CT-MRI CROSS-DOMAIN REGISTRATION FOR BETTER BRAIN SEGMENTATION

Casper Lisager Frandsen
University of Copenhagen
clf@di.ku.dk

Mads Nielsen, Mostafa Mehdipour Ghazi
University of Copenhagen, Pioneer Centre for AI
{madsn,ghazi}@di.ku.dk

August 23, 2024

Abstract. This paper investigates an extension of the SynthMorph network (Hoffmann et al. 2022)[1] with additional capabilities, namely cross-domain registration. The network is originally designed for multi-modal MRI registration, but in this paper we train it to also handle CT to MRI registration. We train the network to do this using only the affine transformation matrices generated by the network. This CT-MRI registration capability could help combine the rapid imaging capabilities of CT scans, and the soft tissue details of MRI scans.

While our findings suggest that the SynthMorph shows promise for cross-modal registration, we believe it will likely need further refinement to be a clinically viable tool. We show that currently the performance does not live up to the standards required in practice yet.

Contents

1 Introduction

Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are widely utilised imaging techniques, used in medical diagnostics. CT scans utilise X-rays to create detailed images, and are particularly effective at visualising bones and internal injuries. These scans provide images of bony structures and are often used in cases including trauma due to the speed at which they can be taken.

MRI scans, on the other hand, use powerful superconducting electromagnets which can produce detailed images of soft tissues such as the brain, which we will be looking at in this paper. MRI is particularly effective for brain scans because it can differentiate between these soft tissues based on their varying water content and relaxation properties. In recent years, segmentation of MRI scans has become very accurate, and can be used to diagnose many health issues. However, MRI faces a few problems; namely that it is time consuming and expensive compared to other types of scan. As such, it is desirable for cheaper and quicker scans such as those made using CT to be automatically registered to an MRI, to leverage the strengths of both imaging methods. The purpose of the project is exactly that. If successful, this would allow such scans to take place in situations where time is of the essence and help medical personnel in taking more informed decisions.

There are many methods for registering images, with a commonly used one being the Elastix framework[2], which uses iterative methods. These methods are quite slow however, and deep learning-based methods such as Synthmorph have been proposed which drastically reduces registration time down to less than a minute on CPU and just a few seconds on GPU[3]. Generally speaking, two different main types of registration exist and are currently used in the field, depending on the use cases, rigid and non-rigid. Rigid registration methods transform images based on constrained affine transformation matrix. The matrix is constrained such that it can only perform a subset of affine transformations, rotation and translation. This type of registration, despite being the "weakest" in terms of how much they can transform is nevertheless sufficient in many medical contexts.

Non-rigid registration includes both deformable and affine transformations. With deformable registration methods, each voxel can be individually transformed, effectively allowing transformations of inputs into completely unrecognisable outputs. Affine registration methods are essentially an extension of the rigid methods. While the affine transformation matrix used in rigid registration is constrained, it is when performing affine registration. This opens up the possibility to both scale, shear and reflect inputs, in addition to the aforementioned rigid capabilities.

Synthmorph consists of several different registration models, including both deformable, rigid and affine models. In this paper, we will be looking at only the rigid and affine versions of the models for registration of intra-subject pairs of CT and MRI scans.

All code used in this paper can be found on Github[1], though it has not as of the time of writing been made into a state usable for others.

2 Methods

2.1 Data

For this project, the SynthRAD dataset[4] will serve as the primary resource. The SynthRAD dataset is a substantial collection of brain and pelvis CT and T1 MRI scans. Only the brain scans are used in this project. The data is also been completely anonymised, with all patient data removed.

The SynthRAD dataset is useful to this work, since it has been thoroughly preprocessed and registered using conventional (rigid) iterative methods. The methods by which this has been done can be read in the corresponding literature. This thorough preprocessing of the data is crucial, as it allows us in this paper to use the registered images as a ground truth for "perfect" registration later in this paper. As such, we will use the SynthRAD dataset for constructing the training, validation, and test sets used throughout this thesis.

It is noted in the original paper[4] that they have concentrated outliers from the norm in this training dataset, but have not provided their test set for public use, due to the fact that it was used in the SynthRAD2023 Grand Challenge. As such, these outliers may mean that the results shown in this paper are not entirely representative. However, this may have proven beneficial for training purposes, as the model should be less susceptible to these types of outliers.

2.2 Models

2.2.1 Registration Network

The basis of the model fine-tuned in this paper is the SynthMorph model developed by Hoffmann et al. (2023)[1]. There are several variants of SynthMorph, including one that generates a dense displacement field. In this paper however, we focus on the variants that produce rigid and affine transformations. Since we are concerned here with pairs of scans from the same patient, any non-rigid deformation is very likely to be incorrect.

[1] https://github.com/Cavtheman/ct_mri_reg/

The original Synthmorph is trained using purely synthetic images, generated from anatomical label maps. The synthetic images are varied significantly during training to cover a wide range of real-world scenarios, including different image contrasts and resolutions. This approach enables the base network to effectively handle various modalities, such as T1 and T2 MRI scans. The goal of our work is to extend this capability beyond just multi-modal compatibility to achieve robust cross-domain performance as well.

The actual network architecture is fully convolutional, utilising eight 3D convolutional layers, which output spatial feature maps separately for the moving and fixed images. The barycentres of these feature maps are then calculated, and the points are used to analytically align the two images with weighted-least-squares.

2.2.2 Segmentation Network

The SynthSeg network, by Billot et al. (2023)[5][6] is also used in this paper. This network is used out of the box however, and is not fine-tuned. Instead, it is utilised to provide a less abstract, and more clinically relevant and interpretable metric for final evaluation of the fine-tuned registration models. The manner in which this is done is detailed in the Evaluation Criteria subsection

The SynthSeg model is a 3D UNet[7] type network, trained on synthetic data like the Synthmorph network. The synthetic data used to train this network is generated based on a corresponding synthetic label map, that is itself generated by a transformation of a label map from the training data. This approach allows the model to generalise well over different image types, including T1 scans as utilised in this paper.

