Functional magnetic resonance imaging progressive deformable registration based on a cascaded convolutional neural network

Qiaoyun Zhu[1,2,3], Guoye Lin[1,2,3], Yuhang Sun[1,2,3], Yi Wu[1,3], Yujia Zhou[1,2,3], Qianjin Feng[1,2,3]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou, China; [2]Guangdong Provincial Key Laboratory of Medical Image Processing, Guangzhou, China; [3]Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University, Guangzhou, China

Correspondence to: Yujia Zhou, PhD; Prof. Qianjin Feng. School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China. Email: zyj.shmily08@gmail.com; 1271992826@qq.com.

Background: Intersubject registration of functional magnetic resonance imaging (fMRI) is necessary for group analysis. Accurate image registration can significantly improve the results of statistical analysis. Traditional methods are achieved by using high-resolution structural images or manually extracting functional information. However, structural alignment does not necessarily lead to functional alignment, and manually extracting functional features is complicated and time-consuming. Recent studies have shown that deep learning-based methods can be used for deformable image registration.

Methods: We proposed a deep learning framework with a three-cascaded multi-resolution network (MR-Net) to achieve deformable image registration. MR-Net separately extracts the features of moving and fixed images via a two-stream path, predicts a sub-deformation field, and is cascaded three times. The moving and fixed images' deformation field is composed of all sub-deformation fields predicted by the MR-Net. We imposed large smoothness constraints on all sub-deformation fields to ensure their smoothness. Our proposed architecture can complete the progressive registration process to ensure the topology of the deformation field.

Results: We implemented our method on the 1000 Functional Connectomes Project (FCP) and Eyes Open Eyes Closed fMRI datasets. Our method increased the peak t values in six brain functional networks to 19.8, 17.8, 15.0, 16.4, 17.0, and 13.2. Compared with traditional methods [i.e., FMRIB Software Library (FSL) and Statistical Parametric Mapping (SPM)] and deep learning networks [i.e., VoxelMorph (VM) and Volume Tweening Network (VTN)], our method improved 47.58%, 11.88%, 18.60%, and 15.16%, respectively.

Conclusions: Our three-cascaded MR-Net can achieve statistically significant improvement in functional consistency across subjects.

Keywords: Functional magnetic resonance imaging (fMRI); deformable image registration; Multi-resolution network (MR-Net); cascaded network

Introduction

Functional magnetic resonance imaging (fMRI) measures brain activity by detecting changes associated with blood flow. The primary imaging of fMRI uses blood-oxygen-level-dependent (BOLD) contrast. Resting-state fMRI (rs-fMRI or R-fMRI) is an fMRI method that measures brain activity by detecting associated blood flow changes. Task-based fMRI measures BOLD signal changes between task-stimulated and control states. FMRI has permeated numerous aspects of neuroscience and is widely used to investigate the brain's structure and functions (1-3). Studies have commonly analyzed fMRI data across a population

of subjects (4-6). When performing group-level statistical analysis and making a population inference, each voxel is assumed to be located in the same anatomical region for all subjects. Therefore, an accurate intersubject registration method, which aligns the group fMRI data to an atlas (7) [e.g., MNI152 atlas (8)], can significantly improve the reliability of a group-level analysis.

Typically, an intersubject registration of fMRI is often achieved with the help of the corresponding structural MRI (sMRI), owing to its high spatial resolution and rich texture information. In particular, deformable image registration establishes the nonlinear spatial correspondence between sMRI and an atlas. After that, the nonlinear spatial correspondence is applied to its corresponding fMRI. The common techniques are Talairach normalization (9), FMRIB's Linear Image Registration Tool (FLIRT) (10), and advanced non-rigid registration methods (11,12). Considerably powerful techniques are guided by the sulcus/gyrus landmarks or curvature maps to align cortical neuroanatomy (13,14).

However, structural alignment does not necessarily lead to functional alignment (15) because functional regions are not consistently located relative to anatomical landmarks (7,16). Some areas (e.g., visual motion area) can vary by more than 2 cm between individuals (17). Accordingly, deformable image registrations based on functional information were later developed to improve brain functions' consistency across subjects (16,18-22). Sabuncu *et al.* (16) used the fMRI time series to index the local functional response profile. After that, the registration of task-related fMRI data was achieved by maximizing each cortical node time series's correspondence. However, this method was based on the assumption that functional signals are synchronized between different subjects. Research has shown that no evident correlation exists between rs-fMRI scanned at different times, even if the stationary subject was in the same position (18). To overcome this limitation, Langs *et al.* (22) proposed a method to achieve image registration by maximizing the similarity of the functional connectivity (FC) patterns at the same spatial position across different subjects.

The FC pattern between the two regions was measured using the correlation coefficient between their functional signals. Conroy *et al.* (18) utilized the whole-brain FC matrix as a descriptor of cortical surface functional information and registered subjects by minimizing the Frobenius norm of the difference between their FC matrices. Jiang *et al.* (20) proposed a functional information

description based on the local FC pattern because the whole-brain FC matrix is sensitive to local interference. The local FC pattern is computed in a small spatial local neighborhood of each voxel to describe the fMRI functional information. Furthermore, Jiang *et al.* (19) proposed a multi-region FC mode. They used a local FC mode to hierarchically guide the registration and gradually increased the size of the local area to capture the FC information on a large scale. However, registration methods based on functional information also have certain limitations.

The preceding traditional methods solve the registration optimization problems by manually extracting and aligning structural/functional features. However, manually extracting the structural/functional features is complicated and time-consuming. Several registration methods (23-25) based on deep convolutional neural networks (CNNs) have emerged in recent years, in which robust structural/ functional corresponding information was automatically extracted, thereby substantially reducing computation time. Chee *et al.* (23) proposed an affine image registration network (AIRNet), which uses the mean square error (MSE) between the predicted and ground truth affine transforms as a loss to train the network. Yang *et al.* (24) predicted the two-dimensional/three-dimensional (2D/3D) deformation field of intersubject brain MR volumes with a UNet-like (25) architecture. They trained the architecture using the ground truth obtained by numerical optimization of the large deformation diffeomorphic metric mapping (26) registration model. However, the ground truth affine matrix/ deformation field was difficult to obtain. Uzunova *et al.* (27) used statistical appearance models to generate ground truth data to solve this problem. However, the simulated data must be sufficiently similar to the clinical data. This challenge motivated several groups to explore unsupervised deep learning registration methods (28-31) with the help of the spatial transformer network (STN) (32). Balakrishnan *et al.* (28,29) proposed a general framework for unsupervised deformable image registration. Kuang *et al.* (30) used a framework inspired by CNN and STN (32) to perform deformable registration of T1-weighted brain MRI volumes. The aforementioned deep learning registration methods have achieved good results in sMRI registration.

