Unsupervised brain MRI image registration based on 3D hybrid ViT and convolutional U-net

Junyao Wang
Wuhan University
Feng Liu (fliuwhu@whu.edu.cn)
Wuhan University
Shi Shu
Wuhan University
Guowei Tao
Wuhan University
Fu Zhou
Wuhan University

Article

Keywords:

Posted Date: May 15th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2898722/v1

Additional Declarations: No competing interests reported.

Unsupervised brain MRI image registration based on 3D hybrid ViT and convolutional U-net

Junyao Wang[1,*], Feng Liu[1,*], Shi Shu[1], Guowei Tao[1], and Fu Zhou[1]

[1]Wuhan University, School of Computer Science, Wuhan, 430072, China
[*]wjyyy1@whu.edu.cn
[*]fliuwhu@whu.edu.cn

ABSTRACT

During a patient's disease progression, MRIs scaned in different times need to be registered before and after. However, the location and structure of tissue inside the human body may change with the growth of illnesses, interfering with the physician's ability to quickly diagnose the progression of the disease. To reach this goal, we proposed a 3D hybrid ViT and convolutional U-net for brain MRI image registration, which achieved a higher dice score than ViT-V-Net and VoxelMorph. In the meantime, we have had an idea of a novel loss function for gray image registration called grad-loss, which concentrates on the difference and gradient at each voxel of the MRI image. Quantitative and qualitative comparison results demonstrate that our model outperforms the previous ViT-based and convolution-based networks and achieved a better dice score of 79.7% in OASIS dataset.

Introduction

In the diagnoses with modern medical techniques, digital images and information help a lot. Image registration works for the evolution of brain tumors, lung nodules, and other kinds of cancer. For the growth assessment of tissue lesions, clinicians usually need to empirically compare CT or MRI images of multiple time points. Therefore, for time series diagnoses, 3D registration helps to compare and analyze the progression of the disease, and then the judgment of the benign and malignant properties of cancer.

Conventional registration methods are usually based on an iterative strategy to continuously optimize affine parameters. And the methods are mostly rigid, so they are called affine registration. However, to align the details of images takes more than just affine information. So deformable registration is needed for further medical and academic use.

With the development of deep networks, traditional iterative algorithms are exposed to be slow and unintelligent. The iterative algorithms start with presetting affine parameters as identity transformation, and then update the affine parameters by minimizing the cost function. When it comes to deep networks, the parameters can be predicted by the training process. Affine transforms can be described by only a few real numbers, whereas a free-form dense deformable field specifies the displacement for each grid point[1,2]. Deformable registration is about dense prediction, which has been developed in image segmentation field. So most of the SOTA deformable registration methods are of encoder-decoder architecture. The encoder is frequently based on an image classification network, also called the backbone, that is pretrained on a large corpus such as ImageNet[3]. The decoder aggregates features from the encoder and converts them to the final dense predictions[4].

After proper registration, the morphological change of details is easier to observe. Brain MRI image registration is not an easy task in practice, due to many influential factors like the inconsistency of scanning devices and the process of disease, which always exist to interfere with the CT image registration[5].

To improve the performance of MRI or CT registration, many novel approaches have been proposed[6,7], which significantly beats traditional methods such as ANTs[8]. Inspired by these algorithms, we propose an improvement of the network structure and an innovation of the loss function. We take advantage of two kinds of bottlenecks, Transformer[9] and convolution[10], to extract more specific information for image registration.

Related works

Registration works are mainly divided by affine registration and deformable registration. Also there are traditional registration and learning-based registration.

For affine registration, there are traditional methods such as Elasitx[11], ANTs[8], and learning-based methods such as AirNet[12] and DLIR[13].