2.3 Artificial Data Generation

In the context of training a network for image registration, it is crucial to have access to relevant pairs of volumes that require registration. However, acquiring unregistered image pairs, particularly those with known and accurate registration transformations, presents significant challenges. To address this, we augment the existing SynthRAD dataset, which consists of 180 preprocessed and registered image pairs, by applying a limited set of transformations. While it may be safe to assume that the fixed scan that will be registered to is centred, we saw negligible difference in the baseline Synthmorph performance on non-centered data. Therefore, we have elected to transform both the fixed and moving volumes, providing better generalisation, and preventing the trained models from degenerating into moving the subject to the centre, regardless of input. For each pair of CT and T1 MRI scans, we output two different pairs of volumes, one pair marked and utilised as the fixed volumes, and one as the moving. For each CT-MRI pair, they are transformed identically, to keep them registered to each other. The fixed and moving pairs are transformed with identical parameters given to the random functions, but the fixed and moving transformations are chosen independently.

The limited set of transformations is used to preserve the clinical relevance of the original scans while introducing just enough variability to train the registration model effectively. This balance is vital to avoid creating synthetic data that might not be representative of actual medical images, which could lead to a model that performs well on our artificial data but fails in clinical practice, or learns irrelevant information.

Both the CT and MRI scans are normalised between 0 and 1 as part of preprocessing. While normalisation of data is common practice, it can be potentially problematic when working with CT scans. Unlike MRI scans, where intensity values are typically relative and vary based on tissue type and scan parameters, CT scans have intensity values corresponding to real physical phenomena. These values can be clinically meaningful and correspond to specific tissue densities, such as bone, air, and soft tissue. However, while the specific values are lost when normalised in this way, the relative values are not.

However, preliminary baseline experiments with the Synthmorph models showed that it performed markedly worse when normalisation was not performed. Given that this normalisation is performed

in the original Synthmorph paper[3], we elected not to pursue this any further. As a result, all results not in the appendix are with normalised CT and T1 MRI scans, unless specifically noted otherwise. These baseline results can be seen in Tables 13-16 in the appendix.

The remainder of the augmentation process involves applying standard rigid and affine transformations, which are varied across four generated datasets to evaluate the performance of different methods and models. The primary objective is to determine whether smaller, more clinically realistic transformations or larger, less realistic ones yield better results for training. Three distinct parameters are varied during training, each with separate components for each spatial dimension. However, not all aspects of the affine transformation space are utilised. Specifically, neither reflection nor scaling is applied.

Reflection is avoided due to the training regimen of the original Synthmorph model, which is trained to register volumes exclusively in the left-inferior-anterior (LIA) orientation. This model limitation was not well-documented, leading to significant challenges and time spent troubleshooting why the model initially failed when applied to the SynthRAD data, which is provided in a left-posterior-superior (LPS) orientation by default. Applying reflection would further exacerbate this issue, rendering the model non-functional. Additionally, given that the brain is not symmetrical in function, reflection could introduce additional inaccuracies when evaluating performance with the SynthSeg model[5][6].

Scaling is also omitted, as the Synthmorph model expects inputs to be standardised to 1mm^3 voxels. Any further scaling, in either direction, would likely degrade the model's performance without providing any practical benefit, as the scans can already be easily scaled to this voxel size. This decision ensures that the augmented data remains both realistic and compatible with the model's expectations.

Thus we are left with rotation, translation, and shear as the primary transformations employed in the augmentation process, though shear is not always applied. Each of these transformations includes components along all three spatial axes. Given that the scans are 3-dimensional, the affine matrices differ from the more commonly encountered 2-dimensional variants. Generalised 3D rotational matrices for each axis are shown below, each taking an angle θ in radians:

$$R_x(\theta_x) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x & 0 \\ 0 & \sin\theta_x & \cos\theta_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_y(\theta_y) = \begin{pmatrix} \cos\theta_y & 0 & \sin\theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_z(\theta_z) = \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 & 0 \\ \sin\theta_z & \cos\theta_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

For simplicity of implementation, these matrices are implemented separately. Since these are linear transformations, calculating the dot product of these matrices to achieve a singular rotation matrix is equivalent to performing each rotation separately. As such, we combine them, rotating around the x-axis first, then the y- and z-axes. For the data generation, values of θ are picked uniformly at random in the ranges ± 0.2 and ± 0.4 radians. Later in this paper, $\theta \pm 0.2$ and $\theta \pm 0.4$ will be used as shorthand to denote these distributions. 0.2 radians corresponds roughly to 11 and 22 degrees respectively, chosen such that ± 0.2 represents a roughly realistic sample, where patient movement could plausibly be the cause of the extra rotation. The other, larger rotations are chosen to determine

whether more variation will help the model generalise better. Since the point of this project is to demonstrate capabilities with CT scans, which may be used for more time-sensitive diagnoses operations, allowing more noise in the form of extra rotations should allow for more robust models. The resulting rotation matrix can be seen below. Just for ease of display here, the bottom row and rightmost column the matrix is removed, as it has no effect on rotation, and can be appended with no issue.

$$
\begin{aligned}
R &= R_z(\theta_z)R_y(\theta_y)R_x(\theta_x) \\
&= \begin{pmatrix} \cos\theta_y \cos\theta_z & \sin\theta_x \sin\theta_y \cos\theta_z - \cos\theta_x \sin\theta_z & \cos\theta_x \sin\theta_y \cos\theta_z + \sin\theta_x \sin\theta_z \\ \cos\theta_y \sin\theta_z & \sin\theta_x \sin\theta_y \sin\theta_z + \cos\theta_x \cos\theta_z & \cos\theta_x \sin\theta_y \sin\theta_z - \sin\theta_x \cos\theta_z \\ -\sin\theta_y & \sin\theta_x \cos\theta_z & \cos\theta_x \cos\theta_y \end{pmatrix}
\end{aligned} \tag{1}
$$

Simple translation is easier, simply corresponding to values in the rightmost column. Similar to rotation, values of t are picked uniformly at random in the ranges ± 20 and ± 40, corresponding to 20 and 40 voxels respectively. Due to the 1mm^3 voxels, this corresponds to and equal amount of millimetres. $T \pm 20$ and $T \pm 40$ will be used as shorthand for this, like with rotation. Like with the rotational values, these values have been picked to represent a roughly realistic sample, and a more extreme example.

