





# MSCAReg-Net: Multi-scale complexity-aware convolutional neural network for deformable image registration

Hu Yu<sup>1</sup>  | Qiang Zheng<sup>1</sup>  | Fang Hu<sup>2</sup> | Chaoqing Ma<sup>1,2</sup>  | Shuo Wang<sup>3,4</sup> | Shuai Wang<sup>5</sup> 

<sup>1</sup>School of Computer and Control Engineering, Yantai University, Yantai, Shandong Province, China

<sup>2</sup>Key Laboratory of Medical Imaging and Artificial Intelligence of Hunan Province, Xiangnan University, Chenzhou, Hunan Province, China

<sup>3</sup>Yantai University Trier College of Sustainable Technology, Yantai University, Yantai, Shandong Province, China

<sup>4</sup>Trier University of Applied Sciences, Trier, Germany

<sup>5</sup>Department of Radiology, Binzhou Medical University Hospital, Binzhou, Shandong Province, China

## Correspondence

Shuai Wang, Department of Radiology, Binzhou Medical University Hospital, Binzhou 256603, China.  
Email: shuashuaizi86@163.com

## Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 61802330, 61802331; Open Project of Key Laboratory of Medical Imaging and Artificial Intelligence of Hunan Province, Xiangnan University, Grant/Award Number: YXZN2022002

## Abstract

Deep learning-based image registration (DLIR) has been widely developed, but it remains challenging in perceiving small and large deformations. Besides, the effectiveness of the DLIR methods was also rarely validated on the downstream tasks. In the study, a multi-scale complexity-aware registration network (MSCAReg-Net) was proposed by devising a complexity-aware technique to facilitate DLIR under a single-resolution framework. Specifically, the complexity-aware technique devised a multi-scale complexity-aware module (MSCA-Module) to perceive deformations with distinct complexities, and employed a feature calibration module (FC-Module) and a feature aggregation module (FA-Module) to facilitate the MSCA-Module by generating more distinguishable deformation features. Experimental results demonstrated the superiority of the proposed MSCAReg-Net over the existing methods in terms of registration accuracy. Besides, other than the indices of Dice similarity coefficient (DSC) and percentage of voxels with non-positive Jacobian determinant ( $|J_\phi| \leq 0$ ), a comprehensive evaluation of the registration performance was performed by applying this method on a downstream task of multi-atlas hippocampus segmentation (MAHS). Experimental results demonstrated that this method contributed to a better hippocampus segmentation over other DLIR methods, and a comparable segmentation performance with the leading SyN method. The comprehensive assessment including DSC,  $|J_\phi| \leq 0$ , and the downstream application on MAHS demonstrated the advances of this method.

## 1 | INTRODUCTION

Deformable image registration is a critical procedure in a number of medical image analysis tasks [1], and widely applied in multi-atlas based image segmentation [2], radiotherapy assessment [3] and computer-assisted surgery [4]. Deformable registration aims to conduct anatomical correspondence between fixed image and moving image by establishing a non-linear transformation. Since conventional medical image registration algorithms [5] suffer complex computation caused by iterative optimization, deep learning approaches have been widely devel-

oped for fast image registration by adopting convolutional neural network (CNN) to learn spatial transformation under supervised [6], weakly supervised [7] or unsupervised learning framework [8]. However, deep learning-based deformable image registration remains a challenging task, especially for images with complicated deformations (small and large deformations) introduced by inter-person variations or disease progression. Besides, the effectiveness of deep learning-based deformable image registration when applied on the downstream medical image analysis tasks is also worthy of investigation regarding balancing the registration accuracy and diffeomorphism.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

## 1.1 | Related work

### 1.1.1 | Single-resolution registration

In single-resolution registration methods, the image-wise [8] or patch-wise [9] strategy is adopted to estimate the deformation field in image registration. Specifically, the typical VoxelMorph method employed an image-wise CNN to capture contextual information and was applied on 3D MRI brain image registration [8]. Ma et al. [10] designed an efficient locality preserving matching method to maintain the local neighbourhood structures of those potential true matches and used it to solve image matching problems. Following this progress, a bidirectional deformation estimation registration network with adaptive feature integration called SuperFusion [11] is proposed for image registration, image fusion, and semantic segmentation tasks. To accelerate the training procedure, Fan et al. [9] employed divide-and-conquer strategy to divide the image into multiple patches for training. To get rid of patches irrelevant to the foreground, Yang et al. [12] proposed to increase the stride of the sliding window to lessen the patch units. Although image-wise or patch-wise registration is effective to capture global contextual information or localized information respectively, each of them is still insufficient in simultaneously perceiving deformations with distinct complexities, which is difficult but desired when using single-resolution registration.

### 1.1.2 | Multi-resolution registration

Multi-resolution registration methods adopted a coarse-to-fine strategy for global and local feature distillation, further advancing the power of registration models in simultaneously aligning small and large deformations. Conventional Symmetric Normalization (SyN) method [5] which is established on the multi-resolution have been widely applied and integrated in a variety of medical image processing toolbox [13]. Inspired by conventional multi-resolution registration, some deep learning-based methods are proposed, including multi-level variational image registration network [14], deep Laplacian Pyramid Image Registration Network (LapIRN) [15] etc. However, the deformation interpolation errors which are inevitable in multi-resolution registration will affect registration accuracy significantly due to its propagation and accumulation, resulting in a limited deformation alignment.

## 1.2 | Contributions

In this study, a multi-scale complexity-aware registration network (MSCAReg-Net) under the single-resolution registration framework was proposed to automatically identify deformations with distinct complexities. The main contributions of our work are summarized as follows:

- (i) MSCAReg-Net: It was conducted by devising the multi-scale complexity-aware module (MSCA-Module) and cas-

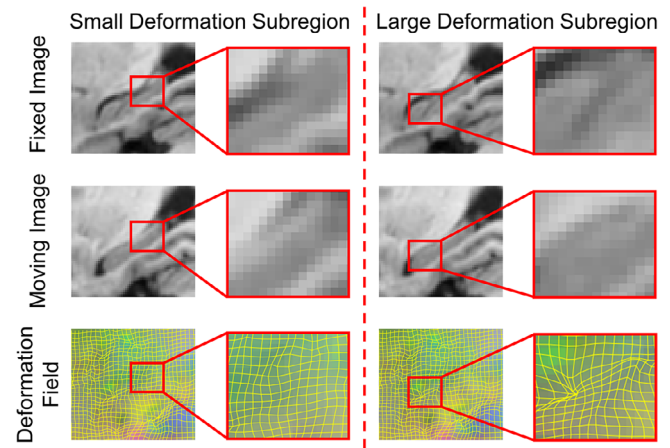
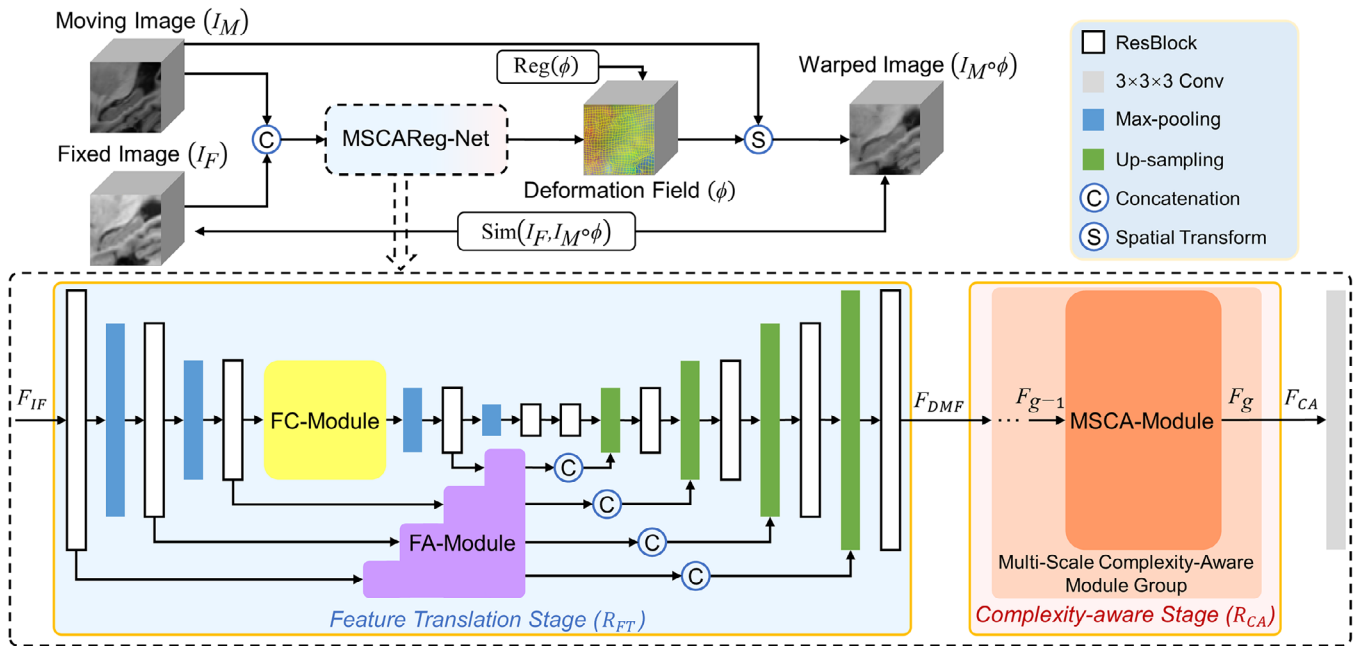


FIGURE 1 Illustration of small and large deformation subregions.

coding it to an upgraded U-Net block. The U-Net block, which integrated a feature calibration module (FC-Module) in the encoder pathway and employed a feature aggregation module (FA-Module) in the skip connection, was devised for better feature representation and deformation learning. The FC-Module and FA-Module merged in the U-Net block further facilitated the MSCA-Module in perceiving small and large deformations by generating more distinguishable deformation features.

- (ii) MSCA-Module: Being constructed by dilated convolutions and attention mechanisms, MSCA-Module is able to capture the deformations with distinct complexities through convolution of different receptive fields, allowing for further optimization of the hard-to-register regions.
- (iii) Diffeomorphic registration (MSCAReg-Net-diff): A diffeomorphic variant of MSCAReg-Net, named MSCAReg-Net-diff, was also accomplished to guarantee the diffeomorphism of the deformation field.
- (iv) Comprehensive assessment of registration performance by applying it on a downstream medical image analysis task: Typically, medical image registration is evaluated by dice similarity coefficient (DSC) and percentage of voxels with non-positive Jacobian determinant  $|J_\phi| \leq 0$  [8, 16]. Other than indices of DSC and  $|J_\phi| \leq 0$ , we performed a comprehensive assessment of registration performance by applying the proposed MSCAReg-Net on multi-atlas hippocampus segmentation (MAHS) to systematically explore the interactive impact of the DSC and diffeomorphism on subsequent medical image analysis task. Of note, the MAHS was chosen for three reasons: ① The hippocampus has been shown to produce large deformations in certain subregions due to individual differences [17]. Furthermore, the hippocampus is of dynamic alterations over progressing from normal control (NC) through mild cognitive impairment (MCI) to Alzheimer's disease (AD) [18], resulting in a complex small and large deformation across the hippocampal subregions (Figure 1). Thus, it can provide evidence for the proposed MSCAReg-Net in perceiving deformations with distinct complexities. ② In the MAHS, image



**FIGURE 2** Illustration of the proposed MSCAReg-Net architecture. The feature translation stage integrated the FC-Module and FA-Module in a U-Net block to facilitate the feature representation. The cascaded complexity-aware stage employed the MSCA-Module to perceive small and large deformations simultaneously. FC-Module = feature calibration module, FA-Module = feature aggregation module, MSCA-Module = multi-scale complexity-aware module.

registration of multi-atlas images to target image is a vital step which significantly affects the subsequent segmentation performance [2]. Therefore, it can be served as a proper medical image analysis task to evaluate image registration in application. © As a typical medical image analysis task, MAHS is well-known and relatively mature, where a variety of label fusion methods including majority voting (MV) [19], non-local patch (NLP) [20], random local binary pattern (RLBP) [21], metric learning (ML) [22], supervised random forests (RF) [23], and RF-SSLP [23] are established on deformable image registration and can be employed for a comprehensive assessment of registration performance.

The remainder of the paper is organized as follows. In Section 2, the proposed registration method is presented. Section 3 is the experimental setup. Section 4 displays the experimental results. In Sections 5 and 6, the discussion and conclusions are given.

## 2 | METHOD

### 2.1 | Multi-scale complexity-aware registration network (MSCAReg-Net)

#### 2.1.1 | Overview

Figure 2 illustrates the architecture of the proposed MSCAReg-Net, which is comprised of a feature translation stage and a complexity-aware stage. Specifically, the feature translation stage was established on a U-Net block to translate the original image features into deformation magnitude features by employing the

FC-Module in the contracting path and the FA-Module in the skip connection. Both modules were to generate more consolidated features for the subsequent complexity-aware stage. As for the U-Net block, the residual block (ResBlock) substituted the conventional convolution operation to enhance the representational capacity, and tackle the gradient vanishing as network deepening [24]. Following to the U-Net block, the MSCA-Module in the complexity-aware stage was cascaded and fed with the deformation magnitude features generated in the feature translation stage, to perceive small and large deformations simultaneously.

In the feature translation stage, U-Net [25] was employed to translate the original image features  $F_{IF}$  into deformation magnitude feature  $F_{DMF}$ . The feature translation function  $R_{FT}(\cdot)$  was defined as follows:

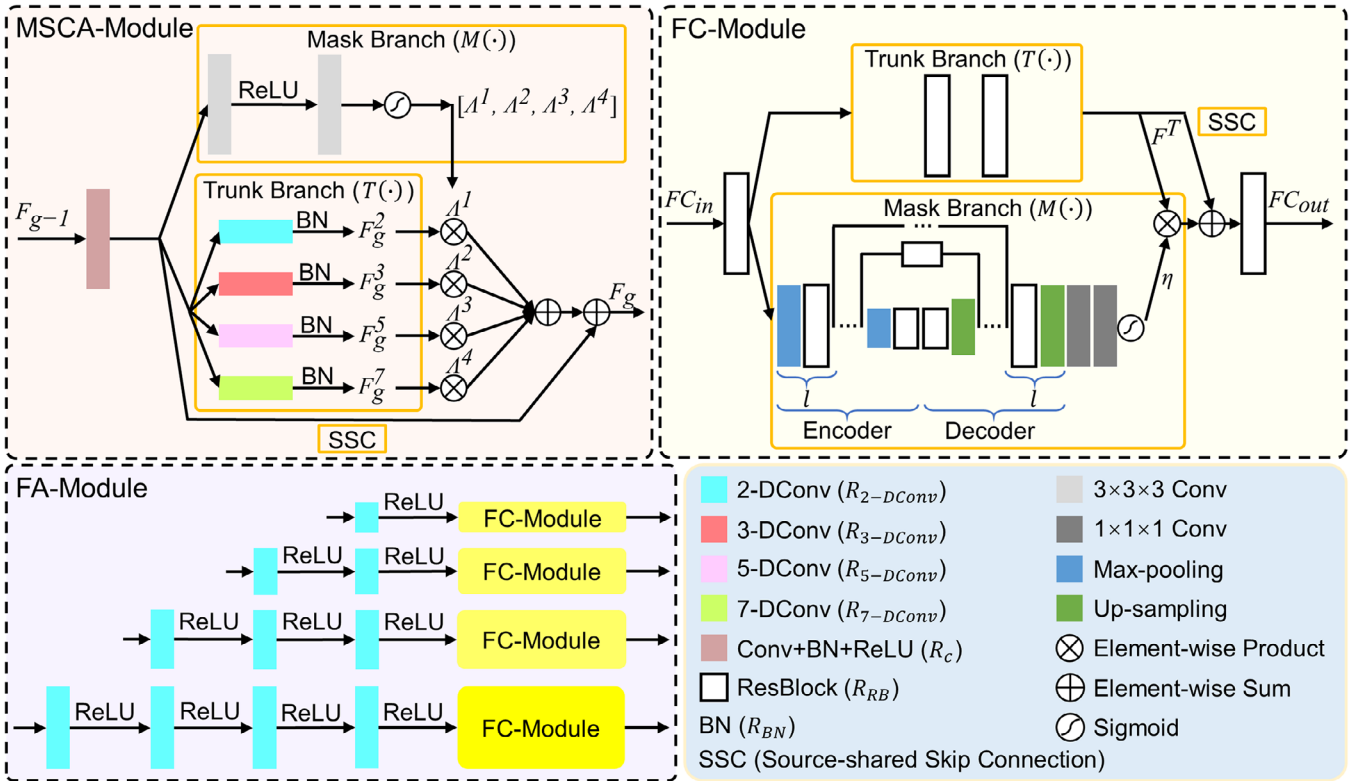
$$F_{DMF} = R_{FT}(F_{IF}) \quad (1)$$

In the complexity-aware stage, a group of MSCA-Modules were designed to optimize the deformation magnitude feature  $F_{DMF}$ , which consisted of  $G = 3$  MSCA-Modules. The MSCA-Module could be used to improve the registration accuracy of image regions with both small and large deformations

$$F_{CA} = R_{CA}(F_{DMF}), \quad (2)$$

where  $R_{CA}(\cdot)$  and  $F_{CA}$  denoted the complexity-aware function and its optimized features. Finally,  $F_{CA}$  was fed to a  $3 \times 3 \times 3$  convolution with stride 1 to generate the deformation field (DF)

$$\phi = R_{DF}(F_{CA}) = R_{MSCAReg-Net}(F_{IF}),$$



**FIGURE 3** Illustration of the multi-scale complexity-aware module (MSCA-Module), feature calibration module (FC-Module), and feature aggregation module (FA-Module). 2-DConv, 3-DConv, 5-DConv and 7-DConv denote dilated convolution with dilation rates of 2, 3, 5 and 7. Conv and BN stand for convolution and batch normalization operations.

where  $R_{DF}(\cdot)$  denotes the convolution operation used to generate the deformation field, the  $R_{MSCAReg-Net}(\cdot)$  denoted the unified function of the proposed MSCAReg-Net.

