Cross-modal MRI registration using U-Net based models

Sreekar Chigurupati, Shalini Kothuru, Venkata Sai Abhigna Devarasetty
Indiana University Bloomington
srchig@iu.edu, skothuru@iu.edu, vdevaras@iu.edu

Abstract

U-Net based deep learning models have been used to solve various medical image registration problems. One such model, VoxelMorph has been successful at solving MRI registration. We postulate that registration-adjacent tasks and cross-modal registration are solvable with minimal changes in the base VoxelMorph architecture. Normalised mutual information as a loss mechanism for U-Net works to solve cross-modal MRI registration. To improve registration performance, the effective receptive field size can be increased by using parallel Convolutional kernels instead of a traditional U-Net. Problems that need intensity and geometric reassignment like susceptibility distortion correction are solvable using the same networks we are using to perform registration.

1 Introduction

Image registration is the process of aligning various images for comparison and analysis. Particularly, deformable registration constitutes a crucial task within the realm of medical imaging studies, representing a subject of continuous research spanning several decades. In deformable registration, the objective is to establish a dense, non-linear correspondence between two images, often exemplified by 3D magnetic resonance (MR) brain scans. This process is beneficial in the medical field and has many applications. It helps doctors to align images precisely from different modalities like MRI, and CT and enhances the diagnosis process of the patient.

Traditional methods for registration usually work by solving an optimization problem for each pair of volumes, where the goal is to align voxels with similar appearance while adhering to specific constraints governing the registration mapping. This pairwise optimization process can prove computationally intensive, resulting in impractical delays.

VoxelMorph [1] is a fast and unsupervised technique that uses a U-Net-based neural network[2]. However, let's see another leading performer in image registration which is the transformer.

VoxelMorph in its unmodified form is able to solve unimodal 3D image registration. We implement this network from scratch and validate the results on the OASIS[4] dataset. We experiment with various losses to make cross-modal registration possible on the same architecture.

Susceptibility distortion correction is a specialized process within the realm of medical image processing, primarily applied to MRI scans. In MRI, the presence of magnetic field variations near tissue-air interfaces or regions with differing magnetic susceptibility can lead to geometric distortions in the acquired images. These distortions manifest as warping and stretching of anatomical structures, which can negatively impact the accuracy of subsequent image analysis and registration tasks. To mitigate this issue, susceptibility distortion correction techniques are employed to reverse or compensate for the distortions, essentially "undistorting" the images. This process typically involves acquiring additional field maps or specialized sequences during MRI acquisition to estimate the magnetic field variations and then using this information to correct the distortions in the original images.