However, these methods have consistently focused on improving the similarity between the warped and fixed images while disregarding the reasonability (with less "folding" issue) of the deformation field. To address the "folding" issue of the deformation field, previous studies have used a straightforward solution that adds a smoothness
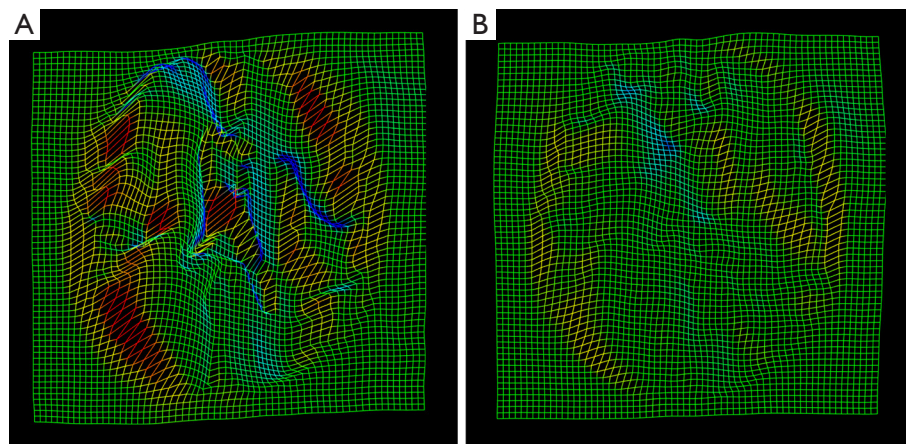
Figure 1 Illustration of two deformation fields using (A) a small smoothness constraint (0.01) and (B) a large smoothness constraint (0.5), respectively.

constraint to penalize the folding area. Although smoothness constraint can avoid the "folding" issue, the registration accuracy may be reduced by 15% or more (30,31). Current deep learning methods set small smoothness constraints to achieve high image similarity; however, the cost is considerably more "folding" and high complexity of the deformation field (*Figure 1A*). Setting a large smoothness constraint will limit the freedom degree of the deformation field (*Figure 1B*), leading to poor image similarity.

To overcome fMRI registration shortcomings based on structural/functional features, we trained a deep learning network directly based on fMRI data. This deep learning method can automatically extract features from the fMRI data, making the registration process faster. We propose a cascaded CNN, namely, a three-cascaded multi-resolution network (MR-Net), to obtain high registration accuracy and ensure the deformation fields' geometrical properties. Moreover, we set a large smoothness constraint (e.g., 0.5) to the deformation field predicted by MR-Net to penalize the unreasonable deformation field. A single MR-Net with large smoothness constraints limits the deformation field's freedom degree, thereby resulting in poor image similarity. Thus, we propose a cascade strategy to successively warp the moving image (I_{moving}) to achieve good registration accuracy. First, we used MR-Net (33) as our subnetwork to extract the robust features of the moving and fixed images (I_{fixed}) via a two-stream path. The MR-Net can relatively ensure that points in I_{moving} can find their matching points in I_{fixed} when a large smoothness constraint is added to the deformation field. Second, we cascaded the MR-Net three times, each taking the current warped and fixed images as

inputs and learning the network parameters of different sub-deformation fields. For example, the first MR-Net takes I_{moving} and I_{fixed} as inputs, and outputs the sub-deformation field $\phi^{(1)}$. After that, I_{moving} can be warped to the warped moving image I_{warped_1} using $\phi^{(1)}$, however, the registration accuracy is low due to a large smoothness constraint. The second MR-Net takes I_{warped_1} and I_{fixed} as inputs, and outputs $\phi^{(2)}$. Thereafter, I_{warped_1} can be warped to I_{warped_2}. By repeating this procedure, I_{moving} can perform a part of the registration in each MR-Net until it finally aligns to the fixed image. The final deformation field can be regarded as the combination of all sub-deformation fields predicted by MR-Net. We imposed large smoothness constraints on all sub-deformation fields to ensure their topology. Lastly, we optimized our three-cascaded MR-Net parameters by measuring the similarity between the warped image of each MR-Net and I_{fixed}. The main contributions of this study are summarized as follows:

(I)   We performed a deep learning deformable image registration method directly based on the fMRI data to address the limitations of methods based on sMRI or functional features;

(II) To obtain a diffeomorphism deformation field, we added large smoothness constraints to the MR-Net. To mitigate poor image similarity caused by large constraints, we propose a cascade strategy to register the moving image to the fixed image progressively.

Methods

Let $I_{moving} \in R^3$ and $I_{fixed} \in R^3$ denote the moving and fixed
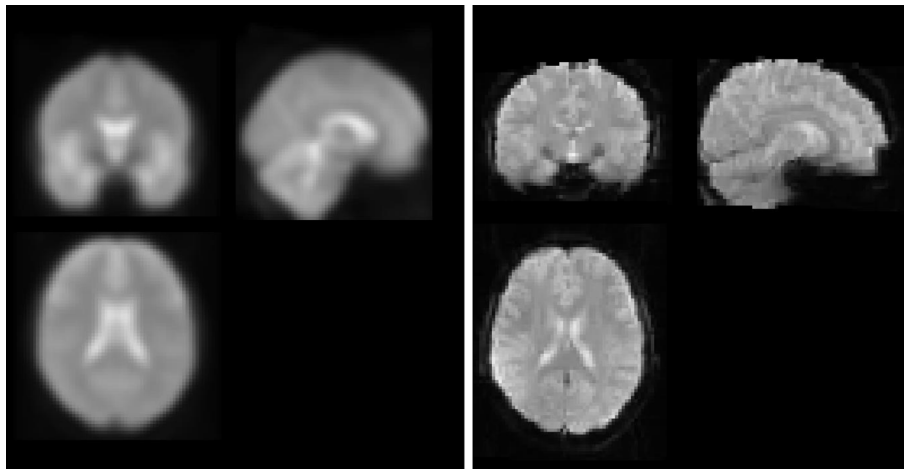
Figure 2 EPI (left) and fMRI (right) spaces. EPI, echo planar imaging; fMRI, functional magnetic resonance imaging.

images, respectively. Deformable image registration aims to find the optimal deformation field ϕ between I_{moving} and I_{fixed} to minimize the energy:

$$\phi = \underset{\phi}{\mathrm{argmin}}\, D\left(I_{moving} \circ \phi, I_{fixed}\right) + \mathrm{R}\left(\phi\right) \qquad [1]$$

where $D(.)$ denotes the similarity term, and $\mathrm{R}(\phi)$ is a regularization term.

Subnetwork

This section describes the particular subnetwork architecture used in our experiments. The traditional deep learning registration network architecture (29,34) simultaneously extracts the moving and fixed images' features by concatenating them. However, the features should be separately extracted between the atlas echo planar imaging (EPI) and fMRI (*Figure 2*) to learn the moving and fixed images (35). Accordingly, we selected a well-designed framework MR-Net (33), as our subnetwork. The MR-Net is a CNN architecture similar to VoxelMorph (VM) (29) with Pyramidal Residual Deformation Field Estimation (PRDFE)-Module. We listed the MR-Net reference (33) to provide further architectural details. The MR-Net constructs two dual-stream network channel paths to extract multiple resolution features of moving and fixed images, similar to a pyramid feature set. Also, the MR-Net uses a feature-warping model to estimate the "residual" deformation fields for each resolution scale. The final output deformation field that warps I_{moving} to I_{fixed} is obtained by weighing the "residual" deformation fields of

all scales. MR-Net can effectively and accurately convert the deformation field from low to high resolution, thereby ensuring that the receptive field of the convolution kernel on the high-resolution scale can cover the corresponding points in I_{moving} and I_{fixed}.