$$
T = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}
$$

Affine shear matrices are similar to the rotational matrices in their makeup. Each non-diagonal value in the rotational part of the matrix corresponds to a shear over a specific axis.However, unlike rotational transformations, shear is not a phenomenon typically encountered in properly calibrated medical scanners. As a result, the values chosen for shear in this study are intentionally smaller. And indeed, for half the datasets generated in this paper, no shear is applied. For the other half, shear is applied in the same fashion as rotation and translation, chosen uniformly at random in the interval ± 0.1. The reasoning behind applying shear to this data, is that we are training both a rigid and affine transformation model to register, and the extra parameters the affine model has to work with may allow for it to more easily converge on useful solutions, even if it does not need to make non-rigid solutions. Nonetheless, both models, rigid and affine, are trained on datasets with and without shear.

$$
S = \begin{pmatrix} 1 & s_{xy} & s_{xz} & 0 \\ s_{yx} & 1 & s_{yz} & 0 \\ s_{zx} & s_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
$$

One problem arises when applying these transformations however, and that is the fact that the rotation is around coordinate $(0, 0, 0)$. In the ideal case, that would be the middle of the scan, but that is not the case here, where $(0, 0, 0)$ is a corner of the volume. To correct for this fact, changes need to be made to the translational matrix, to effectively change the centre of rotation. Note that R in this refers specifically to the form given in Equation 1, without a fourth column or row. $|V_{x,y,z}|$ refers to the size of the scan volume, in each axis, in voxels.

$$\begin{pmatrix} t'_x \\ t'_y \\ t'_z \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} - \left(\begin{pmatrix} 0.5|V_x| \\ 0.5|V_y| \\ 0.5|V_z| \end{pmatrix} - R \begin{pmatrix} 0.5|V_x| \\ 0.5|V_y| \\ 0.5|V_z| \end{pmatrix} \right)$$

$$T' = \begin{pmatrix} 1 & 0 & 0 & t'_x \\ 0 & 1 & 0 & t'_y \\ 0 & 0 & 1 & t'_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The final affine matrix A used to transform the volumes is then calculated by taking the dot product of all these, where R_{full} denotes R with an extra column and row, to be a valid affine matrix:

$$A = S\ R_{full}\ T'$$

Dataset Parameters	Training Set	Validation Set	Test Set
$\theta \pm 0.4$, $T \pm 40$, No Shear	720	180	180
$\theta \pm 0.4$, $T \pm 40$, Shear± 0.1	720	180	180
$\theta \pm 0.2$, $T \pm 20$, No Shear	720	180	180
$\theta \pm 0.2$, $T \pm 20$, Shear± 0.1	720	180	180

Table 1: Size of various Training, Validation, and Test Sets

With all of these transformations, we have generated a training set containing 720 pairs of scans, while validation and test sets each consist of just 180 pairs for each set of generation parameters. With this distribution of data, each training set contains four versions of each original image, each transformed differently, while the test and validation data contain exactly one transformed original.

2.4 Evaluation Criteria

The training process for the Synthmorph model in this paper is not performed in the same was as it was originally trained. Instead of working with completely synthetic images and computing the Dice loss over corresponding label maps, the training performed in this paper uses real, but affinely transformed images as data. This lets us ignore any label maps during training, and use a much simpler loss function. When generating transformed images, both the affine matrix (A) and its inverse (B) used to generate it are saved for later use. Say that a volume in the training set is transformed from its original using an affine matrix A, and its inverse matrix B is generated alongside it. Then the following statements will hold:

$$AB = I$$

$$AB - I = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\|AB - I\| = 0$$

With this setup, given the fixed volume, and the volume transformed by affine matrix A, Synthmorph will then output an affine matrix $\hat{B}$. This means we can then use the $\|A\hat{B} - I\|$ as a measure of how close $\hat{B}$ is to the inverse of A. Importantly, this also means that no output volume needs to be generated, greatly speeding up training time. This approach simplifies training, by focusing on the affine transformations rather than relying on labeled data, while still ensuring that the transformations can be accurately reversed.

An additional similar metric is also used in this paper to evaluate performance, but is not used during training. It is based off of similar principles, leading to the loss function $\|B - \hat{B}\|$.

$$B - B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$\|B - B\| = 0$$

Both of these metrics are quite simple, and make for easy fine-tuning of the Synthmorph models. However, while they model differences and errors in transformations well, they do not account for anatomical details. Therefore, we have also included a variant of the Mutual Information (MI) metric where it is normalised between 0 and 1, which has been demonstrated to be a good metric for registration[8], taking the actual transformed outputs of Synthmorph and comparing to each other.

Mutual Information is a measure based in the field of information theory, which can quantify the amount of information able to be obtained from one random variable through another. In an image registration context, this is done by evaluating how much knowing the intensity distribution of one image reduces uncertainty about the intensity distribution of another. When working with continuous intensity values, these are put in several intensity bins and treated as equivalent. For this, the entropy function is important:

$$H(A) = -\sum_i p(A_i) \log p(A_i)$$

Where $p(A_i)$ denotes the probability of a pixel having intensity A_i, summing over all intensity bins. With these bins, the MI is then calculated using the following formula:

$$\text{MI}(A, B) = H(A) + H(B) - H(A, B)$$

However, this measure can grow potentially infinitely, so we want to normalise it for easier training. This is done as follows:

$$\text{NMI}(A, B) = \frac{2 \cdot \text{MI}(A, B)}{H(A) + H(B)}$$

This Normalised Mutual Information (NMI) is thus one of the metrics that will be used in the rest of this paper to evaluate the performance of the registration. For the purposes of this paper, the NMI is always calculated between the moved and fixed MRI scans. This is also true when evaluating CT-MRI registration, as there is always an MRI counterpart to the transformed CT scan that has been transformed using the same parameters.

The Dice metric is utilised in this paper in much the same manner as the original Synthmorph paper, but is only used for evaluation purposes. However, since the SynthRAD data has not been segmented, we use the SynthSeg model to provide a segmentation of the fixed image, which is then used as the ground truth. When evaluating the registration performance, SynthSeg is also used to segment the moved image. These two segmentations are then used to calculate the Dice score, providing a measure of how well the registration has preserved anatomical structures.