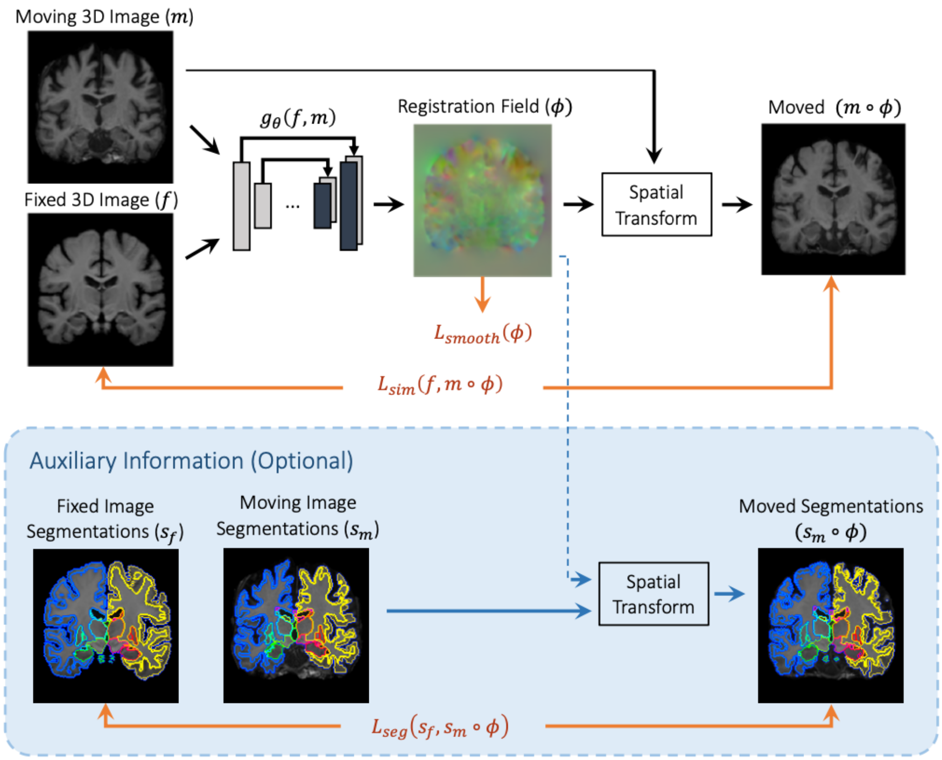

Figure 1: VoxelMorph architecture.

Since, this distortion correction is done by reassignment of geometry and intensity values, it is a good candidate for transfer learning. We train a U-Net using opposite phase-encode MR images to solve distortion correction. This is done on a pretrained VoxelMorph network to utilise the "registration" already learnt.

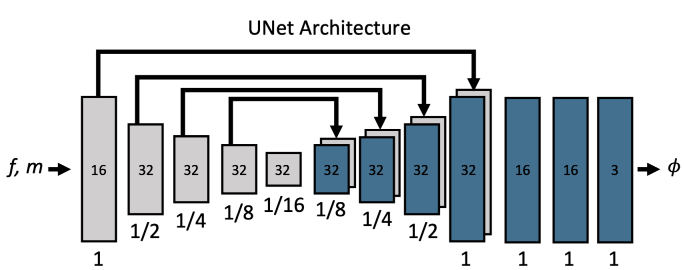

Figure 2: U-Net architecture.

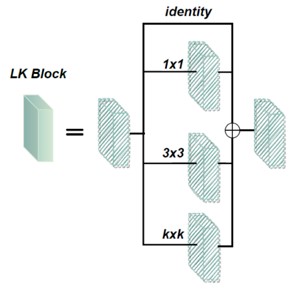

Figure 3: LK block architecture.

To improve registration performance, the effective receptive field size may be increased by adding parallel convolutional blocks negating the need for long-range dependencies typical in these models.

We also have to consider how easy they are to use in real hospitals. U-Net is generally easier to deploy and quick, which is important.

Through our work, we aim to enhance the performance of VoxelMorph for cross-modal MRI image registration and draw a comparison between different architectures. Finally, we aim to enrich the field of image registration by providing an enhanced U-Net architecture that can align images from different modalities efficiently.

2 Experiments

In our study of medical image registration, we have tried and tested different architectures to explore and evaluate various adaptions of the U-Net architecture. We tried 6 different approaches. Each experiment focuses on varying loss functions or network architecture. The details are as follows

2.1 Experiment 1: Original U-Net Architecture in VoxelMorph for Unimodal Registration (T1 to T1)

Architecture overview: In our first experiment, we implemented the original VoxelMorph architecture for the same-modality registration of T1-weighted MRI data. This model uses a U-Net-based architecture followed by a Spatial Transformer. Here is a breakdown of the architecture: Inputs: We feed two types of inputs into the model: 'moving' and 'static' MRI images. The model concatenates these images along their last dimension for processing.

Encoder: Our encoder uses several Conv3D layers with progressively increasing filter sizes (16, 32, 32, 32), followed by a LeakyReLU activation, respectively. The stride set to 2 in these layers downsamples the image, capturing hierarchical features at different levels.

Decoder: We reverse the downsampling in the decoder using UpSampling3D layers, followed by Conv3D layers and LeakyReLU activation. We added skip connections from the encoder at each stage, which helps the model preserve detailed information while upsampling. Output and Deformation Field: The final stages of the model involve Conv3D layers that update the feature maps, producing a deformation field. This field created with kernel initializers for spatial transformations will align the moving image to the static one.