The proposed MSCAReg-Net was carried out under the unsupervised training setting [8]. The fixed and moving images were concatenated into a 2-channel feature map, and the dense deformation field was learned by optimizing the intensity similarity and smoothing constraints. Finally, the optimized deformation field was applied to spatially transform the moving image into the warped image. In the following, the MSCA-Module, FC-Module, and FA-Module in the proposed MSCAReg-Net were described in detail (Figure 3). Of note, both FC-Module and FA-Module were to facilitate deformations feature learning of the U-Net block, further consolidating the identification of deformation complexity by the MSCA-Module.

### 2.1.2 | Multi-scale complexity-aware module (MSCA-Module)

In the convolution procedure, the size of the receptive field is highly associated with deformation feature representation [26]. To leverage small and large receptive fields for identifying deformations with distinct complexities, a common strategy is to adopt pooling operation, which downsamples feature maps to attain receptive fields in different sizes. However, succes-

sive pooling operations will lose precise spatial positions in the feature map, which is not a good practice for medical image registration tasks. This motivates us to design a new module to achieve receptive fields in different sizes.

Inspired by [27, 28], the MSCA-Module in Figure 3 was devised to be aware of deformations with distinct complexities by combining a trunk branch, a mask branch, and a feature source-shared skip connection. The trunk branch was to identify deformations with distinct complexities. The mask branch was to construct the identity mapping. The feature source-shared skip connection was akin to the idea in residual learning [24], which alleviated gradient vanishing and stabilized training processing.

In our study, a total of  $G = 3$  MSCA-Modules were cascaded in the complexity-aware stage, and the  $g$ -th MSCA-Module was formulated as

$$F_g = R_C(F_{g-1}) + R_{MA}(M(R_C(F_{g-1})), T(R_C(F_{g-1}))), \quad (4)$$

where  $F_{g-1}$  and  $F_g$  represented the input and output of the  $g$ -th MSCA-Module.  $R_C(\cdot)$  was a composite operation of convolution, batch normalization (BN), and ReLU function [29] in the beginning of the MSCA-Module, and fed to the subsequent mask branch, trunk branch, and the feature source-shared skip connection. Besides,  $T(\cdot)$  and  $M(\cdot)$  were the functions of the trunk branch and the mask branch, and the outputs of them were integrated by a composite multiplication and



addition operation  $R_{MA}(\cdot, \cdot)$ . The output of  $R_{MA}(\cdot)$  was then added with the output of  $R_C(\cdot)$  by the feature source-shared skip connection, generating the final output  $F_g^r$  of the  $g$ -th MSCA-Module.

In the trunk branch, the dilated convolutions with dilation rates of 2, 3, 5 and 7 were defined to capture the deformations with distinct complexities, thereby generating feature maps with different deformation magnitudes. The trunk branch  $T(\cdot)$  was formulated as

$$F_g^r = T(R_C(F_{g-1})) = R_{BN}(R_{r-DCov}(R_C(F_{g-1}))) \quad (5)$$

$(r = 2, 3, 5, 7),$

where  $R_{BN}(\cdot)$  was BN operation,  $R_{r-DCov}(\cdot)$  denoted the dilated convolution with dilation rate  $r$ .  $R_C(\cdot)$  was a composite operation of convolution, BN, and ReLU function in the beginning of the MSCA-Module and fed to the trunk branch.

In the mask branch, the attention matrices ( $\Lambda^1, \Lambda^2, \Lambda^3$ , and  $\Lambda^4$ ) were learned and multiplied with feature maps of  $F_g^r$  ( $r = 2, 3, 5, 7$ ). The mask branch  $M(\cdot)$  was formulated as

$$\Lambda^i = M(R_C(F_{g-1})) = f(R_{CR}(R_C(F_{g-1}))) (i = 1, 2, 3, 4), \quad (6)$$

where  $R_{CR}(\cdot)$  were convolution and ReLU operations,  $f(\cdot)$  was a sigmoid gating operation.  $R_C(\cdot)$  was a composite operation of convolution, BN, and ReLU function in the beginning of the MSCA-Module and fed to the mask branch.

Since image regions with large deformations are often difficult to align [17], the FC-Module and FA-Module were devised in the feature translation stage to strengthen the deformation representation ability of the U-Net block, thereby further enhancing the ability of MSCA-Module in perceiving complex deformations (Figure 2).

### 2.1.3 | Feature calibration module (FC-Module)

In the feature translation stage, successive pooling operations in the encoder lose precise voxel position information, resulting in representation biases of deformation features. Herein, the FC-Module defined as follows was devised and cascaded in the encoder pathway to calibrate the representation bias of deformation (Figure 3) [30]. The FC-Module was formulated as

$$FC_{out} = R_{RB}((1 + M(R_{RB}(FC_{in}))) \cdot T(R_{RB}(FC_{in}))) \quad (7)$$

The FC-Module was initiated with a ResBlock formulated as  $R_{RB}(FC_{in})$ , where  $FC_{in}$  was the input of the FC-Module. The generated features were then fed to the following trunk branch  $T(\cdot)$  and mask branch  $M(\cdot)$  to obtain the characterized deformation features  $F^T = T(R_{RB}(FC_{in}))$  and to construct the attention mapping  $\eta = M(R_{RB}(FC_{in}))$ , respectively. Through a feature source-shared skip connection, the multiplication of deformation features  $F^T$  and attention mapping  $\eta$  was then added again with deformation features  $F^T$ . The attention mapping  $\eta$  can effectively distinguish deformations with distinct

complexities [17], which facilitates specific optimization of large and small deformations in the deformation feature  $F^T$ . Finally, the output features of the FC-Module  $FC_{out}$  were generated by applying a ResBlock operation  $R_{RB}(\cdot)$  [31].

In the trunk branch, the original deformation features are learned by two residual blocks [24]. In the mask branch, high-level features are extracted by using the encoder-decoder architecture [25] to generate the attention mapping  $\eta$ , and the attention mapping was to boost beneficial features and suppress irrelevant features when multiplied with the deformation features  $F^T$  from trunk branch. Specifically, the encoder-decoder architecture is composed of ResBlock, Max-pooling, Up-sampling, and two  $1 \times 1 \times 1$  convolution layers, and finally passed through a sigmoid gating operation to generate the attention mapping  $\eta$ . Of note, the encoder-decoder architecture employed in the mask branch was adjustable in term of parameter  $l$  according to the feature level it would be placed in the encoder pathway. The parameter  $l$  determined the depth of the encoder-decoder architecture in the mask branch. In the FC-Module, the parameter  $l$  was set to 3 according to the feature level in the encoder pathway that the FC-Module inserted (Figure 2). In the following FA-Module which adopted multiple FC-Modules at different feature levels for skip connection, the parameter  $l$  was set 4, 3, 2 and 1 to satisfy the semantic gap between encoder and decoder pathways.

### 2.1.4 | Feature aggregation module (FA-Module)

In the feature translation stage, the feature map in the encoder was proximate to the original image. In contrary, the feature map in the decoder was adjacent to the deformation field. This led to a large semantic gap [9] that demanded feature aggregation between encoder and decoder pathways. Herein, the FA-Module (Figure 3), established on dilated convolution and FC-Modules, was integrated into the skip connections of the encoder-decoder block in Figure 2 to aggregate the feature maps at different levels and alleviate the semantic gap between encoder and decoder pathways [30]. In the FA-Module, the dilated convolution with a same dilation rate of 2 was adopted, but the number of the dilated convolutions was determined by the feature level in the skip connection. In our study, a total of 10 dilated convolutions was employed in the FA-Module. Due to the gradual reduction of semantic gap with the features going deeper, the amount of dilated convolution decreases from 4 to 1 accordingly. Also, the followed FC-Modules adopted at different levels assigned parameter  $l = 4, 3, 2$  and 1 as mentioned above with the features going deeper.

### 2.1.5 | Loss function

To attain the deformation field  $\phi$ , the following energy function was optimized by maximizing the following formulation:

$$\phi = \underset{\phi}{\operatorname{argmax}} \operatorname{Sim}(I_F, I_M \circ \phi) + \operatorname{Reg}(\phi), \quad (8)$$

where  $\text{Sim}(I_F, I_M \circ \phi)$  defined the similarity between the fixed image  $I_F$  and the warped image  $I_M \circ \phi$ .  $\text{Reg}(\phi)$  refers to the regularization of the deformation field for preserving the smoothness of  $\phi$ . In our study, normalized cross correlation (NCC) [7] was adopted as a similarity metric, which calculated the spatial gradients to assess the smoothness of the deformation field. Therefore, the loss function for the proposed MSCAReg-Net was defined as follows:

$$L(I_F, I_M, \phi) = -NCC(I_F, I_M \circ \phi) + \lambda \|\nabla \phi(p)\|^2, \quad (9)$$

where  $\lambda$  controls the balance between image alignment and deformation field smoothness. Specifically, for  $\nabla \phi(p) = (\frac{\partial \phi(p)}{\partial x}, \frac{\partial \phi(p)}{\partial y}, \frac{\partial \phi(p)}{\partial z})$ , we approximate  $\frac{\partial \phi(p)}{\partial x} \approx \phi((p_x + 1, p_y, p_z)) - \phi((p_x, p_y, p_z))$ , and use similar approximations for  $\frac{\partial \phi(p)}{\partial y}$  and  $\frac{\partial \phi(p)}{\partial z}$ .

## 2.2 | Diffeomorphic deformation (MSCAReg-Net-diff)

In medical image registration, topology preservation and transformation invertibility are critical properties, called diffeomorphism. The diffeomorphism is a smooth and continuous one-to-one mapping with invertible derivatives (i.e. non-zero Jacobian determinant). In this study, we parameterized the MSCAReg-Net into a diffeomorphic deformation variant (MSCAReg-Net-diff) using the stationary velocity field under the Log-Euclidean framework. The MSCAReg-Net-diff was optimized within the space of diffeomorphic maps. Specifically, the diffeomorphic deformation field  $\phi$  was defined as  $\frac{d\phi_t}{dt} = v(\phi^t)$ , subject to  $\phi^{(0)} = Id$  was the identity transformation and  $t$  was time. The integration of the (smooth) stationary velocity field  $v$  was implemented over unit time to obtain the final registration field. Details of the deformation diffeomorphism can be found in the study [16].

## 2.3 | Comprehensive assessment of registration performance by applying on a latter medical image analysis task

DSC and  $|J_\phi| \leq 0$  have been widely used to evaluate the registration performance. For instance, Balakrishnan et al. [8] developed a CNN-based unsupervised registration algorithm VoxelMorph, which used DSC to evaluate the registration accuracy. Dalca et al. [16] introduced a diffeomorphic integration layer to ensure a diffeomorphism in the deformation, and used  $|J_\phi| \leq 0$  to evaluate diffeomorphism. Mok et al. [32] demonstrated that DSC and  $|J_\phi| \leq 0$  were highly interactive during registration, that is, one rose as the other fell. However, many studies typically improve DSC on the premise of ensuring nearly zero  $|J_\phi| \leq 0$ . This raises a new question: How to balance the indices of DSC and  $|J_\phi| \leq 0$  when evaluating the deformable

registration methods particularly in the following medical image analysis tasks?

To investigate this, we conducted a comprehensive assessment of the registration performance not only by the DSC,  $|J_\phi| \leq 0$ , and average running time in seconds (Time) but also applying on a latter medical image analysis task of MAHS. Specifically, seven MAHS methods were employed, including MV [19], NLP [20], RLBP [21], ML [22], RF [23] and RF-SSLP [23]. The parameters associated with the above MAHS methods were set as suggested in each individual study. In addition, nine metrics were used to evaluate the hippocampus segmentation as follows [23]. Given manual segmentation label A and the automated segmentation result B, these metrics were calculated as:

$$\text{DSC} = \frac{2V(A \cap B)}{V(A) + V(B)}, \text{Jaccard} = \frac{V(A \cap B)}{V(A \cup B)},$$

$$\text{Precision} = \frac{V(A \cap B)}{V(B)}, \text{Recall} = \frac{V(A \cap B)}{V(A)},$$

$$\text{MD} = \text{mean}(\min_{e \in \partial A} \min_{f \in \partial B} d(e, f)),$$

$$\text{HD} = \max(H(A, B), H(B, A)), \text{ where,}$$

$$H(A, B) = \max_{e \in \partial A} (\min_{f \in \partial B} d(e, f)),$$

$$\text{ASSD} = (\text{mean}(\min_{e \in \partial A} \min_{f \in \partial B} d(e, f)) + \text{mean}(\min_{e \in \partial B} \min_{f \in \partial A} d(e, f))) / 2,$$

$$\text{RMSD} = \frac{\sqrt{D_A^2 + D_B^2}}{\text{card}\{\partial A\} + \text{card}\{\partial B\}},$$

where  $D_A^2 = \sum_{e \in \partial A} (\min_{f \in \partial B} d(e, f))^2$ .

In the above metrics,  $d(\cdot, \cdot)$  was the Euclidian distance between two points,  $\text{card}\{\cdot\}$  was the cardinality of a set.

## 3 | EXPERIMENTAL SETTINGS

### 3.1 | Data and pre-processing

A total of 135 subjects with brain structural MR images and manual hippocampal segmentation labels were obtained from the ADNI-Harp project ([www.hippocampal-protocol.net](http://www.hippocampal-protocol.net)) [33], including 100 subjects from a preliminary release and 35 subjects from a final release. Of the preliminary release, two subjects (002\_S\_0938, 127\_S\_0259) missed manual segmentation labels at several slices and were excluded. Finally, we obtained 133 subjects, of which, the preliminary release data were split into a training set (80 subjects) and a validation set (18 subjects), the final release data (35 subjects) were served as a testing set.

