You Only Look at Patches: A Patch-wise Framework for 3D Unsupervised Medical Image Registration

No Author Given

Abstract. Medical image registration is a fundamental task for a wide range of clinical procedures. Automatic systems have been developed for image registration, where the majority of solutions are supervised techniques. However, those techniques rely on a large and well-representative corpus of ground truth, which is a strong assumption in the medical domain. To address this challenge, we propose a novel unified unsupervised framework for image registration and segmentation. The highlight of our framework is that patch-based representation is key for performance gain. We first propose a patch-based contrastive strategy that enforces locality conditions and richer feature representation. Secondly, we propose a patch stitching strategy to eliminate artifacts. We demonstrate, through our experiments, that our technique outperforms current state-of-the-art unsupervised techniques.

1 Introduction

Image registration seeks to find a mapping that aligns an unaligned image to a reference one. The estimated spatial mapping aims to best align the anatomical structure of interest. Majority of existing works have been investigated from the classic perspective. Whilst promising performance has been reported, those techniques build upon costly optimisation schemes, which limits their efficiency when using a large volume of data. This limitation has encouraged the fast development of deep learning techniques for medical image registration. A set of techniques have been reported based on supervised learning. However, the need for a well-representative and high-quality ground truth is a strong assumption and hard to obtain in the medical domain. Another set of techniques have been devoted to explore unsupervised techniques e.g. [1,6,5,2]. Existing techniques have proposed several network mechanisms and explicit regularisers, to accommodate a certain level, with the lack of prior knowledge. However, the performance is still limited due to the lack of high-quality prior knowledge.

Our work is motivated by the aforementioned limitation. We argue that better quality prior can be estimated from patches rather than the full image. Medical images have complex anatomical structures, which impose a challenge when estimating an image-to-image mapping. Therefore, our modeling hypothesis is that patch embeddings are more meaningful representation for performance gain. In this work, we introduce a novel unified framework for unsupervised image registration and segmentation, which we call PC-SwinMorph (**P**atch **C**ontrastive Strategy with **S**hifted-**win**dow multi-head self-attention based on Voxel**Morph**). We underline two major highlights of our framework. Firstly, we introduce a patchwise contrastive registration strategy for richer feature representation. Secondly, we propose a patch stitching strategy to address the splitting effect caused by the image patch-based partition. We evaluate our framework using the benchmark dataset LPBA40. We demonstrate through our experimental result that our two patch-based strategies lead to better performance than the state-of-the-art techniques for unsupervised registration and segmentation.

2 Proposed Framework

In this section, we describe the overall workflow of our proposed framework.

Overview Workflow In Fig. 1, our PC-SwinMorph first take the moving and fixed images as inputs. We then generate non-overlap patches from the two input images, and perform patch-level contrastive learning to refine the features (Patch-based Strategy I from Fig. 1). Then the contrasted features are fed into two weight-shared CNN encoders. Followed by a decoder, the features are recursively concatenated and enlarged with skip connections to reconstruct two sets of deformation field patches. We then use a 3D W-MSA and a 3D SW-MSA module [4] to refine and stitch the deformation field patches to obtain the full deformation field (Patch-based Strategy II from Fig. 1). Finally, we wrap the moving image → fixed image, and the fixed image →moving image. After the training registration process, we also adopt the full deformation field to transfer the segmentation mask for fixed masks to obtain the segmentation mask of the moving image. We underline that no masks are used in the training registration process, and they are only used in the testing segmentation stage. Hence, our framework is a unified unsupervised registration and segmentation network.

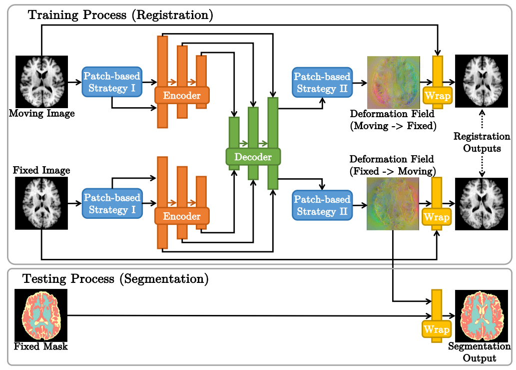

Fig. 1. Workflow of our proposed framework.