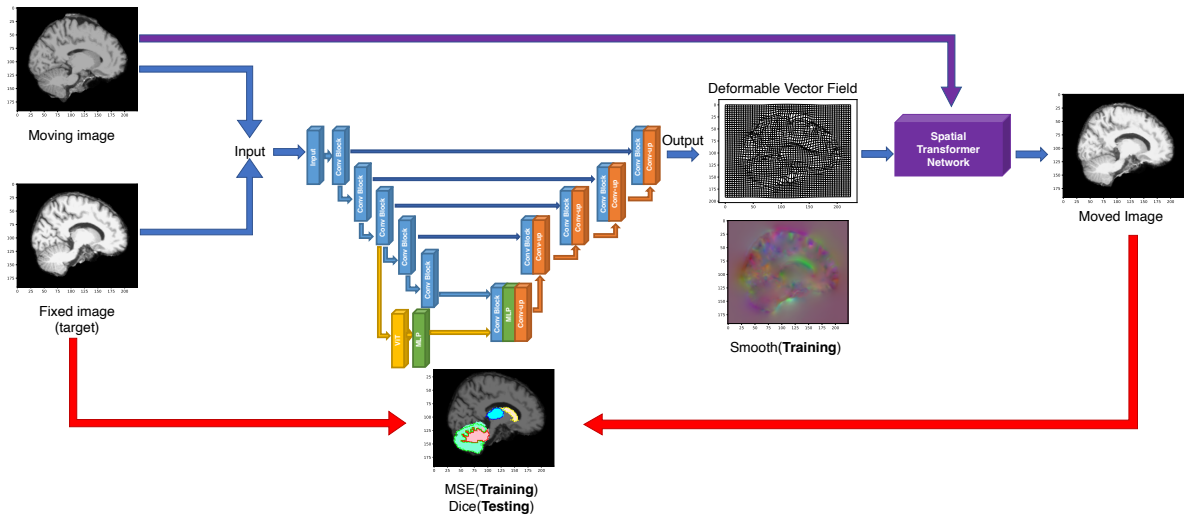

Figure 1. The overall structure of medical image registration

In early works, traditional medical image registration is based on grayscale values of images, and they need tedious parameter adjustment, iterative optimization, and are time-consuming and generally accurate. Beg et al. used the Euler-Lagrange equations to find the minimum incremental cost path on the manifold of diffeomorphisms[14]. Avants et al. proposed a novel symmetric image normalization method (SyN) with cross-correlation to reduce the computation time of registration[15]. Elastix is a collection of parametric intensity-based registration methods, which can solve both affine and deformable problems[11].

Recently, many learning-based approaches are proposed. For supervised method, Cao et al. proposed a method that learns and predicts the deformation field between a reference image and a subject image along with key points[16]. Sokooti et al. trained RegNet with artificially generated DVFs to cast registration as a learning problem[17]. For unsupervised method, VoxelMorph, RCN, VTN and CycleMorph are state-of-the-art approaches[1,6,18], giving several ideas for deformable or nonrigid registration. Besides, adversarial learning methods are also introduced to image registration. By training two networks, one being a generator and the other being a discriminator, a registration estimator and evaluator are built simultaneously[19]. Knowledge distillation and adversarial learning are also used for deformable registration, which significantly strungs the size of registration models[20].

In order to make dense prediction for 3D medical images, U-net was proposed as a useful framework for dealing with medical image segmentation[21]. Then Ferrante et al. proposed a pioneering registration work based on U-net[22], and simultaneously VoxelMorph was proposed by Balakrishnan et al.[6]. After that, transformer based networks such as ViT along with Swin Transformer was introduced to registration tasks[9,23,24]. Following these works, Chen et al. proposed state-of-the-art ViT-v-net and TransMorph[7,25].

Lately, many registration networks have been pushing the envelope and have had a rapid evolution like Swin-VoxelMorph and XMorpher[26,27]. They use novel encoder-decoder structure or attention module to do more with registration, which is a promising field for medical surgical and diagnosing problems.

Method

To demonstrate the overall process of our method, we present the structure of the approach in Fig. 1. We need a moving image(source), a fixed image(target), and combine them as one input of the network. The network generates a deformable vector field(DVF), which warps the moving image to a moved image. At last, we compare the moved image with the target to measure the difference to improve the training or analyse the testing.

When an MRI is transformed into a new one, we need to feed a DVF into the Spatial Transformer Network to direct the transformation[28].

We need to align one moving image(called m) to a fixed image(called f) of the same shape $\mathbf{R}^{H\times W\times L}$. H for height, W for width and L for length of slices. The images are of grayscale and single-channel. To figure out the pairwise relationship of the voxels in two images, we need to predict a transformation function ϕ that warps m to f[25]. To make $m\circ\phi$ closer to f, transformation function ϕ are supposed to be

