PIViT: Large Deformation Image Registration with Pyramid-Iterative Vision Transformer

Tai Ma, Xinru Dai, Suwei Zhang, and Ying Wen(✉)

Shanghai Key Laboratory of Multidimensional Information Processing, School of Communications and Electronic Engineering, East China Normal University, Shanghai 200241, China

ywen@cs.ecnu.edu.cn

Abstract. Large deformation image registration is a challenging task in medical image registration. Iterative registration and pyramid registration are two common CNN-based methods for the task. However, these methods usually consume more parameters and time. Additionally, the existing CNN-based registration methods mainly focus on local feature extraction, limiting their ability to capture the long-distance correlation between image pairs. In this paper, we propose a fast and accurate learning-based algorithm, Pyramid-Iterative Vision Transformer (PIViT), for 3D large deformation medical image registration. Our method constructs a novel pyramid iterative composite structure to solve large deformation problem by using low-scale iterative registration with a Swin Transformer-based long-distance correlation decoder. Furthermore, we exploit pyramid structure to supplement the detailed information of the deformation field by using high-scale feature maps. Comprehensive experimental results implemented on brain MRI and liver CT datasets show that the proposed method is superior to the existing registration methods in terms of registration accuracy, training time and parameters, especially of a significant advantage in running time. Our code is available at https://github.com/Torbjorn1997/PIViT.

Keywords: Medical image registration · convolutional neural networks · image processing

1 Introduction

Deformable image registration is one of the fundamental tasks in computer vision and has been widely used in medical image processing. In recent years, deep learning methods based on convolutional neural networks are widely applied in deformable image registration. Balakrishnan et al. [3] proposed VoxelMorph with

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43999-5_57.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14229, pp. 602–612, 2023.
https://doi.org/10.1007/978-3-031-43999-5_57

a structure similar to Unet and further developed a diffeomorphism implementation of VoxelMorph [8]. Mok et al. [21] proposed SYMNet to achieve accurate diffeomorphic registration by exploiting the cycle consistency of registration. However, when there is a significant difference between the images, it is difficult to learn an accurate deformation field for alignment because large deformation image registration has a high degree of freedom in transformation. Typical registration methods utilize rigid or affine transformation with a low degree of freedom to provide initialized global transformation for large deformation, however, this requires the introduction of additional preprocessing to obtain the corresponding affine matrix [12] [23]. In order to solve the high degree of freedom of large deformation transformation, the end-to-end deformable image registration methods are mainly divided into two types: iterative registration (Fig. 1 (a)) and pyramid registration (Fig. 1 (b)). *(a)* Iterative registration achieves coarse-to-fine image registration by cascading several CNNs, which requires huge GPU memory during training. In addition, iterative registration methods learn separate image features in each iteration, which brings additional computational costs when repeatedly extracting features. Typical iterative registration methods include RCN [28] and LapIRN [22]. *(b)* Pyramid registration achieves coarse-to-fine registration within one iteration by warping feature maps. These methods successively learn feature maps and deformation fields from low to high resolution. Typical pyramid registration methods include Dual-PRNet [14] and NICE-Net [20]. However, current non-iterative registration methods still cannot well solve the image registration problem under the significant differences condition.

Inspired by the capabilities of Transformer in NLP, recent researchers have extended Transformer to computer vision tasks [11] [19] and acquired results that surpass CNNs' in many tasks [17] [27]. Many Transformer-based registration methods have also been proposed for image registration tasks, such as TransMorph [7], Swin-VoxelMorph [30] and XMorpher [26]. Compared with CNN-based methods, Transformer-based methods have achieved better registration results, which illustrates that the global receptive field of Transformer is helpful for image registration.

In this paper, we propose a novel Pyramid-Iterative Vision Transformer (PIViT) by combining Swin Transformer-based long-range correlation decoder and the proposed pyramid-iterative registration framework shown in Fig. 1 (c).