For each MR scan, we performed the N4 correction and affined registration to the Montreal Neurological Institute (MNI) 152 space ( $1 \times 1 \times 1 \text{ mm}^3$ ) with the Advanced Normalization Tools (ANTs) ([www.github.com/ANTsX/ANTs](http://www.github.com/ANTsX/ANTs)) [13]. After aligning to the MNI 152 space, a bounding box ( $60 \times 60 \times 48$ ) entirely covering the hippocampus was defined to

reduce the computational complexity [34]. The hippocampus registration was performed on the hippocampus patch (the above bounding box sized of  $60 \times 60 \times 48$ ), which was same as the registration step adopted in MAHS methods for subsequent comparison.

Since data augmentation could enhance deep learning model, it was also adopted in this study as follows:

- (i) For each subject, the right hippocampus was flipped into the left due to the spatial similarity of the bilateral hippocampus.
- (ii) In the training and validation sets, the fixed image and moving image were selected with an  $n \times (n-1)$  strategy in the left or flipped right hippocampus separately. This procedure generated  $80 \times 79 \times 2 = 12,640$  pairs and  $18 \times 17 \times 2 = 612$  pairs of images for training and validating the MSCAReg-Net. In the test set, the same mutual information (MI) as adopted in MAHS was employed for atlas selection and to pair the test data for a fair comparison on the segmentation performance. Specifically, each left or right hippocampus (fixed images) selected 20 atlases (moving images) according to the MI value, generating  $35 \times 20 \times 2 = 1400$  pairs of images for testing.

### 3.2 | Baseline methods

In this study, the following baseline methods were employed for comparison:

- (i) One conventional registration method SyN [5]: SyN utilizes a 3-level multi-resolution strategy to capture deformations with distinct complexities, which has been demonstrated with the top-performing registration performance among 14 classical non-linear deformation algorithms [35]. In our study, the SyN was implemented with following command: `ANTS 3 -m CC[fixed,moving,1,2] -t SyN[0.25] -r Gauss[3,0] -o output -i 100x100x100 -number-of-affine-iterations 100x100x100 [23]`.
- (ii) Four deep learning-based single-resolution methods: VoxelMorph [8], diffeomorphic variant VoxelMorph-diff [16], SymTrans [36], and SymTrans-diff [36].
- (iii) Two deep learning-based multi-resolution methods: LapIRN and diffeomorphic variant LapIRN-diff [15].

For deep learning-based methods, the official open-source implementation with default parameters was employed for training from the scratch.

### 3.3 | Implementation details of proposed MSCAReg-Net

The proposed MSCAReg-Net was implemented using Pytorch on a single Nvidia GeForce RTX 2080 Ti GPU, which was optimized by the AdamW optimizer with a learning rate  $= 1 \times 10^{-4}$ , mini-batch size 4,  $\lambda = 1$ , and 316k iterations.

## 4 | RESULTS

In this section, we (1) demonstrated the effectiveness of the complexity-aware technique in improving registration performance in Section 4.1; (2) compared the registration performance of MSCAReg-Net and its diffeomorphic variant MSCAReg-Net-diff with baseline methods in Section 4.2; (3) performed a comprehensive evaluation of the proposed registration model by applying it on the subsequent medical image analysis task of MSHS in Section 4.3.

### 4.1 | The effectiveness of the proposed complexity-aware technique

To validate the effectiveness of the complexity-aware technique in perceiving deformations with distinct complexities, we compared the proposed MSCAReg-Net with its two alternatives as follows:

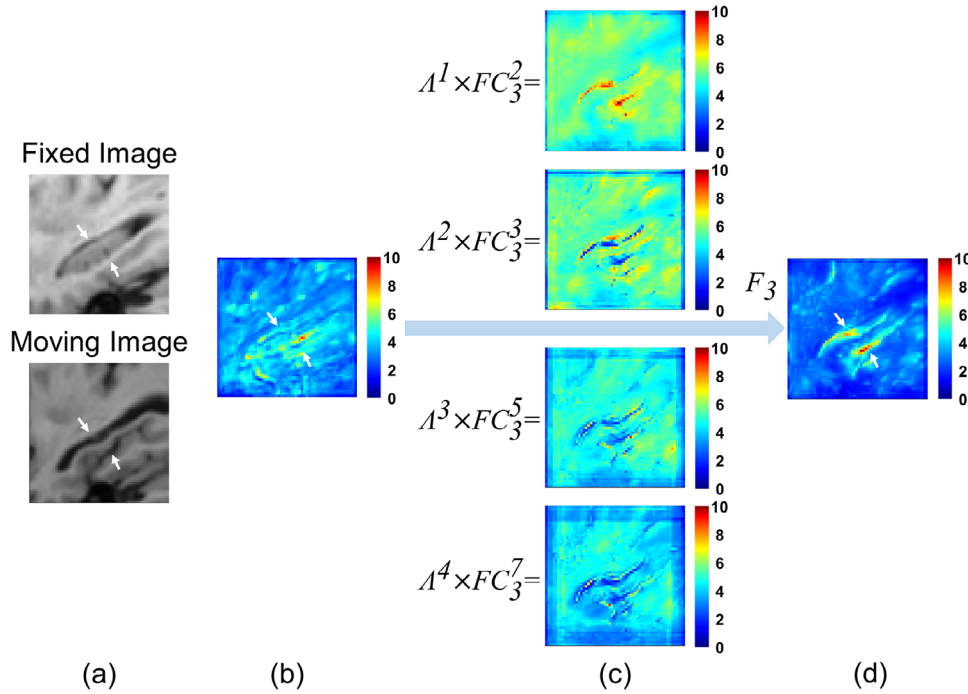
- (i) MSCAReg-Net-1: The U-Net block in Figure 2 but without MSCA-Module, FC-Module and FA-Module.
- (ii) MSCAReg-Net-2: The U-Net block in Figure 2 solely with MSCA-Module, which was to demonstrate the effectiveness of the proposed complexity-aware technique.
- (iii) MSCAReg-Net: The proposed MSCAReg-Net with MSCA-Module, FC-Module and FA-Module simultaneously, which was to demonstrate the contribution of FC-Module and FA-Module in enhancing MSCA-Module in identifying large deformations.

Table 1 summarizes the registration results of left and right hippocampus by comparing MSCAReg-Net-1, MSCAReg-Net-2 and the proposed MSCAReg-Net, indicating that the absence of the designed modules results in lower DSC. Comparing MSCAReg-Net-2 versus MSCAReg-Net-1, the MSCA-Module was demonstrated in improving registration by the complexity-aware technique. Comparing with the MSCAReg-Net-2, the proposed MSCAReg-Net achieved the best performance since the FC-Module and FA-Module facilitated the complexity-aware technique established on MSCA-Module in identifying both small and large deformations.

Figure 4 visualized the advances of the complexity-aware technique by illustrating the corresponding features maps of deformation magnitude. Based on two paired images with large spatial and morphological differences in Figure 4a, Figure 4b,d compared the final output of deformation feature maps generated by the MSCAReg-Net-1 above and the proposed MSCAReg-Net. The high and low values in the features maps correspond to large and small deformation magnitude. The experimental results demonstrated the superiority of the complexity-aware technique in precisely characterizing large and small deformations simultaneously. Figure 4c displayed the deformation feature maps generated by multiplying mask branch and trunk branch in the MSCA-Module, demonstrating the ability of the MSCA-Module in capturing small and large

**TABLE 1** The DSC,  $|J_\phi| \leq 0$ , and time (mean  $\pm$  std) computed over left and right hippocampus using the proposed MSCAReg-Net and its alternatives.

|               | Left hippocampus                   |                                    |                                    | Right hippocampus                  |                                    |                                    |
|---------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
|               | DSC                                | $ J_\phi  \leq 0$ (%)              | Time                               | DSC                                | $ J_\phi  \leq 0$ (%)              | Time                               |
| MSCAReg-Net-1 | 0.768 $\pm$ 0.06                   | 1.774 $\pm$ 0.50                   | <b>0.023 <math>\pm</math> 0.00</b> | 0.781 $\pm$ 0.04                   | 1.646 $\pm$ 0.42                   | <b>0.023 <math>\pm</math> 0.00</b> |
| MSCAReg-Net-2 | 0.782 $\pm$ 0.06                   | 1.793 $\pm$ 0.49                   | 0.106 $\pm$ 0.02                   | 0.794 $\pm$ 0.04                   | 1.665 $\pm$ 0.41                   | 0.106 $\pm$ 0.02                   |
| MSCAReg-Net   | <b>0.790 <math>\pm</math> 0.05</b> | <b>1.681 <math>\pm</math> 0.48</b> | 0.103 $\pm$ 0.00                   | <b>0.802 <math>\pm</math> 0.04</b> | <b>1.549 <math>\pm</math> 0.38</b> | 0.104 $\pm$ 0.00                   |

**FIGURE 4** Illustration of the effectiveness of the MSCA-Module in identifying the deformation with distinct complexities. (a) Input image pair. (b) The feature map of deformation magnitude is derived from the sole U-Net block defined as MSCAReg-Net-1. (c) The feature maps of deformation magnitude are obtained by multiplying mask branch and trunk branch in the MSCA-Module. (d) The feature maps of deformation magnitude generated by MSCA-Module in the proposed MSCAReg-Net.

deformations by employing dilated convolutions with different dilation rates.

Figure 5 compared the feature maps of the MSCAReg-Net generated from the encoder pathway. When the MSCAReg-Net-1 was adopted, the feature map generated in the encoder pathway was exhibited in Figure 5b, indicating an insufficient representation of deformation features. In the proposed MSCAReg-Net, the feature map in Figure 5d generated after the FC-Module in the encoder pathway captured large and small deformations more precisely. The attention map  $\eta$  from the mask branch of the FC-Module was also exhibited in Figure 5c.

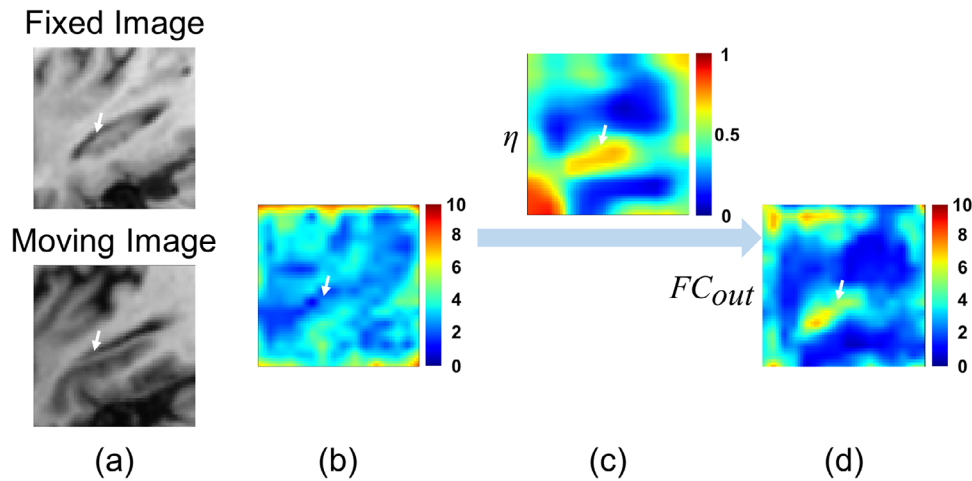
Figure 6 illustrated the training process in terms of NCC loss on training dataset and DSC on validation dataset. Comparing with MSCAReg-Net-1 and MSCAReg-Net-2 defined above, the proposed MSCAReg-Net achieved a more rapid convergence, lower loss value, and higher DSC, which benefited comprehen-

sively from the MSCA-Module, FC-Module, and FA-Module in the deformable image registration task.

## 4.2 | Comparison with baseline methods

The experimental comparison between the proposed MSCAReg-Net and baseline methods were summarized in Table 2. The deep learning-based single-resolution registration method of VoxelMorph and its diffeomorphic variant VoxelMorph-diff achieved relatively low DSC, indicating the difficulty of the single-resolution registration framework in capturing deformation with distinct complexities. Besides, the Transformer-based registration method SymTrans has a slight improvement in terms of DSC compared with VoxelMorph, but is lower than MSCAReg-Net, which is still limited to the



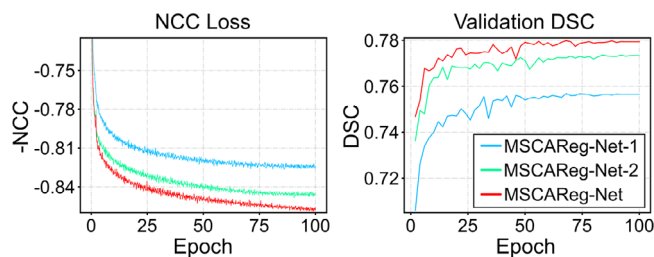


**FIGURE 5** Comparison of feature maps generated in the encoder pathway. (a) Input image pair. (b) The feature map of deformation magnitude generated from the encoder pathway of the MSCAReg-Net-1. (c) The attention map  $\eta$  derived from the mask branch of the FC-Module in the proposed MSCAReg-Net. (d) The deformation magnitude map  $FC_{out}$  generated from the encoder pathway of the proposed MSCAReg-Net.

**TABLE 2** The indices of DSC,  $|J_\phi| \leq 0$ , and time computed over left and right hippocampus of the different registration methods.

|                  | Left hippocampus                   |                                    |                                    | Right hippocampus                  |                                    |                                    |
|------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
|                  | DSC                                | $ J_\phi  \leq 0$ (%)              | Time                               | DSC                                | $ J_\phi  \leq 0$ (%)              | Time                               |
| Affine           | $0.489 \pm 0.12$                   | –                                  | –                                  | $0.522 \pm 0.10$                   | –                                  | –                                  |
| SyN              | $0.787 \pm 0.07$                   | $0.304 \pm 0.17$                   | $18.03 \pm 2.82$                   | $0.796 \pm 0.05$                   | $0.278 \pm 0.15$                   | $18.18 \pm 2.45$                   |
| VoxelMorph       | $0.739 \pm 0.07$                   | $0.701 \pm 0.33$                   | $0.016 \pm 0.00$                   | $0.757 \pm 0.05$                   | $0.650 \pm 0.26$                   | $0.016 \pm 0.00$                   |
| VoxelMorph-diff  | $0.733 \pm 0.07$                   | $0.362 \pm 0.34$                   | $0.031 \pm 0.00$                   | $0.750 \pm 0.05$                   | $0.320 \pm 0.28$                   | $0.031 \pm 0.00$                   |
| SymTrans         | $0.746 \pm 0.07$                   | $0.641 \pm 0.32$                   | $0.030 \pm 0.00$                   | $0.765 \pm 0.05$                   | $0.579 \pm 0.24$                   | $0.026 \pm 0.00$                   |
| SymTrans-diff    | $0.747 \pm 0.07$                   | $0.956 \pm 0.37$                   | $0.020 \pm 0.00$                   | $0.764 \pm 0.05$                   | $0.878 \pm 0.29$                   | $0.021 \pm 0.00$                   |
| LapIRN           | $0.764 \pm 0.06$                   | $4.435 \pm 1.44$                   | <b><math>0.011 \pm 0.00</math></b> | $0.780 \pm 0.04$                   | $4.540 \pm 1.43$                   | <b><math>0.012 \pm 0.00</math></b> |
| LapIRN-diff      | $0.711 \pm 0.08$                   | <b><math>0.240 \pm 0.24</math></b> | $0.014 \pm 0.00$                   | $0.733 \pm 0.06$                   | <b><math>0.260 \pm 0.25</math></b> | $0.015 \pm 0.00$                   |
| MSCAReg-Net      | <b><math>0.790 \pm 0.05</math></b> | $1.681 \pm 0.48$                   | $0.103 \pm 0.00$                   | <b><math>0.802 \pm 0.04</math></b> | $1.549 \pm 0.38$                   | $0.104 \pm 0.00$                   |
| MSCAReg-Net-diff | $0.781 \pm 0.06$                   | $1.255 \pm 0.63$                   | $0.109 \pm 0.00$                   | $0.794 \pm 0.04$                   | $1.081 \pm 0.49$                   | $0.110 \pm 0.00$                   |

Affine: Affine spatial normalization.