$$\text{argmin}_{\phi} MSE(m\circ\phi, f)$$

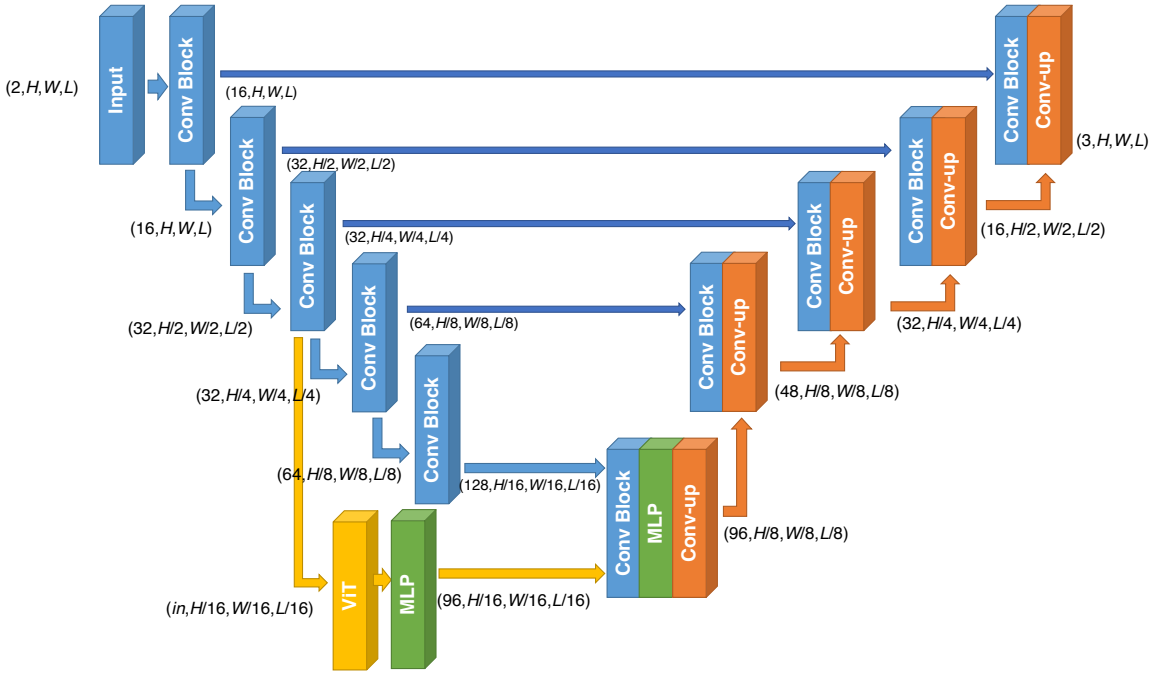

Figure 2. The structure of the hybrid network

In the formula, $\phi = Id + \mathbf{u}, \mathbf{u}$ denotes the flow field of displacement vectors, and Id denotes the identity. Our method is an unsupervised approach. For given m and f, the network generates the flow field $\mathbf{u}$ to align the images.

$m \circ \phi$ is performed via a spatial transformation function[28]. After transforming m by the DVF ϕ, two images are aligned. That is to say, the difference between the two images can be represented as a DVF called ϕ. For this reason, to register or align two images, we just need to generate a ϕ denoting the deformation. And with the assistance of spatial transformation function, ϕ is of the same size as the images. But for 3D images, the vector has a dimension of three, so the shape of ϕ is $\mathbf{R}^{3\times H\times W\times L}$.

Due to the limitation of a single encoder, we mixed Transformer and convolutional structures, to propose a novel hybrid and make it work.

Consequently, the input of the network is the concatenation of f and m, the shape of which is $(2,H,W,L)$. After that, the image is reshaped to 16 channels. Going through several layers, the image is down-sampling into a shape of $(32,H/4,W/4,L/4)$.

By shrinking the shape of the image, it is able to put it into a Vision Transformer[25]. The image is divided into patches of voxels and then be encoded. And at the same time, the down-sampleing still continues.

Because the shape of image patches in Vision Transformer is $(H/16,W/16,L/16)$, in the convolutional layer, the image should be downsampled to a same scale. Then the output of the two encoders will be concatenated. Because our downsample scale is 16, the network can process various shapes of images, as long as the dimension of each shape is a multiple of 16.

Due to the useful skip connection proposed in U-net[21], we apply it to stop the gradient from exploding and disappearing. After upsampling for several times, the shape of the image returns to (H,W,L). But the dimension of deformable vector are three, so the number of final channel is 3, which is shown as Fig. 2.

Then the hybrid network generates a DVF ϕ, we need to put the moving image m along with the ϕ into the spatial transformer network, and finally we can get a transformed image m'.

To enhance the information provided by the image, we have modified the MSE loss to a new one called $WMSE$. The boundaries of each structure in human tissue and organs provide much information about their shape and location. The derivative and differential in the image reflect the significance to the boundaries in different regions. So we introduce the gradient map to our loss function. The higher absolute value of the gradient of a voxel is, the more changes there are around the voxel. Therefore, a weighted factor $(1+\lambda\nabla p)$ is added when calculating the loss function of similarity.

$$MSE(x,y) = \frac{\sum_{p\in y}(x_p - y_p)^2}{|y|}$$

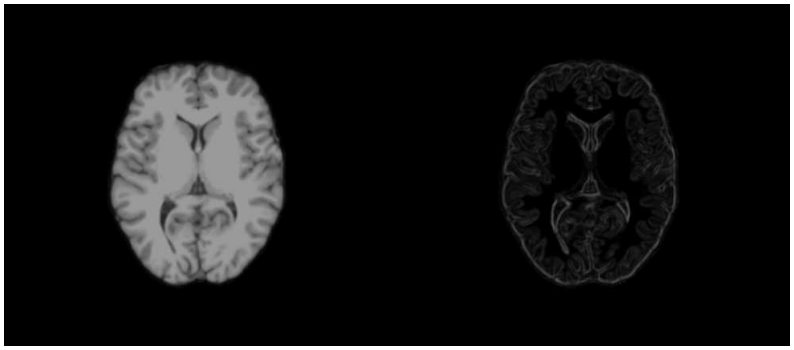

Figure 3. Brain MRI slice and its gradient map(absolute value)