Cascade strategy

The MR-Net can perform a voxel-level 3D medical deformable image registration by using an end-to-end CNN. However, the MR-Net may continue to predict an unreasonable deformation field when the smoothness constraint is small. When we set a large smoothness constraint, the deformation field's freedom degree is limited, thereby resulting in poor image similarity. Therefore, we propose an effective method by cascading the MR-Net for progressive alignment. In each cascade, the MR-Net can predict a sub-deformation field $\phi^{(i)}$. Then, I_{moving} can be warped progressively to I_{warped_i} as follows:

$$I_{warped\_i} = I_{moving} \circ \phi^{(1)} \circ \cdots \circ \phi^{(i)} \qquad [2]$$

Each cascade takes the current warped image I_{warped_i} and fixed images I_{fixed} as inputs, and outputs a sub-deformation field $\phi^{(i)}$. With this cascaded subnetwork strategy, the final predicted deformation field ϕ_{final} can be constructed by the following sub-deformation fields:

$$\phi_{final} = \phi^{(1)} \circ \cdots \circ \phi^{(i)} \circ \cdots \circ \phi^{(n)} \qquad [3]$$

Moreover, the moving image is warped to the final
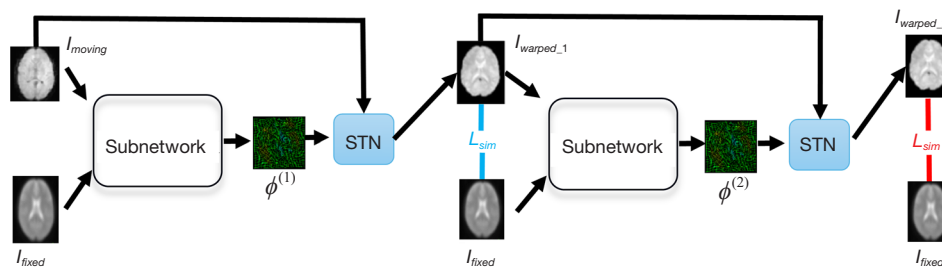
Figure 3 A two-cascaded subnetwork structure. STN, spatial transformer network.

warped image as follows:

$$I_{warped\_final} = I_{moving} \circ \phi_{final} \qquad [4]$$

In each cascade network, we added a large smoothness constraint to each sub-deformation field to maintain its topology. Meanwhile, each cascade can warp I_{moving} progressively to I_{fixed} to achieve good registration performance.

Figure 3 shows a two-cascaded subnetwork structure. The first subnetwork takes I_{moving} and I_{fixed} as inputs and outputs the sub-deformation field $\phi^{(1)}$. After that, the I_{moving} is warped to I_{warped_1} through $\phi^{(1)}$, and the I_{warped_1} is the input of the second subnetwork for further prediction. The second subnetwork predicts $\phi^{(2)}$ to warp I_{warped_1}. Thus, I_{warped_2} is obtained.

Loss function

The final deformation field ϕ_{final} is a composition of all sub-deformation fields $\phi^{(i)}$ predicted by the MR-Net. We added relatively large smoothness constraints to all predicted sub-deformation fields $\phi^{(i)}$ to maintain their topology. A smoothness penalty in the form of l_2-norm was used as the smoothness constraint in our study. To ensure that all I_{warped_i} was similar to I_{fixed}, we added a similarity loss MSE between I_{warped_i} and I_{fixed}.

In our n-cascaded MR-Net, the loss function can be defined as follows:

$$L_{\text{n-cascaded MR-Net}} = \sum_n MSE\left(I_{warped\_i}, I_{fixed}\right) + \lambda_i \sum \left\|\nabla \phi^{(i)}\right\|^2 \quad [5]$$

where λ_i is the regularization parameter corresponding to each sub-deformation field $\phi^{(i)}$.

Dataset

Our experiment used a publicly available 1000 Functional

Connectomes Project (FCP) dataset (http://fcon_1000. projects.nitrc.org). Each subject folder contains a functional image and a T1 structural image. We excluded missing and selected data with a resolution of 3±1 mm through preliminary selection. Eventually, we selected 515 subjects as our dataset, 90% of which were used as the training dataset and 10% of which were used as the testing dataset. The registration of the rs-fMRI data could also be validated based on task-based fMRI data (19). To further evaluate our method, we added an eyes open eyes closed (EOEC) fMRI dataset (http://fcon_1000.projects.nitrc.org/indi/IndiPro. html?tdsourcetag=s_pcqq_aiomsg), which is a task-based fMRI dataset collected by the Beijing Normal University in China, to our testing dataset. This dataset consists of 48 subjects, each with three scans [the first scan participants were instructed to rest with their eyes closed; while the second and third scans were between eyes open (EO) and eyes closed (EC)].

We used the rs-fMRI data processing software package [DPARSF (rfmri.org/DPARSF)] in MATLAB to preprocess all data according to the conventional fMRI processing procedure, including deleting the first 10 time points, slice time correction, head motion correction, and 0.01–0.1 Hz band-pass filtering. Our experiment focused only on the nonlinear transformation part of the registration. Therefore, we used the FLIRT algorithm (13) in FMRIB Software Library (FSL) (36) to linearly register all rs-fMRI images to the MNI space's EPI template. In our experiment, the EPI template is regarded as I_{fixed}. Subsequently, we calculated the average of the rs-fMRI image of each subject at all time points and represented the average image as I_{moving}. All moving images are resampled to 64×64×64 with a 3-mm isotropic voxel size.

Implementation

This study used Keras (https://github.com/fchollet/keras)

with a TensorFlow backend on the NVIDIA GeFORCE RTX 2080 GPU. Our three-cascaded MR-Net was trained for 1,000 epochs with 500 iterations per epoch by using a batch size of four. The optimizer is Adam (37) with a learning rate of $1e^{-4}$. We set the regularization parameter λ of the n-cascade network to a series of values to find the optimal parameter.

We compared our algorithm with the commonly used rs-fMRI traditional registration algorithms, including FSL5.0.8 (36) and SPM12 (https://www.fil.ion.ucl.ac.uk/spm), and the unsupervised deep learning methods VM (29) and Volume Tweening Network (VTN) (38). In particular, we used FLIRT and FMRIB's Nonlinear Image Registration Tool (FNIRT) for registration. In this experiment, we used a 12-degree-of-freedom registration for FLIRT and the FSL standard configuration for FNIRT. The registration was performed using a sum of squared differences (SSD) as the cost function. Trilinear interpolation was used to obtain an improved result. Lastly, registration was performed until the required deformation was achieved [constraint of the minimum acceptable Jacobian value of the deformation (default 0.01)]. We used the SPM_EPI algorithm in the SPM12 toolbox, which involves an affine transform followed by a nonlinear registration of the fMRI image to an EPI atlas. The cost function was the SSD between the atlas and moving image. The number of iterations of nonlinear registration was set to a default value of 16. The interpolation method was 4th degree B-spline interpolation. After that, the warped images were smoothed [6-mm full width at half maximum (FWHM)] for statistical analysis.

Evaluation

Rs-fMRI

The current evaluation of the rs-fMRI registration performance was based on the resting-state brain functional networks' group-level statistical maps. We measured the consistency of the functional brain networks identified by independent component analysis [ICA (21)] between subjects, which was implemented by Group ICA [GIFT (http://icatb.sourceforge.net/groupica.htm)], to evaluate the registration results across subjects. In particular, ICASSO (39) was used to perform 100 times on ICA with different initial settings on the testing dataset, and 20 independent components were generated. After group ICA, back reconstruction was used to restore the individual parts of each subject. We identified and selected six brain networks by calculating the correlation between 20 components and

the corresponding network template (40). These networks included the default mode network (DMN), visual network (VN), central execution network (CEN), sensorimotor network (SMN), right memory network (RMN), and parts of the visual cortex network (VCN).

After the six brain functional networks were identified, we used the following indicators to evaluate the performance of the algorithm:

(I) We performed a one-sample *t*-test on each brain functional network across all subjects to generate a group-level t-map of each functional brain network. A significant, consistent voxel activated in all subjects would appear with a high t-value in the generated t-map. Therefore, a high peak t-value of the t-map indicated a high functional consistency, signifying improved alignment.

(II) We further evaluated the functional consistency of several major components in DMN. We selected the posterior cingulate cortex (PCC), precuneus, Angular_R, and Angular_L covered by the main functional nodes in the DMN. We set a specific threshold to the t-map and calculated the number of voxels that exceeded the statistical threshold. More suprathreshold voxels of the t-map could be detected after using a registration method if the registration method was better.