It should be noted that there are a lot of classes in the SynthSeg output, and as such the Dice score is calculated for each one separately. Doing it this way means that we can read any voxels with the given class a True value, and any other as False, repeating over all classes. Given two volumes, A and B, we can then calculate the Dice score as follows:

$$\text{Dice}(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{2}$$

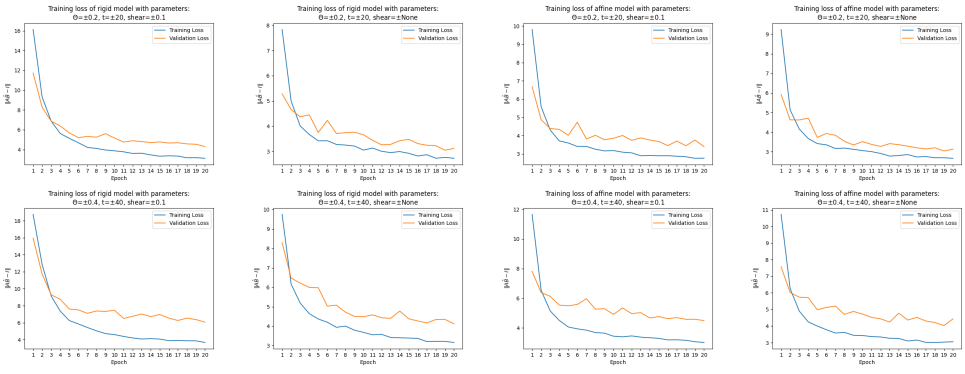

Figure 1: Training loss for fine-tuning both the rigid and affine Synthmorph models, on datasets with various generation parameters.

2.5 Training Process

Training of the of the Synthmorph network in this paper is very much standard, by the book. After the augmented datasets have been generated, the CT and MRI pairs are fed to the model, using the specified $\|A\hat{B} - I\|$ loss, using a learning rate of 1e-4. For each of the datasets, two models are trained for comparison; one rigid and one affine. All of the models are trained on the training dataset detailed in Table 1 for 20 epochs, and validated between each epoch. A checkpoint model is also saved for each epoch, and after training is finished, the model that performed best on the validation dataset is used for further testing and final evaluation. The training process of each of these models can be seen in Figure 1. The models that performed best on the validation set can be seen in Table 2

Dataset Parameters	Epoch	
	Rigid model	Affine model
$\theta \pm 0.4$, $T \pm 40$, No Shear	19	18
$\theta \pm 0.4$, $T \pm 40$, Shear± 0.1	19	19
$\theta \pm 0.2$, $T \pm 20$, No Shear	18	18
$\theta \pm 0.2$, $T \pm 20$, Shear± 0.1	19	19

Table 2: Best performing models for each dataset

3 Experiments and Results

To demonstrate the improvements made here over the capabilities of original Synthmorph model, we have first evaluated its performance on all four test datasets without any fine-tuning. However, for purposes of making the results more medically plausible, most of the models have not been evaluated on the test sets containing images that have been sheared, and metrics have not been calculated for them on the test sets. It will be clearly denoted where the shear test sets have been used.

As is to be expected, Synthmorph without any tuning performs very poorly when trying to register CT to MRI scans. This can be seen in Tables 4 and 6 where both the $\|A\hat{B} - I\|$ and $\|B - \hat{B}\|$ loss values are roughly 2-5 times larger than the corresponding loss values of MRI-MRI registration in Tables 3 and 5.

After training both the affine and rigids versions of the Synthmorph model on our augmented datasets, we find that the network starts outperforming the baseline model. For MRI-MRI registration this can easily be explained via slight overfitting to the dataset at hand. This can be seen in Tables 7 and 9.

For the CT-MRI results we also see significant improvements thanks to the training. The $\|A\hat{B} - I\|$ and $\|B - \hat{B}\|$ loss values can be seen in Tables 8 and 10 to have improved to a point where the CT-MRI registrations perform roughly equivalently to the MRI-MRI registration.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	7.4554	7.4554
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	7.5588	7.5306
$\theta \pm 0.2$, $T \pm 20$, No Shear	3.8292	3.8292
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	4.4211	4.4242

Table 3: MRI-MRI registration results with different transformations using the untuned affine SynthMorph model.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	44.1560	44.1560
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	44.5351	44.1592
$\theta \pm 0.2$, $T \pm 20$, No Shear	42.0576	42.0576
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	43.3569	42.9668

Table 4: CT-MRI registration results with different transformations using the untuned affine SynthMorph model.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	8.8768	8.8768
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	19.1687	19.1985
$\theta \pm 0.2$, $T \pm 20$, No Shear	4.5827	4.5827
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	16.8906	16.8533

Table 5: MRI-MRI registration results with different transformations using the untuned rigid SynthMorph model.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	24.9288	24.9288
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	31.9698	32.7524
$\theta \pm 0.2$, $T \pm 20$, No Shear	23.8434	23.8434
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	29.9778	30.7856

Table 6: CT-MRI registration results with different transformations using the untuned rigid SynthMorph model.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	4.3242	4.3242
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	5.5293	9.5700
$\theta \pm 0.2$, $T \pm 20$, No Shear	3.2817	3.2817
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	4.5198	6.0682

Table 7: MRI-MRI registration results with affine model tuned on the datasets with the shown parameters.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	3.6186	3.6186
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	4.1670	9.1483
$\theta \pm 0.2$, $T \pm 20$, No Shear	2.7113	2.7113
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	3.2596	5.2754

Table 8: CT-MRI registration results with affine model tuned on the datasets with the shown parameters.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	3.4790	3.4790
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	10.4729	12.3017
$\theta \pm 0.2$, $T \pm 20$, No Shear	3.0415	3.0415
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	8.6077	9.3741

Table 9: MRI-MRI registration results with rigid model tuned on the datasets with the shown parameters.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
$\theta \pm 0.4$, $T \pm 40$, No Shear	3.4099	3.4099
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	5.8691	9.6244
$\theta \pm 0.2$, $T \pm 20$, No Shear	2.3015	2.3015
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	4.4737	6.3426

Table 10: CT-MRI registration results with rigid model tuned on the datasets with the shown parameters.

The results shown above are for each dataset, with the model trained on that dataset. While this can give a good overview of the performance of the model, it is not possible to make any statements about their relative performance without evaluating on the same things. In order to evaluate their overall performance we have chosen to evaluate them on just the two test datasets without shearing, with the aim of evaluating the performance on data that is as plausibly realistic as possible. Given that the $\|A\hat{B} - I\|$ and $\|B - \hat{B}\|$ metrics are so similar, we will include only the former. The results of these evaluations can be seen in Tables 11 and 12. We also include the NMI metric here. In both these tables the "Model Version" describes whether the model in question is rigid or affine, and which training data was used to train it.