Grid Generation and Resampling: We generate a regular 3D grid, adjust it according to the deformation field, and use it to resample the moving image, aligning it with the static image. This is done via the Spatial Transformer

Results: Our implementation of the original VoxelMorph architecture successfully registered T1-weighted MRI data within the same modality, having the similar outcomes documented in the reference implementation. We aligned the MRI image pairs using a cross-correlation function as a similarity measure. These results provided a good foundation for further experiments. **The average DICE score over 100 images was 0.828**

2.2 Experiment 2: Original U-Net Architecture for Cross Modality Registration (T1 to T2) with Mutual Information Loss

Architecture : In our second experiment, we retained the architecture used in Experiment 1, the original VoxelMorph model based on a U-Net framework. However, we wanted to try cross-modality registration by taking the alignment of T1 and T2 weighted MRI data. The model used the same concatenation of 'moving' and 'static' images, the encoder-decoder structure with Conv3D layers and LeakyReLU activations, and the skip connections, ensuring detailed feature preservation during the upsampling process.

Loss function : The main change in this experiment was using mutual information as the loss function. Mutual information is a measure used in cross-modality registration, as it gives the common statistical dependence between the two image datasets, making it theoretically suitable for aligning images from different modalities. This is needed as the intensity ranges are different

Results : Despite the theoretical suitability of mutual information for cross-modality registration, our implementation faced challenges. The main issue we observed was the unbounded nature of the loss, which led to improper registration of the T1 and T2 weighted images. This outcome showed us that we could not directly apply the original VoxelMorph architecture for cross-modality, which was made for same-modality registration. The results made us experiment further to get better results for cross-modality registration.

2.3 Experiment 3: Original U-Net Architecture for Cross Modality Registration (T1 to T2) with Normalized Mutual Information Loss

Architecture : We continued to use the U-Net architecture from the original VoxelMorph paper. The focus remained on cross-modality registration, targeting the alignment of T1 and T2 weighted MRI data. This architecture follows the same pattern as in the previous experiments, with an encoder-decoder structure using Conv3D layers, LeakyReLU activations, and skip connections, which are essential for maintaining the integrity of high-resolution features during the image transformation process.

Loss function : This experiment changed by adopting normalized mutual information (NMI) as the loss function. NMI is an enhanced version of the mutual information loss used in experiment 2. It offers a more stable and normalized approach to quantifying the statistical dependence between the two image datasets. By normalizing the mutual information, this loss function aims to provide a more robust and consistent measure for aligning images from different modalities.

Results : The results from this experiment are better than the previous experiment results. Using normalized mutual information as the loss function effectively addressed the unbounded loss encountered earlier in experiment 2. We observed that the images were more accurately registered, indicating that NMI provided a more suitable and stable approach for cross-modality registration in this context. **The images look well registered and acheive an average DICE score of 0.817 for 100 images.**

2.4 Experiment 4: Smaller U-Net Architecture for Low Resolution Images

Architecture : In our fourth experiment, we modified the U-Net architecture to improve its performance with low-resolution images. We reduced the number of layers in both the encoder and decoder of the U-Net framework used in VoxelMorph. This approach was to decrease subsampling within the network to make registration work for images with lower resolutions.

Strategy : Our main goal was to minimize the depth of the network to retain more spatial information. By doing so, the network effectively captures and processes the minor details that might not be lost when using a deeper architecture.

Results : Although this change gave smaller loss values than others, which could not improve the model's performance, the output is not as expected. After the registration process, the images did not align satisfactorily regarding modality. This finding indicates that while reducing the network's depth preserved more spatial information in low-resolution images, it was insufficient for accurate cross-modality registration. These results suggest the need for alternative approaches to address low-resolution image registration in cross-modality scenarios.

