



AMNet: Adaptive multi-level network for deformable registration of 3D brain MR images



Tongtong Che ^a, Xiuying Wang ^{c,*}, Kun Zhao ^a, Yan Zhao ^a, Debin Zeng ^a, Qiongling Li ^b, Yuanjie Zheng ^d, Ning Yang ^e, Jian Wang ^{e,f}, Shuyu Li ^{b,*}

^a School of Biological Science and Medical Engineering, Beihang University, Beijing, 100191, China

^b State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875, China

^c School of Computer Science, The University of Sydney, Sydney, Australia

^d School of Information Science and Engineering, Shandong Normal University, Jinan, 250358, China

^e Department of Neurosurgery, Qilu Hospital of Shandong University and Brain Science Research Institute, Shandong University, Jinan, 250012, China

^f Department of Biomedicine, University of Bergen, Jonas Lies Vei 91, 5009 Bergen, Norway

ARTICLE INFO

Keywords:

Deformable registration
Adaptation
Multi-level
Wavelet transformation

ABSTRACT

Three-dimensional (3D) deformable image registration is a fundamental technique in medical image analysis tasks. Although it has been extensively investigated, current deep-learning-based registration models may face the challenges posed by deformations with various degrees of complexity. This paper proposes an adaptive multi-level registration network (AMNet) to retain the continuity of the deformation field and to achieve high-performance registration for 3D brain MR images. First, we design a lightweight registration network with an adaptive growth strategy to learn deformation field from multi-level wavelet sub-bands, which facilitates both global and local optimization and achieves registration with high performance. Second, our AMNet is designed for image-wise registration, which adapts the local importance of a region in accordance with the complexity degrees of its deformation, and thereafter improves the registration efficiency and maintains the continuity of the deformation field. Experimental results from five publicly-available brain MR datasets and a synthetic brain MR dataset show that our method achieves superior performance against state-of-the-art medical image registration approaches.

1. Introduction

Image registration is fundamental and essential in the fields of computer vision and information fusion. Deformable registration for medical images has been used extensively for image-guided intervention (De Silva et al., 2016; Sarrut, 2006; Yang et al., 2011), motion tracking (Chandrashekara et al., 2003; Yang et al., 2012), patient responses to treatments, and morphometric analysis (Velec et al., 2011; Andersen et al., 2013). It aims to establish spatial correspondence between the “moving” and “fixed” images through searching and computing dense and nonlinear deformation fields. Although many algorithms have been proposed over the past three decades (Sotiras et al., 2013; Shen et al., 2017; Haskins et al., 2020), deformable brain image registration remains a challenging task, especially for images with complicated deformations imposed by the following factors: (1) the intra- or inter-person high anatomical spatially-varying caused by normal aging in the brain (Huang et al., 2021); (2) the intra- or inter-person large topological changes caused by disease progression in the brain (Mok and Chung, 2020).

Traditional registration methods, such as elastic (Bajcsy and Kováčič, 1989; Shen and Davatzikos, 2002), fnirt (Andersson et al., 2008), flirt (Fischer and Modersitzki, 2003), Demons (Thirion, 1998), B-spline (Rueckert et al., 1999) or SyN (Avants et al., 2008), have shown success in general image registration. However, searching the deformable transformation fields with numerical optimization algorithms is considered the major cause of less efficient computation and leads to inefficient registration. Recently, deep-learning-based deformable registration methods (de Vos et al., 2017; Dalca et al., 2018; Balakrishnan et al., 2019; Kim et al., 2021) have been developed, which demonstrated performance improvement over the traditional methods. Although it is often computationally expensive to train a deep-learning-based registration model, once the network is trained, the registration can be completed in seconds and estimates full-size deformation fields in one forward pass (Eppenhof et al., 2019; Haskins et al., 2020). According to the mechanisms for deriving the deformation fields, these deep registration methods can be classified into single-level registration

* Corresponding authors.

E-mail addresses: shuyuli@bnu.edu.cn (S. Li), xiu.wang@sydney.edu.au (X. Wang).

and multi-level registration. The former derives deformation fields from only one resolution level or scale, whereas the latter normally learns deformation based on a coarse-to-fine scheme. Based on the search space, single-level registration models can be further grouped into *image-wise registration* and *patch-wise registration*.

Image-wise single-level registration models (Zhang, 2018; Ito and Ino, 2018; Balakrishnan et al., 2018; Dalca et al., 2018; Duan et al., 2019; Che et al., 2019; Zheng et al., 2021) are often constructed on top of the U-Net model with encoder-decoder architecture. Although features are generated from different layers and scales, only the fully connected layer or the highest-resolution scale is used to derive the deformation fields. Therefore, image-wise registration schemes that directly search the optimal transformations on the original input images (at the highest or finest resolution level), the optimization, without initial estimation on transformation, is often slow and more vulnerable to being trapped into local optimization when dealing with large displacements (Mok and Chung, 2020; Huang et al., 2021). Furthermore, image-wise methods focus more on global contextual information, limiting the network's capacity to capture complex local deformations.

Patch-wise single-level registration models (de Vos et al., 2017; Cao et al., 2018a,b; Liu et al., 2019; Fan et al., 2019; Huang et al., 2021) have been proposed to improve the computational efficiency and the capacity for capturing local deformations. In such a registration framework, the input images are first divided into small patches, and thereafter, the search for deformation fields is restricted within these individual patches for speedy deformation learning. The resulting deformation fields from the patches are then combined into a global deformation field to warp the entire moving image. Although patch-based methods alleviate the burden on memory requirements and can capture more detailed and subtle local deformations, these models may face some major limitations. First, as illustrated in Fig. 1(a), large deformations between the patches may result in dramatic inconsistencies and variations in the anatomical characteristics of the patches in close proximity. Consequently, it is difficult to retain continuity between different patches, which results in discontinuities in the deformation fields. Furthermore, excessive attention focusing on local deformations may lead to being trapped in local optimizations.

To address the limitations of single-level registration, multi-level registration concurrently captures both global and local deformations in a coarse-to-fine manner. Based on network structures, multi-level registration can be broadly classified into *multi-branch registration* and *multi-resolution registration*.

Multi-branch registration models have been designed for dealing with large deformation through multiple task-specific sub-networks in the following three ways. First, affine and deformable image registration frameworks (Sergios et al., 2018; Hu et al., 2018; Shen et al., 2019; de Vos et al., 2019) were proposed to learn global affine transformation parameters and fine deformation fields. Second, some methods have been proposed for coarse-to-fine registration by stacking multiple networks to progressively estimate partial deformations (Sokooti et al., 2017; Zhao et al., 2019a). Third, the multi-branch registration is performed by combining global and local networks for image-wise and patch-wise deformation estimation (Lei et al., 2020; Kim et al., 2021). These methods improved the registration performance for large deformations through multi-step progressive deformation estimation. In patch-based global-local registration framework, it may be challenging to maintain the continuity of the deformation field, mainly because of the splitting and combination of patches in the local sub-network. Furthermore, the complexity of these models with multiple sub-networks poses a tremendous burden on the GPU memory, which subsequently limits the convergence speed and registration stability. In addition, the performance of these networks could be affected by the cumulative errors caused by cascaded propagation (Eppenhof et al., 2019).

Multi-resolution registration models (Fan et al., 2019; Hering et al., 2019; Wang et al., 2018; Lee et al., 2020; Mok and Chung, 2020) have been performed on different resolutions starting from the lowest

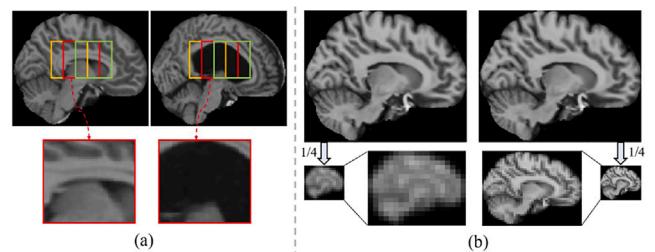


Fig. 1. (a) An example of input image pair with large deformation. The patch selection strategy follows that of Fan et al. (2019). The three different colored boxes represent the three selected adjacent patches with a size of $64 \times 64 \times 64$, and the adjacent patches overlap by 24 voxels. (b) An example of the downsampled image. Left: Gaussian-based downsampling. Right: wavelet-based downsampling (only low-frequency images are shown).

resolution level. Multi-resolution image pyramids can be constructed using different downsampling schemes, such as pooling operations (Lee et al., 2020), Gaussian filtering (Hering et al., 2019), and interpolation operations (Mok and Chung, 2020). Regarding these methods, the global transformations from the lower resolution images are first computed and then used as the initial transformation estimation for the registration focusing on local deformations at higher resolution levels. These methods have the merits of more efficient deformation searching and estimation; however, the downsampling schemes for constructing multi-resolution pyramids inevitably lead to the loss of characteristic information, resulting in weak texture or spatial aliasing in blurred images (such as the left of Fig. 1(b)). How to better use image information and reduce network complexity is yet to be properly explored.

In summary, the current single-level registration methods, especially patch-based methods, have difficulty maintaining the continuity of deformation fields when dealing with large deformations. For the current multi-level registration, there are two major challenges yet to be explored. On the one hand, the complex multi-branch models may strongly rely on computational resources and encounter slow convergence; on the other hand, the current schemes for constructing multi-resolution image pyramids may inherently not be able to fully utilize the image information and hence lead to sub-optimal registration results.

In this paper, we aim to more comprehensively utilize image information to improve registration performance while retaining the continuity of the deformation field when coping with complex deformations and variations. We propose a novel adaptive multi-level network (AMNet) to effectively capture both global and complicated local deformations in one forward pass. The proposed AMNet is constructed using wavelet transformation to enable a more complete use of both low-frequency global information and high-frequency local information. For high-resolution level registration, AMNet progressively adapts the region with updated importance based on its deformation complexity, enabling adaptive focus on large or complex deformation and retaining the continuity of the deformation field. The proposed AMNet has been validated on different datasets and compared with state-of-the-art image registration techniques. The main contributions of our work are summarized as follows:

- The AMNet is designed as a lightweight deep learning model to perform coarse-to-fine registration with one forward pass, avoiding cumulative error propagation as in multi-branch registration networks.
- Our AMNet embraces two levels of adaptation, including (1) information importance adaptation for adjusting registration focus from global context to salient local deformations; (2) network adaptation empowered by an adaptive growth strategy for accelerating the optimization of the deformation fields from low-resolution to high-resolution levels.

