various registration transformations. This experiment demonstrated that using nonrigid registration was essential for the application. The software was also used in the second experiment to register T1-weighted MR and PET images taken from the “retrospective image registration evaluation” (RIRE) project with ground truth

correspondences for the eight corner points of each image. The application of the software to the data demonstrates that the random sampling of voxels off a uniform grid in a target image has been shown to improve the smoothness of the cost function. In the third experiment, the software was applied to CT chest scans of 26 patients taken from a baseline and follow-up scans for a lung cancer screening trial. The mean distance between corresponding points of the warped and target images was used to evaluate the registration accuracy of several multiresolution strategies. This experiment demonstrates that refining the grid associated with the B-spline transformation continued to improve with increasing resolution levels.

The Automatic Non-Rigid Histological Image Registration Challenge (ANHIR) was held to compare the performance of various registration methods on microscopic histological images [3]. The dataset to be registered consists of 481 image pairs of lung lesions, lung lobes, mammary glands, mouse kidney, colon adenocarcinoma, gastric mucosa, human breast, and human kidney. The registration accuracy was evaluated by measuring the relative Target Registration Error (rTRE) between the corresponding landmark positions in the ground-truth and registered images. The methods that performed well in the challenge utilized coarse but robust initial alignment, followed by nonrigid registration. In the first method conducted by Fraunhofer, an affine transformation was found by applying the Gauss-Newton method that optimized the normalized gradient field (NGF) similarity of grayscale versions of the images. A dense, non-rigid transformation was then found using the curvature regularization and Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimization along with the same optimization criterion. In the second method conducted by Chengdu, a rigid, scaling transformation was first estimated on down-sampled images by optimizing the normalized cross-correlation (NCC) using gradient descent. A non-rigid registration represented by B-splines was then found locally using

L-BFGS optimization of the NCC. In a third method conducted by AGH UST, RANSAC was first used to determine an initial similarity or rigid transformation by selecting key-points in each image pair using SIFT, ORB, or SURF. A non-rigid transformation was then found using local affine registration, various versions of the demons algorithms, or a feature-point-based thin-plate spline interpolation.

2.2. MACHINE LEARNING-BASED REGISTRATION APPROACHES

Tang and Scalzo (2016) developed an image registration algorithm to find a transformation that produced a coordinate system of maximal spatial correspondence between 3D MRA and 2D DSA images [6]. The 3D MRA and 2D DSA images were collected from admitted patients diagnosed with symptoms of ischemic stroke at a stroke center. The parameters of the transformation are obtained via Nedler-Mead Simplex algorithm. At each iteration of the optimization, the algorithm first applies an affine transformation using a candidate set of parameters. The algorithm evaluates the quality of the parameter set by comparing the resulting transformed image to a target image using a similarity metric. The similarity metric was learned using Spectral Regression for Discriminant Analysis from corresponding 2D point descriptors between the transformed and target images. The 2D point descriptors were obtained by applying the filtering and thresholding techniques as well as semi-automatic software-assisted segmentation to the original angiographic image data. The root mean square error computed between the same points in the transformed source image, which used the estimated and ground-truth parameters, was utilized to evaluate the method.

De Vos et al. (2017) present a convolutional neural network model consisting of a ConvNet regressor, spatial transformer, and resampler [7]. The ConvNet regressor applies a series of convolutional layers to a concatenated pair of source and target images to output local

deformation parameters. These parameters are used by the spatial transformer to generate a displacement field that allows the resampler to warp the source image to match the target image. The network is trained on randomly sampled batches of source and corresponding target images in an end-to-end fashion to optimize an image similarity metric between the predicted warped and target images. The network was tested not only on handwritten digits but also on slices of cardiac cine MRI scans. The pair of source and target images were corresponding anatomical slices from the same scan of a single patient, but obtained at different points in time of the cardiac cycle. The mean and standard deviation for the Dice score, 95th percentiles of the surface distance (95thSD), and mean absolute surface distance (MAD) were measured to evaluate the approach.

Balakrishnan et al. (2019) used a neural network model, namely VoxelMorph, to parametrize a function that maps a pair of source and target images to a displacement field [8]. The model first computes a displacement field from the source and target images using a Convolutional U-net architecture. The model then warps the source image according to the displacement field using a spatial transformation function. The model is trained to minimize the dissimilarity between the warped image and the target image. The mean-square-error and negative cross correlation were used to determine dissimilarity. Since minimizing the dissimilarity may generate non-smooth, unrealistic displacement fields, a smoothness loss that penalizes local spatial variations was also included in the overall objective function. The model was tested on a large-scale, multi-site, multi-study dataset of T1–weighted brain MRI scans. The Dice score was used to measure the registration accuracy of this approach.

Bhalodia et al. (2019) proposed a cooperative neural network architecture consisting of two interacting convolutional subnetworks [9]. The primary network is tasked with finding a

displacement field between a source and target image, which when applied to the source image, produces a warped image that is aligned to the target image. The purpose of the secondary autoencoder network is to regularize the displacement fields found by the primary network. This is done by reconstructing the displacement field with a bottleneck layer of relatively fewer hidden units. The reconstruction loss from the secondary network as well as registration loss of the primary network constitutes the final objective function which is minimized to encourage the model to restrict displacement fields to a low dimensional manifold. A synthetic dataset of binary shapes was used to demonstrate that the secondary network uses information about the population statistics of the data. Training the model on 2D slices of the Corpus Callosum demonstrates that the displacement fields that were produced lie close to a low-dimensional manifold. The model was also trained on a 3D dataset of appendages. The AE error, Dice coefficient, and Landmark error were used to evaluate the method on these datasets.