2.5 Experiment 5: Increased Kernel Size in U-Net for Same Modality Registration

Implementation of Larger Kernel Sizes : Here, we tried to augment the U-Net architecture for same-modality MRI registration by increasing the kernel sizes [2](and paralleling) in the convolutional layers. This modification aimed to expand the receptive field of each convolutional layer, allowing the network to capture more contextual information from the images.

Loss function : We used normalized cross-correlation as the loss function with the above architecture modification. This choice was driven by the need for a robust and effective measure of similarity between the registered images. This loss function is significant for same-modality scenarios where small differences between images can significantly impact the registration accuracy.

Results : The increased kernel sizes in the U-Net architecture led to a noticeable improvement in the model's ability to register same-modality MRI images. The enlarged receptive field of the convolutional layers enabled the network to understand better and align the broader context within the images. Coupled with the normalized cross-correlation loss function, the model effectively aligned the images, demonstrating an enhanced performance compared to the standard U-Net architecture. These results highlight the potential benefits of larger kernel sizes in convolutional networks for tasks requiring detailed contextual understanding, such as same-modality MRI image registration.

2.6 Experiment 6: Transfer learning on registration adjacent tasks

Susceptibility distortion can be theoretically solved using a registration based approach. This is what we attempted with our transfer learning experiment to showcase the versatility of U-Net architecture.

Architecture : We used the vanilla VoxelMorph but trained it on opposite phase encode images (distortions look opposite due to the nature of MR physics). Pretraining is done on pairs of undistorted Diffusion MR images as the original model was trained on different intensity ranges and the T1 images used have little to no suceptibility distortion.

Loss function : Cross correlation is used as it is a unimodal problem

Results : The U-Net architecture is able to solve susceptibility distortion to a large extent.
This is much faster than the previous physics-based approaches and is novel in its approach. Few methods have used unsupervised learning to solve distortion correction in MR images prior to our attempt.

3 Results and Discussion

We are able to solve both unimodal and multimodal MR image registration using U-Net architecture. The average DICE score achieved was 0.828, 0.817 respectively and is similar to previous optimization based approaches like the ANTs. A sample crossmodal registration result is shown below. The brain shape of a totally diffierent subject in a different constrast is able to match with the static image.

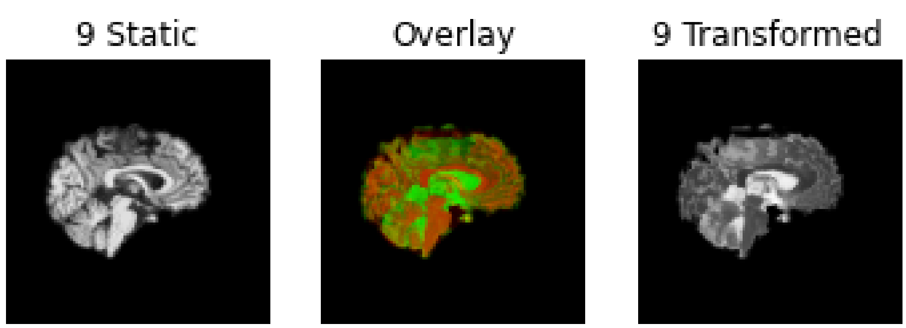

Figure 4: Cross-modal registration

Apart from the core registration results. We tried to solve a distortion correction problem using transfer learning. The resulting model was able to visibly correct distortions. The results are shown in Figure 4.