(III) The intersubject functional network correlation was established. Each subject's six individual functional networks could be obtained after group ICA using back reconstruction. Pearson's correlation coefficient was used to measure the correlation between a specific network of a pair of testing data to assess the specific functional network's alignment performance among individuals in the group. The value of the correlation ranged from 0 to 1. So, if the alignment of the functional brain network across subjects was improved, then the correlation value will approximate 1. We used bar plots with error bars to represent the correlation between the specific brain networks across subjects.

(IV) We further measured the overlap percentage between each subject-specific and group brain network. In particular, we transformed group-level t-map to z-map. A set of binary images of the subject-specific and group brain networks could be generated using a certain threshold. The overlap between each subject-specific and group binary images was computed using the Dice
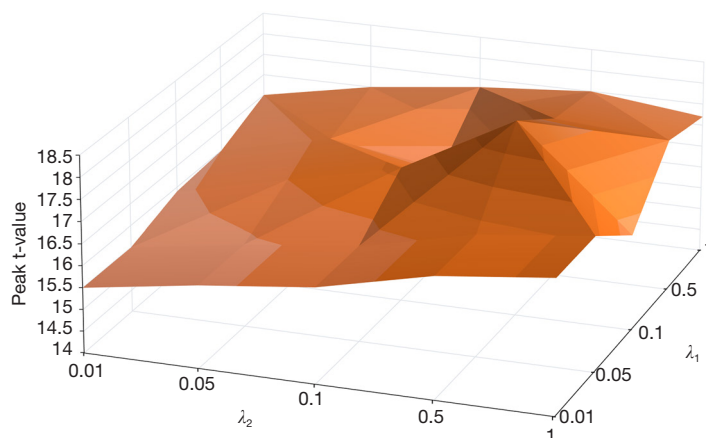
Figure 4 Peak t-value of DMN when we set a series of combinations of regularization parameters λ_1 and λ_2 in the two-cascaded MR-Net. DMN, default mode network; MR-Net, multi-resolution network.

score. A larger overlap percentage indicated better functional consistency across different subjects and a better registration method.

Task-based fMRI

To assess the functionally homologous regions' alignment, we utilized two widely used fMRI metrics for the EOEC dataset. These metrics have been widely applied to detect alterations in brain activities (41,42).

(I)   The amplitude of low-frequency fluctuation [ALFF (43)]. ALFF detects the regional intensity of fluctuations in the BOLD signal, thereby reflecting specific regions' neural activities. We performed a paired *t*-test to detect voxel-level differences in the ALFF maps between EC and EO. A larger difference between EC and EO results in a higher t-value. The differences between EO and EC have been consistently reported in previous studies (44,45). In cases without registration, only a small area showed higher fluctuation in EO than in EC. However, after registration, more areas showed significantly higher EO fluctuation than in EC in the visual cortex, resulting in a larger difference. Therefore, a higher peak t-value in the paired-t map indicated better registration performance.

(II) Regional homogeneity [ReHo (42)]. ReHo can evaluate the similarity between the time series of a given voxel and its nearest neighbors, rapidly mapping the level of regional activities across the whole brain. Our experiment used a cluster size of 27 voxels to include all neighboring voxels adjacent

to a given voxel. We also used a voxel-wise paired *t*-test to reveal the differences between EO and EC. A higher peak t-value in the paired-t map of ReHo indicated improved brain functional network registration performance across subjects.
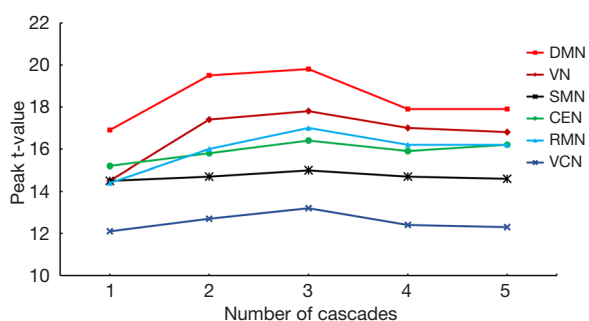
Results

Number of cascades

We gradually increased the number of our cascades to determine the optimal performance. Every parameter regularization item λ_i of each cascaded subnetwork in the experiment was carefully modulated manually. *Figure 4* shows the peak t-value of DMN when we set a series of combinations of regularization parameters λ_1 and λ_2 in the two-cascaded MR-Net. A high peak t-value was obtained when $\lambda_1 = 0.1$ and $\lambda_2 = 0.5$. Therefore, we chose $\lambda_1 = 0.1$ and $\lambda_2 = 0.5$ to train the two-cascaded MR-Net. When training an n-cascaded MR-Net, we set a series of regularization values and combined them in different ways to search for the optimal parameter combination. *Table 1* shows the peak t-value of the n-cascaded MR-Net of the six networks. Also, *Table 1* shows that our cascaded network improved the performance of a limited number of cascades. The number of cascades increases and does not constantly obtain a positive performance improvement. The shallower cascaded network benefits more from this cascade strategy than the deeper cascaded network because the image remains poorly registered. Our experimental results showed that the three-cascaded MR-Net achieved the best results.

**Table 1** Peak t-value of DMN, VN, CEN, SMN, RMN, and VCN of the n-cascaded MR-Net

Number of cascades	DMN	VN	SMN	CEN	RMN	VCN
1	16.9	14.3	14.5	14.9	14.1	12.1
2	19.3	17.4	14.7	15.8	16.0	12.7
3	19.8	17.8	15	16.4	17	13.2
4	17.9	17.0	14.7	15.9	16.2	12.4
5	17.9	16.8	14.6	16.2	16.2	12.3

DMN, default mode network; VN, visual network; SMN, sensorimotor network; CEN, central executive network; RMN, right memory network; VCN, visual cortex network.


Figure 5 Result of the n-cascade MR-Net is plotted to better reflect the trend corresponding to the data in *Table 1*. MR-Net, multi-resolution network; DMN, default mode network; VN, visual network; SMN, sensorimotor network; CEN, central executive network; RMN, right memory network; VCN, visual cortex network.

More than three cascades are likely to affect the composed field's smoothness and deteriorate the registration quality. *Figure 5* plots our results to better illustrate the trend. According to the results of our experiments, we chose a three-cascaded MR-Net for our registration task and regularization parameters λ_1, λ_2, and λ_3 used for each cascade were 0.5, 0.5, and 1, respectively. *Figure 6* shows the sub-deformation field output by each subnetwork after using the three-cascaded MR-Net. Each cascade was allowed to measure a part of the deformation field and could avoid folding. The final deformation field could be decomposed into three sub-deformation fields that can maintain topology by adding reasonable regularization items. The progressive alignment of the moving images can be achieved during testing. The number of cascades causes linear increments to the testing times. *Figure 7* shows the time cost of registering one of the subject fMRI
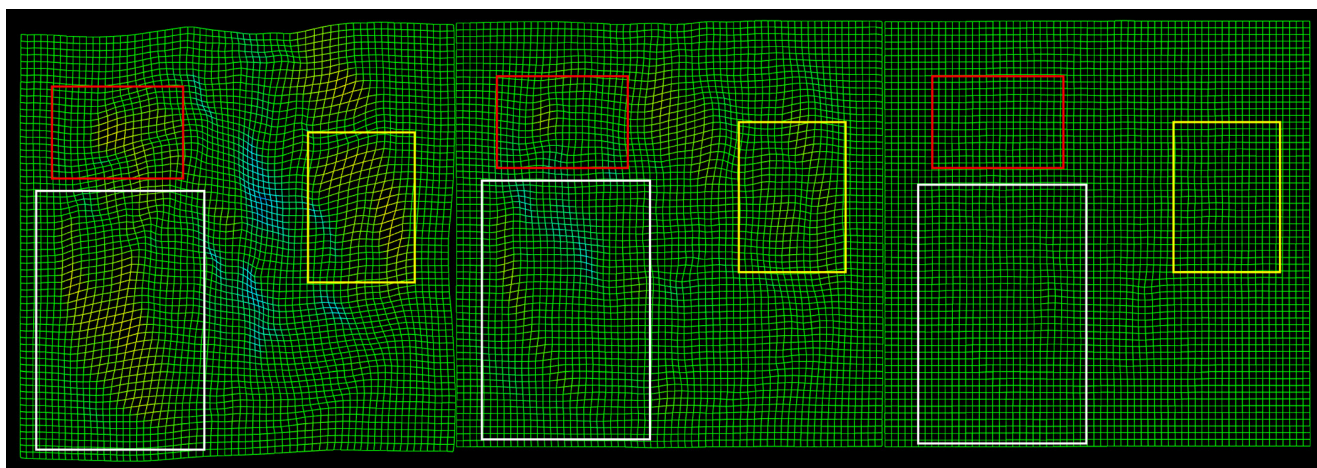