Arar et al. (2020) developed an unsupervised multi-modal image registration algorithm consisting of a spatial transformer network and translation network [10]. During the training procedure, the spatial transformer network first produced a deformation field from the source and target images of different modalities. Two separate training flows are used to produce two warped images using this deformation field. In the first flow, a resampling layer applies the deformation field to the source image to obtain a warped image and is translated to the target modality. In the second flow, the source image is also translated to the target modality and is then resampled according to the deformation field to produce another warped image. The warped images are then compared against the target image to produce a reconstruction loss. An adversarial loss that encouraged the generation of paired-samples to match the real distribution was considered in the overall objective function. The model was trained on a collection of RGB,

IR, and Depth images of banana plants with different growing conditions and phenotypes. The average Euclidean distance between the target points and their matching deformed source points on salient objects and full scene annotations was used to evaluate the accuracy of the approach.

3. METHODS

The standard, advanced, and machine learning-based registration methods are used to automatically align a set of source and target image pairs. The standard approaches used to register the images are the 3D affine transformations from MATLAB and Elastix. The advanced approaches are 3D Demons and 3D B-Spline registration techniques from MATLAB and Elastix, respectively. The VoxelMorph registration algorithm is used as the machine learning-based approach. Since these methods are applied to 3D microscopic dermatological data, the average SSIM is used to evaluate the registration accuracy of the methods.

3.1. MATLAB

The "imreg" command in MATLAB is used to apply a 3D affine transformation to align the source image to the target image. The affine transformation accounts for translation, rotation, scaling, and shearing. The OnePlueOneEvolutionary optimizer is used to maximize the Mattes Mutual Information between the source and target images. At each iteration of the optimizer, the present parameters are perturbed to obtain a set of candidate parameters. If the candidate parameters produce a better result with respect to the optimization criterion, then these parameters are used in the next iteration with perhaps a more aggressive perturbation. If the original parameters return a better result, then these parameters are used instead with relatively less perturbation. During the entirety of the optimization procedure, the Mattes mutual information algorithm uses a single set of pixel locations to estimate a joint probability distribution of pixels between the source and target images. The distribution measures the certainty that the selected pixels between the images have similar values. Hence, high mutual

information corresponds to a low entropy between the marginal distributions of each image. This suggests that there is better alignment between the images [5]. 500 pixel samples and 50 bins are used to compute the mutual information between the source and target images.

The “imregdemons” command in MATLAB is also used to estimate a localized displacement field and apply a 3D nonrigid transformation to the source image. The command uses the non-parametric, diffeomorphic demons algorithm to register the source and target images [5]. In each iteration of the demons algorithm, a small deformation field is sought to minimize the intensity differences between the source and target images in a least-squares sense [11, 12]. Gaussian smoothing is applied to regularize the minimization problem [5]. This efficient, intensity-based registration method involves the minimization of the (mean square error) MSE criterion that uses the efficient second-order minimization (ESM) framework [12]. The “imhistmatch” command is used to correct the illumination differences between the source and target images using histogram matching [5].

3.2. ELASTIX

A 3D affine transformation similar to one in MATLAB is used to register the source and target images [4]. The 3D B-spline deformation from Elastix along with the MI similarity criteria and the adaptive stochastic gradient descent optimizer is used to register the source and target images [3]. The transformation associated with a B-spline deformation consists of a weighted sum of B-spline basis functions placed on a uniform grid of control points [4]. The control points act as parameters of the B-spline deformation. The degree of the deformation depends on the resolution of the grid of control points. A large spacing of control points models global nonrigid deformations, whereas a small spacing of control points models highly local nonrigid deformations [13]. MI measures the extent to which the image intensity in an image relates to the

image intensity in another image. This in turn corresponds to measuring the dependency of the probability density distributions of the intensities of the source and target images. The MI is optimized via gradient descent with an adaptive gain [4].

3.3. VOXELMORPH

A 3D displacement field produced by VoxelMorph is used to register source and target images [8]. VoxelMorph passes a pair of source and target images through several convolutional layers to produce a displacement field. The displacement field is then used to warp the source image using a spatial transformer layer. The unsupervised loss function used to train VoxelMorph consists of an MSE between warped and target images as well as an L2 smoothness regularization on the spatial gradient of the displacement field. This L2 regularization term is used to penalize local spatial variations present in unrealistic displacement fields that result from minimizing the MSE. VoxelMorph is trained and tested on two separate sets of source and target image pairs, respectively. Resampling based on linear interpolation is used to resize the source and target images.