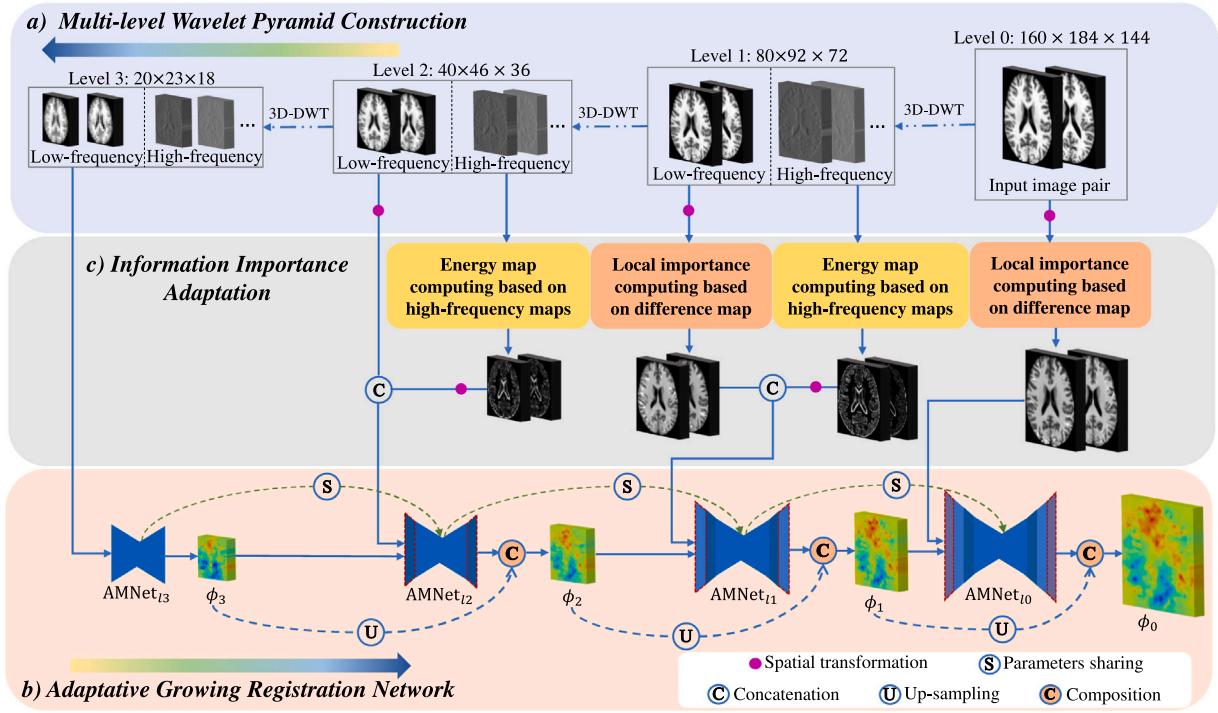


Fig. 2. The proposed AMNet framework consists of three major modules. (a) Multi-level wavelet pyramid construction: From the input moving and fixed images, an image pyramid with four resolution levels is constructed based on 3D discrete wavelet transformation (3D-DWT). (b) Adaptive growing registration network: The registration network from $AMNet_{i3}$ to the final $AMNet_{i0}$ is adaptively deepened with increasing image resolution to adjust the focus on global-local displacements. $AMNet_i$ ($i = 3, 2, 1, 0$) represents the registration network corresponding to the wavelet image of level i . The red dashed lines in the registration network indicate that new network layers are extended at the current level. The green dashed arcs indicate that the current level of the network shares the model parameters from the previous level. (c) Information importance adaptation: The adaptive capabilities of AMNet are mainly reflected in three aspects, including resolution adaptation, frequency adaptation, and importance adaptation for adjusting the registration focus from the global context to local deformations.

- Being constructed with wavelet transform that enables information preservation, our AMNet fully utilizes useful information from low-frequency and high-frequency sub-bands to achieve improved registration performance.

2. Methods

2.1. Related work: wavelet transformation

As mentioned in Luo et al. (2020), wavelet transformation can achieve the same spatial resolution reduction effect as pooling or interpolation operation without any information loss. Wavelets have been proven to be efficient in conventional multi-resolution registration approaches to avoid local minima and alleviate topological changes or foldings (Guariglia and Silvestrov, 2016; Torbati and Ayatollahi, 2019; Gefen et al., 2003; Zavorin and Le Moigne, 2005). More recently, conventional wavelet theory has been incorporated with the advantages of deep learning and applied to image analysis tasks such as super-resolution (Liu et al., 2018), image synthesis (Qu et al., 2020), and disparity estimation (Luo et al., 2020). However, wavelet theory has not been used in deep-learning-based registration. This is mainly because the decomposition of an image into multi-frequency wavelet sub-bands may dramatically increase the complexity of the registration process. For instance, the wavelet transform decomposes a 3D image into eight sub-band images, and the optimal use of these low-frequency and high-frequency sub-band images for registration has yet to be investigated.

2.2. Our method

Given a pair of images, the registration aims to determine a deformation field ϕ that warps a moving image $M \in \mathbb{R}^3$ in accordance to

a fixed image $F \in \mathbb{R}^3$, so that the warped image $M \circ \phi$ spatially aligns with and structurally corresponds to F . To search the global and local deformations for 3D brain MR image registration, we propose an adaptive multi-level network, AMNet, which makes use of wavelet transformation to refine predictions of the deformation field. As illustrated in Fig. 2, the AMNet consists of three modules: (1) multi-level wavelet pyramid construction, (2) information importance adaptation, and (3) adaptive growing registration network.

2.2.1. Multi-level wavelet pyramid construction

For fixed and moving images, F and M , the multi-level image pyramid is constructed by 3D discrete wavelet transformation (3D-DWT) (Ghasemzadeh and Demirel, 2018) and fed to the AMNet at different levels. This allows AMNet to fully utilize the useful complementary information from the multi-frequency and multi-level wavelet sub-band images (see Fig. 3(a)). Each pyramid comprises four resolution levels: level 0 corresponds to the highest resolution or finest level (also the original input images), and level 3 represents the lowest resolution or coarsest level. For each level, there are eight sub-band images including hhh , hhl , lh , hll , lhh , lhl , llh , and lli , where lli represents the low-frequency component or the average global information of the 3D image, whereas the other seven quadrants represent the high-frequency components or the local details in a 3D image. ‘ l ’ and ‘ h ’ represent the low- and high-pass filters, respectively. The low-frequency sub-band is used to learn the global deformation, and seven high-frequency sub-bands from an image are fused into an energy map (EM) to learn the deformation of local details. This allows the model to consider both global and local optimization and reduces the computational complexity. According to Li et al. (2006), the EM at point (x, y, z) in image M or F is obtained using seven high-frequency coefficients as follows:

$$EM = hhh^2(x, y, z) + hhl^2(x, y, z) + hlh^2(x, y, z) + hll^2(x, y, z) + llh^2(x, y, z) + lhl^2(x, y, z) + lll^2(x, y, z) \quad (1)$$

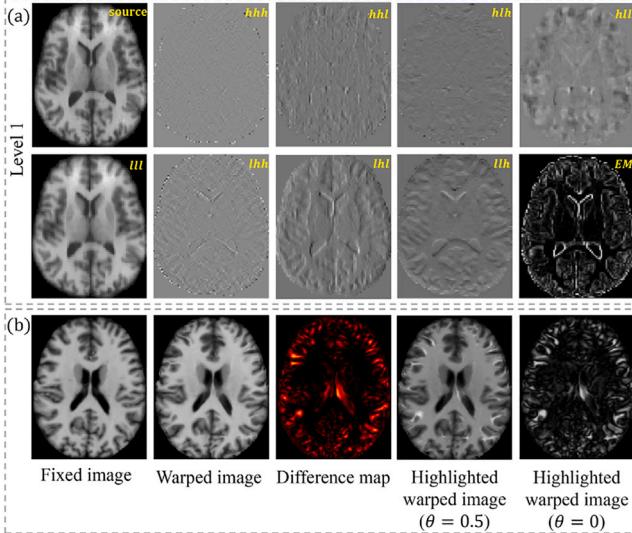


Fig. 3. (a) The first level of 3D-DWT sub-band images. (b) The local importance computing by different coordination parameters at the high-resolution levels.

As shown in Fig. 3(a), the *EM* generated by the seven high-frequency sub-band images reflects the salient features of the original images, particularly the edge features. This impels the registration network to pay extra attention to saliency-aware information for structure alignment. The registration process starts from the coarsest level. The estimates from a coarse level are used to initialize the deformation at the next finer level of the pyramid. This process propagates to the finest level of the pyramid.

2.2.2. Adaptive growing registration network

The AMNet is designed as a lightweight network with an adaptive growing strategy to estimate the deformation field in a multi-level way. As shown in Fig. 2, the registration network from AMNet_{i_3} to the final AMNet_{i_0} is adaptively deepened with increasing image resolution to adjust the focus on global-local displacements. This allows the shallow layers of the network to learn global deformation from low-resolution images and deep layers to learn refined deformation from high-resolution images. More specifically, the shallowest AMNet_{i_3} takes sub-band images at the lowest resolution level (i.e., level 3) as input and outputs the coarsest deformation field ϕ_3 . Subsequently, additional layers are added to the current AMNet_{i_3} to form a deeper AMNet_{i_2} that adapts to higher resolution images and outputs the corresponding deformation field ϕ_2 at level 2. By analogy, the number of network layers is sequentially deepened at the next level until the final AMNet_{i_0} is formed. Specifically, the AMNet_{i_0} is the final network that outputs the finest deformation field ϕ_0 .

2.2.3. Information importance adaptation

The adaptive capabilities of our network are mainly reflected in three aspects: resolution adaptation, frequency adaptation, and local importance adaptation for adjusting the registration focus from global context to local deformations.

First, to achieve fast and accurate image-wise registration, we follow the idea of multi-resolution registration and compute coarse grid solutions that are prolonged and refined on the subsequent high-resolution level. Therefore, our network learns the deformation field using a progressive learning strategy, allowing better adaptivity from low-resolution images with less information to high-resolution images with more detailed information, especially for large deformations.

Second, because the low-frequency sub-bands reflect the global average information of the 3D image, global deformation fields are

obtained from the low-frequency sub-bands. The high-frequency sub-bands with local details are used to obtain local deformation at the tissue boundaries. Therefore, our AMNet firstly learns the global deformation field from low-frequency sub-bands at level 3 and subsequently takes the *EM* (generated by seven high-frequency sub-band images) and the low-frequency sub-band image as dual inputs at levels 2 and 1. This allows our model to better adapt from low-frequency features with global information to high-frequency features with boundary-aware importance.

Third, because a global alignment is performed at low-resolution levels (levels 3 and 2), not all regions have the same importance for registration at high-resolution image levels (levels 1 and 0). The subsequently refined alignment of the image pair at a high-resolution level can be easily affected by regions with large or complex local deformations. Thus, we propose a local importance adaptation mechanism to localize difficult/complex regions while capturing global contextual information. Our AMNet takes the whole image as the input for each level. We exploit the difference map (*DM*) obtained by the warped moving image ($\mathcal{W}_{i-1}^M \circ \hat{\phi}_i$, where the $\hat{\phi}_i$ is upsampled deformation field based on ϕ_i) and fixed image (\mathcal{W}_{i-1}^F) to guide the local optimization of large/complex deformed structures. The *DM* is computed as the intensity difference between the warped moving image and fixed image, and thereafter acts on this image pair. As shown in Fig. 3(b), the highlights in the difference map are the hard-to-register regions that contain complex deformations and require be focused on for further refinement. Subsequently, the *DM* is normalized to [0,1] to reflect the importance level of regions in the image that require further registration. Finally, the input image pair is masked by *DM* as follows:

$$I' = I \cdot \{(1 - \theta)DM + \theta\}, \quad (2)$$

where I represents the input image, and θ is a coordination parameter (as in Li et al. (2019)) to control the impact of the difference map, which is set to 0.5 in our work. As shown in Fig. 3(b), the well-aligned features can be inhibited and the difficult-to-register areas can be highlighted in the input image pair. As such, our AMNet adapts a region with updated importance based on its deformation complexity, and thereafter enables rapid optimization while maintaining the continuity of the deformation field.

