Registration of the T2 and DWI b1200 MRI Sequences of the Prostate

Rohan Sikand rsikand@stanford.edu
Urologic Cancer Innovation Lab
Stanford University School of Medicine
Canary CREST Program 2021

Abstract

Prostate cancer is the most commonly diagnosed cancer as well as the second leading cause of death in men [8]. In the process of diagnosis, prognosis, and treatment planning, Magnetic Resonance Imaging (MRI) is used. Image registration is a computer vision technique that involves transforming the desired anatomical contents of one image (moving image) into the coordinate system of another image (fixed image). In our case, the goal is to align the prostate in a DWI sequence (moving) with the prostate in a T2 sequence (fixed) for the same patient. Specifically, the goal of this research is to develop and test methods for registering the T2 and DWI b1200 sequences of Prostate MRI scans. Within this goal, we test classical approaches such as affine transformations and a deep learning-based deformable approach. We show that using affine transformations only can attain a Dice Coefficient score of 0.709 and a Hausdorff Distance of 11.087. While our research showed promising results for the classical approach, future research is needed regarding the deep learning-based deformable approach.

1. Introduction

Prostate cancer is the most commonly diagnosed cancer as well as the second leading cause of cancer death in men [8]. Magnetic Resonance Imaging (MRI) is used for the diagnosis, prognosis evaluation, and treatment planning in prostate cancer patients. Image registration enables scans taken at different times and with different modalities to be aligned so that medical professionals can better interpret patient scans. That is, medical image registration enables one to be able to view different scans possibly taken at different times all in one. The clinical significance is the ability to view disease progression overtime as well as to view anatomical contents in different modalities in a uniform manner. Additionally, registration can enable the use of scans of multiple modalities to be used in a dataset for a machine learning task which is beneficial since the machine learning models tend to perform better with more data and reduces the chances of overfitting [4].

The goal of this research is to develop and test methods for registering the T2 and DWI b1200 sequences of Prostate MRI scans. Within this goal, we aim to test classical approaches and deep learning-based approaches for this registration task.

1.1 Image Registration

In this section, we will describe the general task of image registration in more detail.

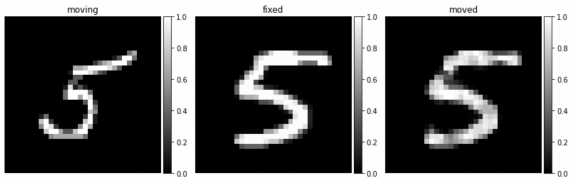

Figure 1: Image registration example showing the co-registration of two handwritten digits. The main contents of the moving image are transformed to fit the geometric space of the fixed image and its contents. This figure was adapted from [1].

Specifically, the goal of medical image registration is to transform the desired anatomical contents of one image (moving image) into the coordinate system and geometric space of another image (fixed image).

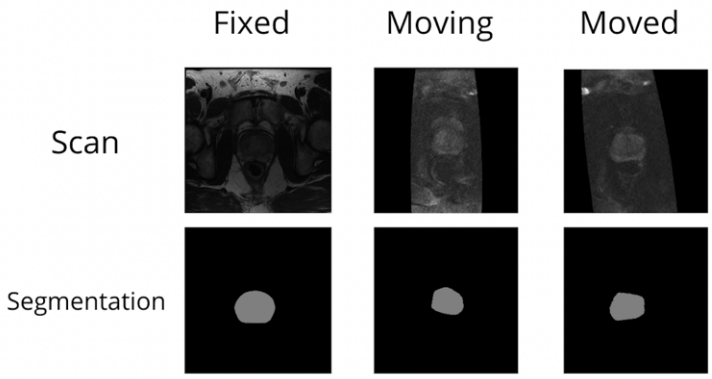

Figure 2: An example of medical image registration. In this case, the goal is to align the prostate in the DWI sequence (moving) with the prostate in the T2 sequence (fixed). Segmentations are also shown for a visualization of this process.

.

So how do we exactly warp the moving image to the fixed image? This is the process that encompasses image registration. In general, there are two main classes of registration: **classical registration** and **learning-based registration**. The gold standard has been classical registration, but in more recent years, with the advent of deep learning, learning-based registration has seen an increase in use [5]. The main

difference between the two is that learning-based registration uses machine learning to learn the transformation whereas classical registration uses mechanical transformations (linear and deformable) to warp the moving image. A similarity between the two approaches is that they both use optimization though in significantly different manners. In classical registration, a similarity metric loss function is optimized over a certain number of iterations for each image pair. Essentially, at each iteration a transformation is made. The loss function value is then calculated. This process repeats until the loss function reaches convergence and is minimized. However, in learning-based registration a loss function is optimized globally across the entire training distribution.