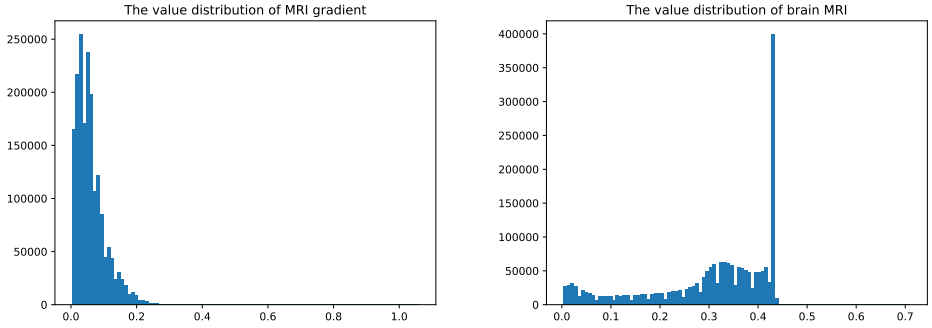

Figure 4. The value distribution of a Brain MRI and its gradient map

$$WMSE(x,y) = \frac{\sum_{p\in y}(1+\lambda|\nabla p|)(x_p - y_p)^2}{|y|}$$

$$\mathscr{L}_{similarity}(x,y) = \frac{\sum_{p\in y}(1+\lambda|\nabla p|)(x_p - y_p)^2}{\sum_{p\in y}(1+\lambda\nabla p)}$$

After adding to the weighted factor, the smoother region contributes less, while the rough part gives the valuable feature. As is shown in the right figure in Fig. 3, the brighter(whiter) the pixel is, the more important it will be.

Considering the difference of value range between the MRI and gradient map, we introduce a parameter λ to balance the influence. The distribution are shown in Fig. 4. If the change in the image matters more, we need to increase the λ weight to enhence the effect of our loss function. Considering the distribution of the absolute gradient value, we may adjust the weight to a proper range to make full use of the image information.

To enforce smoothness in the deformation field, a diffusion regularizer was used. By minimizing $WMSE$, $m\circ\phi$ will be closer to the f , but it may generate a non-smooth ϕ that is not physically realistic[6]. So we introduce a diffusion regularizer on the spatial gradients of displacement $\mathbf{u}$, which is defined as:

$$Diff(\phi) = \sum_{p\in\Omega} \|\nabla\mathbf{u}(p)\|^2$$

$$\mathscr{L}_{smooth}(x,y) = \frac{\sum_{p\in\Omega}\|\nabla\mathbf{u}(p)\|^2}{|y|}$$

, where Ω denotes all the coordinates of the image.

Experiment

Dataset contains IXI dataset, and OASIS dataset[7,29]. IXI dataset has 576 images, in which there are 403 for training, 58 for validation and 115 for testing. The shape of the image is $(224, 192, 160)$. OASIS has 453 images, 353 for training, 50 for