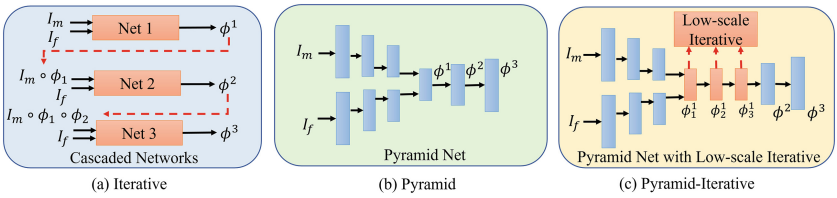

Fig. 1. Network architecture (a) iterative registration, (b) pyramid registration and (c) the proposed pyramid-iterative registration.

Our main contributions of this work are as: (1) We establish a pyramid-iterative registration framework to address large deformation image registration. The framework first extracts feature map pairs via a dual-stream weight-sharing encoder, then performs iterative registration on the low-scale feature space, and finally complements detail information and learns accurate deformation fields during pyramid decoding process. (2) We propose a Swin Transformer-based long-range correlation decoder, which exploits the global receptive field of Swin Transformer on low-scale feature maps to learn high accuracy large deformation fields while maintaining low parameters. (3) Compared with other popular registration methods, the proposed unsupervised end-to-end network is more lightweight and suitable for time-sensitive tasks.

Extensive experiments on 3D brain MRI and liver CT registration tasks demonstrate that PIViT achieves state-of-the-art performance in terms of accuracy but consumes less time and parameters.

2 The Proposed Method

In this section, we first propose a novel pyramid-iterative registration framework to solve large deformation image registration. The pyramid-iterative registration framework combines the advantages of iteration and pyramid registration framework to achieve fast and accurate registration. Then, we introduce a long-range correlation decoder based on Swin Transformer into the iterative registration stage of the proposed framework and utilize the global receptive field of the Swin Transformer to capture global correlations, thereby implementing high accurate and fast registration.

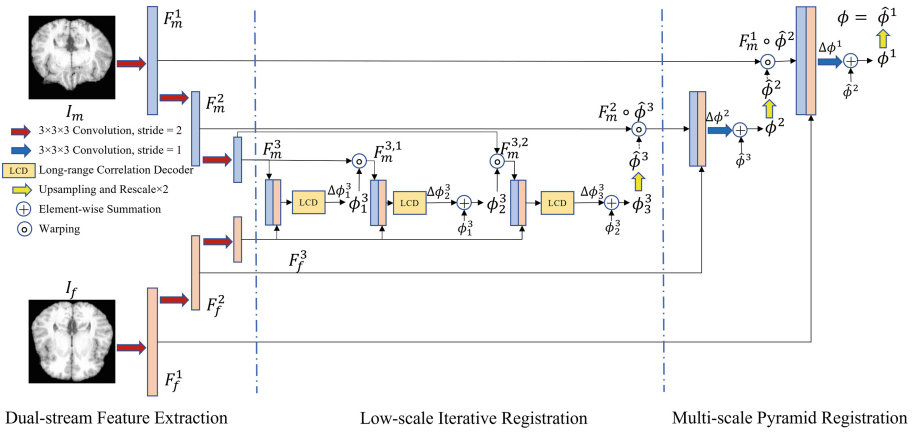

Fig. 2. Overview of the proposed PIViT. The number of pyramid levels N is set as 3 for illustration.

2.1 Pyramid-Iterative Registration Framework

As shown in Fig. 2, the proposed pyramid-iterative registration framework can be divided into three parts: dual-stream feature extraction, low-scale iterative registration and multi-scale pyramid registration.

Dual-Stream Feature Extraction: Similar to pyramid registration network, the proposed framework utilizes a weight-sharing feature encoder to construct feature pyramids for the fixed image I_f and the moving image I_m, respectively. At the i^{th} step ($i \in [1 \cdots N]$), the feature maps of I_f and I_m are formulated as F_f^i and F_m^i, respectively. The weight-sharing feature encoder reuses the same network blocks to extract the feature maps F_f^i and F_m^i without adding parameters or complicating the training process while ensuring that F_f^i and F_m^i are in the same feature space.