3 Experimental Results

In this section, we detail the experimental setup and experimental results to validate our proposed unified unsupervised registration and segmentation framework.

Experimental Setup We evaluate our framework on the publicly available LONI Probabilistic Brain Atlas (LPBA40) dataset [1] using Dice evaluation metrics. For the implementation details regarding the network architecture, data pre-processing, and training and testing schemes, we refer to the reader to [3]. Our code will be publicly available upon the acceptance of this work.

Comparison to the State-of-the-Art Techniques. We compared our technique with recent unsupervised brain segmentation methods, including VoxelMorph [1], DeepTag [6], SYMNet [5], CycleMorph[2]. For a fair comparison, all models use the same backbone, VoxelMorph, which has been fine-tuned to achieve optimal performance. In Fig. 2, the boxplots summarise performance-wise, in terms of the $Dice$ coefficient, the compared SOTA methods, and our PC-SwinMorph. In a closer look at the boxplots, we observe that our method outperforms all other SOTA methods by a large margin on all seven majority anatomical regions for both registration and segmentation tasks. Particularly, for both registration and segmentation tasks, our results report an improvement

[1] https://loni.usc.edu/research/atlases

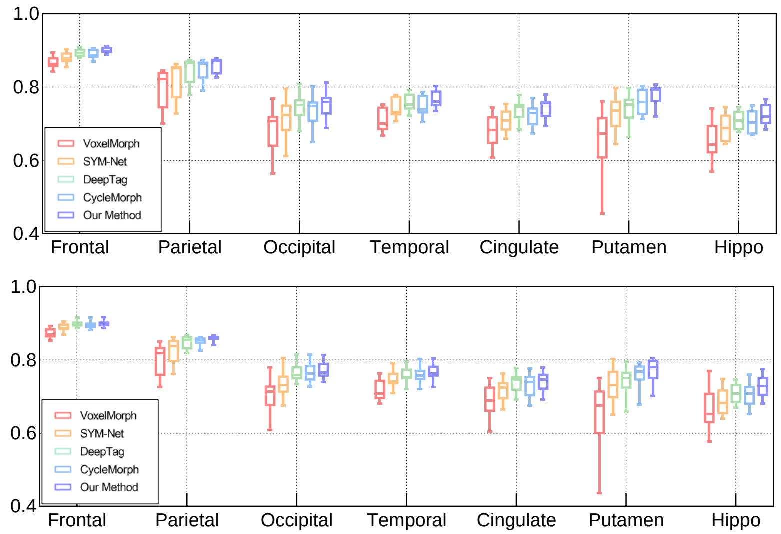

Fig. 2. Boxplots in terms of Dice, per anatomical region, for registration (top) and segmentation (bottom) tasks. The comparison displays our Method (PC-SwinMorph) against SOTA techniques.

of 5.9% compared to VoxelMorph on the average Dice results, and 3.9-4.3% against the other compared SOTA techniques on the average Dice score.

4 Conclusion

We introduced a novel unified unsupervised framework for image registration and segmentation. We showed that patches are crucial for obtaining richer features and preserving anatomical details. Our intuition behind the performance gain of our technique, is that patches can capture not only global but also local spatial structures (more meaningful embeddings). We demonstrated, that at this point in time, our technique reported SOTA performance for both tasks.

References

1. Balakrishnan, G., et al.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
2. Kim, B., et al.: Cyclemorph: Cycle consistent unsupervised deformable image registration. Medical Image Analysis **71**, 102036 (2021)
3. Liu, L., Huang, Z., Liò, P., Schönlieb, C.B., Aviles-Rivero, A.I.: Pc-swinmorph: Patch representation for unsupervised medical image registration and segmentation. arXiv preprint arXiv:2203.05684 (2022)
4. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030 (2021)
5. Mok, T.C., Chung, A.: Fast symmetric diffeomorphic image registration with convolutional neural networks. In: CVPR. pp. 4644–4653 (2020)
6. Ye, M., et al.: Deeptag: An unsupervised deep learning method for motion tracking on cardiac tagging magnetic resonance images. In: CVPR. pp. 7261–7271 (2021)