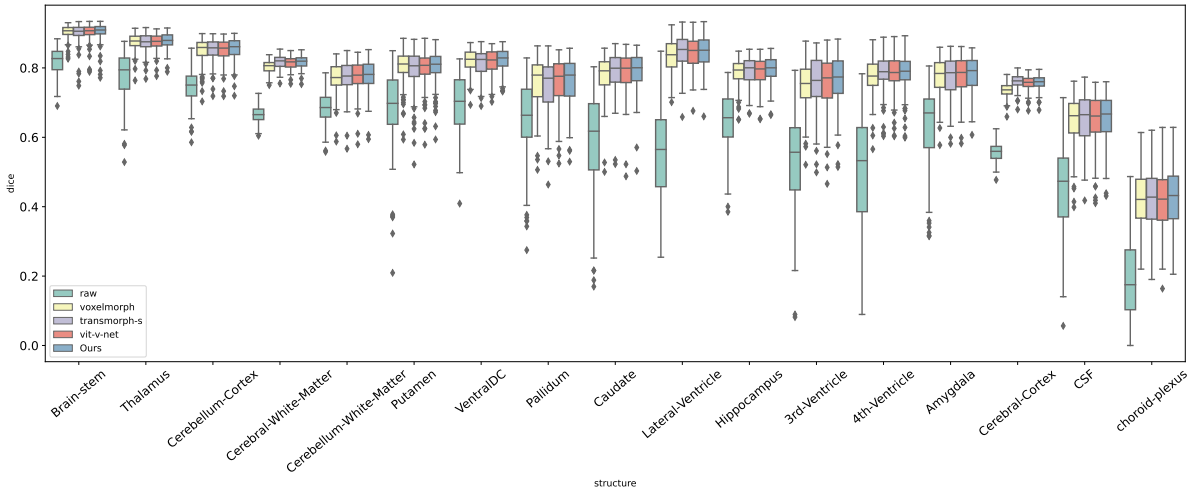

Figure 5. The boxplot of selected anatomical brain structures

validation and 50 for testing. The shape of the image in OASIS is $(192, 224, 160)$. Then the value of images in the datasets are normalized into $[0, 1]$, in order to arrange the data in the same scale.

To measure the similarity of the transformed image and the target image, we calculate the volume overlap for structures using the Dice score[30]. The overlap about structure k of the $m \circ \phi$ and f images are calculated as:

$$Dice_k(m \circ \phi, f) = 2 \cdot \frac{|(m_k \circ \phi) \cap f_k|}{|m_k \circ \phi| + |f_k|}$$

A Dice value close to 1 means that the two parts overlap, and a Dice value close to 0 means that the two parts do not intersect. For each structure, we beat the previous work by a margin of dice score. Because there are many structures, the overall dice value are calculated by averaging all values to get a more fair result. In the average value, we still have advantage over other methods.

$$Dice(m \circ \phi, f) = \frac{\sum_{k \in structure} Dice_k(m \circ \phi, f)}{|structure|}$$

Boxplots in Fig. 5 shows the dice scores of different brain MR substructures using the proposed network and SOTA approaches. We choose the 3 SOTA learning-based approaches that has similar time and memory cost with our method, which are VoxelMorph[6], ViT-V-Net[25], and TransMorph-S[7] respectively.

Beside the overlap metric, we also evaluate the regularity and reality of the deformation fields. Detailedly, we use the Jacobian matrix $J_\phi(p) = \nabla\phi(p) \in \mathbf{R}^{3\times3}$ to represent the local properties of ϕ around voxel p. We count all non-zero voxels for which satisfies $|J_\phi(p)| \leq 0$, where the deformation is not diffeomorphic[31]. In our work, the accuracy is improved while maintaining a low level of the number of voxels satisfying $|J_\phi(p)| \leq 0$.

In Fig. 6, the deformation fields are shown as the warped grids and RGB values. The subfigure(f) in Fig. 6 is the displacement field $\mathbf{u}$, where the spatial direction x, y, and z is normalized and mapped to each of the RGB color channels in range $[0, 1]$, respectively. It means that, the higher the red value in a voxel is, the farther it moves to the x direction.

In Fig. 7, the overlapped regions and the deformation field generated by each model are highlighted, compared with other SOTA methods. It proves that, our method achieved the best result in the labeled regions, maintaining the smoothness and rationality of the deformation field.

The different colors of the boundaries and regions denote different anatomical structures in the brain. In the figures, red represents cerebellum white matter, green for cerebellum cortex, blue for thalamus and yellow for caudate.

We implement the training and testing for registration on NVIDIA RTX 3060 with a memory of 12GB and AMD Ryzen-5 5600G@3.90GHz in Windows 10 Operating System. And we use adam optimizer to run 500 epoches on each network and dataset respectively. Due to the limitation of graphic card, the batchsize of all the training processes are set to 1. In this case, it takes around 12GB while training and 8GB for registration.

In each epoch, we randomly choose 200 pairs of images from training data as (x, y) to train. Firstly, we assign x to be the m, which is the moving image, y to be the f, which is the fixed image. Then the network generates a m' and a flow field which has