3.4. EVALUATING REGISTRATION ACCURACY USING AVERAGE SSIM

The measurement of registration accuracy using pixel-wise error is not highly indicative of perceived similarities. The Structural Similarity Index (SSIM), on the other hand, addresses this shortcoming by taking texture into account in order to compare more complex aspects of two images such as luminance, contrast, and general structures [14]. For this reason, the SSIM is used as a metric of performance to measure registration accuracy. The SSIM of two images x, y is computed using the following formula

$$\boldsymbol{SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\,\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \,\sigma_y^2 + C_2)}}$$

where μ_x is the mean intensity of x, μ_y is the mean intensity of y, σ_x is the standard deviation of x, σ_y is the standard deviation of y, and σ_{xy} is the covariance of x and y. C_1 is a constant used to stabilize the expression $\mu_x^2 + \mu_y^2$ when it is near zero. Similarly, C_2 is a constant used to address the case when $\sigma_x^2 + \sigma_y^2$ is close to zero.

Since the aforementioned methods are applied to 3D microscopic dermatological data, the average SSIM is used to evaluate the methods. The average SSIM is calculated by taking an average across n pairs of corresponding 2D image slices that comprise a pair of 3D images X, Y. The following formula is therefore used to compute the average SSIM

$$\boldsymbol{SSIM_{AVERAGE}} = \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{SSIM}(\boldsymbol{X}[\boldsymbol{i}], \boldsymbol{Y}[\boldsymbol{i}])$$

where $X[i]$ and $Y[i]$ denote the i -th slices of X and Y, respectively.

4. EXPERIMENTAL ANALYSIS

4.1. 3D MICROSOPIC DERMATOLOGICAL DATA

For a given patient case, a set of 3D time-lapsed image stacks of microscopic dermatological data are acquired using Reflectance Confocal Microscopy. Pyramid Elastic Registration [15] is used to align two corresponding images from temporally neighbored image stacks. This procedure is performed so that the image stack acquired before staining can be roughly registered to the image stack acquired after staining. The other images within these stacks are respectively aligned with these two images. As a result, the two image stacks before and after staining are also aligned with each other forming the initially registered input-target image pairs. The dimension of each image in a stack is around 900x900. The images are mono-channeled, and the number of images in each stack can vary.

These image pairs are then used to train a neural network that performs virtual staining. After training, the network performs inference on the input images to generate output images. These output images are aligned with the target images using Pyramid Elastic Registration [15] to achieve a more accurate spatial correspondence between input and target images. Image pairs that consist of the input and output images of the previous registration are used to train another neural network to also perform virtual staining. After training, the network performs inference on the input images to generate output images as was conducted previously.

4.2. COMPARISON OF STANDARD AND ADVANCED REGISTRATION APPROACHES

The target images are aligned to the output images using standard and advanced registration methods from MATLAB and Elastix. The SSIM scores are averaged over all pairs of output and registered target images in their respective stacks to evaluate the registration methods. The optimizer and metric configurations of the MATLAB Affine 3D registration method are initialized with the multi-modal setting. The Initial Radius, Epsilon, Growth Factor, and Maximum Iterations of the optimizer are set to 0.009, 1.5e-4, 1.01, and 300, respectively. The average SSIM for the MATLAB and ELASTIX registration methods are reported in **Table 4.1**. The Gaussian smoothing value associated with the smoothing is set to 1.3 and three multi-resolution image pyramid levels of 500, 400, and 200 iterations each are used in the MATLAB Demons 3D registration. For the first Elastix B-Spline 3D registration (ELASTIX BSPLINE-3D-I), the maximum number of iterations for the convergence of the B-Spline Transform is set at 512. For the second Elastix BSpline 3D Registration, the number of iterations for convergence was set to the default value and an Affine Transform was utilized prior to the BSpline Transform (ELASTIX BSPLINE-3D-II). MATLAB DEMONS-3D obtained the greatest mean average SSIM of 0.61 across the selected cases of patients in **Table 4.1**. **Figure 4.1** shows the target, output, and registered target images for case number 112 layer 24 using the above standard and advanced registration approaches, while **Figure 4.2** shows the target, output, and registered target images for case number 107 layer 21.

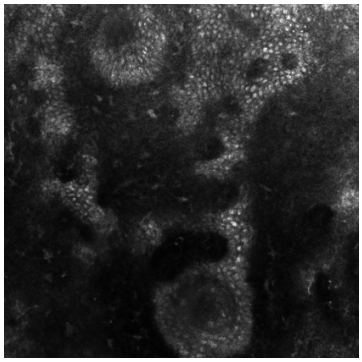
4.3. COMPARISON WITH MACHINE LEARNING-BASED REGISTRATION

The target images are also aligned to the output images using the machine learning-based method VoxelMorph. 100 epochs are used to train VoxelMorph, and it is trained on 100 batches