Low-Scale Iterative Registration: The pyramid-iterative registration uses two different decoding modules at different scales. To capture large deformation, we adopt low-scale feature maps to obtain the coarse distribution of large deformation fields without considering the fine distribution in this paper. Therefore, at the last N^{th} level of feature pyramid, deformation field is predicted from F_f^N and F_m^N multiple times through an iterative structure. Similar to iterative-based registration methods, F_m^N is warped by the predicted deformation field ϕ_t^N, where t is the number of iterations. The warped $F_m^{N,t}$ and F_f^N are used for the next iteration. In the first iteration, the decoder obtains the initial deformation field ϕ_1^N, and in the subsequent iterations, the residual deformation field $\Delta\phi_t^N$ is obtained in each prediction and the updated overall deformation field ϕ_t^N is obtained. This procedure can be formulated as:

$$F_m^{N,t} = F_m^N \circ \phi_t^N, \phi_t^N = \begin{cases} \Delta\phi_t^N, t = 1, \\ \phi_{t-1}^N + \Delta\phi_t^N, t = 2, \cdots, T, \end{cases} \tag{1}$$

where T is the upper limit of iteration, $\circ$ denotes warping the feature map with deformation fields, and $+$ denotes element-wise summation of deformation fields.

Compared with other iterative registration methods, the advantage of iterating only at the N^{th} level is that there is no need to re-extract image features, thus the computational complexity and time consumption of our method can be greatly reduced. This can greatly accelerate the speed of model training and deformation field prediction, and better solve large deformation.

Multi-scale Pyramid Registration: After the implementation of low-scale iterative registration, the deformation field ϕ_T^N is rescaled by a factor of 2 and the rescaled flow $\widehat{\phi}^N$ is obtained. The subsequent process is the same as that of the pyramid registration method. At each level, warped feature $F_m^i \circ \widehat{\phi}^{i+1}$ and fixed feature F_f^i ($i = N-1, \cdots, 1$) are concatenated and the residual deformation field $\Delta\phi^i$ is predicted by 3D convolution. $\Delta\phi^i$ is used to update $\widehat{\phi}^{i+1}$ so as to obtain the deformation field ϕ^i corresponding to the i^{th} layer. ϕ^i is rescaled

by a factor of 2 and warps moving feature F_m^{i-1}. The purpose of introducing multi-scale pyramid registration is to supplement the lack of fine information caused by only using low-scale features in the iterative registration stage. This process is repeated at each level of the feature pyramid until the deformation field is rescaled to the original image resolution. Finally, the pyramid-iterative registration framework obtains the predicted global deformation field.

2.2 Long-Range Correlation Decoder

To capture large deformation at low-scale registration, the study of the decoder is very essential. Therefore, we propose a long-range correlation decoder (LCD) in the iterative registration phase. As shown in Fig. 3, the LCD consists of a Swin transformer-based block and two consecutive convolutions. The Swin transformer-based block models the long-range correlation between F_f^N and $F_m^N \circ \phi_{t-1}^N$ using the self-attention mechanism of the transformer, and then the residual flow field $\Delta\phi_t^N$ is obtained by the convolution block. In order to enhance the information interaction between non-overlapping windows, we adopt the shifted local window attention strategy of the Swin Transformer. The structure of the Swin Transformer-based block is shown in the red frame area in Fig. 3, which consists of shifted window-based self-attention modules (W-SA & SW-SA), followed by a 2-layer MLP. A LayerNorm (LN) layer is applied before each SA and MLP module, and a residual connection is applied after each module.

Current Transformer-based registration methods usually directly migrate the Transformer structure to the 3D image registration task, which leads to a large number of parameters and a remarkably long inference time. In contrast, the proposed PIViT models long-range correlations in low-scale iterative registration with LCD to warp corresponding voxels between feature maps to spatial neighborhoods, thus it is not necessary to use the Transformer on large feature maps at high scales. In addition, LCD also removes position embedding, only uses single-head self-attention and reduces the number of channels. These operations accelerate the speed of PIViT and significantly reduce parameters.