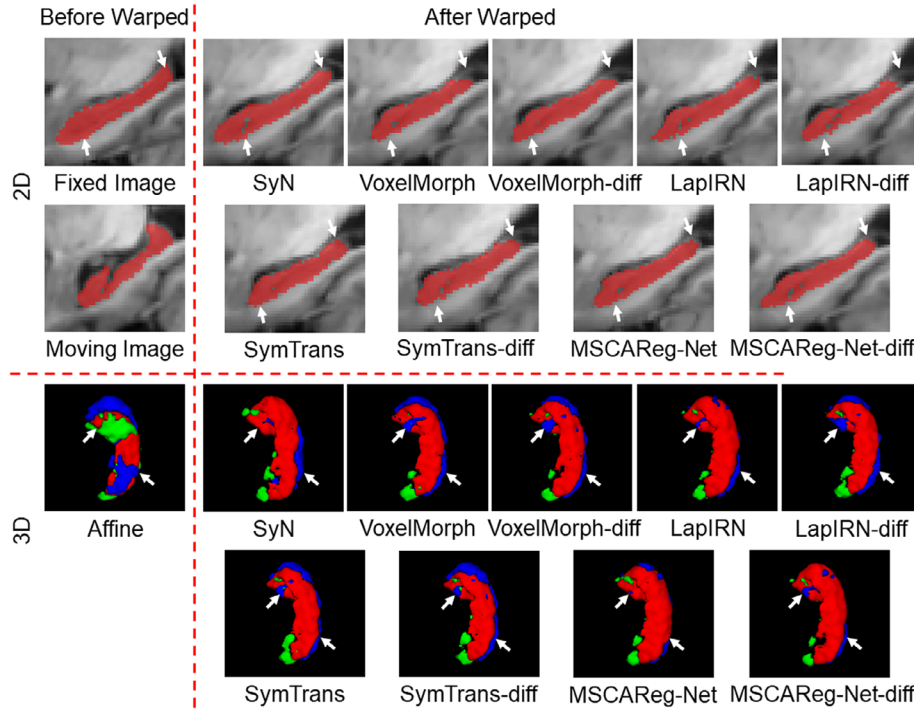


**FIGURE 6** Convergence analysis of the training stage of MSCAReg-Net and its alternatives in term of (left) NCC loss and (right) validation DSC.

single-resolution registration framework. The LapIRN under comparison improved the registration performance by adopting a deep learning-based multi-resolution framework to characterize both large and small deformations, but the performance of its diffeomorphic variant LapIRN decreased sharply.

Despite the complex computation, the conventional registration method of SyN which indeed took a multi-resolution strategy in its implementation achieved a relatively high value of DSC and low value of  $|J_\phi| \leq 0$ , exhibiting a stable and superior performance over VoxelMorph and LapIRN under comparison. Based on the comprehensive comparison with the conventional registration method of SyN, the deep learning-based VoxelMorph, and LapIRN, the proposed MSCAReg-Net achieved the highest DSC value and a moderate value of  $|J_\phi| \leq 0$ , demonstrating the effectiveness of the complexity-aware technique under the single-resolution registration framework in promoting deformable image registration. The 2D and 3D visualized comparison was also exhibited in Figure 7.

In this study, the proposed method achieves an advanced DSC over SyN from Table 2 but the value of  $|J_\phi| \leq 0$  between MSCAReg-Net and SyN exhibits a gap. Therefore, the comparison of the application of the registration methods on the



**FIGURE 7** Example registration results given by SyN, VoxelMorph, VoxelMorph-diff, LapIRN, LapIRN-diff, SymTrans, SymTrans-diff, MSCAReg-Net and MSCAReg-Net-diff. Blue: hand-crafted segmentation, Green: automatic segmentation, Red: overlap between hand-crafted and automatic segmentations.

following medical image analysis task is essential. Therefore, we conducted the following experiments by applying the registration methods under comparison on the subsequent MAHS task to further assess the proposed MSCAReg-Net.

### 4.3 | Applying the MSCAReg-Net on the subsequent multi-atlas hippocampus segmentation task

Table 3 summarizes the left and right hippocampus segmentation results by MAHS framework established on different registration methods. Since the MV segmentation method was directly derived from the wrapped masks of multi-atlas images, the MV segmentation result could directly reflect the DSC performance of registration methods. From Table 2, our MSCAReg-Net achieved the best DSC values over other registration methods, therefore the MV method established on the MSCAReg-Net also obtained the best performance on left and right hippocampus segmentation (Table 3).

For other label fusion methods such as NLP, RLBP, ML, RF and RF-SSLP, they were implemented on a narrowband which was generated by the MV method. Of note that the establishment of the narrowband would affect the subsequent label fusion result greatly. From Table 3, we can find that most diffeomorphic variants of the deep learning-based registration methods (i.e. VoxelMorph-diff, SymTrans-diff and MSCAReg-Net-diff) achieved superior segmentations of hippocampus over their raw version (i.e. VoxelMorph, SymTrans and MSCAReg-Net), demonstrating the positive role of diffeomorphism in

stabilizing the narrowband. Figure 8 displayed the narrowband established on different registration methods. The VoxelMorph, LapIRN and MSCAReg-Net exhibited outliers when conducting the narrowband, while the outliers were not found on their diffeomorphic variants.

Besides, the proposed MSCAReg-Net-diff achieved a competitive performance on the registration and subsequent segmentation task with the conventional but well-performing SyN method. The proposed method achieved an advanced DSC but a relative high  $|J_\phi| \leq 0$  over SyN, but the following hippocampus segmentation performance exhibited comparable. On the left hippocampus from Table 3, the SyN method resulted in slightly better segmentations on most label fusion methods. On the right hippocampus from Table 3, the proposed MSCAReg-Net-diff resulted in slightly better segmentation. Both MSCAReg-Net-diff and SyN led to better segmentations over other registration methods (Table 3). It indicated that although slightly high  $|J_\phi| \leq 0$  in the proposed MSCAReg-Net-diff, the advanced DSC performance over SyN could still help generate competitive segmentation results. The balance between the DSC and  $|J_\phi| \leq 0$  values could only be validated on the following medical image analysis task. Solely boosting the DSC value or decreasing the  $|J_\phi| \leq 0$  might not be approximate in devising deformable registration models.

## 5 | DISCUSSION

In the present study, a multi-scale complexity-aware network defined as MSCAReg-Net was proposed under the

**TABLE 3** Summary of DSC, Jaccard, Precision, Recall, MD, HD, HD95, ASSD and RMSD computed the results over left and right hippocampus of multi-atlas hippocampal segmentation based on the different registration models.

|                  | Left hippocampus    |                     |                     |                     |                     |                     |                     |                     |                     |                     | Right hippocampus   |                     |                     |                     |                     |                     |                     |                     |                     |  |
|------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|--|
|                  | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             |                     |  |
| DSC              | 0.859 ± 0.03        | <b>0.877 ± 0.02</b> | <b>0.883 ± 0.02</b> | 0.883 ± 0.02        | <b>0.885 ± 0.02</b> | <b>0.889 ± 0.02</b> | 0.863 ± 0.02        | <b>0.880 ± 0.01</b> | 0.885 ± 0.01        | 0.885 ± 0.01        | 0.885 ± 0.01        | 0.885 ± 0.01        | 0.863 ± 0.02        | <b>0.880 ± 0.01</b> | 0.885 ± 0.01        | 0.885 ± 0.01        | 0.885 ± 0.01        | 0.885 ± 0.01        | 0.890 ± 0.01        |  |
| VoxelMorph       | 0.818 ± 0.04        | 0.859 ± 0.03        | 0.871 ± 0.02        | 0.874 ± 0.02        | 0.873 ± 0.02        | 0.879 ± 0.02        | 0.831 ± 0.02        | 0.865 ± 0.02        | 0.876 ± 0.02        | 0.878 ± 0.02        | 0.878 ± 0.02        | 0.878 ± 0.02        | 0.831 ± 0.02        | 0.865 ± 0.02        | 0.876 ± 0.02        | 0.878 ± 0.02        | 0.878 ± 0.02        | 0.878 ± 0.02        | 0.884 ± 0.01        |  |
| VoxelMorph-diff  | 0.813 ± 0.04        | 0.861 ± 0.02        | 0.877 ± 0.02        | 0.877 ± 0.02        | 0.879 ± 0.02        | 0.884 ± 0.02        | 0.826 ± 0.02        | 0.867 ± 0.02        | 0.880 ± 0.01        | 0.882 ± 0.01        | 0.882 ± 0.01        | 0.882 ± 0.01        | 0.826 ± 0.02        | 0.867 ± 0.02        | 0.880 ± 0.01        | 0.882 ± 0.01        | 0.882 ± 0.01        | 0.882 ± 0.01        | 0.888 ± 0.01        |  |
| SymTrans         | 0.830 ± 0.04        | 0.865 ± 0.03        | 0.874 ± 0.03        | 0.876 ± 0.03        | 0.876 ± 0.03        | 0.881 ± 0.03        | 0.842 ± 0.03        | 0.871 ± 0.02        | 0.879 ± 0.02        | 0.882 ± 0.02        | 0.882 ± 0.02        | 0.882 ± 0.02        | 0.842 ± 0.03        | 0.871 ± 0.02        | 0.879 ± 0.02        | 0.882 ± 0.02        | 0.880 ± 0.02        | 0.880 ± 0.02        | 0.886 ± 0.02        |  |
| SymTrans-diff    | 0.830 ± 0.04        | 0.867 ± 0.03        | 0.878 ± 0.02        | 0.880 ± 0.02        | 0.879 ± 0.02        | 0.885 ± 0.02        | 0.842 ± 0.02        | 0.872 ± 0.02        | 0.881 ± 0.02        | 0.882 ± 0.02        | 0.882 ± 0.02        | 0.882 ± 0.02        | 0.842 ± 0.02        | 0.872 ± 0.02        | 0.881 ± 0.02        | 0.883 ± 0.02        | 0.882 ± 0.02        | 0.882 ± 0.02        | 0.888 ± 0.02        |  |
| LapIRN           | 0.846 ± 0.03        | 0.874 ± 0.02        | 0.878 ± 0.02        | 0.879 ± 0.02        | 0.878 ± 0.02        | 0.884 ± 0.02        | 0.856 ± 0.02        | 0.877 ± 0.02        | 0.881 ± 0.01        | 0.883 ± 0.02        | 0.881 ± 0.02        | 0.881 ± 0.02        | 0.856 ± 0.02        | 0.877 ± 0.02        | 0.881 ± 0.01        | 0.883 ± 0.02        | 0.881 ± 0.02        | 0.881 ± 0.02        | 0.886 ± 0.01        |  |
| LapIRN-diff      | 0.802 ± 0.06        | 0.859 ± 0.04        | 0.872 ± 0.03        | 0.873 ± 0.04        | 0.872 ± 0.04        | 0.879 ± 0.04        | 0.818 ± 0.03        | 0.867 ± 0.02        | 0.879 ± 0.01        | 0.881 ± 0.02        | 0.880 ± 0.01        | 0.880 ± 0.01        | 0.818 ± 0.03        | 0.867 ± 0.02        | 0.879 ± 0.01        | 0.881 ± 0.02        | 0.880 ± 0.01        | 0.880 ± 0.01        | 0.886 ± 0.01        |  |