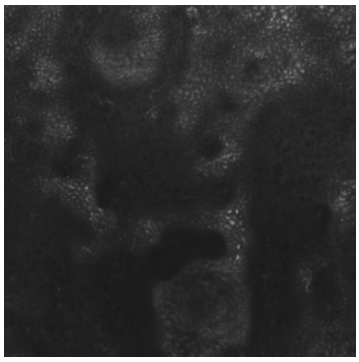
per epoch. The batch size is set to 2. The implementation of VoxelMorph is from https://github.com/voxelmorph/voxelmorph. The number of encoder features for the U-Net are: 8 and 16. The number of decoder features for the U-Net are: 16, 16, and 8. The diffeomorphism option is disabled. The 3D source and target images were resampled to 128x128x128. NVIDIA GeForce GTX 1060 3GB and NVIDIA TITAN RTX are used during the training procedure. VoxelMorph is trained on patient case numbers 29, 30, 32, 33, 34, 35, 36, 37, 39, and 61 and tested on patient case numbers 107, 108, 109, 110, 111, and 112. The regularization weight parameter was set to 0.05. The average SSIM for the VoxelMorph model corresponding to the 58^{th} epoch (ML-BASED VOXEL-MORPH-3D-REGISTRATION) as well as ELASTIX AFFINE-3D, ELASTIX BSPLINE-3D-II, MATLAB DEMONS-3D, and ELASTIX AFFINE-3D are reported in **Table 4.2**. The training and testing losses up to epoch 58 are shown in **Figure 4.5** and **Figure 4.6**, respectively. ML-BASED VOXEL-MORPH-3D-REGISTRATION obtained the greatest average SSIM across the selected cases of patients in **Table 4.2**, thus obtaining the greatest mean average SSIM of 0.72 when compared to the other registration methods. **Figure 4.3** shows the target, output, and registered target images for case number 112 layer 24 using ML-BASED VOXEL-MORPH-3D-REGISTRATION, while **Figure 4.4** shows the target, output, and registered target images for case number 107 layer 21.

5. DISCUSSION AND CONLUSION

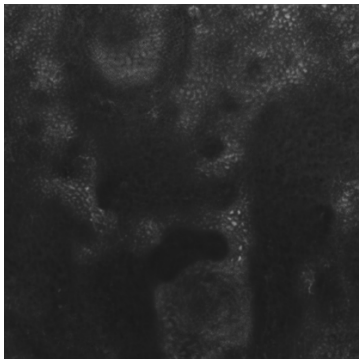
Medical image registration techniques are used to automatically align patients' images. In the thesis, previous literature studies related to standard, advanced, and machine learning-based registration methods are reviewed. Affine, B-spline, and Demons registration methods from MATLAB and Elastix, as well as the robust machine learning-based registration method VoxelMorph are applied to 3D microscopic histological dermatological data. VoxelMorph obtains the greatest registration accuracy across a selected set of patient cases. Therefore, the thesis demonstrates the effectiveness of Machine Learning-based approaches in the registration of skin samples used in virtual histological staining. Such approaches will help dermatologists by reducing the cost and effort involved in invasive techniques like skin biopsies.



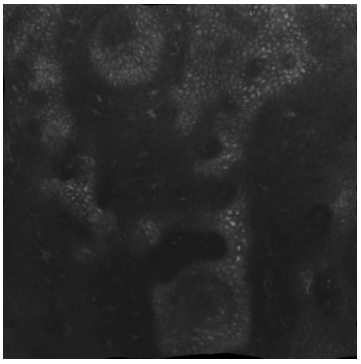
TARGET



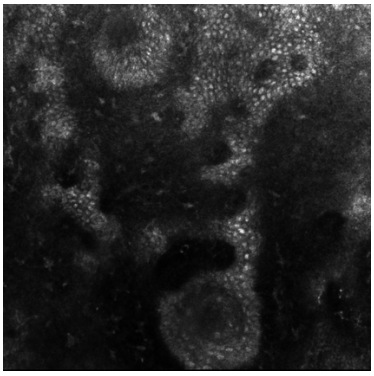
OUTPUT



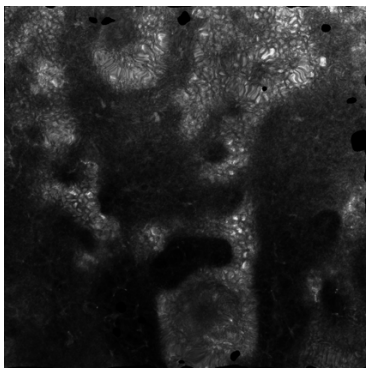
MATLAB AFFINE-3D



MATLAB DEMONS-3D



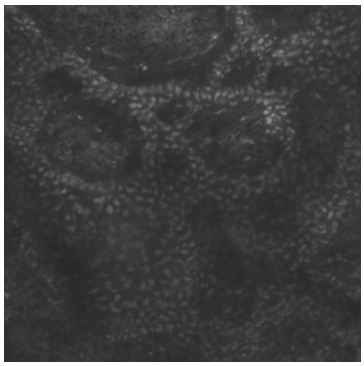
ELASTIX AFFINE-3D



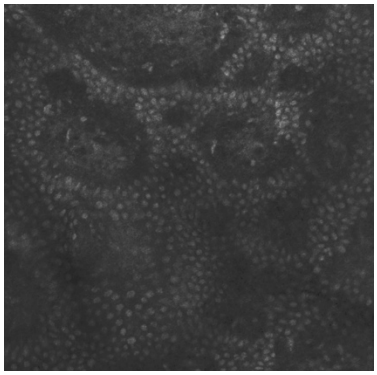
ELASTIX BSPLINE-3D

FIGURE 4.1: PATIENT CASE NUMBER 112 LAYER 24 TARGET, OUTPUT, AND REGISTERED TARGET IMAGES FOR THE STANDARD AND ADVANCED REGISTRATION METHODS