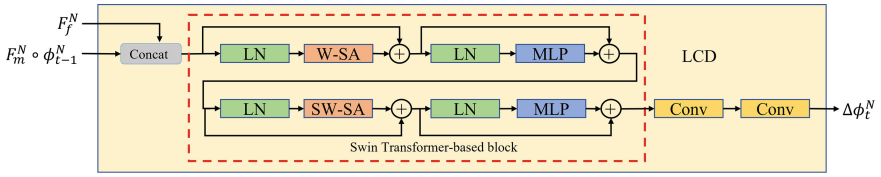

Fig. 3. An illustration of the structure of the proposed long-range correlation decoder (LCD). The red box indicates the Swin Transformer-based block. (Color figure online)

2.3 Loss Function

PIViT is an unsupervised end-to-end registration network. In this section, we design a loss function to train the proposed network. In the final stage of pyramid registration, PIViT obtains the deformation field ϕ between I_m and I_f and the warped image $I_w = I_m \circ \phi$ by using the differential operation based on the spatial transformer network [15]. In order to minimize the difference, we use the normalized cross-correlation (NCC) as a measure of the difference between the warped image I_w and fixed image I_f.

In order to ensure the continuity and smoothness of the deformation field ϕ in space, a regular term on its spatial gradient is introduced. The complete loss function is:

$$L_{I_f,I_m,\phi} = L_{sim} + \lambda L_{smooth} = -NCC(I_f, I_w) + \lambda \sum_{p\in\Omega} \|\nabla\phi(p)\|^2, \qquad (2)$$

where λ is the regularization hyperparameter.

3 Experiments

Data and Pre-processing: We evaluate the performance of PIViT on brain MRI datasets and liver CT datasets. In the experiments, we compare the proposed method with commonly used 3D convolutional registration methods Voxelmorph [3], Dual-PRNet [14], RCN [28], LKU-Net [16], TransMorph [7], Swin-VoxelMorph [30] and NICE-Net [20]. The accuracy of image registration is measured by Dice score [10]. We choose 2303 brain MRI scans from the ABIDE [9], ADHD [5] and ADNI [24] brain MRI datasets for training and LPBA [25] for testing. LPBA dataset contains 40 brain MRI scans with segmentation ground truth of 56 anatomical structures. For liver CT datasets, 1025 scans from MSD [2] and BFH [29] are selected for training and SLIVER [13], LiTS [6] and LSPIG [28] for testing. The images are all resampled to the size of 128×128×128. In order to better verify the effect of each method on large deformation image registration, we do not perform affine pre-alignment process. Atlas-based and scan-to-scan registrations are performed on brain and liver scans, respectively.

Implementation: We set λ to 1 for PIViT to guarantee the smoothness of the deformation field. Algorithm runtimes are computed on an NVIDIA GeForce RTX 3090 GPU and an Intel(R) Xeon(R) Silver 4210R CPU. We implement the model using Keras with a Tensorflow [1] backend and the ADAM [18] optimizer with a learning rate of $1e^{-4}$. The batch size is set as 1 and the networks are trained for 150,000 iterations.

Results: Table 1 shows the comparison of the proposed PIViT and other methods on 4 medical datasets. The Dice score, number of voxels with non-positive Jacobian determinants ($|J_s|_{\leq 0}$), GPU registration time (GRT), CPU registration time (CRT), network parameters and training time per iteration (TPI) of each method are presented.

Table 1. Comparison among VoxelMorph, VoxelMorph-diff, DualPRNet, RCN, TransMorph, Swin-VoxelMorph, LKU-Net, NICE-Net and the proposed PIViT on the LPBA, SLIVER, LiTs and LSPIG datasets. * indicates that the t-test p-value between PIViT and all other methods is less than 0.05.