Figure 6 Sub-deformation field output by each subnetwork after the three-cascaded MR-Net. The rectangular areas show that the sub-deformation field predicted by each cascade is allowed to learn a part of the deformation field. As the depth of the network increases, the subnetwork learns less deformation because the moving image is well aligned. MR-Net, multi-resolution network.
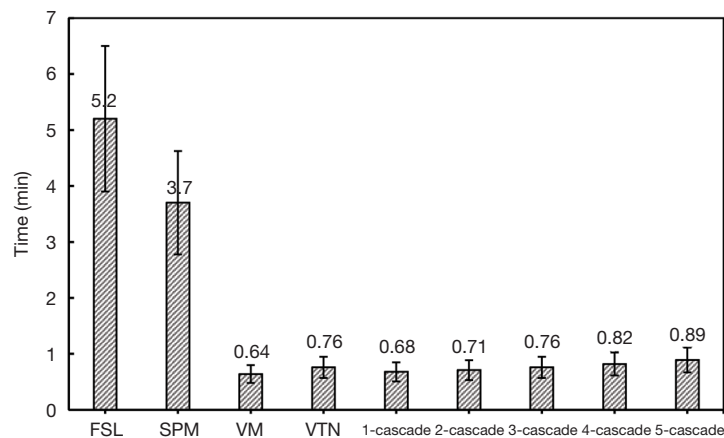
Figure 7 Average calculation time of different registration algorithms (in minutes). FSL, FMRIB Software Library; SPM, Statistical Parametric Mapping; VM, VoxelMorph; VTN, Volume Tweening Network.
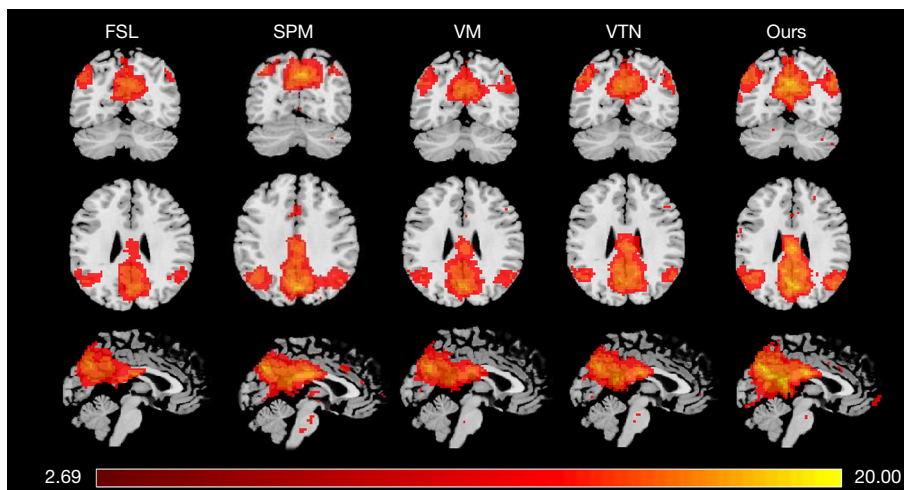
Figure 8 Group-level t maps of DMN with t>2.689 (P<0.01) after registration by FSL, SPM, VM, VTN, and our method. DMN, default mode network; FSL, FMRIB Software Library; SPM, Statistical Parametric Mapping; VM, VoxelMorph; VTN, Volume Tweening Network.

with different registration algorithms.

Evaluation based on rs-fMRI

Figure 8 shows the group-level t-maps of DMN with t>2.689 (P<0.01) by using six different registration algorithms. The t-value of DMN significantly increased using our three-cascaded MR-Net. *Table 2* provides the peak t-values of the group-level t-maps of DMN, VN, CEN, SMN, RMN, and VCN. Our method increased the peak t-values of the six networks to 19.8, 17.8, 15, 16.4, 17, and 13.2. Compared with FSL, SPM, VM, and VTN, our

three-cascaded MR-Net had an improvement of 47.58% {i.e., [(19.8 − 12.4)/12.4 + (17.8 − 9.9)/9.9 + (15 − 10.3)/10.3 + (16.4 − 14.1)/14.1 + (17 − 13.2)/13.2 + (13.2 − 8.5)/8.5]/6 = 47.58%}, 11.88%, 18.60%, and 15.16%, respectively.

We also calculated the suprathreshold voxels of the main components of DMN given three specific thresholds to evaluate our registration algorithm further. *Figure 9* shows the number of suprathreshold voxels using three different thresholds after registration by different methods. The number of suprathreshold voxels in different algorithms gradually decreased with a considerably stringent threshold, whereas they remained highest using our method.

Table 2 Peak t-values of six brain networks with different registration methods

Method	DMN	VN	SMN	CEN	RMN	VCN
FSL	12.4	9.9	10.3	14.1	13.2	8.5
SPM	17.6	14.5	14.6	15.2	14.4	12.3
VM	15.5	13	14.6	14.8	14.1	11.7
VTN	16.4	14.7	14.3	14.9	13.9	11.8
Our method	19.8	17.8	15	16.4	17	13.2

DMN, default mode network; VN, visual network; SMN, sensorimotor network; CEN, central executive network; RMN, right memory network; VCN, visual cortex network; FSL, FMRIB Software Library; SPM, Statistical Parametric Mapping; VM, VoxelMorph; VTN, Volume Tweening Network.
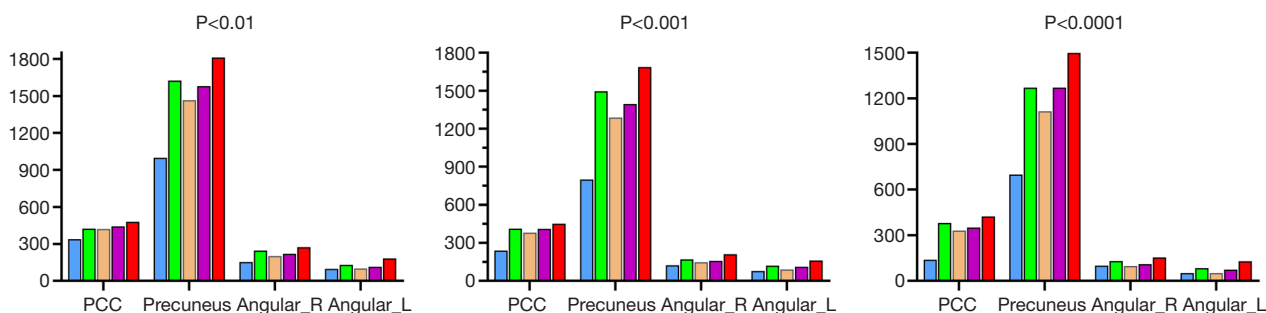
Figure 9 Total number of suprathreshold voxels in PCC, precuneus, Angular_R, and Angular_L after registration via FSL, SPM, VM, VTN, and our method by using different thresholds: t>2.689 (P<0.01), t>3.520 (P<0.001), and t>4.269 (P<0.0001). PCC, posterior cingulate cortex; FSL, FMRIB Software Library; SPM, Statistical Parametric Mapping; VM, VoxelMorph; VTN, Volume Tweening Network.