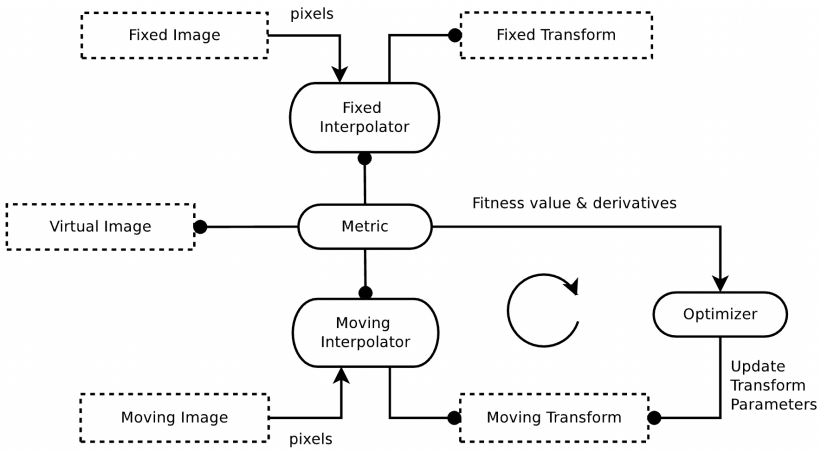

Figure 3: An overview of the classical registration pipeline. This figure is adapted from [11].

2. Methods

2.1 Materials

The entire dataset consisted of T2 volumes, DWI volumes, T2 segmentations corresponding to the T2 volumes, and DWI segmentations corresponding to the DWI volumes. The T2 volumes and segmentations measured 512 x 512 pixels whereas the DWI volumes and segmentations measured 256 x 256 pixels.

For the learning-based approach (described in section 2.3), a few preprocessing steps were taken to ensure the data would fit the parameters of the neural network. Specifically, each MRI volume needed to be of the same shape. All of the DWI b1200 volumes contained 24 frames whereas the T2 volumes contained a varied amount of frames. Thus, all of the T2 volumes were reshaped such that each volume contained exactly 24 frames. Additionally, all volumes were normalized such that all pixel values

are between 0 and 1. Normalization is a common practice in machine learning which helps the neural network learn more useful representations of the data [3].

For the classical approach (described in section 2.2), no preprocessing was performed directly on the volumes themselves. However, since the classical approach requires a fixed volume, moving volume, segmentation of the fixed volumes, and a segmentation of the moving volume, any image pairs that did not have at least one of these components were omitted from the set.

2.2 Classical Registration

For each of the experiments performed in this project, it was determined that the similarity metric, optimizer, and interpolator should be held constant (controls). This was determined by running a few minor experiments which showed that these constants consistently performed the best. This allowed us to focus on the independent variable which was the initialization transformation. There are two main types of transformations: linear and deformable. In this project, we focused solely on linear transformations. The transformations as well as similarity metric, optimizer, and interpolator used in each experiment are denoted in table 2.2.

Experiment	Transformation	Metric	Optimizer	Interpolator
CR-R1	Affine	Mattes MI	Gradient Descent	Linear
CR-R2	Versor	Mattes MI	Gradient Descent	Linear
CR-R3	Versor 3D	Mattes MI	Gradient Descent	Linear
CR-R4	Euler 3D	Mattes MI	Gradient Descent	Linear
CR-R5	Similarity 3D	Mattes MI	Gradient Descent	Linear
CR-R6	Scaled Versor 3D	Mattes MI	Gradient Descent	Linear
CR-R7	Scaled	Mattes MI	Gradient Descent	Linear
CR-R8	Scaled Skew Versor 3D	Mattes MI	Gradient Descent	Linear

2.3 Learning-Based Approach

Learning-based registration involves using machine learning and data to optimize a loss function globally across the entire training distribution as opposed to optimizing a loss function for each image pair. Learning-based registration is a young research area where there is no clear-cut method for training a machine learning for the task of image registration [5] [1]. As such, in this section, we provide an overview of the method we performed for learning-based registration. Specifically, our method used an image synthesis approach instead of a deformable approach where the moved image is artificially generated from scratch based on the learned parameters of a neural network.