| MSCAReg-Net      | <b>0.860 ± 0.03</b> | 0.874 ± 0.02        | 0.874 ± 0.02        | 0.876 ± 0.02        | 0.876 ± 0.02        | 0.881 ± 0.02        | 0.865 ± 0.02        | 0.877 ± 0.02        | 0.877 ± 0.02        | 0.877 ± 0.02        | 0.877 ± 0.02        | 0.865 ± 0.02        | 0.877 ± 0.02        | 0.877 ± 0.02        | 0.877 ± 0.02        | 0.877 ± 0.02        | 0.879 ± 0.02        | 0.879 ± 0.02        | 0.883 ± 0.02        |  |
| MSCAReg-Net-diff | 0.851 ± 0.03        | 0.875 ± 0.02        | 0.883 ± 0.02        | 0.884 ± 0.02        | 0.883 ± 0.02        | 0.888 ± 0.02        | 0.859 ± 0.02        | 0.879 ± 0.01        | 0.887 ± 0.01        | <b>0.887 ± 0.01</b> | <b>0.887 ± 0.01</b> | <b>0.887 ± 0.01</b> | 0.859 ± 0.02        | 0.879 ± 0.01        | <b>0.887 ± 0.01</b> | <b>0.887 ± 0.01</b> | <b>0.887 ± 0.01</b> | <b>0.887 ± 0.01</b> | <b>0.891 ± 0.01</b> |  |
| Jaccard          | 0.754 ± 0.05        | <b>0.782 ± 0.03</b> | <b>0.792 ± 0.03</b> | 0.792 ± 0.03        | <b>0.794 ± 0.03</b> | <b>0.802 ± 0.03</b> | 0.760 ± 0.03        | <b>0.786 ± 0.03</b> | 0.795 ± 0.02        | 0.795 ± 0.02        | 0.795 ± 0.02        | 0.760 ± 0.03        | <b>0.786 ± 0.03</b> | 0.795 ± 0.02        | 0.795 ± 0.02        | 0.795 ± 0.02        | 0.795 ± 0.02        | 0.795 ± 0.02        | 0.803 ± 0.02        |  |
| Syn              | 0.695 ± 0.05        | 0.755 ± 0.04        | 0.773 ± 0.04        | 0.777 ± 0.04        | 0.777 ± 0.04        | 0.785 ± 0.04        | 0.712 ± 0.03        | 0.763 ± 0.03        | 0.780 ± 0.03        | 0.784 ± 0.03        | 0.784 ± 0.03        | 0.712 ± 0.03        | 0.763 ± 0.03        | 0.780 ± 0.03        | 0.784 ± 0.03        | 0.784 ± 0.03        | 0.784 ± 0.03        | 0.784 ± 0.03        | 0.792 ± 0.03        |  |
| VoxelMorph       | 0.688 ± 0.06        | 0.757 ± 0.04        | 0.782 ± 0.03        | 0.783 ± 0.03        | 0.784 ± 0.03        | 0.794 ± 0.03        | 0.705 ± 0.03        | 0.766 ± 0.03        | 0.787 ± 0.02        | 0.789 ± 0.02        | 0.789 ± 0.02        | 0.705 ± 0.03        | 0.766 ± 0.03        | 0.787 ± 0.02        | 0.789 ± 0.02        | 0.789 ± 0.02        | 0.789 ± 0.02        | 0.789 ± 0.02        | 0.799 ± 0.02        |  |
| VoxelMorph-diff  | 0.711 ± 0.06        | 0.763 ± 0.05        | 0.777 ± 0.04        | 0.780 ± 0.04        | 0.780 ± 0.04        | 0.789 ± 0.04        | 0.728 ± 0.04        | 0.771 ± 0.03        | 0.785 ± 0.03        | 0.787 ± 0.03        | 0.787 ± 0.03        | 0.728 ± 0.04        | 0.771 ± 0.03        | 0.785 ± 0.03        | 0.787 ± 0.03        | 0.787 ± 0.03        | 0.787 ± 0.03        | 0.787 ± 0.03        | 0.796 ± 0.03        |  |
| SymTrans         | 0.712 ± 0.05        | 0.766 ± 0.04        | 0.783 ± 0.04        | 0.787 ± 0.03        | 0.785 ± 0.03        | 0.795 ± 0.04        | 0.727 ± 0.04        | 0.773 ± 0.04        | 0.788 ± 0.03        | 0.791 ± 0.03        | 0.789 ± 0.03        | 0.727 ± 0.04        | 0.773 ± 0.04        | 0.788 ± 0.03        | 0.791 ± 0.03        | 0.789 ± 0.03        | 0.789 ± 0.03        | 0.789 ± 0.03        | 0.798 ± 0.03        |  |
| VoxelMorph-diff  | 0.735 ± 0.05        | 0.777 ± 0.03        | 0.783 ± 0.03        | 0.785 ± 0.03        | 0.784 ± 0.03        | 0.793 ± 0.03        | 0.749 ± 0.03        | 0.781 ± 0.03        | 0.787 ± 0.03        | 0.788 ± 0.03        | 0.788 ± 0.03        | 0.749 ± 0.03        | 0.781 ± 0.03        | 0.787 ± 0.03        | 0.787 ± 0.03        | 0.788 ± 0.03        | 0.788 ± 0.03        | 0.788 ± 0.03        | 0.797 ± 0.03        |  |
| LapIRN           | 0.674 ± 0.07        | 0.755 ± 0.05        | 0.776 ± 0.05        | 0.778 ± 0.05        | 0.775 ± 0.05        | 0.786 ± 0.05        | 0.694 ± 0.05        | 0.766 ± 0.03        | 0.786 ± 0.03        | 0.786 ± 0.03        | 0.786 ± 0.03        | 0.694 ± 0.05        | 0.766 ± 0.03        | 0.786 ± 0.03        | 0.786 ± 0.03        | 0.786 ± 0.03        | 0.786 ± 0.03        | 0.786 ± 0.03        | 0.797 ± 0.02        |  |
| MSCAReg-Net      | <b>0.756 ± 0.04</b> | 0.778 ± 0.03        | 0.777 ± 0.04        | 0.781 ± 0.03        | 0.781 ± 0.03        | 0.788 ± 0.03        | 0.763 ± 0.03        | 0.781 ± 0.03        | 0.781 ± 0.03        | 0.785 ± 0.03        | 0.785 ± 0.03        | 0.763 ± 0.03        | 0.781 ± 0.03        | 0.781 ± 0.03        | 0.781 ± 0.03        | 0.785 ± 0.03        | 0.785 ± 0.03        | 0.785 ± 0.03        | 0.792 ± 0.03        |  |
| MSCAReg-Net-diff | 0.741 ± 0.04        | 0.778 ± 0.03        | 0.792 ± 0.03        | <b>0.793 ± 0.03</b> | 0.792 ± 0.03        | 0.800 ± 0.03        | 0.754 ± 0.03        | 0.784 ± 0.02        | <b>0.797 ± 0.02</b> | <b>0.798 ± 0.02</b> | <b>0.797 ± 0.02</b> | 0.754 ± 0.03        | 0.784 ± 0.02        | <b>0.797 ± 0.02</b> | <b>0.798 ± 0.02</b> | <b>0.797 ± 0.02</b> | <b>0.797 ± 0.02</b> | <b>0.797 ± 0.02</b> | <b>0.804 ± 0.02</b> |  |
| Precision        | 0.844 ± 0.04        | <b>0.866 ± 0.04</b> | 0.873 ± 0.03        | 0.876 ± 0.03        | <b>0.878 ± 0.03</b> | <b>0.886 ± 0.03</b> | 0.849 ± 0.04        | 0.870 ± 0.04        | 0.877 ± 0.03        | 0.879 ± 0.03        | 0.880 ± 0.03        | 0.849 ± 0.04        | 0.870 ± 0.04        | 0.877 ± 0.03        | 0.879 ± 0.03        | 0.880 ± 0.03        | 0.880 ± 0.03        | 0.880 ± 0.03        | 0.889 ± 0.03        |  |
| VoxelMorph       | 0.789 ± 0.06        | 0.837 ± 0.05        | 0.850 ± 0.04        | 0.859 ± 0.04        | 0.855 ± 0.04        | 0.865 ± 0.04        | 0.800 ± 0.04        | 0.844 ± 0.03        | 0.857 ± 0.03        | 0.863 ± 0.03        | 0.863 ± 0.03        | 0.800 ± 0.04        | 0.844 ± 0.03        | 0.857 ± 0.03        | 0.863 ± 0.03        | 0.863 ± 0.03        | 0.863 ± 0.03        | 0.863 ± 0.03        | 0.873 ± 0.03        |  |
| VoxelMorph-diff  | 0.797 ± 0.05        | 0.845 ± 0.04        | 0.865 ± 0.03        | 0.870 ± 0.03        | 0.869 ± 0.03        | 0.878 ± 0.03        | 0.806 ± 0.05        | 0.853 ± 0.03        | 0.868 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.806 ± 0.05        | 0.853 ± 0.03        | 0.868 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.882 ± 0.03        |  |
| SymTrans         | 0.805 ± 0.06        | 0.846 ± 0.05        | 0.856 ± 0.04        | 0.864 ± 0.04        | 0.861 ± 0.04        | 0.870 ± 0.04        | 0.819 ± 0.05        | 0.856 ± 0.04        | 0.865 ± 0.04        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.819 ± 0.05        | 0.856 ± 0.04        | 0.865 ± 0.04        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.878 ± 0.03        |  |
| VoxelMorph-diff  | 0.813 ± 0.05        | 0.853 ± 0.04        | 0.866 ± 0.04        | 0.873 ± 0.03        | 0.871 ± 0.04        | 0.879 ± 0.04        | 0.824 ± 0.04        | 0.861 ± 0.04        | 0.870 ± 0.03        | 0.877 ± 0.03        | 0.877 ± 0.03        | 0.824 ± 0.04        | 0.861 ± 0.04        | 0.870 ± 0.03        | 0.877 ± 0.03        | 0.877 ± 0.03        | 0.874 ± 0.03        | 0.874 ± 0.03        | 0.883 ± 0.03        |  |
| SymTrans-diff    | 0.835 ± 0.04        | 0.863 ± 0.04        | 0.864 ± 0.04        | 0.872 ± 0.03        | 0.868 ± 0.03        | 0.877 ± 0.04        | 0.842 ± 0.04        | 0.868 ± 0.03        | 0.870 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.842 ± 0.04        | 0.868 ± 0.03        | 0.870 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.873 ± 0.03        | 0.882 ± 0.03        |  |
| LapIRN           | 0.791 ± 0.06        | 0.850 ± 0.04        | 0.864 ± 0.06        | 0.873 ± 0.03        | 0.868 ± 0.03        | 0.878 ± 0.03        | 0.800 ± 0.05        | 0.855 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.800 ± 0.05        | 0.855 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.868 ± 0.03        | 0.883 ± 0.03        |  |
| LapIRN-diff      | <b>0.845 ± 0.04</b> | 0.862 ± 0.04        | 0.857 ± 0.04        | 0.864 ± 0.04        | 0.861 ± 0.04        | 0.871 ± 0.04        | <b>0.852 ± 0.04</b> | 0.865 ± 0.04        | 0.862 ± 0.04        | 0.862 ± 0.04        | 0.862 ± 0.04        | <b>0.852 ± 0.04</b> | 0.865 ± 0.04        | 0.862 ± 0.04        | 0.862 ± 0.04        | 0.862 ± 0.04        | 0.866 ± 0.03        | 0.866 ± 0.03        | 0.875 ± 0.03        |  |
| MSCAReg-Net      | 0.840 ± 0.04        | 0.865 ± 0.04        | <b>0.873 ± 0.03</b> | <b>0.880 ± 0.03</b> | 0.875 ± 0.03        | 0.884 ± 0.03        | 0.847 ± 0.04        | <b>0.871 ± 0.03</b> | <b>0.878 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | 0.847 ± 0.04        | <b>0.871 ± 0.03</b> | <b>0.878 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.890 ± 0.02</b> |  |
| MSCAReg-Net-diff | 0.840 ± 0.04        | 0.865 ± 0.04        | <b>0.873 ± 0.03</b> | <b>0.880 ± 0.03</b> | 0.875 ± 0.03        | 0.884 ± 0.03        | 0.847 ± 0.04        | <b>0.871 ± 0.03</b> | <b>0.878 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | 0.847 ± 0.04        | <b>0.871 ± 0.03</b> | <b>0.878 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.882 ± 0.03</b> | <b>0.890 ± 0.02</b> |  |