The NMI score shown here is, as mentioned earlier calculated using the MRI transformed alongside the moved CT scan that was used for the registration.

Examples showing the registration process can be seen in Figure 2. This is included to show that despite the NMI scores being in the range of roughly $0.20 - 0.22$, the resulting registration is visually very close, albeit with small differences. These differences can be visualised better in Figure 3, which shows the absolute voxel errors of a single slice.

$\|A\hat{B} - I\|$	Model Version				Baseline
Test Set parameters	Rigid				Rigid
	No Shear		Shear±0.1		
	$\theta \pm 0.2, T \pm 20$	$\theta \pm 0.4, T \pm 40$	$\theta \pm 0.2, T \pm 20$	$\theta \pm 0.4, T \pm 40$	
$\theta \pm 0.2, T \pm 20$, No Shear	2.7420	**2.6641**	3.5810	3.9383	4.5827
$\theta \pm 0.4, T \pm 40$, No Shear	6.8055	3.5053	8.3744	4.8060	8.8768
	Affine				Affine
$\theta \pm 0.2, T \pm 20$, No Shear	2.7112	2.8222	2.9794	2.7970	3.8292
$\theta \pm 0.4, T \pm 40$, No Shear	5.6467	**3.3947**	6.7972	3.6661	7.4554

Table 11: CT-MRI registration results showing the $\|A\hat{B} - I\|$ loss of models trained on the different datasets. Rightmost column is MRI-MRI registration with baseline model for comparison. Bolded numbers denote the model with best performance for the relevant test set. Note that each test set here is spread out over two rows.

Normalised Mutual Information	Model Version				Baseline
Test Set parameters	Rigid				Rigid
	No Shear		Shear±0.1		
	$\theta \pm 0.2, T \pm 20$	$\theta \pm 0.4, T \pm 40$	$\theta \pm 0.2, T \pm 20$	$\theta \pm 0.4, T \pm 40$	
$\theta \pm 0.2, T \pm 20$, No Shear	0.2462	0.2464	0.2434	0.2426	0.2182
$\theta \pm 0.4, T \pm 40$, No Shear	0.2430	**0.2451**	0.2349	0.2369	0.2137
	Affine				Affine
$\theta \pm 0.2, T \pm 20$, No Shear	0.2475	0.2477	0.2476	**0.2478**	0.2055
$\theta \pm 0.4, T \pm 40$, No Shear	0.2441	0.2442	0.2444	0.2444	0.2017

Table 12: CT-MRI registration results showing the MI score, for models trained on the different datasets. Rightmost column is MRI-MRI registration with baseline model for comparison. Bolded numbers denote the model with best performance for the relevant test set. Note that each test set here is spread out over two rows.

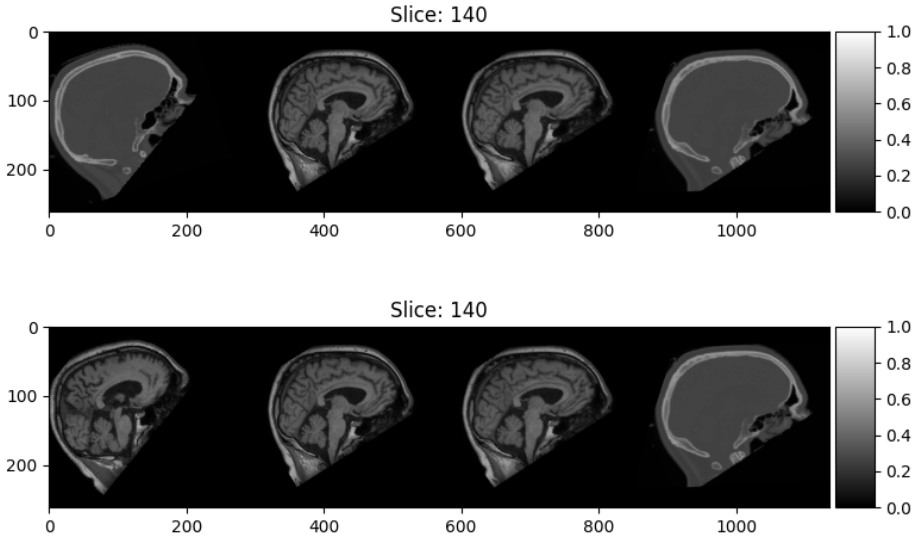

Figure 2: Examples showing a 2D slice of a registered image. From left to right: moving, fixed, moved MRI, moved CT. The moving images have been subject to the exact same transform, for ease of comparison. The moved images in the top row have been moved according to the affine transform found by registering the moving CT to the fixed MRI. The bottom row is transformed according to the moving MRI.

Our use of the SynthSeg for segmentation outputs a volume with 32 different classes. This is used as a tool to evaluate the clinical viability of the registration. An example showing a slice of segmented brains can be seen in Figure 4.

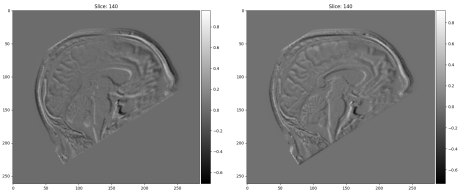

Figure 3: Example showing showing the same 2D slice as in Figure 2. Left image is CT-MRI registered, right is MRI-MRI registered. This shows the absolute differences between the fixed and moved MRI images.

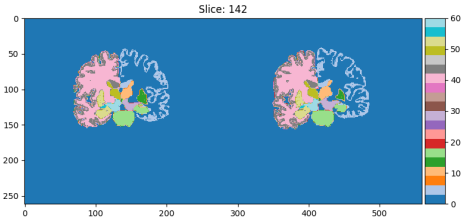

Figure 4: Example showing a 2D slice of the segmentation. Left is "ground truth" segmentation of the fixed volume. Right is segmentation of moved image.