2.3. Network details

The architecture of AMNet and its detailed configurations can be found in Fig. 4. It undergoes four stages (from level 3 to level 0) of adaptive growth: the first three stages are trained with the wavelet sub-band images, and the last stage is trained with the original images.

For level 3, the AMNet_{i_3} concatenates only the low-frequency images at the lowest resolution (level 3) $\{\mathcal{W}_3^F, \mathcal{W}_3^M\}$ into a 2-channel 3D volume (with a size of $20 \times 23 \times 18$) and generates a coarse deformation field ϕ_3 by capturing global contextual information. The AMNet_{i_3} is a shallow U-Net-style architecture that comprises an encoder-decoder with skip connections. The residual blocks are utilized in the encoder and decoder paths, and each block contains two $3 \times 3 \times 3$ 3D convolution layers with stride 1. Besides, one strided convolution layer with a stride of 2 is employed in the encoder path to halve the spatial dimensions. The transposed convolution layer is included in the decoder to up-sample low-dimensional feature maps. Also, we add the input and output layers at both sides of the current level. The input layer includes one $3 \times 3 \times 3$ convolution layer with stride 1 to match the number of feature maps on the left side of the level. The output layer includes two consecutive $3 \times 3 \times 3$ convolution layers with stride 1, followed by SoftSign activation to output the target deformation field. Each convolution layer is followed by a leaky rectified linear unit (LeakyReLU) activation (Maas et al., 2013). Note that the number of channels in each layer has been marked in Fig. 4.

For level $i < 3$, the AMNet_{i_1} is first deepened by adding two residual blocks and up/down sampling operations to adapt to higher

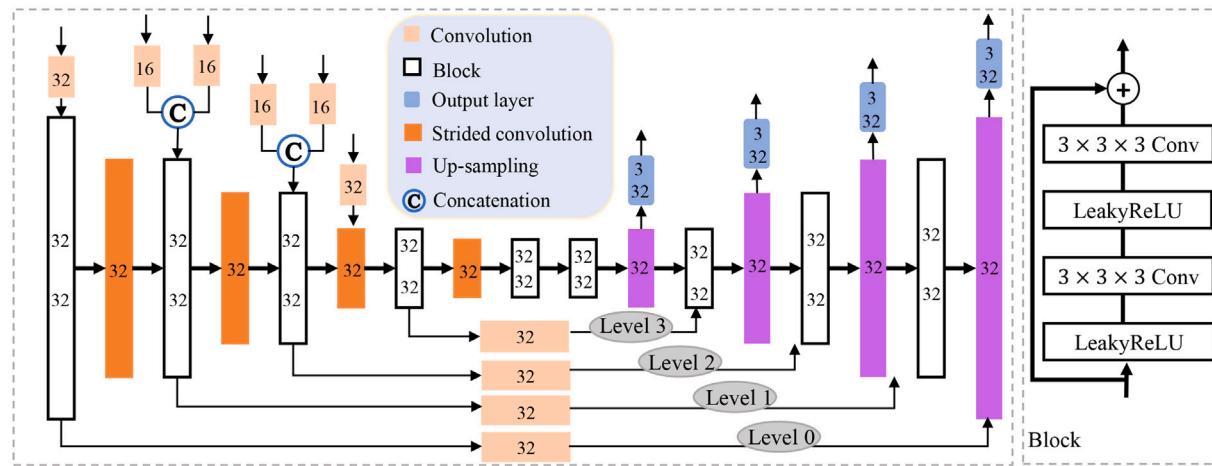


Fig. 4. The detailed network architecture of AMNet. The numbers in each box represent the number of feature channels at each layer, where each block and output layer contains two convolution layers.

Algorithm 1 Adaptive multi-level registration network.

Input: a pair of images F and M ;
Output: deformation field ϕ ;

- 1: Create 4-level wavelet pyramid, with finest level $i = 0$ and $i = 3$ coarsest;
- 2: **for** level $i = 3, 2, 1, 0$ **do**
- 3: **Training** AMNet_{l_i}
- 4: **if** level $i = 3$ **then**
- 5: $\phi_i = \text{AMNet}_{l_i}(III_i^M, III_i^F)$
- 6: **end if**
- 7: **if** level $i = 2$ **then**
- 8: $\phi_i = \text{AMNet}_{l_i}(III_i^M \circ \hat{\phi}_{i+1}, III_i^F, EM_i^M \circ \hat{\phi}_{i+1}, EM_i^F, \hat{\phi}_{i+1})$
- 9: **end if**
- 10: **if** level $i = 1$ **then**
- 11: $\phi_i = \text{AMNet}_{l_i}((III_i^M \circ \hat{\phi}_{i+1}) \cdot DM, (III_i^F \circ \hat{\phi}_{i+1}) \cdot DM, EM_i^M \circ \hat{\phi}_{i+1}, EM_i^F, \hat{\phi}_{i+1})$
- 12: **end if**
- 13: **if** level $i = 0$ **then**
- 14: $\phi_i = \text{AMNet}_{l_i}(M \cdot DM, F \cdot DM, \hat{\phi}_{i+1})$
- 15: **end if**
- 16: $\hat{\phi}_i = \text{Upsample}(\phi_i)$
- 17: **end for**

resolution images. Thereafter, new input and output layers are added at the current level, and the input and output layers of the previous level are removed. Besides, we compose the output deformation field ϕ_i of the current level with the upsampled deformation field $\hat{\phi}_{i+1}$ from the previous level to obtain the final ϕ_i . The composition is implemented by warping $\hat{\phi}_{i+1}$ with ϕ_i and adding the result with ϕ_i (Kim et al., 2021; Vercauteren et al., 2009). In particular, the inputs are different for each level. AMNet_{l_2} contains two input layers (with a size of $40 \times 46 \times 36$), one takes the concatenated upsampled deformation field $\hat{\phi}_3$ and low-frequency sub-band images $\{III_2^M \circ \hat{\phi}_3, III_2^F\}$ as a 5-channel input, and the other takes the concatenated energy maps $\{EM_2^M \circ \hat{\phi}_3, EM_2^F\}$ as a 2-channel input. Similarly, AMNet_{l_1} contains two input layers (with a size of $80 \times 92 \times 72$), one takes the concatenated upsampled deformation field $\hat{\phi}_2$ and low-frequency sub-band images with local importance $\{(III_1^M \circ \hat{\phi}_2), (III_1^F)\}$ as a 5-channel input, and the other takes the concatenated energy maps $\{EM_1^M \circ \hat{\phi}_2, EM_1^F\}$ as a 2-channel input. For the level 0, the upsampled deformation field $\hat{\phi}_1$ and the original images with local importance $\{(M \circ \hat{\phi}_1), (F)\}$ are input to AMNet_{l_0} , and outputs the final deformation field ϕ_0 (with a size of $160 \times 184 \times 144$). It is important to note that the higher resolution feature map after downsampling has the same size as the lower resolution image of the previous level.

2.4. Training strategy

Our AMNet is a level-by-level progressive training scheme, which is trained in an unsupervised manner. The training starts from AMNet_{l_3} with the coarsest image resolution ($20 \times 23 \times 18$), which harnesses the least number of network layers. When the similarity measure at a level (AMNet_{l_i}) converged, the training proceeded to the next level ($\text{AMNet}_{l(i-1)}$), and additional network layers are extended on both sides of AMNet_{l_i} . The network AMNet_{l_i} is used as a pre-trained model of ($\text{AMNet}_{l(i-1)}$) and the model parameters of AMNet_{l_i} (except for the input and output layers) are shared by $\text{AMNet}_{l(i-1)}$. This training strategy progresses all the way up to AMNet_{l_0} with the original image resolution ($160 \times 184 \times 144$). Besides, along with the AMNet_{l_i} adaptively grows, coarse-to-fine registration is performed in one forward pass, and the deformation fields are forward propagated between different levels in three ways. (1) The deformation field ϕ_i obtained from AMNet_{l_i} is upsampled ($\hat{\phi}_i$) by interpolation as the input for $\text{AMNet}_{l(i-1)}$. (2) The upsampled deformation field $\hat{\phi}_i$ is used as the initial deformation of the level ($i - 1$) by warping the moving images (i.e., low-frequency wavelet sub-band image III_{i-1}^M and energy map EM_{i-1}^M) (Jaderberg et al., 2015). (3) Final deformation field ϕ_{i-1} at level ($i - 1$) is obtained by composing the upsampled deformation field $\hat{\phi}_i$ with the output of $\text{AMNet}_{l(i-1)}$. The composition is implemented by warping $\hat{\phi}_i$ with ϕ_{i-1} and adding the result with ϕ_{i-1} (Vercauteren et al., 2009; Kim et al., 2021).

2.5. Losses and objectives

For level i , we employ the local normalized cross-correlation (NCC) (Balakrishnan et al., 2018) as the similarity metric to compare the alignment between the warped image and its corresponding fixed image. The loss function for each level is defined as follows:

$$\mathcal{L}_{sim} = -[NCC(III_i^M \circ \phi_i, III_i^F) + \lambda_e NCC(EM_i^M \circ \phi_i, EM_i^F)], \quad (3)$$

where NCC is the local normalized cross-correlation, $(III_i^M \circ \phi_i, III_i^F)$ denotes the low-frequency wavelet sub-band images at level i , $(EM_i^M \circ \phi_i, EM_i^F)$ denotes the energy maps at level i , and λ_e is the weight parameter. λ_e is set to 0 at levels 3 and 0; otherwise, it is set to 1. The image similarity is also calculated on energy maps, ensuring that the tissue boundary could be aligned more accurately than when only matching the image similarity in the low-frequency intensity image.

In addition to similarity guidance, we evaluate the smoothness of the deformation field by computing the spatial gradients (Balakrishnan et al., 2019). The loss function for each level is defined as follows:

$$\mathcal{L}_{smo} = \sum_{p \in \Omega} \|\nabla \phi_i(p)\|_2^2, \quad (4)$$

where $\nabla \phi_i(p)$ denotes the gradient of the deformation field ϕ_i at level i at the voxel p . However, the balance between the image alignment and deformation field smoothness is difficult to control because all pixels in an image pair have different degrees of deformation. The registration results will be inaccurate when a large weight is assigned to the smoothness constraint because of excessive regularization. Otherwise, a small weight will improve registration accuracy, but it leads to folding in the deformation field. In previous studies, the anti-folding constraint was used to dynamically control the smoothness weight (Zhang, 2018; Huang et al., 2021). Herein, we also use such an anti-folding constraint in the objective function as:

$$\mathcal{L}_{ant} = \left[\sum_{d \in \{x,y,z\}} \delta(\Delta \phi_i^d(p)) \|\Delta \phi_i^d(p)\|_2^2 \right] \cdot \|\nabla \phi_i(p)\|_2^2, \quad (5)$$

where $\Delta \phi_i^d(p)$ is the Laplacian operation of the deformation field along the d th direction at voxel location p , and $\delta(\cdot)$ is an indicator function used to localize the voxels with folding. In other words, if $\Delta \phi_i^d(p) < 0$, $\delta(\Delta \phi_i^d(p)) = 1$; and $\delta(\Delta \phi_i^d(p)) = 0$, otherwise.