Method	LPBA		SLIVER	LiTs	LSPIG	GRT ↓	CRT ↓	Params ↓	TPI ↓
	Dice ↑	$\lvert J_s\rvert_{\leq 0}$ ↓	Dice ↑	Dice ↑	Dice ↑				
VoxelMorph (CVPR 2018)	55.3 ± 6.3	17353.1	73.0 ± 6.7	67.3 ± 8.3	69.0 ± 8.8	**0.05 s**	0.73 s	**312.7K**	**0.14 s**
VoxelMorph-diff (MICCAI 2019)	60.1 ± 4.9	**0.0**	82.8 ± 6.2	80.1 ± 7.1	76.8 ± 6.5	0.08 s	0.99 s	319.7K	0.22 s
DualPRNet (MICCAI 2019)	58.3 ± 4.9	5446.9	79.1 ± 5.6	77.0 ± 6.5	71.5 ± 7.5	0.11 s	1.59 s	581.7K	0.28 s
RCN (ICCV 2019)	67.6 ± 2.6	6559.3	80.0 ± 6.6	73.7 ± 8.5	68.8 ± 7.3	0.35 s	2.20 s	938.7K	0.36 s
TransMorph (MIA 2022)	56.7 ± 5.6	24588.1	88.2 ± 4.1	78.7 ± 11.6	81.6 ± 5.3	0.10 s	1.99 s	45675.0K	0.86 s
Swin-VoxelMorph (MICCAI 2022)	58.1 ± 5.2	15857.6	77.1 ± 6.1	69.2 ± 10.5	70.6 ± 6.5	0.12 s	6.74 s	26573.9K	1.18 s
LKU-Net (MICCAI 2022)	58.9 ± 5.5	2215.7	81.4 ± 5.1	77.3 ± 7.4	73.9 ± 6.8	0.10 s	1.27 s	2037.4K	0.18 s
NICE-Net (MICCAI 2022)	68.1 ± 2.4	8061.6	87.1 ± 5.0	82.8 ± 6.8	79.6 ± 6.4	0.11 s	1.06 s	1033.7K	0.25 s
PIViT	$\mathbf{70.1 \pm 1.4}$*	697.0	$\mathbf{90.9 \pm 3.7}$*	$\mathbf{86.9 \pm 5.3}$*	$\mathbf{84.7 \pm 4.7}$	0.07 s	**0.32 s**	420.8K	**0.14 s**

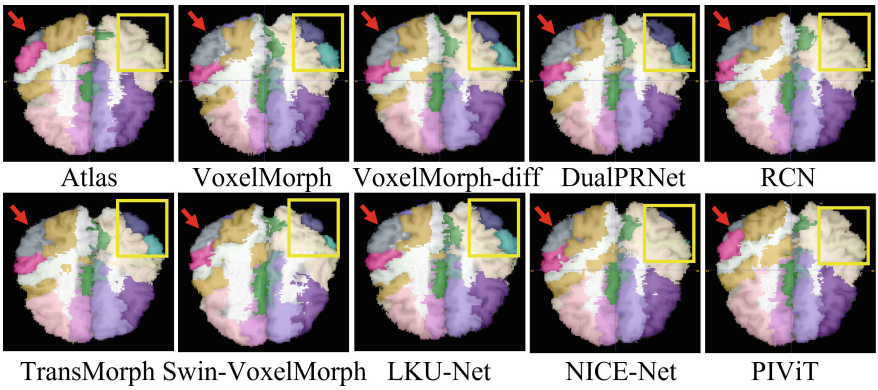

Fig. 4. Visualization of comparative registration results.

As shown in Table 1, VoxelMorph, VoxelMorph-diff, TransMorph, Swin-VoxelMorph and LKU-Net all get low Dice scores, indicating that these single-stream registration methods are difficult to solve large deformation. However, the iterative registration method RCN and the pyramid registration method NICE-Net obtain relatively good Dice scores, indicating that both iterative and pyramid registration methods can be useful to deal with large deformation. However, the Dice score of the proposed PIViT combining their advantages surpasses that of VoxelMorph by 14.8%, and the improvements compared to RCN and NiceNet also reach 2.5% and 2.0% on LPBA, respectively. On 3 liver datasets, compared with VoxelMorph, PIViT achieves 17.9%, 19.6% and 15.7% improvements, while compared with NICE-Net, PIViT achieves 3.8%, 4.1% and 5.1% improvements, respectively. The experiments indicate that the proposed PIViT implements large deformation fine registration better than other methods.