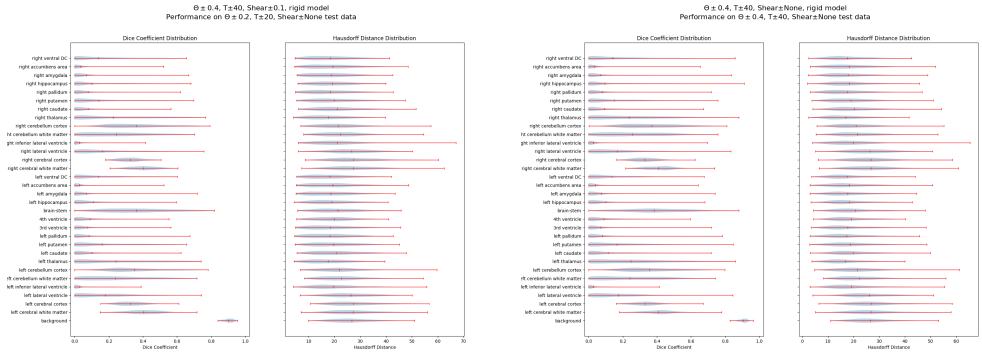

Figure 5: Dice scores and Hausdorff distances of the SynthSeg segmentations of moved images using one of the best performing registration models.

4 Discussion

In this paper, we have successfully demonstrated the capability of the Synthmorph network to register not only various MRI modalities to each other, but with extra training, also CT scans. However, we have been unable to reproduce their success in clinically relevant registration. Specifically, their paper reports a Dice score of over 0.6 on white matter, grey matter and cerebrospinal fluid across various forms of multi-modal registration.[3] In our experiments however, we were only ever able to achieve a mean Dice score of 0.3312 with their model, even after fine-tuning on the SynthRAD data.

This is likely due to the fact that they calculate this metric differently than in this paper. In their paper, they use an overlapping, one-hot encoded segmentation, meaning that any one voxel can be segmented as multiple classes. This makes any small mistakes that could otherwise potentially significantly decrease the score less punishing. The formula they use to calculate their "soft" Dice can be seen in Equation 3[3], formulated as a loss function. In this paper, we use a non-overlapping segmentation of the volumes, and calculate the Dice score using the formula shown in Equation 2.4.

$$\mathcal{L}(T_\theta, s_m, s_f) = -\frac{2}{|J|} \sum_{j \in J} \sum_{x \in \Omega} \frac{(s_m|_j \circ T_\theta(x) \times s_f|_j(x))}{(s_m|_j \circ T_\theta(x) + s_f|_j(x))} \tag{3}$$

However, the poor overall segmentation performance shown in Figure 5 would seem to indicate that the overall performance is still not all that good, with mean Hausdorff distances for nearly all classes being at or above 20.

4.1 Limitations

This paper has demonstrated that Synthmorph can achieve roughly the same performance on CT-MRI registration as it does with just MRI-MRI registration. However, this is not a guarantee that the model can generalise to other, different types of scans such as PET scans, without further training. This has not been tested, but given the models baseline performance on CT-MRI registration, it is safe to assume.

5 Future Work

5.1 Validation of model multi-modal capabilities

The dataset in this paper consists only of T1 MRI scans, and while this is one of the modalities used in the original Synthmorph paper, we have not shown that the new capabilities demonstrated here also extend to the various other modalities of MRI scans. Evaluating this would require finding a new dataset. Given the capabilities of Synthmorph already, it would not come as a surprise that this extension will generalise.

5.2 Expansion of model capabilities

Given that this paper has shown the potential for the Synthmorph network to be trained to work in more than just the original MRI, further investigation should explore whether this capability can be extended to more types of scans.

5.3 CT-MRI transformation

Having achieved successful cross-domain registration, the logical next step would be to try and convert between CT and MRI scans. This has been demonstrated previously to be possible by Lyu & Wang 2022[9]. However, methods based on diffusion have been shown to sometimes exhibit the phenomenon known as "hallucinations"[10], where the model seems to make up things that look plausible, but have no basis in reality. This is of course an issue when dealing with medical images that could be used for diagnosis and treatment of patients. If a diffusion model hallucinates when generating an image in this context, it could lead to misdiagnoses and incorrect treatments. Creating a network that doesn't hallucinate would be a big step forward in that context.

6 Conclusion

In this paper, we successfully created a method of automatically registering CT scans to T1 MRI scans, demonstrating that the multi-modal capabilities of the Synthmorph network can be extended beyond the various modalities of MRI scans to CT scans as well. Through various evaluation methods, we have also shown that this extra capability did not seem to come at the expense of poorer MRI to MRI registration performance.

However, we have also demonstrated that the Synthmorph network itself it likely not yet good enough to be of use in clinical settings in practice.

References

[1] Malte Hoffmann, Benjamin Billot, Douglas N Greve, Juan Eugenio Iglesias, Bruce Fischl, and Adrian V Dalca. Synthmorph: learning contrast-invariant registration without acquired images. *IEEE Transactions on Medical Imaging*, 41(3):543–558, 2022.

[2] Stefan Klein, Marius Staring, Keelin Murphy, Max A. Viergever, and Josien P. W. Pluim. elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010. doi:10.1109/TMI.2009.2035616.

[3] Malte Hoffmann, Andrew Hoopes, Bruce Fischl, and Adrian V. Dalca. Anatomy-specific acquisition-agnostic affine registration learned from fictitious images. In Olivier Colliot and Ivana Išgum, editors, *Medical Imaging 2023: Image Processing*, volume 12464, page 1246402. International Society for Optics and Photonics, SPIE, 2023. doi:10.1117/12.2653251. URL https://doi.org/10.1117/12.2653251.

[4] Adrian Thummerer, Erik van der Bijl, Arthur Galapon, Joost J. C. Verhoeff, Johannes A. Langendijk, Stefan Both, Cornelis (Nico) A. T. van den Berg, and Matteo Maspero. Synthrad2023 grand challenge dataset: Generating synthetic ct for radiotherapy. *Medical Physics*, 50(7):4664–4674, June 2023. ISSN 2473-4209. doi:10.1002/mp.16529. URL http://dx.doi.org/10.1002/mp.16529.

[5] Benjamin Billot, Douglas N. Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V. Dalca, and Juan Eugenio Iglesias. Synthseg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis*, 86:102789, 2023. ISSN 1361-8415. doi:10.1016/j.media.2023.102789.