Accordingly, the loss function of our proposed AMNet for deformable registration is formulated as follows:

$$\mathcal{L} = \mathcal{L}_{sim} + \lambda_s \mathcal{L}_{smo} + \lambda_a \mathcal{L}_{ant}, \quad (6)$$

where the parameters (i.e., λ_s and λ_a) are determined via grid search, and finally best-performing combination (i.e., $\lambda_s = 1$ and $\lambda_a = 1$) are selected. The influence of different parameters on registration results is discussed in Section 4.

The scheme of AMNet is summarized in Algorithm 1.

3. Experimental settings

3.1. Datasets

We performed experiments on five publicly available datasets (OASIS (Shattuck et al., 2008), LONI LPBA40, IBSR18, CUMC12, and MGH10 (Klein et al., 2009a)) and a synthetic dataset (LPBA40_{syn}). The OASIS dataset includes 416 subjects aged from 18 to 96, and 100 of the included subjects over the age of 60 have been clinically diagnosed with very mild to moderate Alzheimer's disease (AD).¹ LPBA40, IBSR18, CUMC12, and MGH10 datasets from the Continuous Registration Challenge (CRC),² which comprise T1-weighted MR brain images from 40, 18, 12 and 10 human subjects, respectively. The synthetic dataset is

obtained by warping LPBA40 with artificial transformations. There are 150 nonlinear transformations artificially created using a B-spline model, which contains deformations of different degrees (i.e., 20%, 50%, and 100%).

Training datasets. To improve the generalization ability of the model, the training images are derived from two-center datasets with large deformations (LONI LPBA40 and OASIS). We randomly chose 30 images (30×29 image pairs) from LPBA40 and 70 images (70×69 image pairs) from OASIS as the training data. Besides, the synthetic images corresponding to the 30 images in LPBA40 (30×150 image pairs) were also used for training.

Testing datasets. We have evaluated our method on all six datasets mentioned above to verify the universality of the model. Specifically, the testing datasets include the remaining 10 images of LPBA40, 50 images randomly selected from OASIS (different from the training data), and all samples in IBSR18, CUMC12, and MGH10. For each testing dataset, the images were registered with each other, resulted in 90, 2450, 306, 132, 90, and 1500 image pairs for LPBA40, OASIS, IBSR18, CUMC12, MGH10, and LPBA40_{syn}, respectively. We computed the average Dice score on the labeled anatomical ROIs, which includes 56, 35, 96, 128, 106, and 56 anatomical structures for LPBA40, OASIS, IBSR18, CUMC12, MGH10, and LPBA40_{syn}, respectively.

Data preprocessing. The standard preprocessing steps were performed on all subjects. Skull stripping was performed using FreeSurfer (Fischl, 2012). Acquisition details were different for different datasets and we resampled all volumes after center cropping unnecessary areas around the brain to a $160 \times 184 \times 144$ grid with a resolution of 1 mm for each dimension. Subsequently, histogram matching and intensity normalization were performed to standardize the intensity range [0, 1]. Besides, all the datasets were affine registered using ANTs toolkit (Avants et al., 2008). Finally, 35 subcortical anatomical structures were segmented using FreeSurfer for the final OASIS evaluation. For other datasets, we used the official segmentation maps delineated by experts.

3.2. Implementation details

The AMNet was implemented using PyTorch (Paszke et al., 2017). We set the initial learning rate as $1e-4$ and then it was multiplied by 0.5 every 50k iterations after the first 60k iterations for each level. We employed an Adam optimizer (Kingma and Ba, 2014) and trained the network for 4, 4, 6 and 6 epochs in levels 1, 2, 3, and 4, respectively. The network was implemented using a single NVIDIA TITAN Xp and server with an Intel(R) Xeon(R) CPU (v4 @2.40 GHz).

3.3. Evaluation metrics

Dice score (DSC). We first evaluated our method using the Dice score of the subcortical segmentation maps. The DSC of two anatomical structures F^k and M^k between the warped moving and the fixed images is formulated as:

$$DSC = 2 \cdot \frac{|F^k \cap (M \circ \phi)^k|}{|F^k| + |(M \circ \phi)^k|}. \quad (7)$$

If a deformation field ϕ represents accurate correspondences, the regions in F and $M \circ \phi$ that correspond to the same anatomical structure should overlap well. The closer the Dice value is to 1, the better the overlap of the two images, that is, the better the registration performance.

Average Symmetric Surface Distance (ASSD). We also evaluated the surface distance between the predicted and the reference segmentation maps. The ASSD is defined as:

$$ASSD = \frac{1}{|M_S| + |F_S|} \left(\sum_{x \in M_S} d(x, F_S) + \sum_{y \in F_S} d(y, M_S) \right), \quad (8)$$

where x and y are points on the boundary of the predicted surface M_S and the reference surface F_S , respectively.

¹ <https://www.oasis-brains.org>.

² <https://continuousregistration.grand-challenge.org>.

Table 1

Quantitative evaluation of different registration methods on six datasets. DSC indicates the registration accuracy (the higher, the better). $|JD| \leq 0$ represents the average percentage of folding voxels in the deformation field (the lower, the better). ASSD is the surface distance between the predicted surface and the reference surface (the lower, the better).

Datasets	Metrics	SyN	Demons	B-spline	VM	CM	LapIRN	HDAR	AMNet
OASIS	DSC	0.726 (± 0.016)	0.724 (± 0.033)	0.711 (± 0.083)	0.731 (± 0.019)	0.769 (± 0.021)	0.764 (± 0.017)	0.755 (± 0.031)	0.789 (± 0.014)
	$ JD \leq 0$	0.000 (± 0.000)	0.721 (± 0.093)	0.677 (± 0.073)	2.175 (± 0.157)	1.752 (± 0.074)	0.010 (± 0.007)	1.814 (± 0.152)	0.198 (± 0.069)
	ASSD	1.034 (± 0.658)	1.052 (± 0.704)	1.106 (± 0.820)	1.012 (± 0.693)	0.823 (± 0.531)	0.826 (± 0.341)	0.937 (± 0.422)	0.821 (± 0.274)
LPBA40	DSC	0.729 (± 0.014)	0.725 (± 0.031)	0.723 (± 0.074)	0.717 (± 0.018)	0.747 (± 0.015)	0.738 (± 0.009)	0.736 (± 0.021)	0.755 (± 0.012)
	$ JD \leq 0$	0.000 (± 0.000)	0.604 (± 0.075)	0.461 (± 0.053)	1.872 (± 0.105)	1.370 (± 0.083)	0.008 (± 0.002)	1.703 (± 0.349)	0.172 (± 0.066)
	ASSD	1.074 (± 0.519)	1.093 (± 0.657)	1.104 (± 0.638)	1.082 (± 0.725)	0.877 (± 0.428)	0.892 (± 0.373)	0.988 (± 0.513)	0.869 (± 0.309)
IBSR18	DSC	0.570 (± 0.027)	0.571 (± 0.042)	0.566 (± 0.069)	0.567 (± 0.030)	0.579 (± 0.039)	0.581 (± 0.012)	0.582 (± 0.027)	0.590 (± 0.013)
	$ JD \leq 0$	0.000 (± 0.000)	0.472 (± 0.155)	0.328 (± 0.094)	2.163 (± 0.319)	1.791 (± 0.102)	0.016 (± 0.008)	2.766 (± 0.315)	0.322 (± 0.104)
	ASSD	1.136 (± 0.572)	1.139 (± 0.526)	1.145 (± 0.433)	1.153 (± 0.632)	1.092 (± 0.507)	1.054 (± 0.418)	1.071 (± 0.434)	1.057 (± 0.472)
CUMC12	DSC	0.623 (± 0.018)	0.636 (± 0.046)	0.629 (± 0.051)	0.658 (± 0.025)	0.665 (± 0.020)	0.665 (± 0.016)	0.671 (± 0.024)	0.671 (± 0.019)
	$ JD \leq 0$	0.000 (± 0.000)	0.439 (± 0.084)	0.527 (± 0.075)	1.794 (± 0.201)	1.033 (± 0.197)	0.023 (± 0.011)	3.932 (± 0.673)	0.337 (± 0.105)
	ASSD	1.127 (± 0.614)	1.115 (± 0.529)	1.122 (± 0.740)	1.109 (± 0.592)	1.094 (± 0.371)	1.034 (± 0.617)	1.044 (± 0.582)	0.973 (± 0.502)
MGH10	DSC	0.647 (± 0.021)	0.644 (± 0.028)	0.639 (± 0.042)	0.653 (± 0.019)	0.664 (± 0.026)	0.663 (± 0.014)	0.659 (± 0.034)	0.672 (± 0.016)
	$ JD \leq 0$	0.000 (± 0.000)	0.583 (± 0.117)	0.332 (± 0.086)	1.821 (± 0.196)	1.142 (± 0.105)	0.019 (± 0.006)	4.376 (± 0.581)	0.283 (± 0.088)
	ASSD	1.054 (± 0.421)	1.063 (± 0.523)	1.124 (± 0.681)	1.022 (± 0.439)	1.083 (± 0.531)	0.951 (± 0.364)	1.107 (± 0.503)	0.911 (± 0.361)
LPBA40 _{syn}	DSC	0.655 (± 0.019)	0.653 (± 0.021)	0.645 (± 0.076)	0.682 (± 0.023)	0.687 (± 0.034)	0.697 (± 0.016)	0.687 (± 0.028)	0.711 (± 0.018)
	$ JD \leq 0$	0.000 (± 0.000)	0.625 (± 0.097)	0.503 (± 0.102)	2.312 (± 0.207)	1.270 (± 0.194)	0.033 (± 0.010)	1.941 (± 0.204)	0.287 (± 0.041)
	ASSD	1.021 (± 0.532)	1.031 (± 0.501)	1.193 (± 0.570)	0.967 (± 0.621)	0.941 (± 0.452)	0.928 (± 0.463)	0.960 (± 0.701)	0.881 (± 0.475)

Jacobian determinant (JD). We then quantitatively evaluated the topology (smoothness) of the obtained deformation fields using JD. The quality of the deformation field is quantified by computing the percentage of voxels with a non-positive Jacobian determinant ($|JD| \leq 0$). The smaller the $|JD| \leq 0$, the better the smoothness of the deformation field.

4. Results

In this section, we compared the proposed method with seven state-of-the-art methods on five public-available datasets and a synthetic dataset.

4.1. Comparisons with the state-of-the-art methods

To evaluate the performance of our proposed method, comparisons with several state-of-the-art deformable registration algorithms are performed, including three traditional registration algorithms (i.e., symmetric normalization registration (denoted as SyN) (Avants et al., 2008), demons algorithm (denoted as Demons) (Thirion, 1998), fast free-form deformations with b-splines (denoted as B-spline) (Rueckert et al., 1999)), a deep-learning-based single-level registration method VoxelMorph (denoted as VM) (Balakrishnan et al., 2019), and three deep-learning-based multi-level registration methods (i.e., Laplacian pyramid image registration network (denoted as LapIRN) (Mok and Chung, 2020), global-local multi-scale registration framework CycleMorph (denoted as CM) (Kim et al., 2021), and difficulty-aware hierarchical registration algorithms (denoted as HDAR) (Huang et al., 2021)).

SyN (Avants et al., 2008) - a widely used registration method implemented in the ANTs software package (Avants et al., 2011). We used cross-correlation (CC) with a sampling radius set to 4 for SyN registration. We used multi-resolution optimization with four scales and {2000, 1000, 1000, 50} iterations, and the standard deviation for Gaussian smoothing at each level was set to {3, 2, 1, 0}. The flow standard deviation to smooth the gradient field was set to 3.