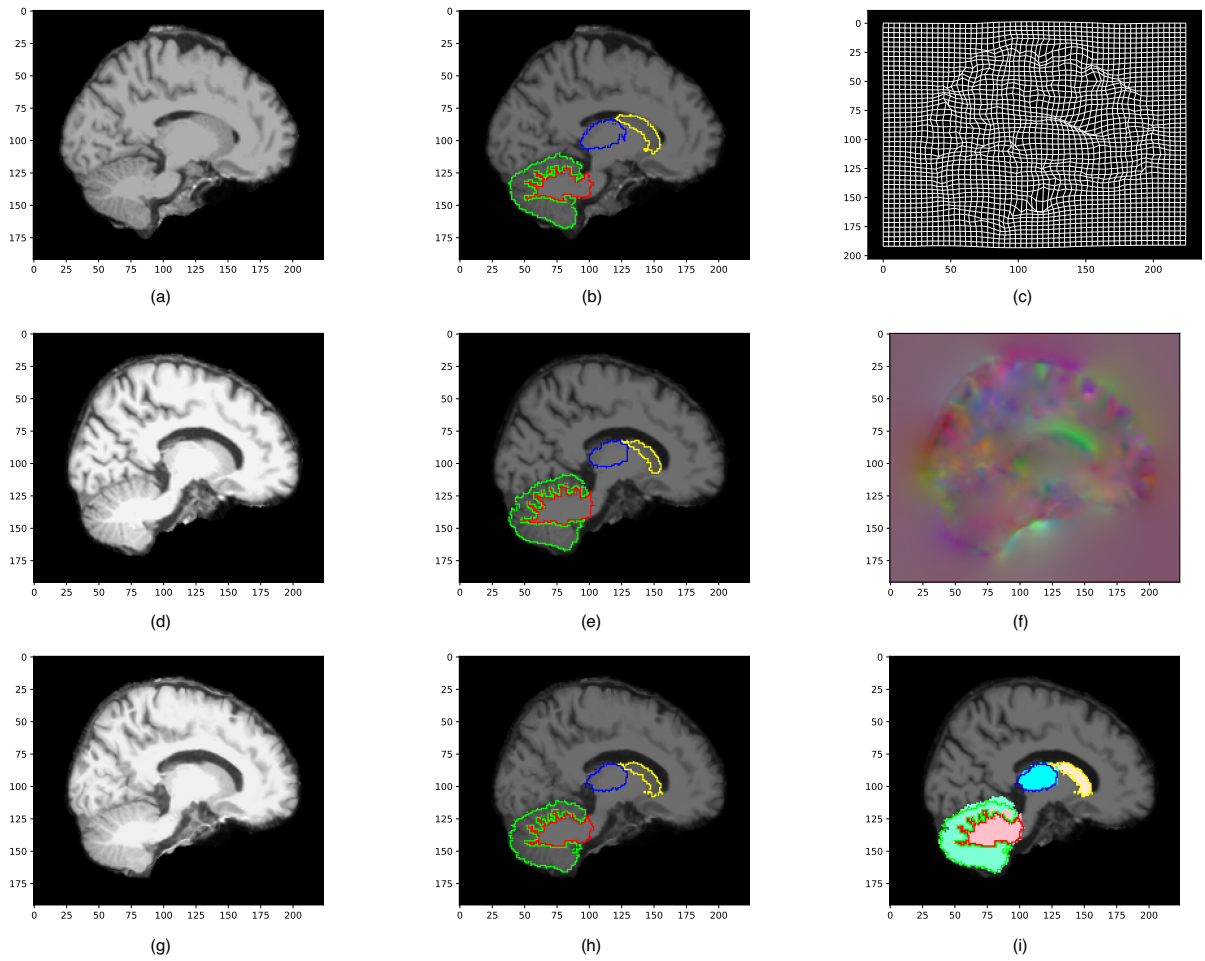

Figure 6. An example of experiment. (a)Slice of moving image (b) Structure of moving image (c) The warp grid of the DVF (d) Slice of fixed image (e) Structure of fixed image (f) The RGB representation of DVF (g) Slice of moved image (h) Structure of moved image (i) Overlap of moved image and fixed image(the boundary denotes the structure of moved image, the area denotes the structre of fixed image)

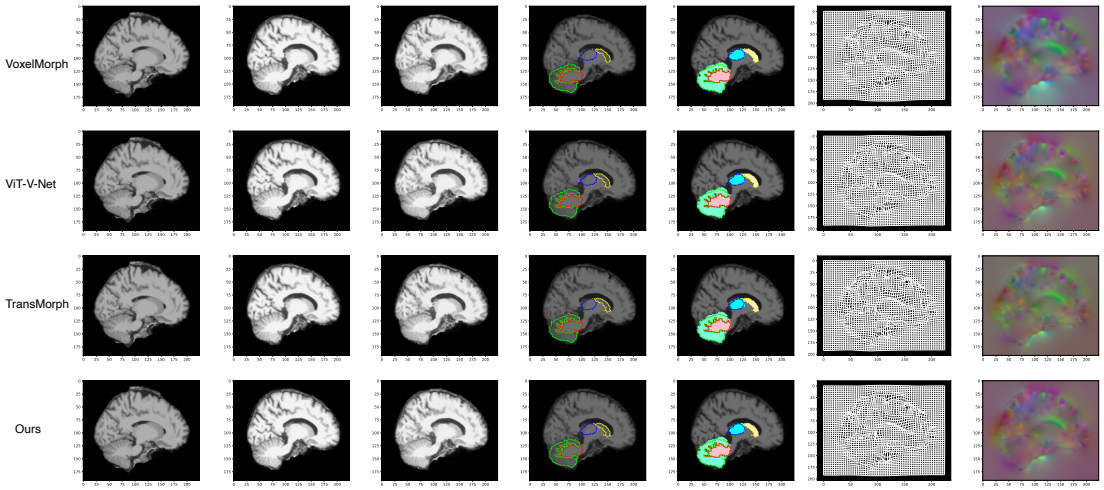

Figure 7. An example of comparison experiment. Column 1 is the moving image m. Column 2 is the fixed image f. Column 3 is the moved/warped image m'. Column 4 is the segmentation of each represented structure in the moved image. Column 5 shows the different between the fixed image and the moved image. Column 6 is the warp grid of each DVF. Column 7 is the RGB representation of the DVFs.