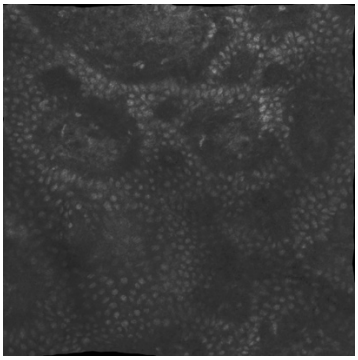
TARGET



OUTPUT

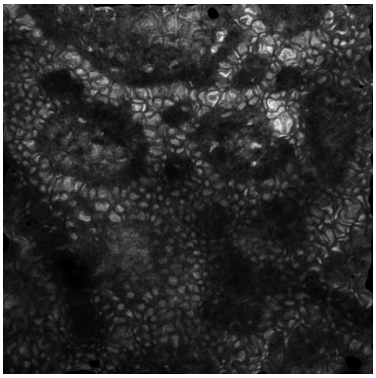


MATLAB AFFINE-3D



MATLAB DEMONS-3D

ELASTIX AFFINE-3D



ELASTIX BSPLINE-3D

FIGURE 4.2: PATIENT CASE NUMBER 107 LAYER 21 TARGET, OUTPUT, AND REGISTERED TARGET IMAGES FOR THE STANDARD AND ADVANCED REGISTRATION METHODS

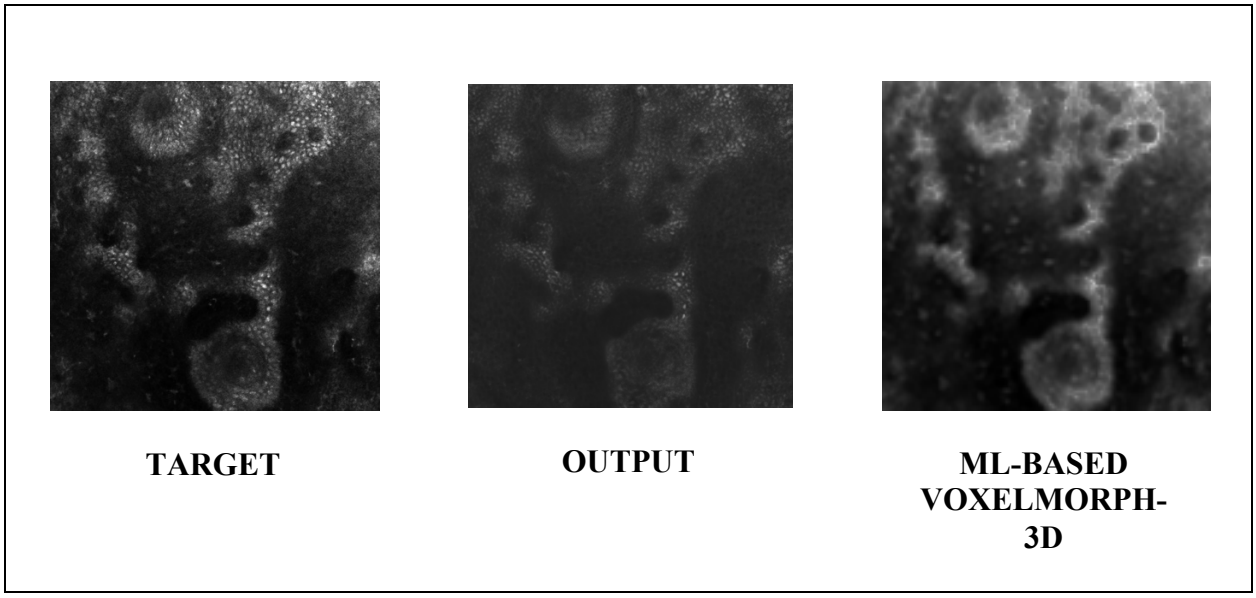


FIGURE 4.3: PATIENT CASE NUMBER 112 LAYER 24 TARGET, OUTPUT, AND REGISTERED TARGET IMAGES FOR THE MACHINE LEARNING-BASED VOXELMORPH-3D REGISTRATION METHOD