1. Though, it is worth noting that the unsupervised deformable approach described in VoxelMorph [1] is arguably the most common given the applicability of it. We experimented with a similar approach described in section 2.3.1

The first step in the learning-based registration pipeline included preprocessing the data for the image registration network. In this step, we concatenated the fixed and moving image samples for each pair. The concatenated tensor was subsequently flattened into one dimension so that the neural network can take it in as input.

The next step was to construct the model architecture used in the experiment. The model used was the standard Multilayer Perceptron. The base Multilayer Perceptron is constructed in a manner in which the input layer contains the amount of neurons that is the length of the flattened concatenated tensor made in the previous step (i.e. the length of the image pair flattened). The output layer contains the length of the fixed image flattened. The resulting neuron values when the model is evaluated will be the pixels for the moved image. Figure 2.3 shows a visualization of this approach.

Note that in other learning-based image registration approaches [1], it is common to have the neural network output a deformation field first instead of the moved image. The deformation field then tells how each pixel should be moved (it is a vector at the pixel level so both the displacement and the direction is detailed) in the moving image. A spatial transformation layer [6] is applied during this "warping" process. We forego this idea here and instead try to learn the moved image directly which means that the output of the neural network is the moved image itself.

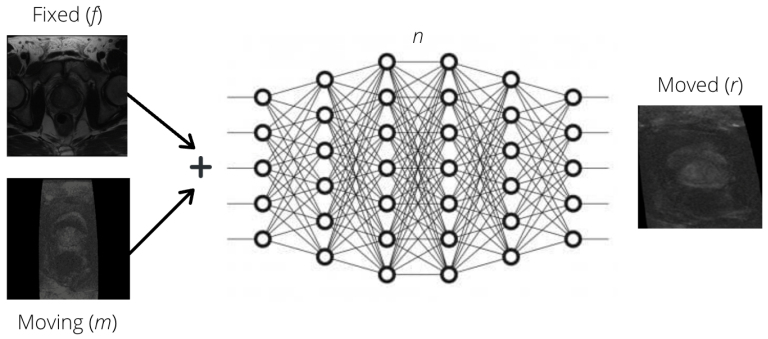

Figure 4: Visualization of the learning-based image registration model. For each image pair, the fixed sample (f) and the moving sample (m) are concatenated together forming $f + m$. This result is then evaluated by the network n, $n(f + m)$, to obtain the moved output (r).

Importantly, during training, a regression-based loss that measures the similarity between the fixed image and the output is used. Which loss function to use for registration is a research topic of its own and is not well standardized like other machine learning tasks such as classification [5]. As such, we tested several common regression-based loss functions for each model architecture. It is common to test how well an image registration method performs via the Dice coefficient calculation. Thus, this was one of the several loss functions used. The other ones used included Cosine similarity, Mean Squared Error, and Mean Absolute Error.

To register a new pair of images, the trained network evaluates the new image pair (as shown in figure 2.3) by first concatenating then subsequently flattening the image pair. This new flattened, concatenated vector $(f + m)$ is then evaluated by the network n. The following equation describes how to get the moved output r:

$$r = n(f + m)$$

Importantly, evaluating the network on a new image pair takes significantly less time and computational resources to co-register the images than with the classical approach which requires an optimization process from scratch for each new image pair.

2.3.1 Deformable Learning-Based Approach

We also tried to implement a deformable image registration network that utilizes the spatial transformation layer [6]. This network was trained in an unsupervised manner. The spatial transformation layer consists of using the output of the localization neural network (n) which regresses a $2x3$ affine transformation matrix to construct a deformation field which is used to warp the moving image (m) producing the moved image (r). The overall pipeline of the network is shown in figure 2.3.1. Despite the addition of the spatial transformation layer, the network was unable to learn anything useful as most outputs were random pixel values or simply blank. As such, the outputs were not evaluated in the results. It is also worth noting that this approach did not use segmentation masks and therefore is not needed when registering a new pair of MRI volumes.