[6] Benjamin Billot, You Colin, Magdamo Cheng, Sudeshna Das, and Juan Eugenio Iglesias. Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. *Proceedings of the National Academy of Sciences (PNAS)*, 120(9):1–10, 2023. doi:10.1073/pnas.2216399120.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL http://arxiv.org/abs/1505.04597.

[8] Navdeep Kanwal, Shweta Jain, and Paramjeet Kaur. *Evaluating Robustness for Intensity Based Image Registration Measures Using Mutual Information and Normalized Mutual Information*, pages 73–81. Springer International Publishing, Cham, 2020. ISBN 978-3-030-48118-6. doi:10.1007/978-3-030-48118-6_7. URL https://doi.org/10.1007/978-3-030-48118-6_7.

[9] Qing Lyu and Ge Wang. Conversion between ct and mri images using diffusion and score-matching models, 2022. URL https://arxiv.org/abs/2209.12104.

[10] Sumukh K Aithal, Pratyush Maini, Zachary C Lipton, and J Zico Kolter. Understanding hallucinations in diffusion models through mode interpolation. *arXiv e-prints*, pages arXiv–2406, 2024.

7 Appendix

7.1 All Dice scores and Hausdorff distances

7.2 Synthmorph baselines, with centered fixed image

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
No Rotation, No Translation, No Shear	0	0
$\theta \pm 0.4$, $T \pm 40$, No Shear	9.2585	9.2585
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	9.8315	12.3868
Normalised, $\theta \pm 0.4$, $T \pm 40$, No Shear	7.4767	7.4767
Normalised, $\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	7.4014	10.3401
$\theta \pm 0.2$, $T \pm 20$, No Shear	5.8611	5.8611
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	6.1733	7.3278
Normalised, $\theta \pm 0.2$, $T \pm 20$, No Shear	4.1896	4.1896
Normalised, $\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	4.3918	5.6838

Table 13: MR-MR registration results with different transformations using the baseline affine SynthMorph model.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
No Rotation, No Translation, No Shear	81.4608	81.4608
$\theta \pm 0.4$, $T \pm 40$, No Shear	66.2298	66.2298
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	90.0000	90.9563
Normalised, $\theta \pm 0.4$, $T \pm 40$, No Shear	44.1781	44.1781
Normalised, $\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	45.2697	45.9374
$\theta \pm 0.2$, $T \pm 20$, No Shear	87.5443	87.5443
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	269.8159	276.1376
Normalised, $\theta \pm 0.2$, $T \pm 20$, No Shear	42.0669	42.0669
Normalised, $\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	43.9500	43.9333

Table 14: CT-MR registration results with different transformations using the baseline affine SynthMorph model.

Generation Parameters	$\|A\hat{B} - I\|$	$\|B - \hat{B}\|$
No Rotation, No Translation, No Shear	0	0
$\theta \pm 0.4$, $T \pm 40$, No Shear	7.5015	7.5015
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	17.7726	18.3158
Normalised, $\theta \pm 0.4$, $T \pm 40$, No Shear	8.8842	8.8842
Normalised, $\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	19.0528	19.1625
$\theta \pm 0.2$, $T \pm 20$, No Shear	4.5461	4.5461
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	17.1753	17.1819
Normalised, $\theta \pm 0.2$, $T \pm 20$, No Shear	4.6545	4.6545
Normalised, $\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	17.3669	17.2396

Table 15: MR-MR registration results with different transformations using the baseline rigid SynthMorph model.

Generation Parameters	$\|A\hat{A}^{-1} - I\|$	$\|B - \hat{B}\|$
No Rotation, No Translation, No Shear	60.5308	60.5308
$\theta \pm 0.4$, $T \pm 40$, No Shear	54.8180	54.8180
$\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	61.0844	62.0519
Normalised, $\theta \pm 0.4$, $T \pm 40$, No Shear	24.9550	24.9550
Normalised, $\theta \pm 0.4$, $T \pm 40$, Shear $= \pm 0.1$	32.4703	33.4180
$\theta \pm 0.2$, $T \pm 20$, No Shear	50.7994	50.7994
$\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	56.2187	57.1268
Normalised, $\theta \pm 0.2$, $T \pm 20$, No Shear	23.7351	23.7351
Normalised, $\theta \pm 0.2$, $T \pm 20$, Shear $= \pm 0.1$	31.8721	32.9091

Table 16: CT-MR registration results with different transformations using the baseline rigid SynthMorph model.