(Continues)

TABLE 3 (Continued)

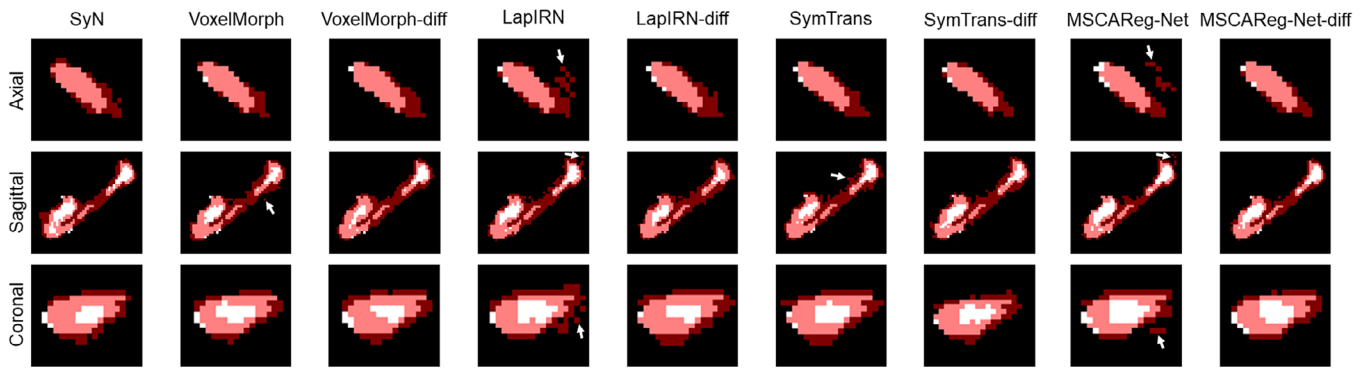
|                  | Left hippocampus    |                     |                     |                     |                     |                     |                     |                     |                     |                     | Right hippocampus   |                     |                     |                     |                     |                     |                     |                     |                     |  |
|------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|--|
|                  | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             |                     |  |
| Recall           | 0.879 ± 0.05        | <b>0.892 ± 0.03</b> | <b>0.896 ± 0.03</b> | <b>0.893 ± 0.03</b> | 0.894 ± 0.03        | 0.894 ± 0.03        | 0.882 ± 0.05        | <b>0.892 ± 0.04</b> | 0.896 ± 0.03        | 0.894 ± 0.04        | 0.894 ± 0.04        | 0.894 ± 0.03        | 0.882 ± 0.05        | <b>0.892 ± 0.04</b> | 0.896 ± 0.03        | 0.894 ± 0.04        | 0.894 ± 0.04        | 0.894 ± 0.04        | 0.894 ± 0.03        |  |
| VoxelMorph       | 0.857 ± 0.07        | 0.887 ± 0.04        | 0.895 ± 0.03        | 0.891 ± 0.04        | <b>0.895 ± 0.03</b> | <b>0.896 ± 0.03</b> | 0.870 ± 0.06        | 0.890 ± 0.04        | <b>0.898 ± 0.04</b> | 0.890 ± 0.04        | <b>0.897 ± 0.04</b> | <b>0.898 ± 0.04</b> | 0.870 ± 0.06        | 0.890 ± 0.04        | <b>0.898 ± 0.04</b> | <b>0.895 ± 0.04</b> | <b>0.897 ± 0.04</b> | <b>0.898 ± 0.04</b> | <b>0.898 ± 0.04</b> |  |
| VoxelMorph-diff  | 0.840 ± 0.09        | 0.881 ± 0.05        | 0.892 ± 0.04        | 0.887 ± 0.04        | 0.891 ± 0.04        | 0.893 ± 0.03        | 0.855 ± 0.07        | 0.885 ± 0.05        | 0.896 ± 0.04        | 0.891 ± 0.04        | 0.894 ± 0.04        | 0.896 ± 0.04        | 0.855 ± 0.07        | 0.885 ± 0.05        | 0.896 ± 0.04        | 0.892 ± 0.04        | 0.894 ± 0.04        | 0.896 ± 0.04        | 0.896 ± 0.04        |  |
| SymTrans         | 0.863 ± 0.07        | 0.888 ± 0.04        | 0.894 ± 0.04        | 0.891 ± 0.04        | 0.893 ± 0.04        | 0.895 ± 0.03        | 0.873 ± 0.06        | 0.889 ± 0.05        | 0.896 ± 0.04        | 0.891 ± 0.04        | 0.894 ± 0.04        | 0.896 ± 0.04        | 0.873 ± 0.06        | 0.889 ± 0.05        | 0.896 ± 0.04        | 0.894 ± 0.04        | 0.895 ± 0.04        | 0.897 ± 0.04        | 0.897 ± 0.04        |  |
| SymTrans-diff    | 0.854 ± 0.07        | 0.884 ± 0.05        | 0.893 ± 0.04        | 0.889 ± 0.04        | 0.889 ± 0.04        | 0.893 ± 0.03        | 0.865 ± 0.06        | 0.886 ± 0.05        | 0.895 ± 0.04        | 0.892 ± 0.04        | 0.892 ± 0.04        | 0.895 ± 0.04        | 0.865 ± 0.06        | 0.886 ± 0.05        | 0.895 ± 0.04        | 0.892 ± 0.04        | 0.892 ± 0.04        | 0.895 ± 0.04        | 0.895 ± 0.04        |  |
| LapIRN           | 0.864 ± 0.07        | 0.888 ± 0.04        | 0.894 ± 0.03        | 0.889 ± 0.04        | 0.891 ± 0.03        | 0.893 ± 0.03        | 0.875 ± 0.05        | 0.889 ± 0.04        | 0.895 ± 0.04        | 0.891 ± 0.04        | 0.892 ± 0.04        | 0.895 ± 0.04        | 0.875 ± 0.05        | 0.889 ± 0.04        | 0.895 ± 0.04        | 0.891 ± 0.04        | 0.892 ± 0.04        | 0.893 ± 0.04        | 0.893 ± 0.04        |  |
| LapIRN-diff      | 0.826 ± 0.10        | 0.873 ± 0.07        | 0.884 ± 0.06        | 0.878 ± 0.06        | 0.880 ± 0.06        | 0.883 ± 0.05        | 0.844 ± 0.07        | 0.883 ± 0.05        | 0.893 ± 0.04        | 0.888 ± 0.04        | 0.889 ± 0.04        | 0.893 ± 0.04        | 0.844 ± 0.07        | 0.883 ± 0.05        | 0.893 ± 0.04        | 0.888 ± 0.04        | 0.889 ± 0.04        | 0.892 ± 0.04        | 0.892 ± 0.04        |  |
| MSCAReg-Net      | <b>0.879 ± 0.05</b> | 0.890 ± 0.04        | 0.894 ± 0.03        | 0.891 ± 0.03        | 0.894 ± 0.03        | 0.894 ± 0.03        | <b>0.883 ± 0.05</b> | 0.892 ± 0.04        | 0.895 ± 0.04        | 0.896 ± 0.04        | 0.896 ± 0.04        | 0.895 ± 0.04        | <b>0.883 ± 0.05</b> | 0.892 ± 0.04        | 0.895 ± 0.04        | 0.894 ± 0.04        | 0.896 ± 0.04        | 0.895 ± 0.04        | 0.895 ± 0.04        |  |
| MSCAReg-Net-diff | 0.868 ± 0.06        | 0.887 ± 0.04        | 0.895 ± 0.03        | 0.890 ± 0.04        | 0.893 ± 0.03        | 0.894 ± 0.03        | 0.877 ± 0.05        | 0.889 ± 0.04        | 0.897 ± 0.04        | 0.893 ± 0.04        | 0.894 ± 0.04        | 0.897 ± 0.04        | 0.877 ± 0.05        | 0.889 ± 0.04        | 0.897 ± 0.04        | 0.893 ± 0.04        | 0.894 ± 0.04        | 0.894 ± 0.04        | 0.894 ± 0.04        |  |
| MD               | <b>0.300 ± 0.07</b> | <b>0.255 ± 0.04</b> | 0.267 ± 0.07        | 0.267 ± 0.08        | 0.262 ± 0.07        | <b>0.240 ± 0.04</b> | 0.303 ± 0.07        | <b>0.265 ± 0.05</b> | 0.273 ± 0.06        | 0.270 ± 0.06        | 0.269 ± 0.06        | 0.273 ± 0.06        | 0.303 ± 0.07        | <b>0.265 ± 0.05</b> | 0.273 ± 0.06        | 0.270 ± 0.06        | 0.269 ± 0.06        | 0.273 ± 0.06        | 0.273 ± 0.06        |  |
| VoxelMorph       | 0.399 ± 0.09        | 0.290 ± 0.06        | 0.287 ± 0.06        | 0.276 ± 0.06        | 0.276 ± 0.06        | 0.251 ± 0.05        | 0.371 ± 0.09        | 0.288 ± 0.07        | 0.296 ± 0.07        | 0.283 ± 0.07        | 0.284 ± 0.07        | 0.296 ± 0.07        | 0.371 ± 0.09        | 0.288 ± 0.07        | 0.296 ± 0.07        | 0.283 ± 0.07        | 0.284 ± 0.07        | 0.284 ± 0.07        | 0.260 ± 0.06        |  |
| VoxelMorph-diff  | 0.435 ± 0.15        | 0.292 ± 0.06        | 0.285 ± 0.08        | 0.283 ± 0.09        | 0.278 ± 0.07        | 0.249 ± 0.05        | 0.399 ± 0.09        | 0.290 ± 0.06        | 0.285 ± 0.06        | 0.277 ± 0.06        | 0.277 ± 0.06        | 0.285 ± 0.06        | 0.399 ± 0.09        | 0.290 ± 0.06        | 0.285 ± 0.06        | 0.277 ± 0.06        | 0.277 ± 0.06        | 0.252 ± 0.05        | 0.252 ± 0.05        |  |
| SymTrans         | 0.376 ± 0.09        | 0.281 ± 0.06        | 0.283 ± 0.06        | 0.278 ± 0.07        | 0.275 ± 0.06        | 0.250 ± 0.05        | 0.353 ± 0.09        | 0.283 ± 0.07        | 0.292 ± 0.07        | 0.285 ± 0.07        | 0.281 ± 0.07        | 0.292 ± 0.07        | 0.353 ± 0.09        | 0.283 ± 0.07        | 0.292 ± 0.07        | 0.285 ± 0.07        | 0.281 ± 0.07        | 0.256 ± 0.06        | 0.256 ± 0.06        |  |
| SymTrans-diff    | 0.385 ± 0.10        | 0.281 ± 0.06        | 0.278 ± 0.06        | 0.269 ± 0.06        | 0.277 ± 0.07        | 0.247 ± 0.05        | 0.364 ± 0.09        | 0.284 ± 0.09        | 0.286 ± 0.07        | 0.284 ± 0.09        | 0.284 ± 0.09        | 0.286 ± 0.07        | 0.364 ± 0.09        | 0.284 ± 0.09        | 0.286 ± 0.07        | 0.277 ± 0.07        | <b>0.208 ± 0.06</b> | <b>0.255 ± 0.06</b> | <b>0.255 ± 0.06</b> |  |
| LapIRN           | 0.349 ± 0.10        | 0.269 ± 0.05        | 0.276 ± 0.06        | 0.273 ± 0.06        | 0.271 ± 0.05        | 0.248 ± 0.05        | 0.328 ± 0.08        | 0.277 ± 0.06        | 0.286 ± 0.07        | 0.281 ± 0.07        | 0.280 ± 0.08        | 0.286 ± 0.07        | 0.328 ± 0.08        | 0.277 ± 0.06        | 0.286 ± 0.07        | 0.281 ± 0.07        | 0.280 ± 0.08        | 0.258 ± 0.06        | 0.258 ± 0.06        |  |
| LapIRN-diff      | 0.487 ± 0.24        | 0.320 ± 0.18        | 0.317 ± 0.23        | 0.310 ± 0.20        | 0.320 ± 0.24        | 0.289 ± 0.22        | 0.427 ± 0.12        | 0.295 ± 0.07        | 0.292 ± 0.07        | 0.289 ± 0.08        | 0.292 ± 0.07        | 0.292 ± 0.07        | 0.427 ± 0.12        | 0.295 ± 0.07        | 0.292 ± 0.07        | 0.289 ± 0.08        | 0.292 ± 0.07        | 0.259 ± 0.06        | 0.259 ± 0.06        |  |
| MSCAReg-Net      | 0.305 ± 0.07        | 0.269 ± 0.05        | 0.278 ± 0.06        | 0.273 ± 0.05        | 0.269 ± 0.05        | 0.254 ± 0.05        | <b>0.302 ± 0.07</b> | 0.275 ± 0.06        | 0.292 ± 0.07        | 0.283 ± 0.07        | 0.280 ± 0.06        | 0.292 ± 0.07        | <b>0.302 ± 0.07</b> | 0.275 ± 0.06        | 0.292 ± 0.07        | 0.283 ± 0.07        | 0.280 ± 0.06        | 0.263 ± 0.06        | 0.263 ± 0.06        |  |
| MSCAReg-Net-diff | 0.333 ± 0.08        | 0.264 ± 0.04        | <b>0.265 ± 0.05</b> | <b>0.260 ± 0.05</b> | <b>0.260 ± 0.05</b> | 0.240 ± 0.04        | 0.318 ± 0.07        | 0.268 ± 0.05        | <b>0.269 ± 0.06</b> | <b>0.266 ± 0.05</b> | 0.265 ± 0.05        | <b>0.269 ± 0.06</b> | 0.318 ± 0.07        | 0.268 ± 0.05        | <b>0.269 ± 0.06</b> | <b>0.266 ± 0.05</b> | 0.265 ± 0.05        | <b>0.247 ± 0.05</b> | <b>0.247 ± 0.05</b> |  |
| HD               | 3.585 ± 0.97        | 3.432 ± 0.91        | 3.478 ± 1.04        | 3.428 ± 1.12        | 3.456 ± 0.13        | 3.354 ± 0.94        | 3.722 ± 0.84        | 3.482 ± 0.73        | 3.268 ± 0.61        | 3.271 ± 0.60        | 3.298 ± 0.71        | 3.268 ± 0.61        | 3.722 ± 0.84        | 3.482 ± 0.73        | 3.268 ± 0.61        | 3.271 ± 0.60        | 3.298 ± 0.71        | 3.215 ± 0.64        | 3.215 ± 0.64        |  |
| VoxelMorph       | 3.984 ± 1.11        | 3.619 ± 1.07        | 3.491 ± 0.92        | 3.600 ± 1.10        | 3.490 ± 0.88        | 3.464 ± 0.90        | 4.079 ± 0.81        | 3.804 ± 0.90        | 3.503 ± 0.73        | 3.678 ± 1.07        | 3.688 ± 1.06        | 3.503 ± 0.73        | 4.079 ± 0.81        | 3.804 ± 0.90        | 3.503 ± 0.73        | 3.678 ± 1.07        | 3.688 ± 1.06        | 3.607 ± 0.88        | 3.607 ± 0.88        |  |
| VoxelMorph-diff  | 3.845 ± 1.02        | 3.538 ± 0.96        | 3.455 ± 1.09        | 3.575 ± 1.07        | 3.447 ± 1.03        | 3.436 ± 0.04        | 3.856 ± 0.88        | 3.602 ± 0.78        | 3.349 ± 0.60        | 3.371 ± 1.89        | 3.484 ± 0.94        | 3.349 ± 0.60        | 3.856 ± 0.88        | 3.602 ± 0.78        | 3.349 ± 0.60        | 3.371 ± 1.89        | 3.484 ± 0.94        | 3.321 ± 0.82        | 3.321 ± 0.82        |  |
| SymTrans         | 3.767 ± 0.94        | 3.675 ± 1.11        | <b>3.357 ± 1.96</b> | 3.534 ± 1.14        | 3.528 ± 1.06        | 3.537 ± 0.96        | 3.870 ± 0.82        | 3.558 ± 0.80        | 3.394 ± 0.72        | 3.507 ± 0.96        | 3.371 ± 0.99        | 3.394 ± 0.72        | 3.870 ± 0.82        | 3.558 ± 0.80        | 3.394 ± 0.72        | 3.507 ± 0.96        | 3.371 ± 0.99        | 3.235 ± 0.72        | 3.235 ± 0.72        |  |
| SymTrans-diff    | 3.721 ± 0.95        | 3.494 ± 0.95        | 3.409 ± 0.93        | 3.464 ± 1.01        | 3.414 ± 1.03        | 3.367 ± 0.89        | 3.738 ± 0.71        | 3.513 ± 0.73        | 3.291 ± 0.79        | 3.377 ± 0.86        | <b>3.229 ± 0.79</b> | 3.291 ± 0.79        | 3.738 ± 0.71        | 3.513 ± 0.73        | 3.291 ± 0.79        | 3.377 ± 0.86        | <b>3.229 ± 0.79</b> | <b>3.155 ± 0.61</b> | <b>3.155 ± 0.61</b> |  |
| LapIRN           | 3.756 ± 1.09        | 3.437 ± 0.91        | 3.399 ± 0.90        | 3.371 ± 1.86        | 3.412 ± 0.83        | 3.233 ± 0.70        | 3.687 ± 0.80        | 3.583 ± 0.88        | 3.419 ± 0.76        | 3.467 ± 1.89        | 3.337 ± 0.82        | 3.419 ± 0.76        | 3.687 ± 0.80        | 3.583 ± 0.88        | 3.419 ± 0.76        | 3.467 ± 1.89        | 3.337 ± 0.82        | 3.254 ± 0.77        | 3.254 ± 0.77        |  |
| LapIRN-diff      | 4.108 ± 1.19        | 3.649 ± 1.16        | 3.485 ± 1.10        | 3.582 ± 1.86        | 3.630 ± 1.25        | 3.544 ± 1.14        | 3.946 ± 0.87        | 3.609 ± 0.85        | 3.565 ± 0.87        | 3.511 ± 1.10        | 3.406 ± 1.09        | 3.565 ± 0.87        | 3.946 ± 0.87        | 3.609 ± 0.85        | 3.565 ± 0.87        | 3.511 ± 1.10        | 3.406 ± 1.09        | 3.316 ± 0.92        | 3.316 ± 0.92        |  |
| MSCAReg-Net      | 3.691 ± 0.98        | 3.410 ± 0.98        | 3.381 ± 0.85        | 3.372 ± 0.85        | <b>3.283 ± 1.81</b> | 3.257 ± 0.83        | <b>3.616 ± 0.79</b> | 3.557 ± 0.95        | 3.593 ± 0.83        | 3.650 ± 0.98        | 3.504 ± 0.86        | 3.593 ± 0.83        | <b>3.616 ± 0.79</b> | 3.557 ± 0.95        | 3.593 ± 0.83        | 3.650 ± 0.98        | 3.504 ± 0.86        | 3.377 ± 0.82        | 3.377 ± 0.82        |  |
| MSCAReg-Net-diff | <b>3.526 ± 0.96</b> | <b>3.337 ± 0.82</b> | 3.441 ± 0.91        | <b>3.282 ± 1.76</b> | 3.342 ± 0.92        | <b>3.223 ± 0.77</b> | 3.627 ± 0.79        | <b>3.416 ± 0.77</b> | <b>3.245 ± 0.62</b> | <b>3.254 ± 0.65</b> | 3.240 ± 0.65        | <b>3.245 ± 0.62</b> | 3.627 ± 0.79        | <b>3.416 ± 0.77</b> | <b>3.245 ± 0.62</b> | <b>3.254 ± 0.65</b> | 3.240 ± 0.65        | 3.209 ± 0.55        | 3.209 ± 0.55        |  |

(Continues)



TABLE 3 (Continued)