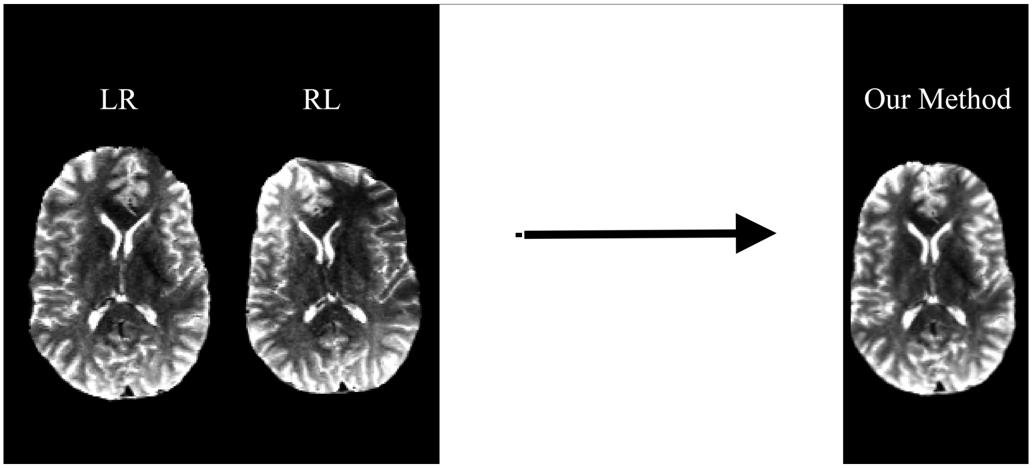

Figure 5: Susceptibility distortion correction

Since, there is no metric to evaluate distortion correction, we have extracted the magnitude of Gaussian gradient (as edge information) from structural images that are undistorted and overlaid them on the corrected image. The contours are in alignment visibly across the cortical folds.

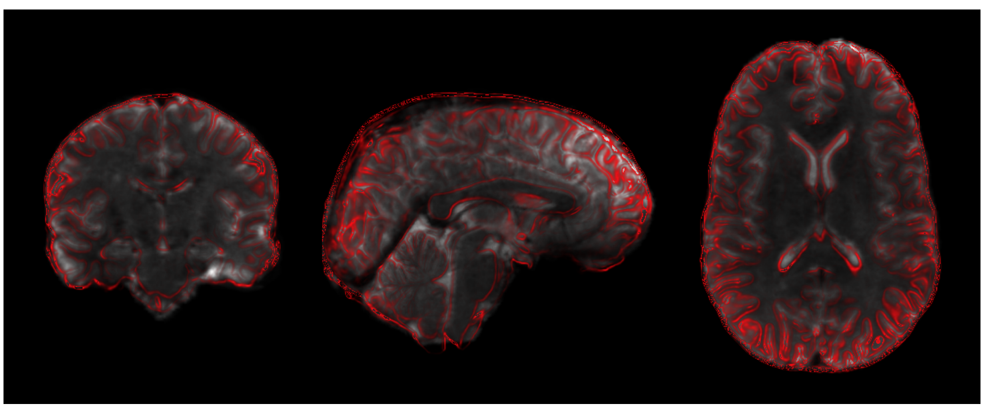

Figure 6: Contours are in alignment

4 Conclusion

In our comprehensive exploration of medical image registration, we focused on enhancing the U-Net-based model through six different experiments. These range from utilizing the original architecture for unimodal registration to adapting it for cross-modality registration challenges, including modifications for low-resolution image processing and increasing kernel sizes. We successfully perform uni-modal and cross-modal registration using a U-Net based architecture.

We presented a novel method to correct susceptibility distortion without the need for an opposite phase-encode acquisition.

It does so while achieving a high speedup for typical workflows and is able to closely match qualitatively the correction by state-of-the-art methods. Faster correction methods like these will enable widespread use of dMRI in the clinical setting where accurate shape is needed for critical decisions like treatment planning.

References

[1] 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*.

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer, Cham

[3] Huang, Hao, et al. "Correction of B0 susceptibility induced distortion in diffusion-weighted images using large-deformation diffeomorphic metric mapping." Magnetic resonance imaging 26.9 (2008): 1294-1302

[4] LaMontagne, Pamela J., et al. "OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease." MedRxiv (2019): 2019-12.

Code

The code for uni and multimodal registration is in "multimodal.ipynb". The code for Inference is in "infer.ipynb". The code for Transfer learning is in "transfer.ipynb". The code for the large kernel variant of U-Net is in "largekernel-train.ipynb".