We calculated the correlations of the six brain networks at the individual level. *Figure 10* depicts the correlations between individuals using a bar plot with an error bar. We conducted a one-way analysis of variance (ANOVA) to demonstrate whether our registration method exhibited a significant improvement. In the one-way ANOVA [95% confidence interval (CI)], the P value was below the 0.01 significance level. After that, we performed a multiple pairwise-comparison to further demonstrate that our method had markedly improved the correlations between individuals in each specific brain functional network.

We computed the overlap percentages of the specific functional network after using different registration methods. We applied different thresholds to evaluate the functional consistency across subjects after using different methods. *Figure 11* illustrates the overlap percentages for different registration methods. The overlap percentages of all methods decreased when given a more stringent

threshold; however, our method remained the highest. One-way ANOVA (95% CI) and multiple pairwise-comparison were applied to our result. As expected, there were statistically significant differences between our method and the others.

Evaluation based on task-fMRI

Paired *t*-tests were performed for the EOEC dataset using the ALFF and ReHo evaluations after registration by FSL, SPM, VM, VTN, and our method. *Table 3* and *Figure 12* show the results using the different registration methods. We could detect higher peak t values of the ALFF and ReHo paired-t maps in some regions in the visual cortex, including the bilateral middle occipital gyrus (P<0.01). Compared with FSL, SPM, VM, and VTN, our three-cascaded MR-Net had an improvement of 83.54% (i.e., [(6.98 – 4.02)/4.02 + (8.57 – 4.43)/4.43]/2 = 83.54%), 23.89%, 55.43%, and 40.59%.
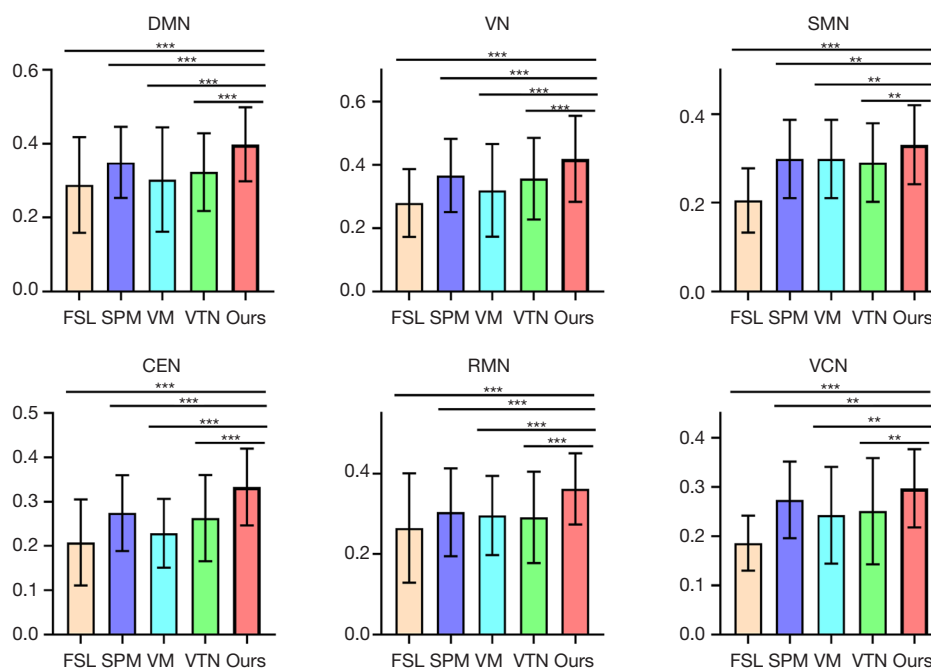
Figure 10 Bar plot with the error bar of the distribution of intersubject functional network correlation for six networks DMN, VN, SMN, CEN, RMN, and VCN. [*** indicates significant improvement via one-way repeated-measures ANOVA (P<0.001); ** indicates significant improvement via one-way repeated-measures ANOVA (P<0.01)]. DMN, default mode network; VN, visual network; SMN, sensorimotor network; CEN, central executive network; RMN, right memory network; VCN, visual cortex network; ANOVA, analysis of variance; FSL, FMRIB Software Library; SPM, Statistical Parametric Mapping; VM, VoxelMorph; VTN, Volume Tweening Network.

Discussion

A common preprocessing challenge associated with group-level fMRI analysis is to register multiple subjects into a standard space. Presently, registration using the EPI template in the MNI space as the fixed image is a common technique used to achieve spatial consistency among multiple subjects. Accurate image registration can significantly improve the statistical analytical results.

This study proposed a deep learning deformable image registration network structure three-cascaded MR-Net to register a group of subjects. We used MR-Net as our subnetwork, as it considers the situation when the receptive field cannot cover the corresponding points in moving and fixed images. In particular, the accuracy of registration depends on its ability to find the corresponding points in moving and fixed images, regardless of the type of deep learning registration methods. A correct displacement can be output only when the convolution kernel's receptive field can cover the corresponding points in the moving and fixed images. For example, if an image is ×16 down-

sampled, a kernel with a size of 3×3 can cover size of 48×48 receptive fields and easily find the corresponding points. At present, unsupervised registration network architectures make a straightforward prediction for the deformation field between intensity images (29,34) based on the UNet structure. However, previous studies (29,34) only warp the moving image in the original spatial resolution. Therefore, a 3×3 convolution kernel may be unable to cover the corresponding points that are distant in the moving and fixed images when aligning images at the original resolution.