Demons (Thirion, 1998) - a classical non-parametric registration framework. We ran Demons via SimpleITK (Lowe kamp et al., 2013) and obtained the best results when the number of iterations is 300 and the standard deviation of the Gaussian smoothing for the displacement field is set to 1.2. We used MSE as the similarity measure.

B-spline (Rueckert et al., 1999) - a free-form deformation algorithm with cubic b-splines integrated in Elastix (Klein et al., 2009b). We ran B-spline with the parameters in accordance to the settings in Zhao et al. (2019b).

VM (Balakrishnan et al., 2019) - a famous deep-learning-based registration network, which takes the whole image as input and outputs a dense deformation field at a single level. For a fair comparison, we trained the VM from scratch using their official implementation, and the best parameters were determined via grid search.

LapIRN (Mok and Chung, 2020) - a multi-level diffeomorphic registration algorithm based on Laplacian pyramid architecture, which utilizes three identical CNN-based registration networks to mimic the registration with the multi-resolution scheme. We trained the LapIRN from scratch using their official implementation, and the best parameters were determined via grid search.

HDAR (Huang et al., 2021) - a recent multi-level patch-based registration model, which is based on an attention mechanism to identify hard-to-register regions automatically. We trained the HDAR from

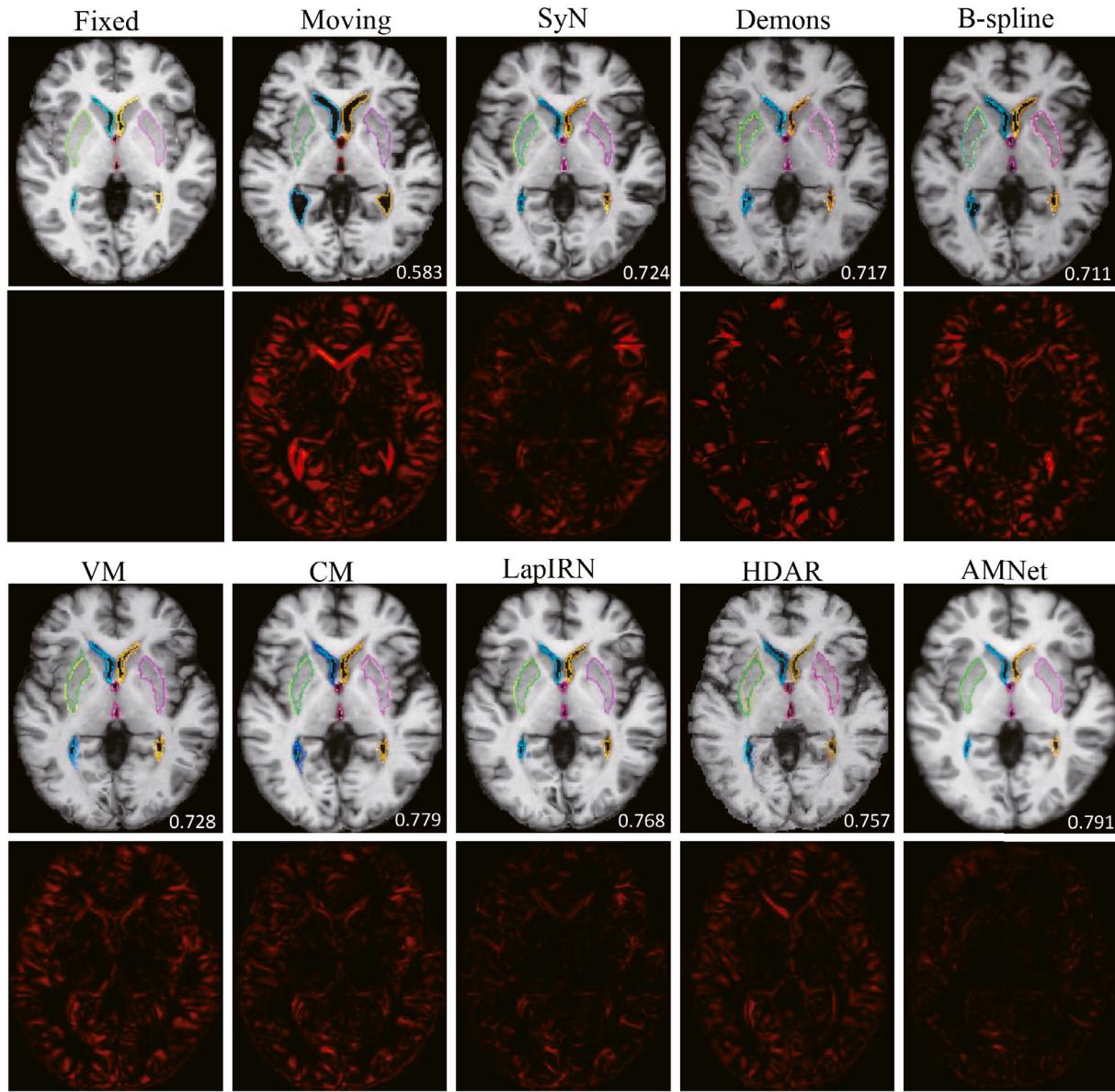


Fig. 5. Qualitative registration results by eight different methods for one example from OASIS dataset. We overlaid the boundaries of several anatomical structures with a lighter color for target contours and a darker color for moving/warped contours (blue: left lateral ventricle, yellow: right lateral ventricle, green: left putamen, purple: right putamen, red: third ventricle). The second and fourth rows show corresponding heat maps of the absolute difference with respect to the fixed image. The values in the lower right corner of moving/warped images are the Dice score (DSC) between the moving/warped image and the fixed image.

scratch using their official implementation online and set the same parameters as the original paper (also the best parameters).

CM (Kim et al., 2021) - a recent global-local multi-scale registration framework, which combines global and local networks for image-wise and patch-wise deformation estimation. We trained the CM from scratch using their official implementation and set the learning rate as $1e-4$ for global and local networks for better results, whereas for other parameters, we used the same values as the brain MRI registration of CM.

Table 1 reports the quantitative results over all ROIs of the different datasets compared with state-of-the-art methods. Our AMNet achieved the best registration accuracy over all the comparing methods. By comparing the Dice score from traditional registration methods on the OASIS dataset, AMNet improved performance by 9% compared to SyN and Demons, and by 11% compared to B-spline. In terms of $|JD| \leq 0$, although the conventional registration method SyN and the diffeomorphic LapIRN obtain smaller $|JD| \leq 0$, they actually sacrifice DSC and ASSD. We compared the classical deep registration method VM,

our AMNet yielded higher accuracy with few folding on all datasets, which reflects the effectiveness of multi-level registration. Additionally, patch-based multi-level registration methods, CM and HDAR, are used for comparison. Our method outperforms them in terms of both registration accuracy and smoothness of deformation fields, although HDAR used similar local importance for large deformations. This is mainly because we take the whole image information at different resolution levels into account, rather than local information limited to the patches is utilized in the registration process. In short, AMNet can better balance the accuracy of registration and the smoothness of the deformation field, which captures large and small deformations in the full field of view at all levels.

Fig. 5 shows one example prediction by different methods on OASIS. We can observe that the results obtained by AMNet are most close to the fixed image, especially in the marked structure regions. The superior performance of AMNet indicates that the energy map with high-frequency information can provide useful guidance to further refine the registration at the structural edges. In addition, from the

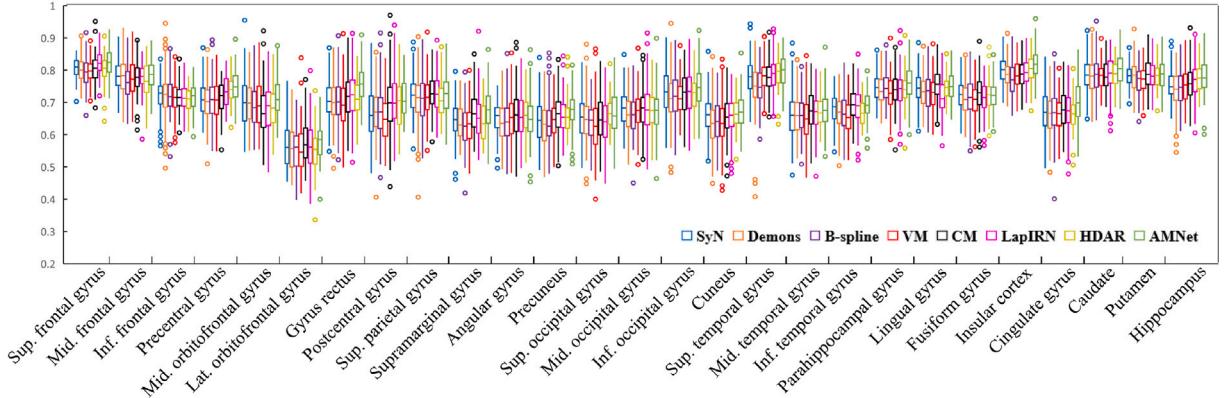


Fig. 6. Box plots of DSC for 54 labeled anatomical structures (excluding the cerebellum and brainstem) of the 10 validation subjects from the LPBA40 dataset. The scores for left and right brain structures are averaged into one score.

Table 2

Quantitative results from AMNet with and without energy map in the brain dataset OASIS.

	DSC	$ JD \leq 0$	ASSD
Without EM	0.783 ± 0.021	0.242 ± 0.308	0.827 ± 0.309
With EM	0.789 ± 0.014	0.198 ± 0.069	0.821 ± 0.274

registration results obtained by VM, there are discontinuities (similar to cracks) at the edges of some structures, which are caused by the poor smoothness of the deformation field. In contrast, AMNet, which considers updated local importance and anti-folding, can achieve smooth deformation field prediction.

Fig. 6 reports the statistics of DSC (mean and variance) of all methods over 54 brain structures by propagating the deformation field to the segmentation labels. Our AMNet performs better than SyN (except for angular gyrus, fusiform gyrus, and putamen), especially for difficult regions such as the hippocampus. This indicates that the AMNet is effective in handling complex deformations. Also, AMNet achieved the highest scores for almost all anatomical structures compared to Demons, B-spline, and VM. Besides, our method gave statistically significant improvements ($p < 0.05$, paired t -tests) in terms of DSC for 40, 38 and 48 out of 54 ROIs over CM, LapIRN, and HDAR, respectively, and comparable results for the other ROIs.

4.2. Ablation analysis

4.2.1. Analysis of information importance adaptability

To evaluate the performance of the proposed network with information importance adaptation, we designed two variants of AMNet by removing energy map (*EM*) computing and local importance computing.

First, we removed the “Energy Map Computing” phase to evaluate the contribution of energy map to our method. The network without *EM* took only low-frequency sub-band images (“III”) at different resolutions and λ_e was set to 0 in Eq. (3). As shown in Table 2, we can observe that the higher DSC and fewer folding voxels are obtained by AMNet with *EM* on OASIS testing dataset, and Fig. 7 shows that AMNet without *EM* performs poorly on the edges of the anatomical structure (see the red box). This indicates that the network can learn more detailed and difficult edge deformations under the guidance of high-frequency energy maps.