In addition to the superior Dice score, another advantage of the proposed PIViT is its fast and lightweight registration. Table 1 shows that the parameters, training and registration time of PIViT are close to that of VoxelMorph, far less than those of RCN and NICE-Net. These properties of PIViT make it easier to train and more suitable for time-sensitive tasks. Compared to other Transformer-based methods, the proposed PIViT has orders of magnitude optimization in parameters, which is because we only use the Transformer block at low scales, and LCD is tuned and optimized for 3D image registration tasks. What's more, although there is no additional constraint on the diffeomorphism of the deformation field, the deformation field obtained by PIViT has better diffeomorphism properties than those obtained by other methods except VoxelMorph-diff.

The visualization result of the experiment on the LPBA dataset is shown in Fig. 4. Obviously, most of the methods produce severe misregistration in the yellow regions of Fig. 4, due to the existence of large deformation. Compared with the registration results of RCN and NICE-Net, the proposed method achieves better alignment on the fine structure, which can be seen in the areas indicated by the red arrow in Fig. 4. It can be seen, since PIViT focuses on lightweight and fast registration of large deformations, its effectiveness on fine registration tasks is still somewhat weak.

Table 2. Dice score vs. different decoders and number of iterations.

t	Decoder Type											
	CNN				GRU				LCD			
	LPBA ↑	SLIVER ↑	Params ↓	TPI ↓	LPBA ↑	SLIVER ↑	Params ↓	TPI ↓	LPBA ↑	SLIVER ↑	Params ↓	TPI ↓
1(pyramid)	67.2	87.7	224.6K	0.12 s	68.3	88.4	413.7K	0.11 s	69.3	88.7	291.0K	0.12 s
2	68.2	88.7	278.7K	0.12 s	69.4	89.7	413.7K	0.12 s	69.8	90.4	355.4K	0.13 s
3	**68.9**	89.5	332.7K	0.12 s	69.5	90.2	413.7K	0.14 s	**70.1**	**90.9**	420.8K	0.14 s
4	68.8	**90.3**	386.7K	0.12 s	**70.0**	90.7	413.7K	0.14 s	70.0	90.8	486.1K	0.16 s
5	**68.9**	90.1	440.8K	0.13 s	69.9	**91.2**	413.7K	0.16 s	70.0	**90.9**	551.5K	0.19 s

Number of Iterations and Decoder Type: In this section, we explore how the number of iterations and decoder type of block in the long-range correlation decoder affect the registration performance. We select three different blocks, i.e. LCD, CNN and GRU [4], to perform iterative decoding and predict low-scale deformation fields. In order to verify the effectiveness of iterative registration and how the number of iterations affects the registration effect, we performed 1 to 5 iterations for each decoder.

The Dice scores corresponding to different decoders and number of iterations are shown in Table 2, t represents the time of iteration. Experiments are performed on LPBA and SLIVER datasets. Obviously, among the three decoders, LCD gets the best registration results in a limited number of iterations. When the number of iterations is 1, the proposed structure degenerates into pyramid registration, and when the number of iterations is greater than or equal to 2, the pyramid-iteration structure is used. Obviously, the registration accuracy is

greatly improved compared with the pyramid structure when the number of iterations is 2. On this task, when the number of iterations reaches 3, the registration accuracy tends to be stable, which indicates that the large deformation has been basically captured. Compared with the GRU block commonly used in optical flow tasks, LCD requires fewer iterations to converge, which verifies that LCD can better capture long-distance correlation and learn accurate flow.

4 Conclusion

In this paper, we propose an unsupervised pyramid-iterative vision Transformer (PIViT) for large deformation image registration. PIViT is an iterative and pyramid composite framework to achieve fine registration of large deformable images by iterative registration of low-scale feature maps and pyramid feature supplementation on high-scale feature maps. Furthermore, in the iterative decoding stage, a Swin Transformer-based long-range correlation decoder is introduced to capture the long-distance dependencies between feature maps, which further improves the ability to handle large deformation. Experiments on brain MRI scans and liver CT scans demonstrate that our method can accurately register 3D large deformation medical images. Furthermore, our method has significant advantages in terms of parameters and time, which can make it more suitable for time-sensitive tasks.