The MR-Net can address the aforementioned issue. It trains a pyramidal feature descriptor to extract the moving and fixed features at multi-scales via a two-stream path. After that, the MR-Net constructs a warping module that can warp the moving features at each scale to the fixed features. In this manner, corresponding features with large deformation can be covered by a small kernel size's receptive field at the low-resolution scale. Therefore, the moving image can easily find the fixed image's corresponding points, even with a small kernel size, when the moving image is down-sampled numerous times. Therefore, we used the
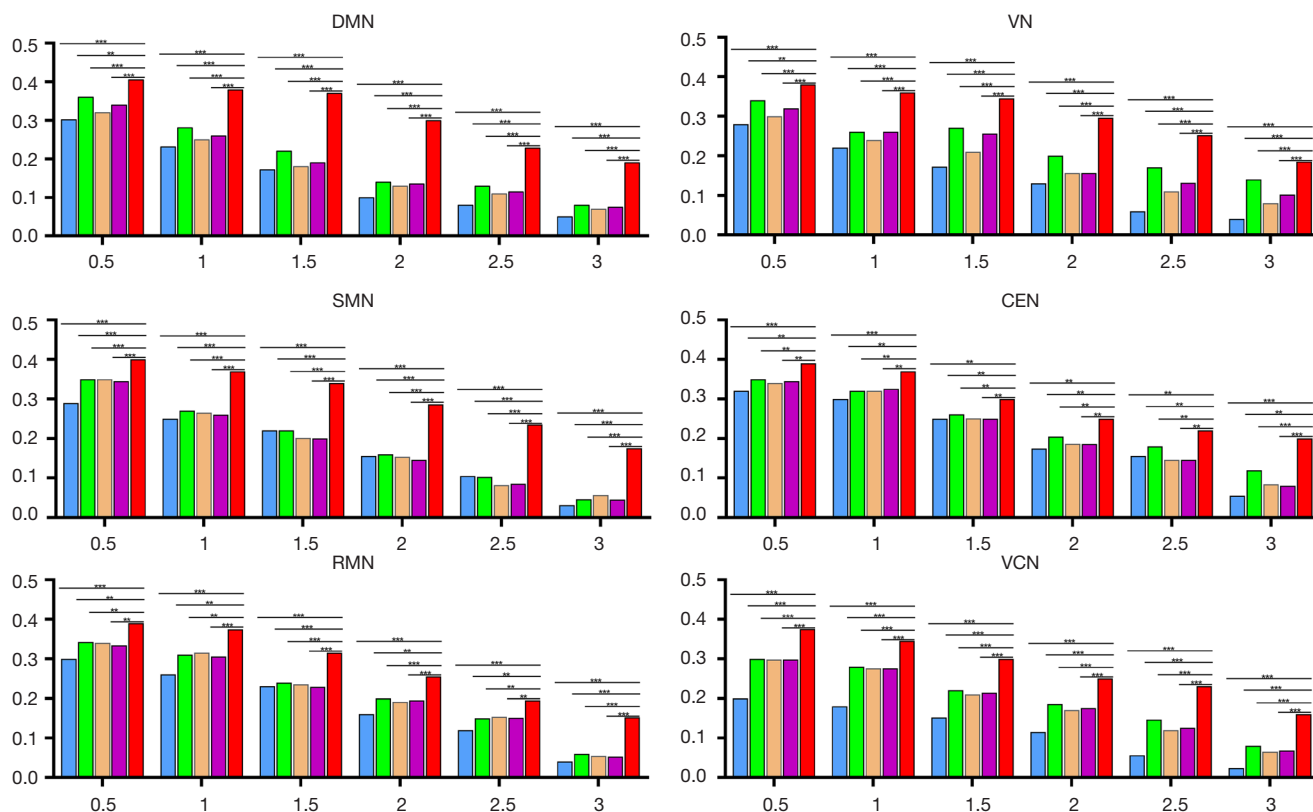
Figure 11 Overlap between each subject-specific and group brain networks with different thresholds after using varying registration methods. [*** indicates significant improvement via one-way repeated-measures ANOVA (P<0.001); ** indicates significant improvement via one-way repeated-measures ANOVA (P<0.01)]. DMN, default mode network; VN, visual network; SMN, sensorimotor network; CEN, central executive network; RMN, right memory network; VCN, visual cortex network; ANOVA, analysis of variance.

Table 3 Peak t-values in the ALFF and ReHo paired-t maps with different registration methods

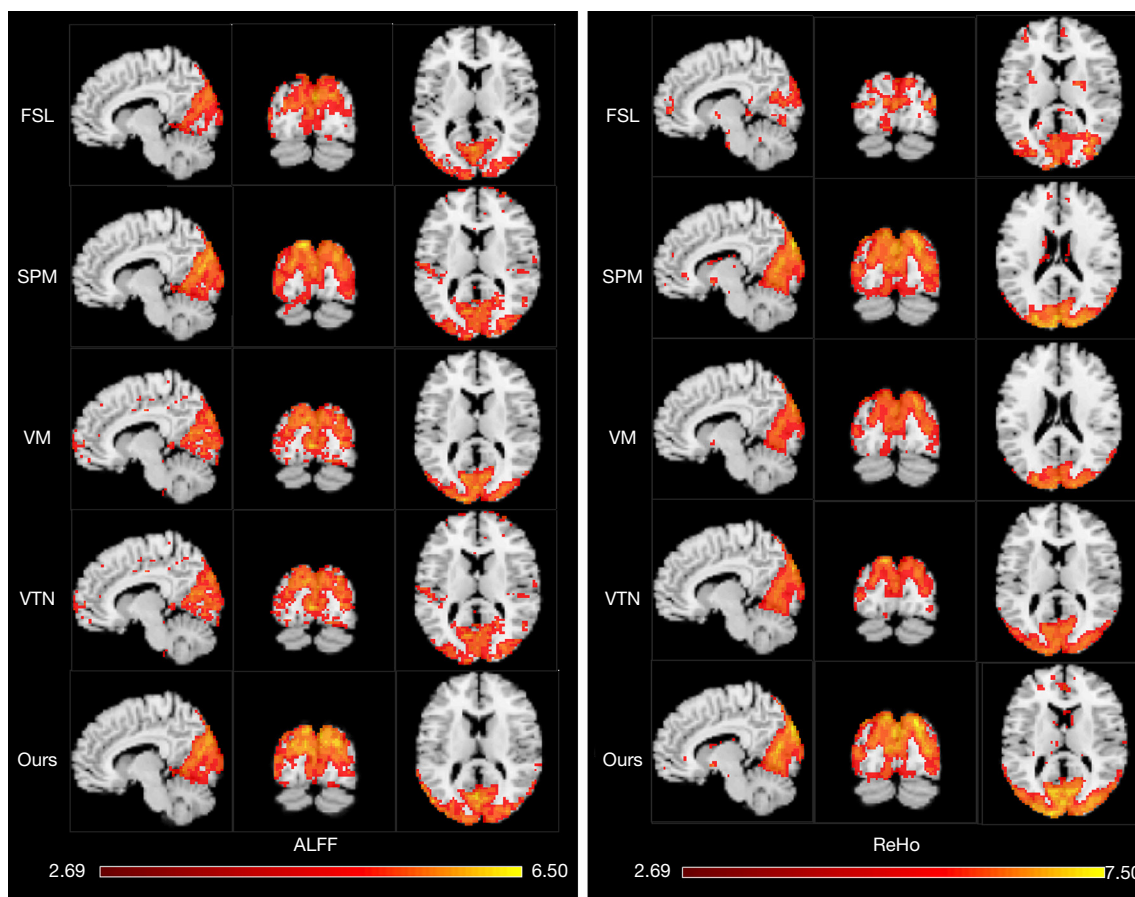| Evaluation metrics	FSL	SPM	VM	VTN	Our method
ALFF	4.02	5.53	4.24	4.86	6.98
ReHo	4.43	7.05	5.86	6.23	8.57

ALFF, amplitude of low-frequency fluctuation; ReHo, regional homogeneity; FSL, FMRIB Software Library; SPM, Statistical Parametric Mapping; VM, VoxelMorph; VTN, Volume Tweening Network.

MR-Net as our subnetwork and cascaded it three times.

To evaluate our algorithm, we compared our method with commonly used traditional fMRI registration algorithms, such as FSL and SPM, and advanced deep learning algorithms, such as VM and VTN. We extracted six brain functional networks to evaluate our registration performance and performed a one-sample *t*-test on each network. Our algorithm obtained the highest peak t-value. To further evaluate our method, we further used a task-based fMRI dataset (EOEC dataset). The peak t-values in the ALFF and ReHo paired-t maps were the highest. The experiments showed that our proposed three-cascaded MR-Net could achieve a good deformable image registration performance. However, the cascade strategy is limited after increasing the number of cascades by more than three times. Future research expects to further improve the deformable registration performance by exploring an effective method to deepen the cascaded network.

Figure 12 Paired *t*-test for the EOEC dataset using ALFF and ReHo evaluations after registration by FSL, SPM, VM, VTN, and our method. EOEC, eyes open eyes closed; ALFF, amplitude of low-frequency fluctuation; ReHo, regional homogeneity; FSL, FMRIB Software Library; SPM, Statistical Parametric Mapping; VM, VoxelMorph; VTN, Volume Tweening Network.

Conclusions

This study proposed a three-cascaded MR-Net for fMRI deformable registration to achieve progressive fMRI registration while ensuring the deformation fields' reasonability. In our experiment, six brain networks (i.e., DMN, VN, CEN, SMN, RMN, and VCN) of a group of subjects were analyzed using group-level statistical t-maps after registration. The experimental results showed that our three-cascaded MR-Net improved the group-level analysis and achieved advanced performance compared with the traditional fMRI registration methods (i.e., FSL and SPM) and deep learning frameworks (i.e., VM and VTN).