Second, we removed the “Local Importance Computing” phase to evaluate the contribution of local importance to our method. The variant without local importance learns the transformation of global images with the same importance. Table 3 lists quantitative results obtained by AMNet with and without local importance for images with deformations at varying degrees, in terms of registration accuracy and

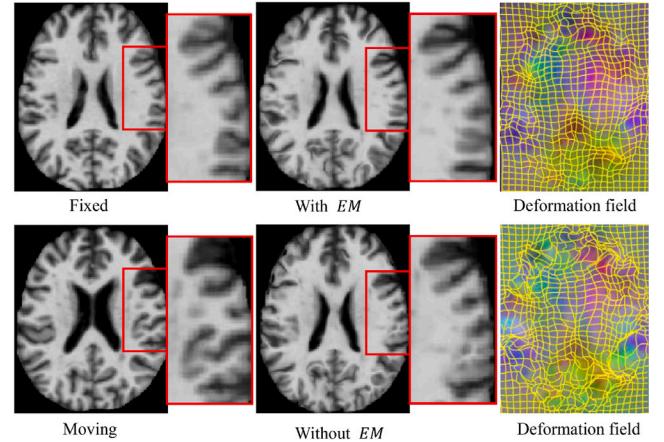


Fig. 7. Visualization of an example from AMNet with and without energy map (*EM*) in the brain dataset OASIS. The red box shows the remarkable parts. The last column shows the deformation grid obtained at the highest resolution.

deformation smoothness. These results demonstrate that the absence of local importance computing results in lower DSCs and higher $|JD| \leq 0$ on the overall testing datasets.

4.2.2. Analysis of adaptive growth strategy

To investigate the effectiveness of such an adaptive growth strategy, we designed a variant of AMNet, AMNet_{static}, in which all resolution levels harnessed the same network depth (layers). The convergence curves of the models, with and without adaptive growth network, are shown in Fig. 8. We show the convergence process of the total loss after determining the optimal parameters. It can be observed that the performance of AMNet saturated rapidly during training, particularly at lower resolution levels. We also conducted experiments to quantitatively compare the DSC of these two methods to prove our analysis, as shown in Table 4. Under the same number of iterations, AMNet_{static} performs significantly worse than AMNet from the coarsest level (level 3). The registration accuracy at the coarse level gradually affects the registration at the subsequent levels, eventually leading to sub-optimal results. Overall, AMNet improved its performance in terms of both convergence speed and registration accuracy, benefiting from an adaptive growth network.

4.2.3. Analysis of multi-level image-wise registration strategy

To determine whether the wavelet-based multi-level image-wise registration strategy in AMNet can ensure the continuity of the deformation field and registration accuracy, we designed a patch-based

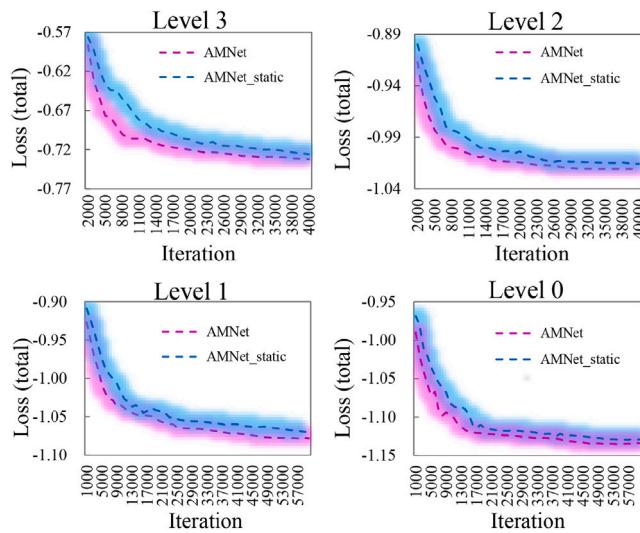
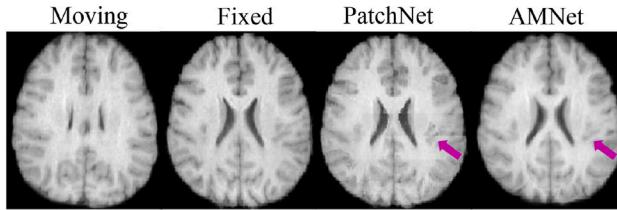
Table 3

Results for synthetic LPBA40 with 20%, 50%, and 100% deformation degree, respectively.

	Without local importance			With local importance		
	DSC	$ JD \leq 0$	ASSD	DSC	$ JD \leq 0$	ASSD
20%	0.709 ± 0.012	0.254 ± 0.115	0.781 ± 0.312	0.712 ± 0.011	0.256 ± 0.107	0.764 ± 0.183
50%	0.698 ± 0.104	0.364 ± 0.101	0.918 ± 0.376	0.705 ± 0.013	0.302 ± 0.088	0.884 ± 0.381
100%	0.684 ± 0.014	0.480 ± 0.237	1.054 ± 0.573	0.696 ± 0.011	0.296 ± 0.163	0.962 ± 0.417

Table 4DSC for synthetic LPBA40 from AMNet and its variant AMNet_{static} in each level.

Method	AMNet _{static}	AMNet
Level 3	0.824 ± 0.011	0.882 ± 0.007
Level 2	0.763 ± 0.008	0.801 ± 0.015
Level 1	0.706 ± 0.014	0.714 ± 0.012
Level 0	0.694 ± 0.013	0.709 ± 0.017

**Fig. 8.** The training curves for the total loss from AMNet and its variant AMNet_{static} in each level.**Fig. 9.** Visualization of an example from AMNet and its variant PatchNet in the brain dataset LPBA40. The purple arrows indicate the remarkable parts.

single-level registration variant named PatchNet. In detail, we followed the current mainstream patch division method (Fan et al., 2019), in which the $64 \times 64 \times 64$ patches on the highest resolution image (original image) were used as input for the network.

From Fig. 9, we can find that the results obtained by PatchNet have obvious grid-like artifacts along the edges of the patches. This is because it is difficult to ensure the continuity of the patch edges when patch-based deformation is fused into the final whole-image transformation. Besides, patch-based training is prone to fall into the local optimum, resulting in sub-optimal registration results. By contrast, our method with a multi-level registration strategy takes the whole image at different resolutions as input, and the optimization from global to local is performed such that the information is fully utilized. In addition, the multi-level image-wise registration strategy makes the

network better coordinated with large deformations, thus reducing errors and difficulties caused by one-step registration.

4.2.4. Analysis of parameter setting

For the proposed AMNet, the weighting parameters in Eq. (6) control the relative contributions of similarity and regularization losses. The setting of parameters may significantly affect the registration performance. In this section, we conducted experiments based on the grid search by fixing λ_s to investigate the influence of λ_a and fixing λ_a to investigate the influence of λ_s . Fig. 10 shows how the performance of registration on all testing datasets is influenced by the weighting parameters. We checked different λ_s values in {0.1, 1, 1.5, 2} and different λ_a values from 0.01 to 10 with a step size of 10. From the results, it can be observed that when a smaller or larger weight is assigned to the loss, the registration accuracy is degraded. With an increase in λ_a and λ_s , the folding voxels in the deformation field gradually decrease, whereas the registration accuracy does not improve. Considering the dice score (DSC) and the percentage of non-positive Jacobian determinant ($|JD| \leq 0$), the best result is achieved when $\lambda_s = 1$, and $\lambda_a = 1$. This indicates that the smoothness constraints (i.e., \mathcal{L}_{ant} and \mathcal{L}_{smo}) assigned with an appropriate weight can effectively balance the smoothness of the deformation field and the registration accuracy.

4.3. Computational speed and complexity

First, we compared the testing time of the proposed AMNet with that of other baseline methods. The running times are approximately the same across datasets, so we test them on LPBA40_{syn}, with a 12-GB NVIDIA TITAN-Xp GPU and a 2.4-GHz Intel Xeon E5-2680 v4 CPU. For a more direct comparison, we calculated the runtime of image registration using a CPU and GPU for the deep-learning-based methods. The average processing time per image pair for our method and its counterparts are reported in Table 5. SyN takes more than 800 s, and even fast registration methods Demons and B-spline require more than 100 s, whereas the deep-learning-based methods require less than 35 s on the CPU, and in comparison, our AMNet requires 17 s. When registering a new image pair, the required iterative optimization in the classical registration methods is the major reason for the computational cost. Furthermore, AMNet shows a huge boost when implemented on a GPU and requires the least amount of time (~0.35 s). This is mainly due to the fact that our AMNet performs coarse-to-fine registration in one forward pass through the forward propagation of the deformation fields. These results of AMNet demonstrate that the combined multi-level optimization strategy and adaptive ability provide a good balance of efficiency and accuracy.

Second, we evaluated the model complexity in terms of the number of parameters and floating point operations (FLOPs) from five deep-learning-based methods. The number of parameters in the model is related to the number of channels and the depth of the network. FLOPs are related to the number of channels, depth of the network, and size of the image input to the network. These methods, as shown in Table 6, can be implemented in single-scale (i.e., VM (Balakrishnan et al., 2019)) or multi-scale (i.e., CM (Kim et al., 2021), LapIRN (Mok and Chung, 2020), HDAR (Huang et al., 2021), and AMNet) fashion. The model parameters of CM are captured from both global and local networks. Table 6 displays the number of parameters and FLOPs when our AMNet has the most number of layers at the highest resolution level. In Table 6, CM and HDAR have lower FLOPs because of the

Table 5

Average computational times (in seconds) of different registration algorithms for registering each pair of images in LPBA40_{syn}.

Method	SyN	Demons	B-spline	VM	CM	LapIRN	HDAR	AMNet
CPU	847 ± 86	153 ± 28	121 ± 13	19 ± 4	31 ± 6	17 ± 3	22 ± 4	17 ± 2
GPU	-	-	-	0.581 ± 0.086	1.183 ± 0.092	0.369 ± 0.071	0.840 ± 0.067	0.351 ± 0.057

Table 6

The complexity of the five deep-learning-based methods. “Parameters” is the number of parameters, “FLOPs” is the floating point operations, “Levels” represents the number of image resolutions (or scales) the model was trained on, and the corresponding input image size at each level is shown in “Input size”.

	VM	CM	LapIRN	HDAR	AMNet
Parameters	0.396M	1.446M	0.924M	17.635M	0.823M
FLOPs	142.082G	218.684G	507.599G	0.551G	417.124G
Levels	1	2	3	3	4
Input size	160 × 184 × 144	160 × 192 × 224	160 × 184 × 144	56 × 56 × 56	160 × 184 × 144
-	-	64 × 64 × 64	80 × 92 × 72	56 × 56 × 56	80 × 92 × 72
-	-	-	40 × 46 × 36	56 × 56 × 56	40 × 46 × 36
-	-	-	-	-	20 × 23 × 18

patch-based inputs. In addition, compared with other multi-level-based methods (i.e., CM, LapIRN, and HDAR), our AMNet involves fewer parameters. This is mainly because AMNet_{li} is a level-by-level progressive growing network and the model parameters at the lower resolution level are shared by the higher resolution level. Compared with the simple single-level-based method VM (Balakrishnan et al., 2019), our model complexity is not the best, however, the registration performance (accuracy and smoothness) far exceeds the VM (see Table 1). This is because the VM used the simplest U-Net network, resulting in low complexity but sacrificing performance.