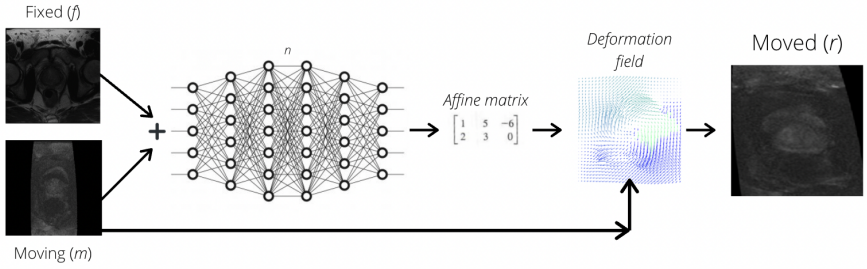

Figure 5: Visualization of the deformable learning-based image registration model that utilizes the spatial transformation layer.

3. Results

The metric calculations (Dice coefficient and Hausdorff distance) for each of the classical registration experiments described in section 2.2 are shown in table 3. Each experiment was denoted by an abbreviation and the corresponds to the abbreviations

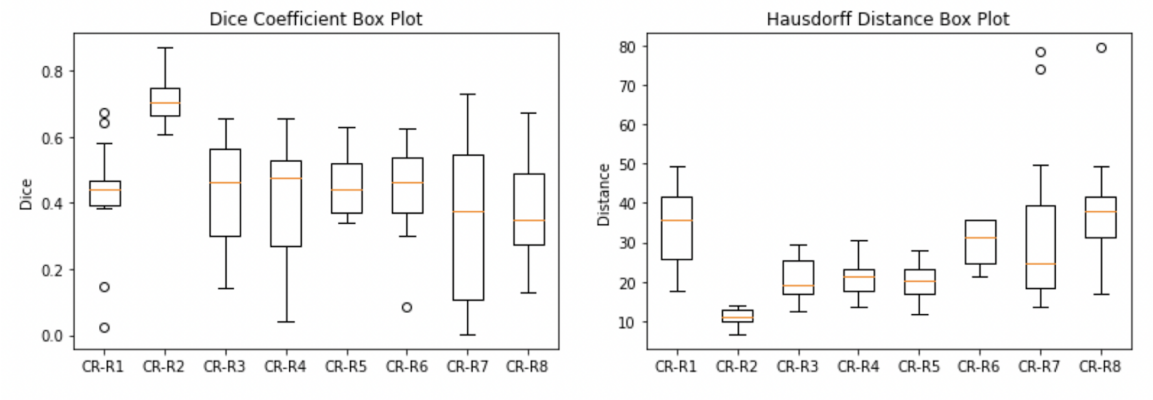

Figure 6: Box plots of Dice Coefficients and Hausdorff Distances for each experiment.

listed in the table. These metrics were calculated between the segmentation of the moved volume and the segmentation of the fixed volume.

Experiment	Dice Coefficient	Hausdorff Distance
CR-R1	0.424	33.437
CR-R2	**0.709**	**11.087**
CR-R3	0.430	20.225
CR-R4	0.414	20.831
CR-R5	0.457	20.051
CR-R6	0.427	29.899
CR-R7	0.354	34.053
CR-R8	0.378	38.864

Table 1: Average scores for the metric calculations of each experiment. Each method described in the respective experiment was tested on the same subset of the data. From here, the average for each metric is calculated.

The metric calculations were also plotted on a Box plot for both the Dice coefficient scores and Hausdorff distances. These plots are shown in figure 3. An example output produced from a given image pair is shown in figure 3.

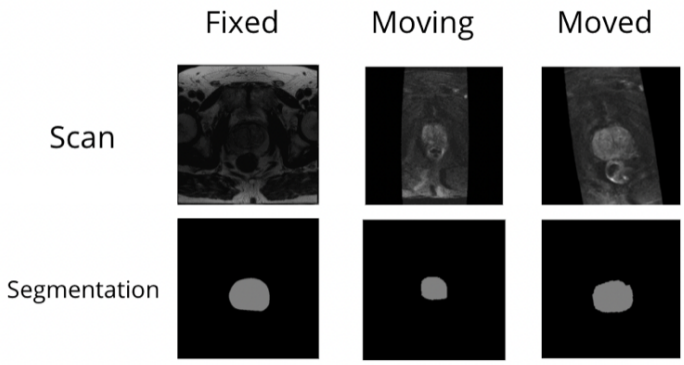

Figure 7: An example output produced via the classical approach. This particular example was produced in experiment **CR-R2**. The figure shows the fixed volume, moving volume, and moved volume as well as a corresponding segmentation of the Prostate for each volume.