|                  | Left hippocampus    |                     |                     |                     |                     |                     |                     |                     |                     |                     | Right hippocampus   |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |  |
|------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|--|
|                  | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             | MV                  | NLP                 | RLBP                | ML                  | RF                  | RF-SSLP             |                     |                     |                     |                     |                     |                     |  |
| HD95             | 1.435 ± 0.56        | 1.332 ± 0.53        | 1.241 ± 0.62        | 1.249 ± 0.60        | 1.192 ± 0.57        | <b>1.139 ± 0.27</b> | 1.464 ± 0.34        | <b>1.295 ± 0.31</b> | 1.160 ± 0.24        | 1.139 ± 0.21        | 1.130 ± 0.19        | <b>1.142 ± 0.20</b> | 1.882 ± 0.56        | 1.532 ± 0.56        | 1.256 ± 0.41        | 1.280 ± 0.44        | 1.233 ± 0.41        | 1.301 ± 0.50        | 1.785 ± 0.35        | 1.542 ± 0.33        | 1.200 ± 0.27        | 1.241 ± 0.30        | 1.199 ± 0.28        | 1.226 ± 0.31        |  |
| VoxelMorph       | 1.857 ± 0.55        | 1.482 ± 0.45        | 1.263 ± 0.54        | 1.348 ± 0.57        | 1.265 ± 0.54        | 1.198 ± 0.28        | 1.755 ± 0.34        | 1.469 ± 0.31        | 1.139 ± 0.22        | 1.151 ± 0.21        | 1.177 ± 0.20        | 1.174 ± 0.22        | 1.722 ± 0.46        | 1.471 ± 0.46        | 1.178 ± 0.32        | 1.225 ± 0.32        | 1.188 ± 0.32        | 1.236 ± 0.40        | 1.675 ± 0.33        | 1.465 ± 0.29        | 1.171 ± 0.25        | 1.206 ± 0.25        | 1.184 ± 0.24        | 1.217 ± 0.25        |  |
| SymTrans         | 1.707 ± 0.38        | 1.430 ± 0.44        | 1.189 ± 0.27        | 1.172 ± 0.24        | 1.180 ± 0.26        | 1.170 ± 0.30        | <b>1.361 ± 0.34</b> | 1.427 ± 0.33        | 1.156 ± 0.26        | 1.203 ± 0.26        | 1.135 ± 0.24        | 1.181 ± 0.25        | 1.565 ± 0.46        | 1.342 ± 0.45        | 1.206 ± 0.34        | 1.215 ± 0.35        | 1.168 ± 0.30        | 1.199 ± 0.37        | 1.533 ± 0.33        | 1.324 ± 0.32        | 1.156 ± 0.26        | 1.197 ± 0.28        | 1.156 ± 0.25        | 1.200 ± 0.27        |  |
| LapIRN           | 2.016 ± 0.78        | 1.537 ± 0.75        | 1.324 ± 0.86        | 1.369 ± 0.86        | 1.298 ± 0.92        | 1.347 ± 0.94        | 1.850 ± 0.49        | 1.481 ± 0.35        | 1.192 ± 0.25        | 1.217 ± 0.38        | 1.139 ± 0.21        | 1.172 ± 0.23        | 1.565 ± 0.46        | 1.342 ± 0.45        | 1.206 ± 0.34        | 1.215 ± 0.35        | 1.168 ± 0.30        | 1.199 ± 0.37        | 1.533 ± 0.33        | 1.324 ± 0.32        | 1.156 ± 0.26        | 1.197 ± 0.28        | 1.156 ± 0.25        | 1.200 ± 0.27        |  |
| LapIRN-diff      | 1.390 ± 0.46        | <b>1.305 ± 0.43</b> | 1.241 ± 0.42        | 1.245 ± 0.40        | 1.243 ± 0.41        | 1.248 ± 0.43        | 1.452 ± 0.30        | 1.331 ± 0.34        | 1.231 ± 0.34        | 1.255 ± 0.33        | 1.231 ± 0.33        | 1.255 ± 0.33        | 2.016 ± 0.78        | 1.537 ± 0.75        | 1.324 ± 0.86        | 1.369 ± 0.86        | 1.298 ± 0.92        | 1.347 ± 0.94        | 1.850 ± 0.49        | 1.481 ± 0.35        | 1.192 ± 0.25        | 1.217 ± 0.38        | 1.139 ± 0.21        | 1.172 ± 0.23        |  |
| MSCAReg-Net      | 1.479 ± 0.43        | 1.363 ± 0.41        | <b>1.130 ± 0.26</b> | <b>1.158 ± 0.24</b> | <b>1.146 ± 0.24</b> | 1.141 ± 0.26        | 1.467 ± 0.27        | 1.359 ± 0.26        | <b>1.094 ± 0.18</b> | <b>1.118 ± 0.19</b> | <b>1.106 ± 0.18</b> | 1.151 ± 0.21        | 1.390 ± 0.46        | <b>1.305 ± 0.43</b> | 1.241 ± 0.42        | 1.245 ± 0.40        | 1.243 ± 0.41        | 1.248 ± 0.43        | 1.452 ± 0.30        | 1.331 ± 0.34        | 1.231 ± 0.34        | 1.255 ± 0.33        | 1.231 ± 0.33        | 1.255 ± 0.33        |  |
| ASSD             | 0.359 ± 0.08        | <b>0.315 ± 0.07</b> | 0.288 ± 0.05        | 0.292 ± 0.05        | <b>0.285 ± 0.05</b> | <b>0.278 ± 0.04</b> | 0.364 ± 0.06        | <b>0.321 ± 0.05</b> | 0.292 ± 0.04        | 0.296 ± 0.04        | 0.293 ± 0.04        | 0.286 ± 0.04        | 0.359 ± 0.08        | <b>0.315 ± 0.07</b> | 0.316 ± 0.07        | 0.316 ± 0.07        | 0.312 ± 0.07        | 0.306 ± 0.07        | 0.459 ± 0.08        | 0.366 ± 0.06        | 0.319 ± 0.05        | 0.317 ± 0.06        | 0.312 ± 0.05        | 0.304 ± 0.05        |  |
| VoxelMorph       | 0.480 ± 0.12        | 0.368 ± 0.09        | 0.316 ± 0.07        | 0.316 ± 0.07        | 0.312 ± 0.06        | 0.293 ± 0.05        | 0.475 ± 0.08        | 0.364 ± 0.06        | 0.304 ± 0.04        | 0.306 ± 0.05        | 0.303 ± 0.04        | 0.294 ± 0.04        | 0.499 ± 0.13        | 0.366 ± 0.08        | 0.305 ± 0.06        | 0.312 ± 0.06        | 0.303 ± 0.05        | 0.293 ± 0.05        | 0.475 ± 0.08        | 0.364 ± 0.06        | 0.304 ± 0.04        | 0.306 ± 0.05        | 0.303 ± 0.04        | 0.294 ± 0.04        |  |
| VoxelMorph-diff  | 0.443 ± 0.10        | 0.352 ± 0.08        | 0.308 ± 0.06        | 0.310 ± 0.06        | 0.307 ± 0.06        | 0.300 ± 0.06        | 0.425 ± 0.08        | 0.350 ± 0.06        | 0.310 ± 0.05        | 0.310 ± 0.06        | 0.305 ± 0.05        | 0.298 ± 0.05        | 0.443 ± 0.10        | 0.352 ± 0.08        | 0.308 ± 0.06        | 0.310 ± 0.06        | 0.307 ± 0.06        | 0.300 ± 0.06        | 0.425 ± 0.08        | 0.350 ± 0.06        | 0.310 ± 0.05        | 0.310 ± 0.06        | 0.305 ± 0.05        | 0.298 ± 0.05        |  |
| SymTrans         | 0.443 ± 0.09        | 0.348 ± 0.07        | 0.300 ± 0.06        | 0.299 ± 0.05        | 0.300 ± 0.05        | 0.290 ± 0.05        | 0.429 ± 0.07        | 0.350 ± 0.06        | 0.304 ± 0.05        | 0.305 ± 0.05        | 0.301 ± 0.05        | 0.294 ± 0.05        | 0.397 ± 0.09        | 0.322 ± 0.06        | 0.298 ± 0.06        | 0.300 ± 0.05        | 0.296 ± 0.05        | 0.289 ± 0.05        | 0.385 ± 0.07        | 0.327 ± 0.05        | 0.304 ± 0.05        | 0.305 ± 0.06        | 0.302 ± 0.05        | 0.296 ± 0.05        |  |
| SymTrans-diff    | 0.397 ± 0.09        | 0.322 ± 0.06        | 0.298 ± 0.06        | 0.300 ± 0.05        | 0.296 ± 0.05        | 0.289 ± 0.05        | 0.385 ± 0.07        | 0.327 ± 0.05        | 0.304 ± 0.05        | 0.305 ± 0.06        | 0.302 ± 0.05        | 0.296 ± 0.05        | 0.533 ± 0.17        | 0.372 ± 0.12        | 0.323 ± 0.14        | 0.325 ± 0.12        | 0.328 ± 0.15        | 0.316 ± 0.13        | 0.498 ± 0.10        | 0.360 ± 0.06        | 0.308 ± 0.05        | 0.310 ± 0.06        | 0.309 ± 0.05        | 0.297 ± 0.05        |  |
| LapIRN-diff      | 0.533 ± 0.17        | 0.372 ± 0.12        | 0.323 ± 0.14        | 0.325 ± 0.12        | 0.328 ± 0.15        | 0.316 ± 0.13        | 0.498 ± 0.10        | 0.360 ± 0.06        | 0.308 ± 0.05        | 0.310 ± 0.06        | 0.309 ± 0.05        | 0.297 ± 0.05        | 0.533 ± 0.17        | 0.372 ± 0.12        | 0.323 ± 0.14        | 0.325 ± 0.12        | 0.328 ± 0.15        | 0.316 ± 0.13        | 0.498 ± 0.10        | 0.360 ± 0.06        | 0.308 ± 0.05        | 0.310 ± 0.06        | 0.309 ± 0.05        | 0.297 ± 0.05        |  |
| MSCAReg-Net      | <b>0.359 ± 0.07</b> | 0.320 ± 0.06        | 0.310 ± 0.07        | 0.310 ± 0.06        | 0.305 ± 0.06        | 0.301 ± 0.06        | <b>0.359 ± 0.06</b> | 0.326 ± 0.06        | 0.318 ± 0.06        | 0.316 ± 0.06        | 0.312 ± 0.06        | 0.306 ± 0.06        | <b>0.359 ± 0.07</b> | 0.320 ± 0.06        | 0.310 ± 0.07        | 0.310 ± 0.06        | 0.305 ± 0.06        | 0.301 ± 0.06        | 0.378 ± 0.06        | 0.328 ± 0.04        | <b>0.290 ± 0.04</b> | <b>0.293 ± 0.04</b> | <b>0.290 ± 0.04</b> | <b>0.285 ± 0.04</b> |  |
| MSCAReg-Net-diff | 0.388 ± 0.08        | 0.324 ± 0.06        | <b>0.288 ± 0.04</b> | <b>0.289 ± 0.04</b> | 0.286 ± 0.04        | 0.281 ± 0.04        | 0.378 ± 0.06        | 0.328 ± 0.04        | <b>0.290 ± 0.04</b> | <b>0.293 ± 0.04</b> | <b>0.290 ± 0.04</b> | <b>0.285 ± 0.04</b> | 0.388 ± 0.08        | 0.324 ± 0.06        | <b>0.288 ± 0.04</b> | <b>0.289 ± 0.04</b> | 0.286 ± 0.04        | 0.281 ± 0.04        | 0.378 ± 0.06        | 0.328 ± 0.04        | <b>0.290 ± 0.04</b> | <b>0.293 ± 0.04</b> | <b>0.290 ± 0.04</b> | <b>0.285 ± 0.04</b> |  |
| RMSD             | <b>0.669 ± 0.13</b> | <b>0.620 ± 0.12</b> | 0.589 ± 0.12        | 0.595 ± 0.12        | 0.584 ± 0.11        | 0.572 ± 0.08        | 0.674 ± 0.08        | <b>0.619 ± 0.08</b> | 0.581 ± 0.06        | 0.585 ± 0.06        | 0.580 ± 0.06        | 0.574 ± 0.05        | <b>0.669 ± 0.13</b> | <b>0.620 ± 0.12</b> | 0.589 ± 0.12        | 0.595 ± 0.12        | 0.584 ± 0.11        | 0.572 ± 0.08        | 0.674 ± 0.08        | <b>0.619 ± 0.08</b> | 0.581 ± 0.06        | 0.585 ± 0.06        | 0.580 ± 0.06        | 0.574 ± 0.05        |  |
| SyN              | 0.813 ± 0.16        | 0.686 ± 0.14        | 0.616 ± 0.11        | 0.623 ± 0.11        | 0.612 ± 0.11        | 0.610 ± 0.12        | 0.783 ± 0.10        | 0.680 ± 0.08        | 0.619 ± 0.07        | 0.622 ± 0.09        | 0.611 ± 0.08        | 0.604 ± 0.08        | 0.813 ± 0.16        | 0.686 ± 0.14        | 0.616 ± 0.11        | 0.623 ± 0.11        | 0.612 ± 0.11        | 0.610 ± 0.12        | 0.783 ± 0.10        | 0.680 ± 0.08        | 0.619 ± 0.07        | 0.622 ± 0.09        | 0.611 ± 0.08        | 0.604 ± 0.08        |  |
| VoxelMorph       | 0.823 ± 0.17        | 0.677 ± 0.12        | 0.608 ± 0.11        | 0.622 ± 0.13        | 0.604 ± 0.10        | 0.593 ± 0.09        | 0.788 ± 0.09        | 0.670 ± 0.07        | 0.594 ± 0.06        | 0.600 ± 0.07        | 0.594 ± 0.06        | 0.584 ± 0.06        | 0.823 ± 0.17        | 0.677 ± 0.12        | 0.608 ± 0.11        | 0.622 ± 0.13        | 0.604 ± 0.10        | 0.593 ± 0.09        | 0.788 ± 0.09        | 0.670 ± 0.07        | 0.594 ± 0.06        | 0.600 ± 0.07        | 0.594 ± 0.06        | 0.584 ± 0.06        |  |
| VoxelMorph-diff  | 0.762 ± 0.13        | 0.663 ± 0.12        | 0.601 ± 0.09        | 0.616 ± 0.11        | 0.605 ± 0.09        | 0.598 ± 0.09        | 0.740 ± 0.09        | 0.656 ± 0.08        | 0.603 ± 0.07        | 0.609 ± 0.08        | 0.594 ± 0.07        | 0.588 ± 0.07        | 0.762 ± 0.13        | 0.663 ± 0.12        | 0.601 ± 0.09        | 0.616 ± 0.11        | 0.605 ± 0.09        | 0.598 ± 0.09        | 0.740 ± 0.09        | 0.656 ± 0.08        | 0.603 ± 0.07        | 0.609 ± 0.08        | 0.594 ± 0.07        | 0.588 ± 0.07        |  |
| SymTrans         | 0.758 ± 0.12        | 0.654 ± 0.11        | 0.594 ± 0.08        | 0.596 ± 0.07        | 0.598 ± 0.10        | 0.582 ± 0.08        | 0.740 ± 0.09        | 0.655 ± 0.08        | 0.592 ± 0.07        | 0.598 ± 0.07        | 0.587 ± 0.06        | 0.582 ± 0.06        | 0.758 ± 0.12        | 0.654 ± 0.11        | 0.594 ± 0.08        | 0.596 ± 0.07        | 0.598 ± 0.10        | 0.582 ± 0.08        | 0.740 ± 0.09        | 0.655 ± 0.08        | 0.592 ± 0.07        | 0.598 ± 0.07        | 0.587 ± 0.06        | 0.582 ± 0.06        |  |
| SymTrans-diff    | 0.713 ± 0.13        | 0.624 ± 0.10        | 0.592 ± 0.09        | 0.597 ± 0.09        | 0.587 ± 0.08        | 0.581 ± 0.09        | 0.695 ± 0.09        | 0.626 ± 0.08        | 0.596 ± 0.07        | 0.601 ± 0.08        | 0.591 ± 0.07        | 0.587 ± 0.07        | 0.713 ± 0.13        | 0.624 ± 0.10        | 0.592 ± 0.09        | 0.597 ± 0.09        | 0.587 ± 0.08        | 0.581 ± 0.09        | 0.695 ± 0.09        | 0.626 ± 0.08        | 0.596 ± 0.07        | 0.601 ± 0.08        | 0.591 ± 0.07        | 0.587 ± 0.07        |  |
| LapIRN           | 0.871 ± 0.24        | 0.695 ± 0.20        | 0.636 ± 0.26        | 0.642 ± 0.23        | 0.642 ± 0.28        | 0.632 ± 0.26        | 0.817 ± 0.13        | 0.667 ± 0.09        | 0.601 ± 0.07        | 0.610 ± 0.10        | 0.604 ± 0.08        | 0.588 ± 0.07        | 0.871 ± 0.24        | 0.695 ± 0.20        | 0.636 ± 0.26        | 0.642 ± 0.23        | 0.642 ± 0.28        | 0.632 ± 0.26        | 0.817 ± 0.13        | 0.667 ± 0.09        | 0.601 ± 0.07        | 0.610 ± 0.10        | 0.604 ± 0.08        | 0.588 ± 0.07        |  |
| LapIRN-diff      | 0.673 ± 0.11        | 0.622 ± 0.10        | 0.607 ± 0.10        | 0.609 ± 0.10        | 0.600 ± 0.10        | 0.598 ± 0.10        | <b>0.667 ± 0.08</b> | 0.629 ± 0.08        | 0.615 ± 0.09        | 0.618 ± 0.09        | 0.607 ± 0.08        | 0.603 ± 0.08        | 0.673 ± 0.11        | 0.622 ± 0.10        | 0.607 ± 0.10        | 0.609 ± 0.10        | 0.600 ± 0.10        | 0.598 ± 0.10        | <b>0.667 ± 0.08</b> | 0.629 ± 0.08        | 0.615 ± 0.09        | 0.618 ± 0.09        | 0.607 ± 0.08        | 0.603 ± 0.08        |  |
| MSCAReg-Net      | 0.694 ± 0.11        | 0.624 ± 0.10        | <b>0.581 ± 0.07</b> | <b>0.581 ± 0.06</b> | <b>0.576 ± 0.07</b> | <b>0.571 ± 0.07</b> | 0.681 ± 0.07        | 0.625 ± 0.06        | <b>0.576 ± 0.06</b> | <b>0.580 ± 0.06</b> | <b>0.575 ± 0.05</b> | <b>0.572 ± 0.05</b> | 0.694 ± 0.11        | 0.624 ± 0.10        | <b>0.581 ± 0.07</b> | <b>0.581 ± 0.06</b> | <b>0.576 ± 0.07</b> | <b>0.571 ± 0.07</b> | 0.681 ± 0.07        | 0.625 ± 0.06        | <b>0.576 ± 0.06</b> | <b>0.580 ± 0.06</b> | <b>0.575 ± 0.05</b> | <b>0.572 ± 0.05</b> |  |
| MSCAReg-Net-diff |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |  |



**FIGURE 8** Example narrowband established by MV method but on different registration models. The regions with outliers on the narrowband are pointed out by white arrows.

single-resolution registration framework by devising the FC-Module, FA-Module and MSCA-Module. The FC-Module and FA-Module merged in the U-Net block facilitated the MSCA-Module in perceiving small and large deformations by generating more distinguishable features. The diffeomorphic variant MSCAReg-Net-diff was also devised to guarantee the diffeomorphism of the deformation field. Experimental results indicated that the proposed complexity-aware technique in the proposed MSCAReg-Net outperformed the conventional SyN, deep learning-based VoxelMorph and LapIRN in terms of registration accuracy. The registration methods under comparisons were also applied on a downstream medical image analysis task of MAHS. Experimental results demonstrated that both MSCAReg-Net-diff and SyN achieved better hippocampus segmentations over other registration methods (Table 3), and the proposed MSCAReg-Net-diff contributed to a competitive performance on the hippocampus segmentation with the conventional but best-performing SyN method.