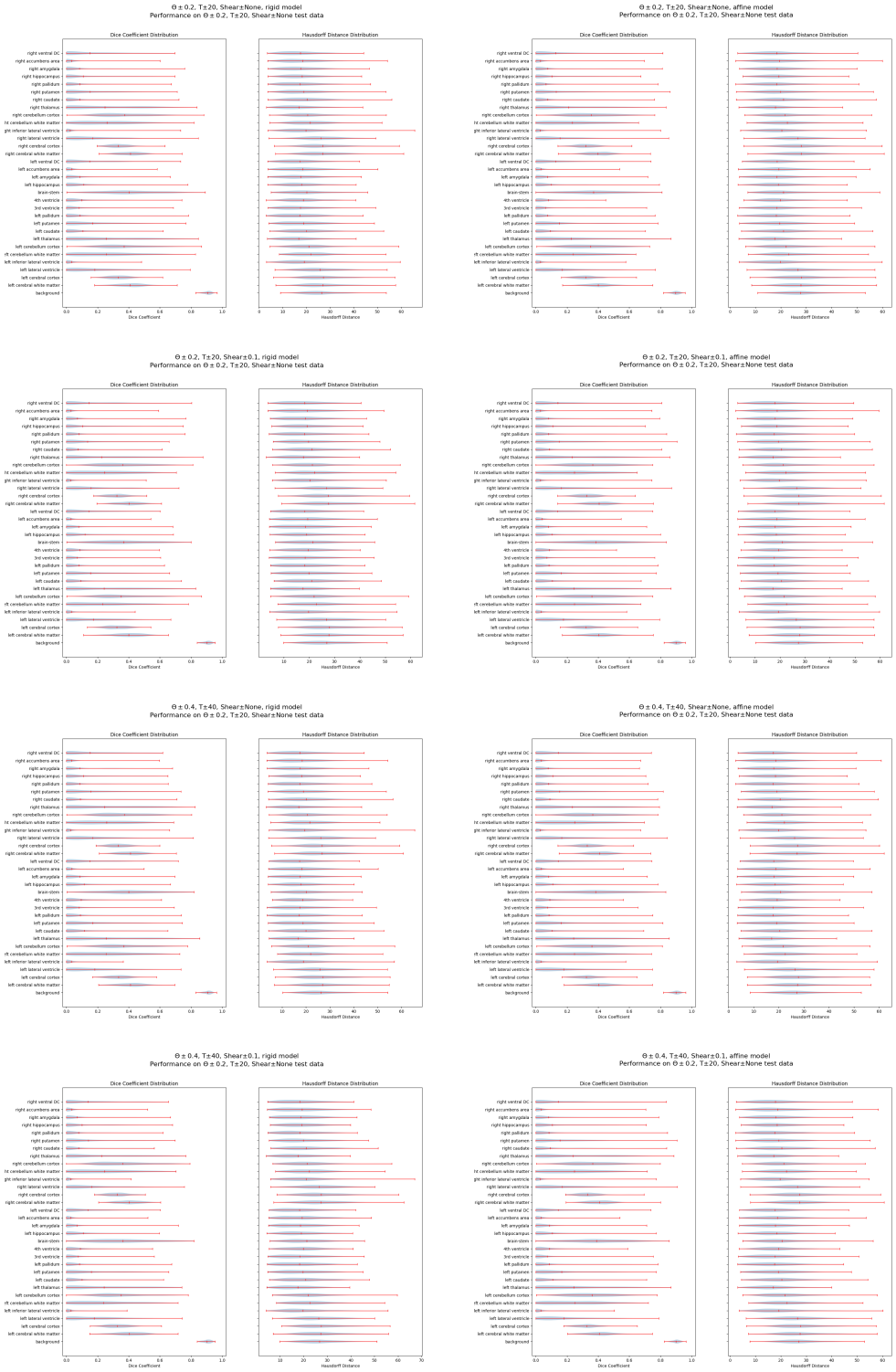

Figure 6: Dice scores and Hausdorff distances of the SynthSeg segmentations on test sets with parameters $\theta \pm 0.2$ and T$\pm$20.

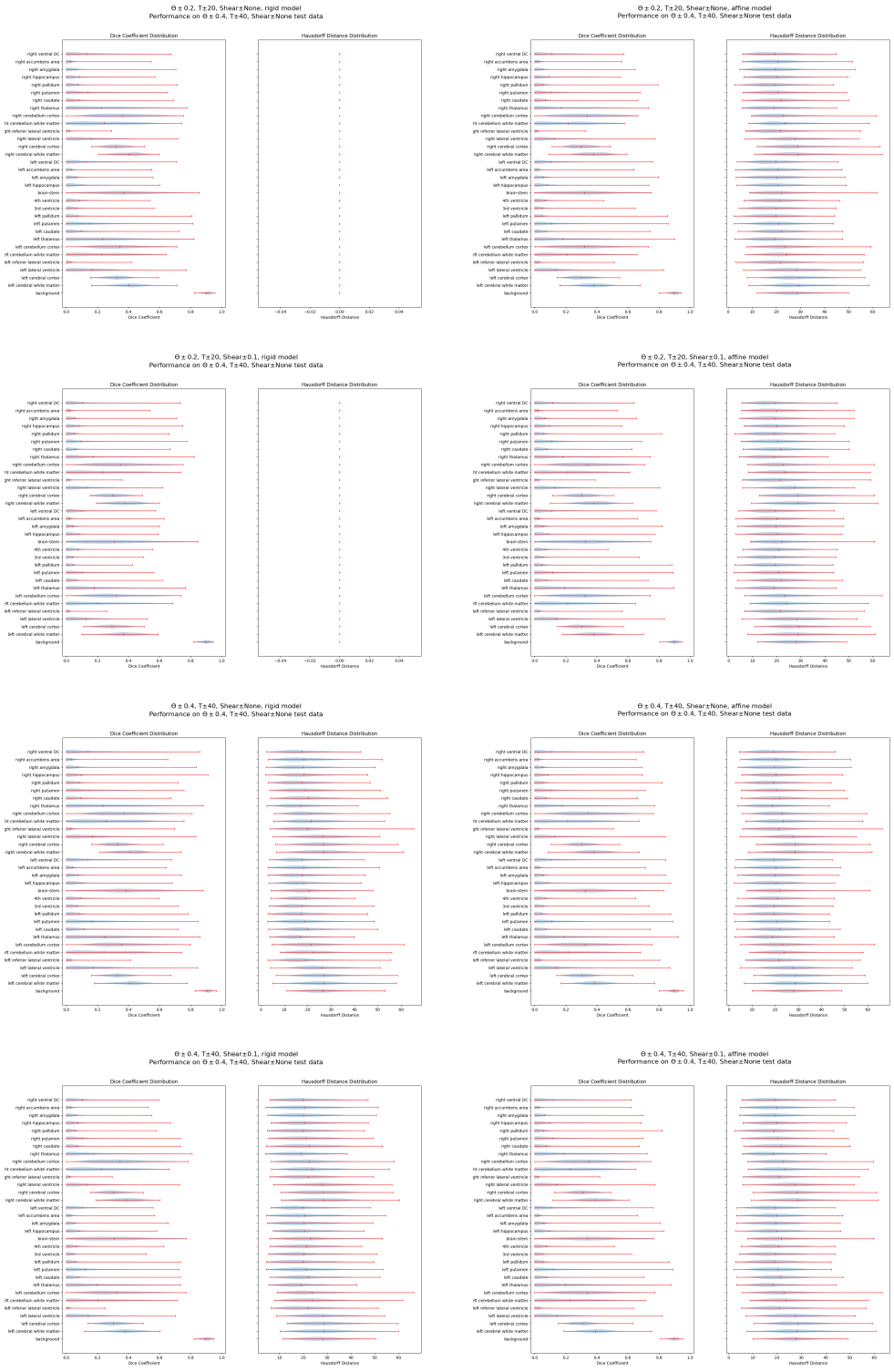

Figure 7: Dice scores and Hausdorff distances of the SynthSeg segmentations on test sets with parameters $\theta \pm 0.4$ and T$\pm$40. An error in the SynthSeg model causes some of these to output a Hausdorff distance of 0.