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi. org/10.21037/qims-20-1289). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons

References

1. Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, Loughead JW, Gur RC, Langleben DD. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Neuroimage 2005;28:663-8.

2. Kahnt T, Chang LJ, Park SQ, Heinzle J, Haynes JD. Connectivity-based parcellation of the human orbitofrontal cortex. J Neurosci 2012;32:6240-50.

3. Huang YL, Zhou JL, Jiang YM, Zhang ZG, Zhao W, Han D, He B. Assessment of lumbar paraspinal muscle activation using fMRI BOLD imaging and T2 mapping. Quant Imaging Med Surg 2020;10:106-15.

4. Fan Y, Liu Y, Wu H, Hao Y, Liu H, Liu Z, Jiang T. Discriminant analysis of functional connectivity patterns on Grassmann manifold. Neuroimage 2011;56:2058-67.

5. Wagner DD, Kelley WM, Heatherton TF. Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. Cereb Cortex 2011;21:2788-96.

6. Lu W, Dong K, Cui D, Jiao Q, Qiu J. Quality assurance of human functional magnetic resonance imaging: a literature review. Quant Imaging Med Surg 2019;9:1147-62.

7. Ng B, Hamarneh G, Abugharbieh R. Modeling brain activation in fMRI using group MRF. IEEE Trans Med Imaging 2012;31:1113-23.

8. AG N. The Montreal Neurological Institute. Can Med Assoc J 1934;31:548-9.

9. Laitinen L. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. By Jean Talairach and Pierre Tournoux. Translated by Mark Rayport. Stuttgart-New York: Georg Thieme Verlag, 1988:122. (Stuttgart: Georg Thieme Verlag; New York: Thieme Medical Publishers, Inc.). 1989.

10. Fischer B, Modersitzki J. FLIRT: A flexible image registration toolbox. In: International Workshop on Biomedical Image Registration. Heidelberg: Springer, 2003: 261-70.

11. Beg MF, Miller MI, Trouvé A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comput Vis 2005;61:139-57.

12. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. Neuroimage 2009;45:S61-72.

13. Fischl B. FreeSurfer. Neuroimage 2012;62:774-81.

14. Pantazis D, Joshi A, Jiang J, Shattuck DW, Bernstein LE, Damasio H, Leahy RM. Comparison of landmark-based and automatic methods for cortical surface registration. Neuroimage 2010;49:2479-93.

15. Çetin MS, Khullar S, Damaraju E, Michael AM, Baum SA, Calhoun VD. Enhanced disease characterization through multi network functional normalization in fMRI. Front Neurosci 2015;9:95.

16. Sabuncu MR, Singer BD, Conroy B, Bryan RE, Ramadge PJ, Haxby JV. Function-based intersubject alignment of human cortical anatomy. Cereb Cortex 2010;20:130-40.

17. Watson JD, Myers R, Frackowiak RS, Hajnal JV, Woods RP, Mazziotta JC, Shipp S, Zeki S. Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cereb Cortex 1993;3:79-94.

18. Conroy BR, Singer BD, Haxby JV, Ramadge PJ. fMRI-based inter-subject cortical alignment using functional connectivity. Adv Neural Inf Process Syst 2009;22:378-86.

19. Jiang D, Du Y, Cheng H, Jiang T, Fan Y. Groupwise spatial normalization of fMRI data based on multi-range functional connectivity patterns. Neuroimage 2013;82:355-72.

20. Jiang D, Jiang T, Fan Y. Groupwise fMRI registration using multi-range functional connectivity patterns. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). IEEE, 2012:1763-6.

21. Khullar S, Michael AM, Cahill ND, Kiehl KA, Pearlson G, Baum SA, Calhoun VD. ICA-fNORM: spatial normalization of fMRI data using intrinsic group-ICA networks. Front Syst Neurosci 2011;5:93.

22. Langs G, Golland P, Tie Y, Rigolo L, Golby AJ. Functional geometry alignment and localization of brain areas. Adv Neural Inf Process Syst 2010;1:1225-33.

23. Chee E, Wu Z. AIRNet: Self-supervised affine registration for 3D medical images using neural networks. arXiv preprint, 2018. arXiv:1810.02583.

24. Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: Fast predictive image registration - A deep learning approach. Neuroimage 2017;158:378-96.

25. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In:

International Conference on Medical image computing and computer-assisted intervention. Cham: Springer, 2015:234-41.

26. Glaunès J, Qiu A, Miller MI, Younes L. Large deformation diffeomorphic metric curve mapping. Int J Comput Vis 2008;80:317-36.

27. Uzunova H, Wilms M, Handels H, et al. editors. Training CNNs for Image Registration from Few Samples with Model-based Data Augmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017.

28. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018:9252-60.

29. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: A learning framework for deformable medical image registration. IEEE Trans Med Imaging 2019. [Epub ahead of print]. doi: 10.1109/TMI.2019.2897538.

30. Kuang D, Schmah T. FAIM—A ConvNet method for unsupervised 3D medical image registration. In: International Workshop on Machine Learning in Medical Imaging. Cham: Springer, 2019:646-54.

31. Chen J, Li Y, Du Y, Frey EC. Generating anthropomorphic phantoms using fully unsupervised deformable image registration with convolutional neural networks. Med Phys 2020;47:6366-80.

32. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. arXiv preprint, 2015. arXiv:1506.02025.

33. Zhou Y, Pang S, Cheng J, Sun Y, Wu Y, Zhao L, Liu Y, Lu Z, Yang W, Feng Q. Unsupervised deformable medical image registration via pyramidal residual deformation fields estimation. arXiv preprint, 2020. arXiv:2004.07624..

34. Fan J, Cao X, Yap PT, Shen D. BIRNet: Brain image registration using dual-supervised fully convolutional networks. Med Image Anal 2019;54:193-206.

35. Du B, Liao J, Turkbey B, Yan P. Multi-Task Learning for Registering Images With Large Deformation. IEEE J

Biomed Health Inform 2021;25:1624-33.

36. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage 2012;62:782-90.

37. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint, 2014. arXiv:1412.6980.

38. Zhao S, Lau T, Luo J, Chang EI, Xu Y. Unsupervised 3D end-to-end medical image registration with volume tweening network. IEEE J Biomed Health Inform 2020;24:1394-404.

39. Himberg J, Hyvarinen A. Icasso: software for investigating the reliability of ICA estimates by clustering and visualization. In: Toulouse: 2003 IEEE XIII Workshop on Neural Networks for Signal Processing (IEEE Cat. No. 03TH8718). IEEE, 2003:259-68.

40. Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF. Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A 2009;106:13040-5.

41. Wu T, Long X, Zang Y, Wang L, Hallett M, Li K, Chan P. Regional homogeneity changes in patients with Parkinson's disease. Hum Brain Mapp 2009;30:1502-10.

42. Liu L, Rao AA, Talavage TM. Regional approach to fMRI data analysis using hemodynamic response modeling. In: San Jose: Computational Imaging V. International Society for Optics and Photonics, 2007;6489:648917.

43. Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. AJNR Am J Neuroradiol 2001;22:1326-33.

44. Liu D, Dong Z, Zuo X, Wang J, Zang Y. Eyes-open/eyes-closed dataset sharing for reproducibility evaluation of resting state fMRI data analysis methods. Neuroinformatics 2013;11:469-76.

45. Yuan LX, Wang JB, Zhao N, Li YY, Ma Y, Liu DQ, He HJ, Zhong JH, Zang YF. Intra- and inter-scanner reliability of scaled subprofile model of principal component analysis on ALFF in resting-state fMRI under eyes open and closed conditions. Front Neurosci 2018;12:311.