## 5.1 | Registration network architecture and framework

The size of the receptive field is highly associated with the identification of deformation complexity, and simultaneously perceiving deformations with distinct complexities can significantly improve the registration performance [26], [17]. To investigate this, the MSCA-Module was devised to achieve different receptive fields by defining dilated convolutions with different dilation rates, thereby facilitating representation learning for deformations with distinct complexities. Besides, the dilated convolution could effectively avoid the loss of precise position information of each voxel caused by pooling operation. The deformation magnitude maps in Figure 4 showed that small receptive field was beneficial to characterize local subtle deformation, whereas large receptive field provides more spatial information to learn regions with large positional bias. Furthermore, the deformation magnitude maps with (Figure 4d) and without (Figure 4b) MSCA-Module showed that the complexity-aware technique is beneficial for representation

learning of deformations with distinct complexities. The DSC scores in Table 1 indicated that the MSCA-Module could significantly improve the registration performance. It well proved that the complexity-aware technique could simultaneously and effectively perceive deformations with distinct complexities under single-resolution registration framework.

The improvement of registration performance was often limited by the representation capability of the deformation features. Conventional U-Net was typically used in image registration algorithms to learn deformation features [8]. However, the deformation representation capability of conventional U-Net was insufficient, and the semantic gap of feature aggregation between encoder and decoder was also a fatal defect in registration tasks [9]. Considering the above problems, FC-Module was designed to calibrate the representation weights of deformations with distinct complexities, thereby consolidating the identification ability of MSCA-Module for displacement features. The attention map  $\eta$  in Figure 5 showed that FC-Module could identify regions with large spatial position difference between fixed image and moving image. Furthermore, FA-Module was akin to the gap-fill mechanism [9], which effectively alleviated the semantic gap of feature aggregation. The results in Table 1 also demonstrated that facilitating the representation ability of U-Net in terms of deformation complexity could effectively improve registration accuracy.

Besides, this study demonstrated that a single-resolution registration framework integrating complexity-aware techniques was also capable of simultaneously perceiving both small and large deformations. The single-resolution registration framework can avoid the interpolation error and huge memory consumption of the multi-resolution registration framework, further highlighting its usability and scalability in image registration. It is noteworthy that the complexity-aware technique effectively alleviated the difficulty of the single-resolution registration framework in aligning image regions with large deformations (Table 1, Figures 4, and 5), and can be easily extended to whole brain imaging. Furthermore, our method outperformed existing single- and multi-resolution registration methods (especially the best-performing conventional method SyN) in terms of registration accuracy, highlighting the

superiority of the single-resolution registration framework integrating complexity-aware techniques. Although the runtime of the deep learning-based baseline methods is slightly faster than the proposed MSCAReg-Net, they sacrifice more registration accuracy. Specifically, the average runtime of MSCAReg-Net over left and right hippocampus is 0.104 s, which is only 0.088, 0.076 and 0.092 s slower than VoxelMorph, SymTrans and LapIRN, respectively, but the registration accuracy of MSCAReg-Net is significantly improved by a margin of 6.4%, 5.3% and 3.1% of DSC, respectively. Importantly, the runtime of all deep learning-based registration methods does not exceed 1 s.

## 5.2 | Further evaluation of registration in subsequent applications

Application comparisons of registration in subsequent medical image analysis tasks are essential. To the best of our knowledge, we are the first to apply the registration algorithm into subsequent applications to evaluate the registration performance. Previous registration studies have typically used metrics such as DSC and  $|J_\phi| \leq 0$  to evaluate registration performance [8], [16]. In this study, the registration algorithm was comprehensively evaluated by employing a MAHS application. The segmentation results in Table 3 showed that the proposed MSCAReg-Net and MSCAReg-Net-diff contributed to a better hippocampus segmentation over other deep learning-based single- or multi-resolution registration methods under comparison, and a comparable segmentation performance with the conventional but leading SyN method.

Although the registration accuracy (such as DSC) and diffeomorphism (such as  $|J_\phi| \leq 0$ ) reflect the registration performance in different aspects, their priorities may be dependent on different applications. Regarding the hippocampus segmentation, MSCAReg-Net contributed to the best hippocampus segmentation by the MV method since the MV method was solely and highly independent on the registration accuracy of DSC. However, the other label fusion methods such as NLP, RLBP, ML, RF, and RF-SSLP achieved the best segmentation based on the diffeomorphic variants (Table 3), demonstrating the diffeomorphism could help stabilize the establishment of the narrowband, further facilitating the subsequent segmentation. The narrowband visualization results in Figure 8 showed that the narrowband established by the registration method without diffeomorphism had more outliers, highlighting the positive role of diffeomorphism in terms of stabilizing the narrowband. Besides, the proposed method achieved an advanced DSC but a relative high  $|J_\phi| \leq 0$  than SyN, but the following hippocampus segmentation exhibited a comparable performance, demonstrating the increase of the registration accuracy of DSC could help offset the increase of the  $|J_\phi| \leq 0$  in segmentation to some extent as long as the registration performance of the diffeomorphic variant did not decrease sharply such as LapIRN-diff (Table 2). Therefore, the balance between the DSC and  $|J_\phi| \leq 0$  values need to be validated on the following medical image analysis task.

Solely boosting the DSC value or decreasing the  $|J_\phi| \leq 0$  might not be approximate in devising deformable registration algorithms.

## 5.3 | Limitation

Although the priority between DSC and  $|J_\phi| \leq 0$  was discussed in this study using MAHS as an example, a further investigation is still needed in more and various medical applications. However, it needs to be pointed out that further exploration of the balance between DSC and  $|J_\phi| \leq 0$  need to be established on relatively mature medical image analysis tasks such as MAHS adopted in this study. As mentioned before, it might be not a good choice when image registration is not a vital step and cannot greatly affect the latter medical image analysis task.

## 6 | CONCLUSION

In the present study, a multi-scale complexity-aware network defined as MSCAReg-Net was proposed under the single-resolution registration framework, demonstrating the effectiveness of the complexity-aware technique in promoting deformable image registration. Specifically, the FC-Module and the FA-Module were devised and integrated into a U-Net block to promote learning of deformation magnitude features in the feature translation stage. In the subsequent complexity-aware stage, the MSCA-Module which was established on the complexity-aware technique and the optimized deformation features generated by the FC-Module and the FA-Module, was cascaded to perceive small and large deformations simultaneously. The diffeomorphic variant MSCAReg-Net-diff was also accomplished to ensure the deformation diffeomorphism. Experimental results demonstrated the superiority of the proposed MSCAReg-Net over the existing single- and multi-resolution registration methods (especially the best-performing SyN) in terms of registration accuracy. Besides, other than the indices of DSC and  $|J_\phi| \leq 0$ , a comprehensive evaluation of the registration performance was carried out by applying the MSCAReg-Net on a latter medical image analysis task of MAHS. Experimental results demonstrated that the proposed MSCAReg-Net and MSCAReg-Net-diff contributed to a better hippocampus segmentation over other deep learning-based single- or multi-resolution registration methods under comparison, and a comparable segmentation performance with the conventional but leading SyN method. The comprehensive assessment including DSC,  $|J_\phi| \leq 0$ , and the latter application on a specific medical image analysis task demonstrated the advances of the proposed MSCAReg-Net in medical image deformation registration.

## AUTHOR CONTRIBUTIONS

**Hu Yu:** Conceptualization; Methodology; Validation; Formal analysis; Investigation; Data Curation; Software; Visualization; Writing—Original Draft. **Qiang Zheng:** Conceptualization; Methodology; Formal analysis; Investigation; Writing—Review

and Editing. **Fang Hu**: Methodology; Writing—Review and Editing. **Chaoqing Ma**: Writing—Review and Editing. **Shuo Wang**: Writing—Review and Editing. **Shuai Wang**: Conceptualization; Methodology; Formal analysis; Investigation; Resources; Data Curation; Writing—Review and Editing; Supervision; Project administration; Funding acquisition.

## ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (grant number 61802330, 61802331), Open Project of Key Laboratory of Medical Imaging and Artificial Intelligence of Hunan Province, Xiangnan University (grant number YXZN2022002).

## CONFLICT OF INTEREST STATEMENT

The authors declare that no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Experimental data that support the findings of this study are available at the website provided for database: ADNI (<http://adni.loni.usc.edu>), including brain MR images and hippocampal segmentation labels of 135 subjects from the ADNI-HarP project ([www.hippocampal-protocol.net](http://www.hippocampal-protocol.net)). Additional data related to this paper may be requested from the authors after reasonable application. The study was approved by the institutional review boards of the participating institutions.

## ORCID

Hu Yu  <https://orcid.org/0000-0002-1381-2791>

Qiang Zheng  <https://orcid.org/0000-0002-7853-8033>

Chaoqing Ma  <https://orcid.org/0000-0001-5084-7689>

Shuai Wang  <https://orcid.org/0000-0002-9721-6590>

## REFERENCES

- Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. *IEEE Trans. Med. Imaging* 32(7), 1153–1190 (2013)
- Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46(3), 726–738 (2009)
- Lee, C., Langen, K.M., Lu, W., Haimerl, J., Schnarr, E., Ruchala, K.J., Olivera, G.H., Meeks, S.L., Kupelian, P.A., Shellenberger, T.D.: Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration. *Int. J. Radiat. Oncol. Biol. Phys.* 71(5), 1563–1571 (2008)
- Collins, J.A., Weis, J.A., Heiselman, J.S., Clements, L.W., Simpson, A.L., Jarnagin, W.R., Miga, M.I.: Improving registration robustness for image-guided liver surgery in a novel human-to-phantom data framework. *IEEE Trans. Med. Imaging* 36(7), 1502–1510 (2017)
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12(1), 26–41 (2008)
- Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: Svf-Net: Learning deformable image registration using shape matching. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2017)*. pp. 266–274. Springer, Cham, Quebec (2017)
- Balakrishnan, G., Zhao, A., Sabuncu, M.R., Gutttag, J., Dalca, A.V.: Voxel-morph: A learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* 38(8), 1788–1800 (2019)
- Balakrishnan, G., Zhao, A., Sabuncu, M.R., Gutttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake (2018)
- Fan, J., Cao, X., Yap, P.-T., Shen, D.: Birnet: Brain image registration using dual-supervised fully convolutional networks. *Med. Image Anal.* 54, 193–206 (2019)
- Ma, J., Zhao, J., Jiang, J., Zhou, H., Guo, X.: Locality preserving matching. *Int. J. Comput. Vision* 127, 512–531 (2019)
- Tang, L., Deng, Y., Ma, Y., Huang, J., Ma, J.: Superfusion: A versatile image registration and fusion network with semantic awareness. *IEEE/CAA J. Autom. Sin.* 9(12), 2121–2137 (2022)
- Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration—A deep learning approach. *NeuroImage* 158, 378–396 (2017)
- Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage* 54(3), 2033–2044 (2011)
- Hering, A., van Ginneken, B., Heldmann, S.: Mlvrnet: Multilevel variational image registration network. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2019)*. pp. 257–265. Springer, Cham, Shenzhen (2019)
- Mok, T.C., Chung, A.C.: Large deformation diffeomorphic image registration with laplacian pyramid networks. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2020)*. pp. 211–221. Springer, Cham, Lima (2020)
- Dalca, A.V., Balakrishnan, G., Gutttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2018)*. pp. 729–738. Springer, Cham, Granada (2018)
- Huang, Y., Ahmad, S., Fan, J., Shen, D., Yap, P.-T.: Difficulty-aware hierarchical convolutional neural networks for deformable registration of brain mr images. *Med. Image Anal.* 67, 101817 (2021)
- Shi, F., Liu, B., Zhou, Y., Yu, C., Jiang, T.: Hippocampal volume and asymmetry in mild cognitive impairment and alzheimer's disease: Meta-analyses of mri studies. *Hippocampus* 19(11), 1055–1064 (2009)
- Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr, C.R.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage* 21(4), 1428–1442 (2004)
- Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954 (2011)
- Zhu, H., Cheng, H., Fan, Y.: Random local binary pattern based label learning for multi-atlas segmentation. In: *Medical Imaging 2015: Image Processing*. (2015)
- Zhu, H., Cheng, H., Yang, X., Fan, Y., Alzheimer's Disease Neuroimaging Initiative: Metric learning for multi-atlas based segmentation of hippocampus. *Neuroinformatics* 15(1), 41–50 (2017)
- Zheng, Q., Wu, Y., Fan, Y.: Integrating semi-supervised and supervised learning methods for label fusion in multi-atlas based image segmentation. *Front. Neuroinf.* 12, 69 (2018)
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas (2016)
- Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*. pp. 234–241. Springer, Cham, Munich (2015)
- Duan, L., Yuan, G., Gong, L., Fu, T., Yang, X., Chen, X., Zheng, J.: Adversarial learning for deformable registration of brain mr image using a multi-scale fully convolutional network. *Biomed. Signal Process. Control* 53, 101562 (2019)
- Qin, Y., Kamnitsas, K., Ancha, S., Nanavati, J., Cottrell, G., Criminisi, A., Nori, A.: Autofocus layer for semantic segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2018)*. pp. 603–611. Springer, Cham, Granada (2018)
- Liu, B., Zheng, Q., Zhao, K., Li, H., Ma, C., Wu, S., Tong, X.: Hpscseg-Net: Hippocampus segmentation network integrating autofocus attention



- mechanism and feature recombination and recalibration module. In: *International Conference on Image and Graphics*. (2021)
29. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: *International Conference on Machine Learning (ICML)*. Haifa (2010)
  30. Yu, H., Zheng, Q., Zhao, K., Li, H., Ma, C., Wu, S., Tong, X.: Hpreg-Net: Unsupervised U-Net integrating dilated convolution and residual attention for hippocampus registration. In: *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Zhuhai (2021)
  31. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2017)
  32. Mok, T.C., Chung, A.: Fast symmetric diffeomorphic image registration with convolutional neural networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle (2020)
  33. Boccardi, M., Bocchetta, M., Morency, F.C., Collins, D.L., Nishikawa, M., Ganzola, R., Grothe, M.J., Wolf, D., Redolfi, A., Pievani, M.: Training labels for hippocampal segmentation based on the eadc-adni harmonized hippocampal protocol. *Alzheimer's Dementia* 11(2), 175–183 (2015)
  34. Hao, Y., Wang, T., Zhang, X., Duan, Y., Yu, C., Jiang, T., Fan, Y., Initiative, A.S.D.N.: Local label learning (Ll) for subcortical structure segmentation: Application to hippocampus segmentation. *Hum. Brain Mapp.* 35(6), 2674–2697 (2014)
  35. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage* 46(3), 786–802 (2009)
  36. Ma, M., Xu, Y., Song, L., Liu, G.: Symmetric transformer-based network for unsupervised image registration. *Knowledge-Based Syst.* 257, 109959 (2022)

**How to cite this article:** Yu, H., Zheng, Q., Hu, F., Ma, C., Wang, S., Wang, S.: MSCAReg-Net: Multi-scale complexity-aware convolutional neural network for deformable image registration. *IET Image Process.* 18, 839–855 (2024). <https://doi.org/10.1049/ipr2.12988>