Table 1. The registration result of IXI dataset in different models

model	memory	parameter	dice	% of $\lvert J_\phi \rvert \leq 0$
Affine(before registration)	N/A	N/A	0.5250±0.055	N/A
Voxelmorph	14125MB	0.31M	0.6760±0.035	0.65 ± 0.28
ViT-V-Net	11969MB	30.10M	0.6816±0.036	0.89 ± 0.27
Transmorph-Small	13650MB	11.21M	0.6816±0.037	1.02 ± 0.29
Ours	12052MB	32.58M	**0.6855±0.048**	0.95 ± 0.27

transformed m to m'. After that, we exchange x and y for f and m respectively, and do the same thing. In each experiment, there are 500 epoches to be executed.

To calculate the dice, we try to register the test set pairwisely. But due to the large amount of results, we draw the following chart from part of them. The data in the table below contains the result of registration pairwisely. To represent the practicality of our model, we have chosen anatomical structures from part of brain, considering various features of different networks.

The results in Table1 and Table2 show that, in the same scale of memory and parameter cost, our work outstands and beat other methods by a margin, especially in some anatomical structures such as Brain stem and VentralDC.

The loss function contains $\mathscr{L}_{similarity}$ and $\mathscr{L}_{smooth}$. In our experiment, the weight of the two parts is set as

$$\mathscr{L} = \mathscr{L}_{similarity} + \beta \mathscr{L}_{smooth}$$

According to previous work like VoxelMorph, β is set to 0.02[6]. And it is confirmed that 0.02 is of proper level. The λ parameter in the loss function $\mathscr{L}_{similarity}$ is set to 20, according to the distribution of absolute gradient value shown in Fig. 4 and our fintuning process. The change and convergence of loss function in each experiment are shown in Fig. 8. We trained our method and other methods under the same loss function. Our method converges faster than other methods and gets a better

Table 2. The registration result of OASIS dataset in different models

model	memory	dice	% of $\lvert J_\phi \rvert \leq 0$
Affine(before registration)	N/A	0.583±0.067	N/A
Voxelmorph	13973MB	0.772±0.029	0.58±0.22
ViT-V-Net	10895MB	0.789±0.030	0.87±0.24
Transmorph-Small	13648MB	0.794±0.031	1.33±0.31
Ours	10953MB	**0.797±0.042**	0.86±0.24

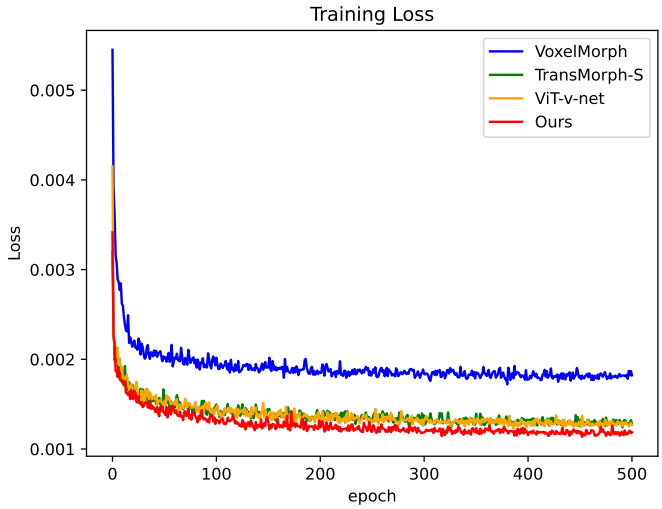

Figure 8. The loss function during training($\lambda = 20$)

result.

Conclusion

In this paper, we proposed a 3D hybrid encoder structure to improve the performance of unsupervised brain MRI registration. The mixed part is in the bottleneck of our network, so the addition of parameter and memory are limited, which is resource-friendly. This work helps surgeons to look into the growth of tissue and disease and make more precise diagnosis and treatment plans. As shown in the tables, we beat TransMorph-S, which needs similar memory and executing time, and achieved 79.7% in dice score on OASIS dataset, 68.6% on IXI dataset.

Beside the network structure, we also mentioned a novel loss function to extract more information from the MRI image, which has a great potential in medical image processing.