5. Discussion

The experimental validations demonstrated that our AMNet is able to (1) improve the convergence speed of the model, (2) capture deformations at various degrees of complexity, and retain the continuity of the deformation field, and (3) achieve high-performance registration.

5.1. Adaptive growth strategy

Our AMNet with an adaptive growth strategy contributes to improving the convergence speed of the model and registration accuracy. This finding has been validated by the experiments on the LPBA40 dataset (see Fig. 8 and Table 4). On the one hand, AMNet requires fewer iterations to achieve convergence in comparison to AMNet_{static} at each level. This is because the depth of the network is adaptive to the scale of the input image, which effectively avoids the problem of slow model convergence caused by the redundancy of the network layer. On the other hand, AMNet achieves higher registration accuracy in comparison to AMNet_{static}, which means that the adaptive network learns deformation fields at multiple levels in one forward pass and avoids cumulative errors caused by multiple branches.

5.2. Information importance adaptation

Our AMNet is able to capture deformations of varying complexity and retain the continuity of the deformation field, which is mainly attributed to the characteristics of information importance adaptation, including resolution adaptation, frequency adaptation, and local importance adaptation. (1) The contribution of resolution adaptation has been justified by the experiments on six datasets. As shown in Table 1, the single-level-based method VM (Balakrishnan et al., 2019) obtained the deformation field from one resolution and produced lower DSCs with higher negative JD values, which means that the large deformation is hard to correct due to lacking initial transformation estimation for optimization. In comparison, with the help of resolution adaptation, our AMNet improved registration performance by capturing the deformation field from low-resolution images to high-resolution images. (2)

Table 7

Quantitative evaluation of AMNet on the testing datasets with and without artificial homogeneous regions. “W/o” represents the testing images without artificially homogeneous regions, and “W” represents the testing images with artificially homogeneous regions.

	DSC	JD ≤ 0	ASSD
W/o	0.783 ± 0.014	0.281 ± 0.107	0.825 ± 0.182
W/	0.783 ± 0.011	0.287 ± 0.091	0.826 ± 0.202

The contribution of frequency adaptation has been justified by the ablation experiment, which removed the “Energy Map Computing” phase (see Fig. 7). With high-frequency energy map guidance, AMNet can adaptively learn the accurate deformation of the structure edge at high-resolution levels. (3) The contribution of local importance adaptation has been justified by the ablation experiment, which removed the “Local Importance Computing” phase (see Table 3). With local importance guidance, AMNet can adaptively focus on difficult-to-register regions with updated importance based on the deformation complexity at the high-resolution level and achieve high-performance registration for images with various degrees of deformation. Interestingly, as observed from Table 3, as the degree of image warping increases, the adaptive AMNet with local importance has a stronger learning capacity for large deformations and generates smoother deformation fields with fewer folding voxels.

Furthermore, we conducted additional validation to validate the effectiveness of information importance adaptation in challenging cases, including artificial homogeneous regions, large deformations caused by Alzheimer’s disease, and large deformations caused by craniotomy. First, the figure on the left in Fig. 11 shows the visualization of our method for artificial homogeneous regions and large deformations caused by Alzheimer’s disease. In detail, homogeneous regions are artificially added to twelve images from the OASIS testing dataset. Five of the images are from patients with Alzheimer’s disease, and seven are from healthy subjects. As illustrated in the oval areas on the left in Fig. 11, the ellipses with different colors represent different homogeneous regions. As shown in the difference map on the left in Fig. 11, there is no obvious gray-value difference in these homogeneous regions but with anatomical heterogeneity (see the oval areas in moving and fixed segmentation maps on the left in Fig. 11). From the visualization of the registered image on the left in Fig. 11, these homogeneous and large-displacement regions are successfully aligned with the corresponding structures in the fixed image by our adaptive multi-level framework AMNet. The quantitative results in Table 7 demonstrate that our registration model is robust with homogeneous regions and large deformations on the images. Besides, the figure on the right in Fig. 11 shows the visualization of our method for one case study, which is from images with large deformations caused by craniotomy.

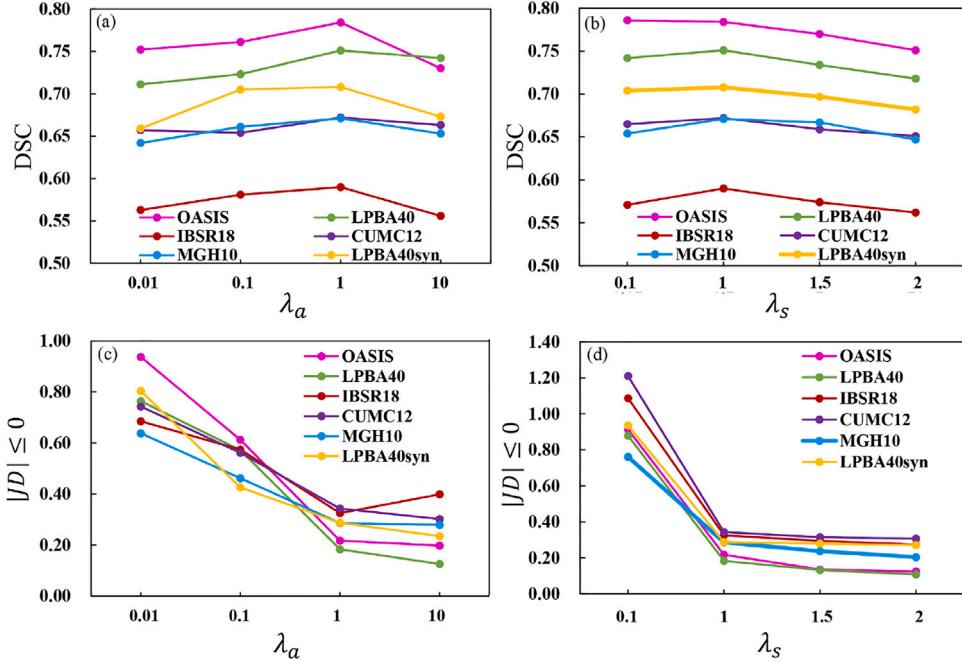


Fig. 10. The results are produced by the AMNet under different weighting parameters. (a) The DSC by changing λ_a when fixing λ_s to 1. (b) The DSC by changing λ_s when fixing λ_a to 1. (c) The percentage of non-positive Jacobian determinants by changing λ_a when fixing λ_s to 1. (d) The percentage of non-positive Jacobian determinants by changing λ_s when fixing λ_a to 1.

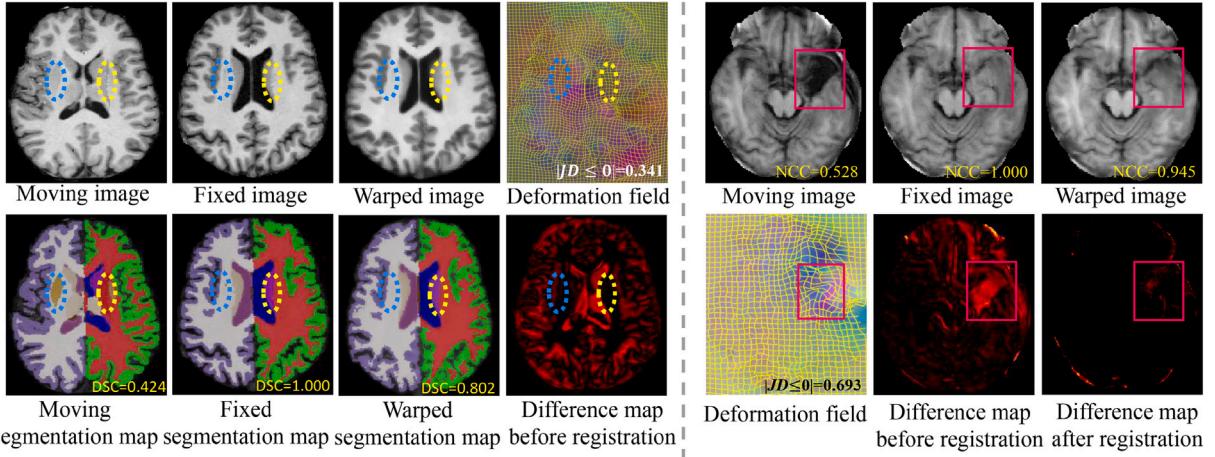


Fig. 11. Left panel: Registration visualization of images with artificial homogeneous regions and large deformations caused by disease progression. The yellow and blue ellipses represent the two homogeneous regions, respectively. Right panel: Registration visualization of images with large deformations caused by craniotomy. The red boxes marked the regions of the tumor. Difference maps are obtained from the absolute difference between the moving/warped image and the fixed image.

The fixed image is obtained from pre-craniotomy, and the moving image is obtained from post-craniotomy. Since there is no anatomical structure correspondence between the tumor (pre-craniotomy) and resection cavity (post-craniotomy), Dice is not suitable for evaluating the registration accuracy for this circumstance. We used normalized cross-correlation (NCC) (Balakrishnan et al., 2018) to evaluate the registration accuracy of this case. We can observe that the regions with the large mass effects in the pre-craniotomy image and the resection cavities in the post-craniotomy image are still registered successfully. In a word, our method is able to perform the accurate registration for challenging images with different complexity benefits from the joint action of the three adaptive mechanisms (i.e., resolution adaptation, frequency adaptation, and local importance adaptation).

5.3. Multi-level registration based on wavelet transform

Our AMNet achieved high-performance registration by comprehensively utilizing image information based on wavelet transformation. The results in Table 1 demonstrated the effectiveness of the multi-level registration using the wavelet pyramid. Compared with the state-of-the-art patch-wise multi-scale methods (Lei et al., 2020; Huang et al., 2021; Kim et al., 2021) related to our work, they have achieved attractive performance in large or complex deformation tasks. Kim et al. (2021) and Lei et al. (2020) designed the multi-scale registration frameworks by concatenating an image-wise global network and a patch-wise local network. Huang et al. (2021) presented the difficulty-aware multi-scale model, including three sub-networks for image-level coarse registration and patch-level fine registration to automatically

identify hard-to-register regions. However, these methods leveraged a cascade framework with multiple deep networks to learn global-to-local contextual information and optimized the local deformation field in a patch-based manner. Such bulky network structures and the splitting and combination of patches jeopardized the convergence speed of the model and the continuity of the deformation field. Furthermore, by comparing with the patch-based ablation experiment, PatchNet, AMNet always considered the full field of view at all levels, which achieved better continuity, as shown in Table 1 and Fig. 9. In other words, it can be confirmed that local deformations optimized on the whole image can effectively improve the registration accuracy of large/difficult deformations and ensure the continuity of the deformation field at all levels. Compared with the image-wise multi-resolution registration method, LapIRN (Mok and Chung, 2020), our AMNet achieved better registration results owing to the information fidelity characteristics of the wavelet transformation. LapIRN performed slightly better than our method in terms of the smoothness of the deformation field. This is because the diffeomorphic constraint of LapIRN excessively controls the smoothness of the deformation field, but results in sub-optimal registration accuracy. This also fully demonstrates that our method can simultaneously achieve high-performance registration and the smoothness of the deformation field.

5.4. Future work

As we effectively evaluated the registration performance of the proposed AMNet, the major future work is to realize the more challenging registration task for 3D brain images based on AMNet to enhance the universality of our method. Specifically, we will focus on (1) the registration of multi-modal images (e.g., CT and PET); and (2) the registration of images containing pathologies (e.g., obtained by pre-, intra-, and post-craniotomy).