From these calculations, it is clear that experiment **CR-R2**, using the **Versor** transformation performed the best for this task.

The learning-based approach described in section 2.3 suffered from changes in intensity. Specifically, the intensity of the pixel values was skewed compared to the original fixed scan. This is fairly visual and several results from the learning-based approach are shown in figure 3. The most likely reason that this occurs is that the neural network must regress the pixel values from scratch. Additionally, the neural network does not utilize a Spatial Transformation Layer [6] and foregoes the warping process using in deformable image registration.

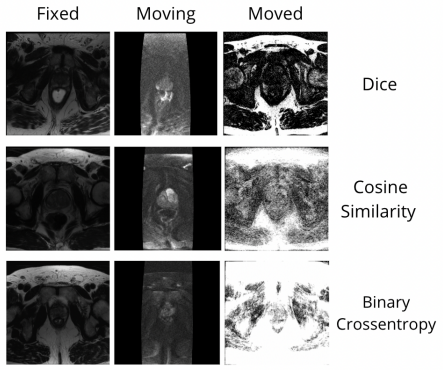

Figure 8: A select number of resultant scans from the learning-based experiments. The differences between the three image groups are that the moving and fixed scan for each image row is different (for visual purposes) and that a different loss function was used. This visual illuminates the changes in the moved image pixel intensities when the loss functions is changed.

For the deformable learning-based approach that uses the spatial transformation layer (described in section 2.3.1), the network was unable to produce useful outputs as most outputs were random pixel values, highly distorted, or simply blank. We believe this is due to a flaw in the implementation of the network in the code.

Since both learning-based approaches suffered from inherent flaws and produced insufficient outputs, we did not calculate Dice coefficient scores or Hausdorff distances for these approaches.

4. Discussion

Future research and methods would include improving the learning-based method for this task. Specifically, constructing the neural network such that it outputs a deformation field (which was foregone in the method described in section 2.3) to warp the image would help preserve the intensities of the pixels in the moving image. Moreover, other deep learning techniques such as Generative Adverserial Networks [4], Neural Style Transfer [2], and Neural Radiance Fields [7] can be explored. In fact, CycleGAN, a type of Generative Adverserial Network, was able to preserve pixel intensity relationships between two multimodal images for a different task (image translation) unlike our method described in section 2.3 [9, 10]. Other classical approaches can be explored as well including classical deformable image registration. However, this would require a large amount of computational resources. Overall, we show here that using a classical approach involving only affine transformations is able to produce a statistically relevant result.

Acknowledgements

I would like to thank The Canary Center at Stanford Medicine for inviting me to participate in the CREST program. I would like to acknowledge my faculty mentor Dr. Geoffrey Sonn as well as my lab mentor Dr. Richard Fan. In addition, I would like to acknowledge the following members of the lab that helped me conduct my research: Mirabela Rusu, Ph.D., Ahmed Alsinian, Ph.D., Simon John Christoph Soerenson, and Wei Shao, PhD.

References

[1] Guha Balakrishnan et al. "Voxelmorph: a learning framework for deformable medical image registration". In: *IEEE transactions on medical imaging* 38.8 (2019), pp. 1788–1800.

[2] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. "A neural algorithm of artistic style". In: *arXiv preprint arXiv:1508.06576* (2015).

[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[4] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems* 27 (2014).

[5] Grant Haskins, Uwe Kruger, and Pingkun Yan. "Deep learning in medical image registration: a survey". In: *Machine Vision and Applications* 31.1 (2020), pp. 1–18.

[6] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. "Spatial transformer networks". In: *Advances in neural information processing systems* 28 (2015), pp. 2017–2025.

[7] Ben Mildenhall et al. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *ECCV*. 2020.

[8] James Mohler et al. "Prostate cancer". In: *Journal of the National Comprehensive Cancer Network* 8.2 (2010), pp. 162–200.

[9] Snehashis Roy et al. "MR to CT registration of brains using image synthesis". In: *Medical Imaging 2014: Image Processing*. Vol. 9034. International Society for Optics and Photonics. 2014, p. 903419.

[10] Christine Tanner et al. "Generative adversarial networks for mr-ct deformable image registration". In: *arXiv preprint arXiv:1807.07349* (2018).

[11] Ziv Yaniv et al. "SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research". In: *Journal of digital imaging* 31.3 (2018), pp. 290–303.