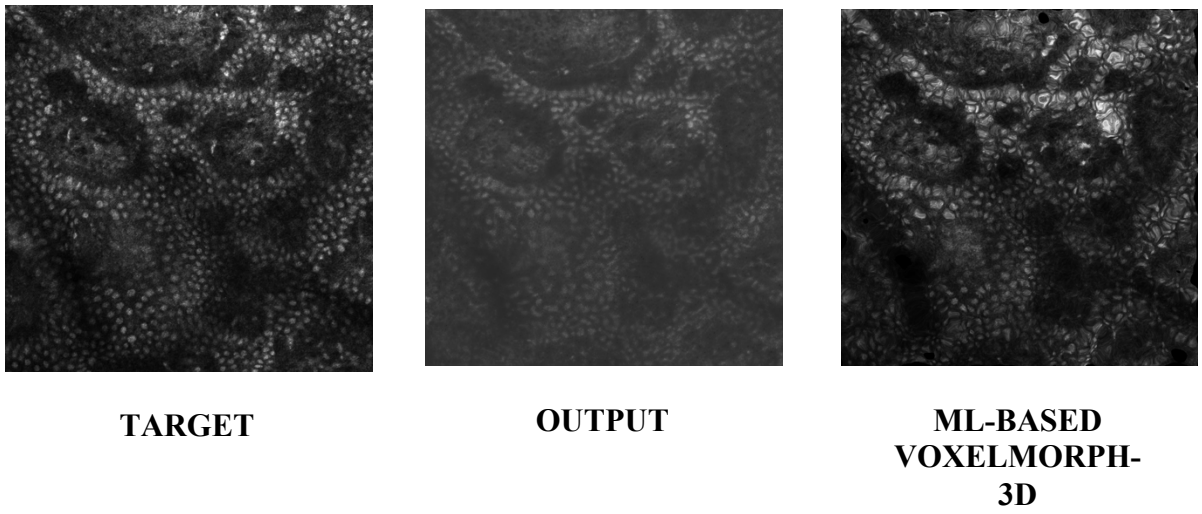


FIGURE 4.4: PATIENT CASE NUMBER 107 LAYER 21 TARGET, OUTPUT, AND REGISTERED TARGET IMAGES FOR THE MACHINE LEARNING-BASED VOXELMORPH-3D REGISTRATION METHOD