Considering the hybrid structure, our further research will focus on multi-organ registration. Different encoders are good at handling different input data, so a single encoder may not have a general practical perceptions. A hybrid one can improve the robustness and precision of registration for different types of images.

However, the training period is a little long for 3D images. We are persuing a way to significantly reduce the training time and registration time. Still, the memory cost of state-of-the-art registration approaches are high, to run it in parallel needs more memory and computational resources, which is a problem to solve.

Data Availability

All the brain MRI scans used in this article were obtained from publicly available datasets, which are included in this published article "TransMorph: Transformer for Unsupervised Medical Image Registration"[7]. The referred article processed the source data and made the converted datasets available.

The source of IXI dataset: https://brain-development.org/ixi-dataset/

The source of OASIS dataset: https://www.oasis-brains.org/

References

1. 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*, 10600–10610 (2019).

2. Zhao, S. *et al.* Unsupervised 3d end-to-end medical image registration with volume tweening network. *IEEE journal biomedical health informatics* **24**, 1394–1404 (2019).

3. Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis. (IJCV)* **115**, 211–252, DOI: 10.1007/s11263-015-0816-y (2015).

4. Ranftl, R., Bochkovskiy, A. & Koltun, V. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12179–12188 (2021).

5. Zhou, F., Luo, F., Kong, R., Chen, Y.-P. P. & Feng, L. Mmar-net: a multi-stride and multi-resolution affine registration network for ct images. In *The Asia Pacific Bioinformatics Conference (APBC)* (2023).

6. Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J. & Dalca, A. V. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging* **38**, 1788–1800 (2019).

7. Chen, J. *et al.* Transmorph: Transformer for unsupervised medical image registration. *Med. image analysis* **82**, 102615 (2022).

8. Avants, B. B., Tustison, N., Song, G. *et al.* Advanced normalization tools (ants). *Insight j* **2**, 1–35 (2009).

9. Vaswani, A. *et al.* Attention is all you need. *Adv. neural information processing systems* **30** (2017).

10. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *science* **313**, 504–507 (2006).

11. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging* **29**, 196–205 (2009).

12. Chee, E. & Wu, Z. Airnet: Self-supervised affine registration for 3d medical images using neural networks. *arXiv preprint arXiv:1810.02583* (2018).

13. De Vos, B. D. *et al.* A deep learning framework for unsupervised affine and deformable image registration. *Med. image analysis* **52**, 128–143 (2019).

14. Beg, M. F., Miller, M. I., Trouvé, A. & Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. journal computer vision* **61**, 139–157 (2005).

15. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. image analysis* **12**, 26–41 (2008).

16. Cao, X. *et al.* Deformable image registration using a cue-aware deep regression network. *IEEE Transactions on Biomed. Eng.* **65**, 1900–1911 (2018).

17. Sokooti, H. *et al.* Nonrigid image registration using multi-scale 3d convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, 232–239 (Springer, 2017).

18. Kim, B. *et al.* Unsupervised deformable image registration using cycle-consistent cnn. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, 166–174 (Springer, 2019).

19. Yan, P., Xu, S., Rastinehad, A. R. & Wood, B. J. Adversarial image registration with application for mr and trus image fusion. In *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9*, 197–204 (Springer, 2018).

20. Tran, M. Q. *et al.* Light-weight deformable registration using adversarial learning with distilling knowledge. *IEEE transactions on medical imaging* **41**, 1443–1453 (2022).

21. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241 (Springer, 2015).

22. Ferrante, E., Oktay, O., Glocker, B. & Milone, D. H. On the adaptability of unsupervised cnn-based deformable image registration to unseen image domains. In *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9*, 294–302 (Springer, 2018).

23. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

24. Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022 (2021).

25. Chen, J., He, Y., Frey, E. C., Li, Y. & Du, Y. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. *arXiv preprint arXiv:2104.06468* (2021).

26. Zhu, Y. & Lu, S. Swin-voxelmorph: A symmetric unsupervised learning model for deformable medical image registration using swin transformer. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, 78–87 (Springer, 2022).

27. Shi, J. *et al.* Xmorpher: Full transformer for deformable medical image registration via cross attention. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, 217–226 (Springer, 2022).

28. Jaderberg, M., Simonyan, K., Zisserman, A. *et al.* Spatial transformer networks. *Adv. neural information processing systems* **28** (2015).

29. Marcus, D. S. *et al.* Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *J. cognitive neuroscience* **19**, 1498–1507 (2007).

30. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302 (1945).

31. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).

Acknowledgements

This research has been supported by the National Natural Science Foundation of China (62172309).

Author contributions statement

J.W. conceived this paper and proposed the main idea. S.S. conducted the experiment(s), F.L., F.Z. and G.T. analysed the results and gave advice. All authors reviewed the manuscript.

Additional information

The code is accessible in https://github.com/wjy-yy/Hybrid-net.