Acknowledgments. This work was supported in part by the National Nature Science Foundation of China (62273150), Shanghai Natural Science Foundation (22ZR1421000), Shanghai Municipal Science and Technology Committee of Shanghai Outstanding Academic Leaders Plan (21XD1430600), the Science and Technology Commission of Shanghai Municipality (14DZ2260800).

References

1. Abadi, M., et al.: Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265–283 (2016)
2. Antonelli, M., et al.: The medical segmentation decathlon. arXiv preprint arXiv:2106.05735 (2021)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
4. Ballas, N., Yao, L., Pal, C., Courville, A.: Delving deeper into convolutional networks for learning video representations. arXiv preprint arXiv:1511.06432 (2015)
5. Bellec, P., Chu, C., Chouinard-Decorte, F., Benhajali, Y., Margulies, D.S., Craddock, R.C.: The neuro bureau ADHD-200 preprocessed repository. Neuroimage **144**, 275–286 (2017)
6. Bilic, P., et al.: The liver tumor segmentation benchmark (LITS). arXiv preprint arXiv:1901.04056 (2019)

7. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: TransMorph: transformer for unsupervised medical image registration. Med. Image Anal. **82**, 102615 (2022)
8. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Med. Image Anal. **57**, 226–236 (2019)
9. Di Martino, A., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry **19**(6), 659–667 (2014)
10. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
11. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
12. He, Y., et al.: Geometric visual similarity learning in 3D medical image self-supervised pre-training (2023). https://doi.org/10.48550/ARXIV.2303.00874. https://arxiv.org/abs/2303.00874
13. Heimann, T., et al.: Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans. Med. Imaging **28**(8), 1251–1265 (2009)
14. Hu, X., Kang, M., Huang, W., Scott, M.R., Wiest, R., Reyes, M.: Dual-stream pyramid registration network. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 382–390. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_43
15. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
16. Jia, X., Bartlett, J., Zhang, T., Lu, W., Qiu, Z., Duan, J.: U-Net vs TransFormer: is U-Net outdated in medical image registration? In: Lian, C., Cao, X., Rekik, I., Xu, X., Cui, Z. (eds.) MLMI 2022. LNCS, vol. 13583, pp. 151–160. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-21014-3_16
17. Jiang, S., Campbell, D., Lu, Y., Li, H., Hartley, R.: Learning to estimate hidden motions with global motion aggregation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9772–9781 (2021)
18. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
20. Meng, M., Bi, L., Feng, D., Kim, J.: Non-iterative coarse-to-fine registration based on single-pass deep cumulative learning. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 88–97. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_9
21. Mok, T.C., Chung, A.: Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4644–4653 (2020)
22. Mok, T.C.W., Chung, A.C.S.: Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: Martel, L., et al. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 211–221. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_21
23. Mok, T.C., Chung, A.: Affine medical image registration with coarse-to-fine vision transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20835–20844 (2022)

24. Mueller, S.G., et al.: Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). Alzheimer's Dement. **1**(1), 55–66 (2005)
25. Shattuck, D.W., et al.: Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage **39**(3), 1064–1080 (2008)
26. Shi, J., et al.: XMorpher: full transformer for deformable medical image registration via cross attention. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 1346, pp. 217–226. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_21
27. Xu, H., Zhang, J., Cai, J., Rezatofighi, H., Tao, D.: GMFlow: learning optical flow via global matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8121–8130 (2022)
28. Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al.: Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10600–10610 (2019)
29. Zhao, S., Lau, T., Luo, J., Eric, I., Chang, C., Xu, Y.: Unsupervised 3D end-to-end medical image registration with volume Tweening network. IEEE J. Biomed. Health Inform. **24**(5), 1394–1404 (2019)
30. Zhu, Y., Lu, S.: Swin-VoxelMorph: a symmetric unsupervised learning model for deformable medical image registration using Swin transformer. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention - MICCAI 2022, vol. 13436, pp. 78–87. Springer, Cham (2022)