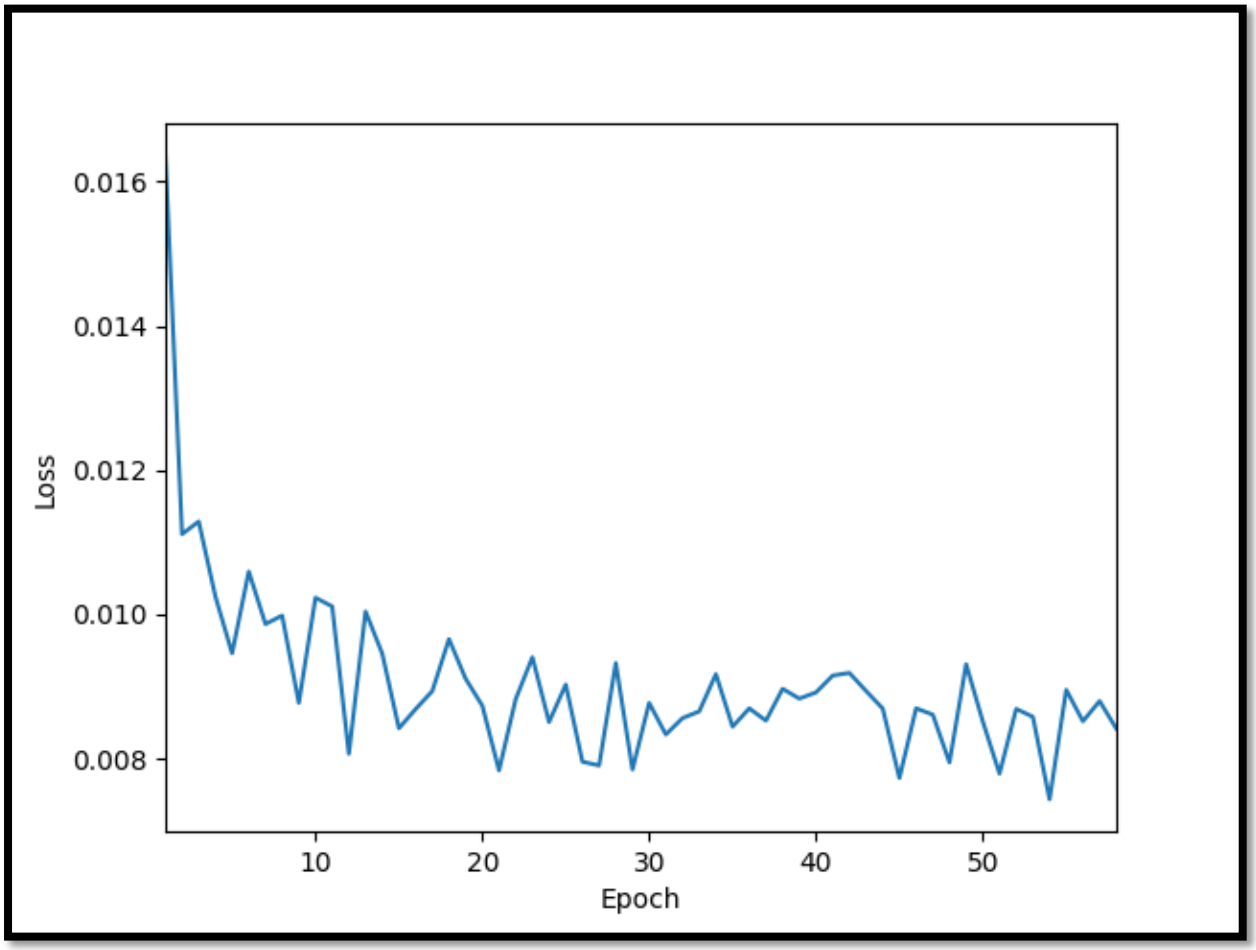


FIGURE 4.5: TRAINING LOSS FOR MACHINE LEARNING-BASED VOXELMORPH-3D REGISTRATION

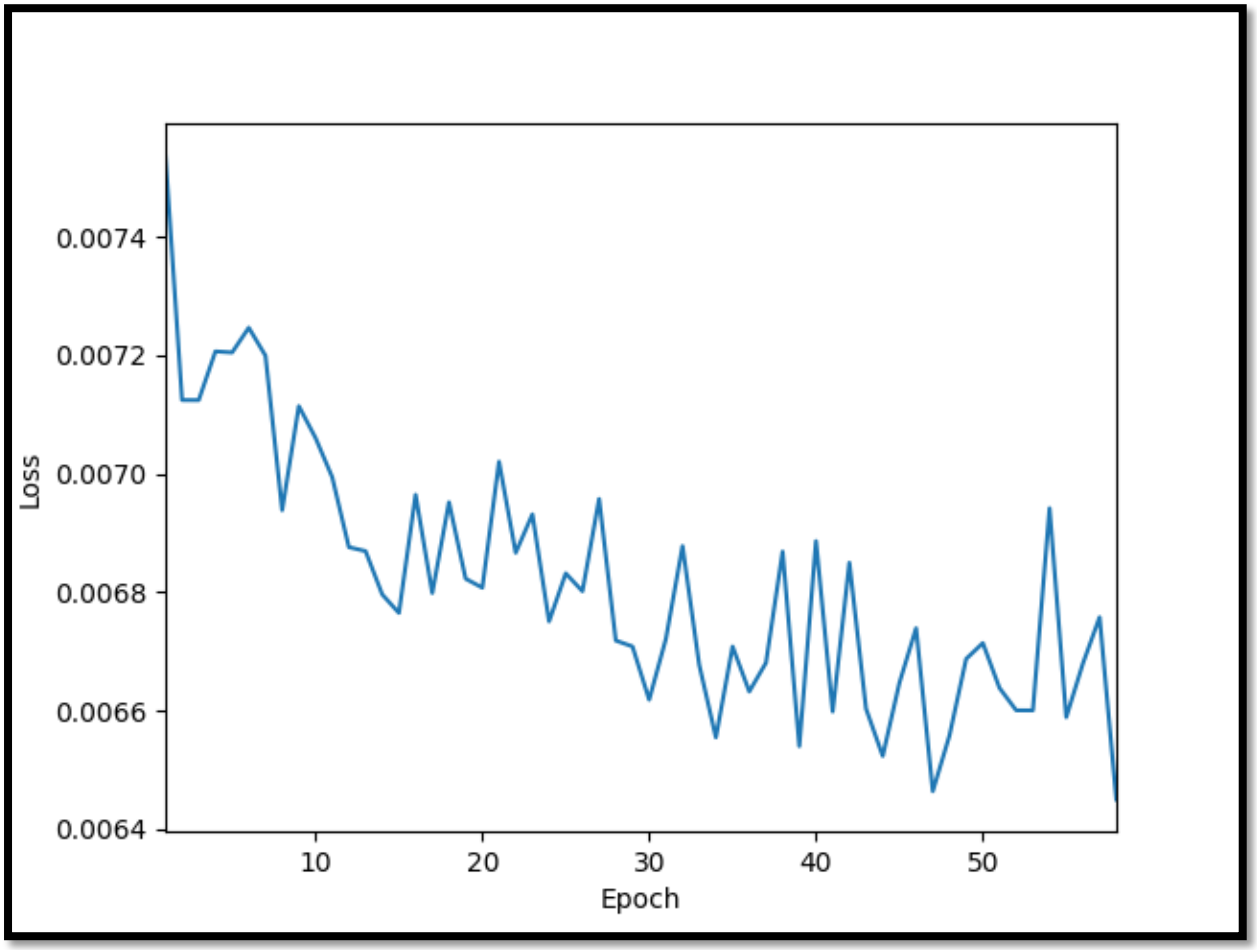


FIGURE 4.6: TESTING LOSS FOR MACHINE LEARNING-BASED VOXELMORPH-3D REGISTRATION

CASE NUMBER (patient)	ORIGINAL	MATLAB AFFINE-3D	MATLAB DEMONS-3D	ELASTIX AFFINE-3D	ELASTIX BSPLINE-3D-I	ELASTIX BSPLINE-3D-II
107	0.44	0.44	0.59	0.44	0.43	0.43
108	0.40	0.40	0.63	0.40	0.39	0.39
109	0.39	0.39	0.63	0.40	0.34	0.37
110	0.37	0.37	0.53	0.40	0.42	0.41
111	0.44	0.44	0.55	0.40	0.41	0.40
112	0.49	0.49	0.70	0.48	0.46	0.46
MEAN	0.42	0.42	0.61	0.42	0.41	0.41

TABLE 4.1: THE COMPARISON OF STANDARD AND ADVANCED REGISTRATION APPROACHES UTILIZING AVERAGE SSIM FOR VARIOUS PATIENTS

CASE NUMBER (patient)	ORIGINAL	ELASTIX AFFINE-3D	ELASTIX BSPLINE-3D-II	MATLAB DEMONS-3D	ML-BASED VOXELMORPH-3D
107	**0.44**	**0.44**	**0.43**	**0.59**	**0.67**
108	**0.40**	**0.40**	**0.39**	**0.63**	**0.67**
109	**0.39**	**0.40**	**0.37**	**0.63**	**0.72**
110	**0.37**	**0.40**	**0.41**	**0.53**	**0.67**
111	**0.44**	**0.40**	**0.40**	**0.55**	**0.79**
112	**0.49**	**0.48**	**0.46**	**0.70**	**0.81**
MEAN	**0.42**	**0.42**	**0.41**	**0.60**	**0.72**