6. Conclusion

In this paper, we have introduced an effective adaptive registration network, AMNet, for deformations with various degrees of complexity. With our novel adaptation mechanism, the proposed AMNet adaptively transfers focus from global deformation to hard-to-register local deformation based on information importance adaptation, and accelerates the convergence for network training with an adaptive increase in the network depth. Constructed with wavelets, our AMNet can more completely utilize image information from different frequencies and resolutions, and more importantly, is capable of better maintaining the continuity of the deformation field. The experimental results demonstrated the superior registration performance of our AMNet on six testing datasets with large/complex deformations. In addition, our method can maintain performance when handling varying degrees of deformation, whereas the comparison methods encountered a large performance degradation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [Nos. 81972160 and 81622025], and the Startup Funds for Talents at Beijing Normal University, China.

References

- Andersen, E.S., Noe, K.Ø., Sørensen, T.S., Nielsen, S.R.K., Fokdal, L., Paludan, M., Lindegaard, J.C., Tanderup, K., 2013. Simple DVH parameter addition as compared to deformable registration for bladder dose accumulation in cervix cancer brachytherapy. *Radiat. Oncol.* 107 (1), 52–57.
- Andersson, J., Smith, S., Jenkinson, M., 2008. FNIRT–FMRIB’s non-linear image registration tool. *Hum. Brain Mapp.* 2008.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12 (1), 26–41.
- Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54 (3), 2033–2044.
- Bajcsy, R., Kovačič, S., 1989. Multiresolution elastic matching. *Comput. Vis. Graph. Image Process.* 46 (1), 1–21.
- Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2018. An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9252–9260.
- Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: a learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* 38 (8), 1788–1800.
- Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., Shen, D., 2018a. Deep learning based inter-modality image registration supervised by intra-modality similarity. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 55–63.
- Cao, X., Yang, J., Zhang, J., Wang, Q., Yap, P.-T., Shen, D., 2018b. Deformable image registration using a cue-aware deep regression network. *IEEE Trans. Biomed. Eng.* 65 (9), 1900–1911.
- Chandrashekara, R., Rao, A., Sanchez-Ortiz, G.I., Mohiaddin, R.H., Rueckert, D., 2003. Construction of a statistical model for cardiac motion analysis using nonrigid image registration. In: Biennial International Conference on Information Processing in Medical Imaging. Springer, pp. 599–610.
- Che, T., Zheng, Y., Sui, X., Jiang, Y., Cong, J., Jiao, W., Zhao, B., 2019. DGR-net: Deep groupwise registration of multispectral images. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 706–717.
- Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2018. Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 729–738.
- De Silva, T., Uneri, A., Ketcha, M., Reuangamornrat, S., Kleinszig, G., Vogt, S., Aygun, N., Lo, S., Wolinsky, J., Siewersden, J., 2016. 3D–2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch. *Phys. Med. Biol.* 61 (8), 3009.
- Duan, L., Yuan, G., Gong, L., Fu, T., Yang, X., Chen, X., Zheng, J., 2019. Adversarial learning for deformable registration of brain MR image using a multi-scale fully convolutional network. *Biomed. Signal Process. Control* 53, 101562.
- Epnenhof, K.A., Lafarge, M.W., Veta, M., Pluim, J.P., 2019. Progressively trained convolutional neural networks for deformable image registration. *IEEE Trans. Med. Imaging* 39 (5), 1594–1604.
- Fan, J., Cao, X., Yap, P.-T., Shen, D., 2019. Birnet: Brain image registration using dual-supervised fully convolutional networks. *Med. Image Anal.* 54, 193–206.
- Fischer, B., Modersitzki, J., 2003. FLIRT: A flexible image registration toolbox. In: International Workshop on Biomedical Image Registration. Springer, pp. 261–270.
- Fischl, B., 2012. Freesurfer. *Neuroimage* 62 (2), 774–781.
- Gefen, S., Tretiak, O., Nissanov, J., 2003. Elastic 3-D alignment of rat brain histological images. *IEEE Trans. Med. Imaging* 22 (11), 1480–1489.
- Ghasemzadeh, A., Demirel, H., 2018. 3D discrete wavelet transform-based feature extraction for hyperspectral face recognition. *IET Biom.* 7 (1), 49–55.
- Guariglia, E., Silvestrov, S., 2016. Fractional-wavelet analysis of positive definite distributions and wavelets on $D'(C)$. In: Engineering Mathematics II. Springer, pp. 337–353.
- Haskins, G., Kruger, U., Yan, P., 2020. Deep learning in medical image registration: a survey. *Mach. Vis. Appl.* 31 (1), 1–18.
- Hering, A., van Ginneken, B., Heldmann, S., 2019. Mlvirnet: Multilevel variational image registration network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 257–265.
- Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Noble, J.A., Barratt, D.C., Vercauteren, T., 2018. Label-driven weakly-supervised learning for multimodal deformable image registration. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 1070–1074.
- Huang, Y., Ahmad, S., Fan, J., Shen, D., Yap, P.-T., 2021. Difficulty-aware hierarchical convolutional neural networks for deformable registration of brain MR images. *Med. Image Anal.* 67, 101817.
- Ito, M., Ino, F., 2018. An automated method for generating training sets for deep learning based image registration. In: BIOIMAGING. pp. 140–147.
- Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* 28.
- Kim, B., Kim, D.H., Park, S.H., Kim, J., Lee, J.-G., Ye, J.C., 2021. CycleMorph: Cycle consistent unsupervised deformable image registration. *Med. Image Anal.* 71, 102036.

- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al., 2009a. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46 (3), 786–802.
- Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2009b. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* 29 (1), 196–205.
- Lee, G.M., Seo, K.D., Song, H.J., Park, D.G., Ryu, G.H., Sagong, M., Park, S.H., 2020. Unsupervised learning model for registration of multi-phase ultra-widefield fluorescein angiography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 201–210.
- Lei, Y., Fu, Y., Wang, T., Liu, Y., Patel, P., Curran, W.J., Liu, T., Yang, X., 2020. 4D-CT deformable image registration using multiscale unsupervised deep learning. *Phys. Med. Biol.* 65 (8), 085003.
- Li, S., Peng, J., Kwok, J.T., Zhang, J., 2006. Multimodal registration using the discrete wavelet frame transform. In: 18th International Conference on Pattern Recognition (ICPR'06), Vol. 3. IEEE, pp. 877–880.
- Li, L., Xu, M., Liu, H., Li, Y., Wang, X., Jiang, L., Wang, Z., Fan, X., Wang, N., 2019. A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE Trans. Med. Imaging* 39 (2), 413–424.
- Liu, X., Jiang, D., Wang, M., Song, Z., 2019. Image synthesis-based multi-modal image registration framework by using deep fully convolutional networks. *Med. Biol. Eng. Comput.* 57 (5), 1037–1048.
- Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W., 2018. Multi-level wavelet-CNN for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 773–782.
- Lowekamp, B.C., Chen, D.T., Ibáñez, L., Blezek, D., 2013. The design of simpleitk. *Front. Neuroinform.* 7, 45.
- Luo, C., Li, Y., Lin, K., Chen, G., Lee, S.-J., Choi, J., Yoo, Y.F., Polley, M.O., 2020. Wavelet synthesis net for disparity estimation to synthesize dslr calibre bokeh effect on smartphones. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2407–2415.
- Maas, A.L., Hannun, A.Y., Ng, A.Y., 2013. Rectifier nonlinearities improve neural network acoustic models. In: Proc. Icml, Vol. 30. Citeseer, p. 3.
- Mok, T.C., Chung, A.C., 2020. Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 211–221.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch.
- Qu, L., Zhang, Y., Wang, S., Yap, P.-T., Shen, D., 2020. Synthesized 7T MRI from 3T MRI via deep learning in spatial and wavelet domains. *Med. Image Anal.* 62, 101663.
- Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* 18 (8), 712–721.
- Sarrut, D., 2006. Deformable registration for image-guided radiation therapy. *Z. Med. Phys.* 16 (4), 285–297.
- Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39 (3), 1064–1080.
- Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21 (11), 1421–1439.
- Shen, Z., Han, X., Xu, Z., Niethammer, M., 2019. Networks for joint affine and non-parametric image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4224–4233.
- Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248.
- Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B.P., Işgum, I., Staring, M., 2017. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 232–239.
- Sotiras, A., Davatzikos, C., Paragios, N., 2013. Deformable medical image registration: A survey. *IEEE Trans. Med. Imaging* 32 (7), 1153–1190.
- Stergiou, C., Mihir, S., Maria, V., Guillaume, C., Marie-Pierre, R., Stavroula, M., Nikos, P., 2018. Linear and deformable image registration with 3d convolutional neural networks. In: Image Analysis for Moving Organ, Breast, and Thoracic Images. Springer, pp. 13–22.
- Thirion, J.-P., 1998. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med. Image Anal.* 2 (3), 243–260.
- Torbati, N., Ayatollahi, A., 2019. A transformation model based on dual-tree complex wavelet transform for non-rigid registration of 3D MRI images. *Int. J. Wavelets Multiresolut. Inf. Process.* 17 (04), 1950025.
- Velev, M., Moseley, J.L., Eccles, C.L., Craig, T., Sharpe, M.B., Dawson, L.A., Brock, K.K., 2011. Effect of breathing motion on radiotherapy dose accumulation in the abdomen using deformable registration. *Int. J. Radiat. Oncol. Biol. Phys.* 80 (1), 265–272.
- Vercauteran, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* 45 (1), S61–S72.
- de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Işgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* 52, 128–143.
- de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Işgum, I., 2017. End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, pp. 204–212.
- Wang, C., Ren, Q., Qin, X., Yu, Y., 2018. Adaptive diffeomorphic multiresolution demons and their application to same modality medical image registration with large deformation. *Int. J. Biomed. Imaging* 2018.
- Yang, X., Akbari, H., Halig, L., Fei, B., 2011. 3D non-rigid registration using surface and local salient features for transrectal ultrasound image-guided prostate biopsy. In: Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling, Vol. 7964. International Society for Optics and Photonics, p. 79642V.
- Yang, X., Ghafourian, P., Sharma, P., Salman, K., Martin, D., Fei, B., 2012. Nonrigid registration and classification of the kidneys in 3D dynamic contrast enhanced (DCE) MR images. In: Medical Imaging 2012: Image Processing, Vol. 8314. International Society for Optics and Photonics, p. 83140B.
- Zavorin, I., Le Moigne, J., 2005. Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery. *IEEE Trans. Image Process.* 14 (6), 770–782.
- Zhang, J., 2018. Inverse-consistent deep networks for unsupervised deformable image registration. arXiv preprint arXiv:1809.03443.
- Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al., 2019a. Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10600–10610.
- Zhao, S., Lau, T., Luo, J., Eric, I., Chang, C., Xu, Y., 2019b. Unsupervised 3D end-to-end medical image registration with volume tweening network. *IEEE J. Biomed. Health Inf.* 24 (5), 1394–1404.
- Zheng, Y., Sui, X., Jiang, Y., Che, T., Zhang, S., Yang, J., Li, H., 2021. SymReg-GAN: Symmetric image registration with generative adversarial networks. *IEEE Trans. Pattern Anal. Mach. Intell..*