TABLE 4.2: THE COMPARISON OF STANDARD AND ADVANCED REGISTRATION WITH MACHINE LEARNING-BASED VOXELMORPH-3D UTILIZING AVERAGE SSIM FOR VARIOUS PATIENTS

REFERENCES

[1] J. Michael Fitzpatrick. The Retrospective Image Registration Evaluation Project. Retrieved October 10, 2020 from https://www.insight-journal.org/rire/index.php.

[2] Rivenson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., Günaydın, H., Zuckerman, J. E., Chong, T., Sisk, A. E., Westbrook, L. M., Wallace, W. D., & Ozcan, A. (2019). Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. Nature Biomedical Engineering, 3(6), 466–477. https://doi.org/10.1038/s41551-019-0362-y. Retrieved October 31, 2020 from https://www.researchgate.net/publication/331506635_Virtual_histological_staining_of_unlabelled_tissue-autofluorescence_images_via_deep_learning.

[3] J. Borovec *et al*., "ANHIR: Automatic Non-Rigid Histological Image Registration Challenge," in *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3042-3052, Oct. 2020, doi: 10.1109/TMI.2020.2986331. Retrieved November 13, 2020 from https://ieeexplore.ieee.org/document/9058666.

[4] S. Klein, M. Staring, K. Murphy, M.A. Viergever and J.P.W. Pluim, "elastix: a toolbox for intensity based medical image registration," IEEE Transactions on Medical Imaging, vol. 29, no. 1, pp. 196 - 205, January 2010. Retrieved October, 31, 2020 from https://elastix.lumc.nl/marius/downloads/2010_j_TMI.pdf.

[5] MATLAB and Statistics Toolbox Release 2020b, The MathWorks, Inc., Natick, Massachusetts, United States. Retrieved October 31, 2020 from https://www.mathworks.com/.

[6] Tang A., Scalzo F. (2016) Similarity Metric Learning for 2D to 3D Registration of Brain Vasculature. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2016. Lecture

Notes in Computer Science, vol 10072. Springer, Cham. https://doi.org/10.1007/978-3-319-50835-1_1. Retrieved October 31, 2020 from https://link.springer.com/chapter/10.1007/978-3-319-50835-1_1.

[7] de Vos, B. D., Berendsen, F. F., Viergever, M. A., Staring, M., & Išgum, I. (2017). End-to-End Unsupervised Deformable Image Registration with a Convolutional Neural Network. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 204–212). Springer International Publishing. https://doi.org/10.1007/978-3-319-67558-9_24. Retrieved October 31, 2020 from https://arxiv.org/abs/1704.06065.

[8] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., & Dalca, A. V. (2019). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging, 38(8), 1788–1800. https://doi.org/10.1109/tmi.2019.2897538. Retrieved October 31, 2020 from https://arxiv.org/abs/1809.05231.

[9] Bhalodia, Riddhish, et al. "A Cooperative Autoencoder for Population-Based Regularization of CNN Image Registration." ArXiv.org, 19 Aug. 2019, arxiv.org/abs/1908.05825. Retrieved October 31, 2020 from https://arxiv.org/pdf/1908.05825.pdf.

[10] Arar, Moab, et al. "Unsupervised Multi-Modal Image Registration via Geometry Preserving Image-to-Image Translation." ArXiv.org, 18 Mar. 2020, arxiv.org/abs/2003.08073. Retrieved October 31, 2020 from https://openaccess.thecvf.com/content_CVPR_2020/papers/Arar_Unsupervised_Multi-Modal_Image_Registration_via_Geometry_Preserving_Image-to-Image_Translation_CVPR_2020_paper.pdf.

[11] Jean-Philippe Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, Elsevier, 1998, 2 (3), pp.243—260. Retrieved November 13, 2020 from https://hal.inria.fr/inria-00615088/document.

[12] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration.. *NeuroImage*, Elsevier, 2009, 45 (1), Supplement 1, Pages S61-S72. Retrieved November 13, 2020 from https://www.hal.inserm.fr/inserm-00349600/document.

[13] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," in *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712-721, Aug. 1999, doi: 10.1109/42.796284. Retrieved November 13, 2020 from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=796284.

[14] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004, doi: 10.1109/TIP.2003.819861. Retrieved December 4, 2020 from http://www.cns.nyu.edu/pub/lcv/wang03-preprint.pdf.

[15] Rivenson, Y., Ceylan Koydemir, H., Wang, H., Wei, Z., Ren, Z., Günaydın, H., Zhang, Y., Göröcs, Z., Liang, K., Tseng, D., & Ozcan, A. (2018). Deep Learning Enhanced Mobile-Phone Microscopy. ACS Photonics, 5(6), 2354–2364. https://doi.org/10.1021/acsphotonics.8b00146. Retrieved November 19, 2020 from https://pubs.acs.org/doi/10.1021/acsphotonics.8b00146.