Medical Image Analysis 71 (2021) 102036

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

CycleMorph: Cycle consistent unsupervised deformable image registration☆

Boah Kim[a], Dong Hwan Kim[b], Seong Ho Park[c], Jieun Kim[d], June-Goo Lee[e], Jong Chul Ye[a,*]

[a] Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
[b] Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[c] Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea
[d] Smart Car R&D Division, AI-Bigdata R&D Center, Korea Automotive Technology Institute (KATECH), Republic of Korea
[e] Department of Convergence Medicine, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 20 August 2020
Revised 26 February 2021
Accepted 8 March 2021
Available online 12 March 2021

Keywords:
Cycle consistency
Image registration
Deep learning
Unsupervised learning

ABSTRACT

Image registration is a fundamental task in medical image analysis. Recently, many deep learning based image registration methods have been extensively investigated due to their comparable performance with the state-of-the-art classical approaches despite the ultra-fast computational time. However, the existing deep learning methods still have limitations in the preservation of original topology during the deformation with registration vector fields. To address this issues, here we present a cycle-consistent deformable image registration, dubbed CycleMorph. The cycle consistency enhances image registration performance by providing an implicit regularization to preserve topology during the deformation. The proposed method is so flexible that it can be applied for both 2D and 3D registration problems for various applications, and can be easily extended to multi-scale implementation to deal with the memory issues in large volume registration. Experimental results on various datasets from medical and non-medical applications demonstrate that the proposed method provides effective and accurate registration on diverse image pairs within a few seconds. Qualitative and quantitative evaluations on deformation fields also verify the effectiveness of the cycle consistency of the proposed method.

1. Introduction

Image registration is one of the fundamental tasks in medical imaging, since the shape of anatomical structures in images vary due to the disease progress, patient motion, breathing, etc. For example, radiologists often diagnose the liver tumor with multiphase contrast enhanced CT (CECT) images (Kim et al., 2011), but the images at different temporal phases are usually different in their shape and image contrast as shown in Fig. 1.

Although conventional image registration methods (Christensen and Johnson, 2001; Leow et al., 2005; Ashburner, 2007; Beg et al., 2005; Avants et al., 2008; Klein et al., 2009) have been studied to address this using a variational framework that solves an optimization problem for each image pair to be aligned with similar appearance, these approaches usually suffer from extensive computation and long registration time.

Recently, deep learning approaches have been developed due to their fast runtime over conventional methods for image registration. Given source and target images, deep neural networks are trained to generate deformation fields corresponding to the input image pair, so that it enables significant fast registration (Onofrey et al., 2013; Zhang et al., 2008; Yang et al., 2017; Rohé et al., 2017; Sokooti et al., 2017). Nowadays, these methods have been evolved to unsupervised learning methods that do not require ground-truth deformation fields (Krebs et al., 2018; Balakrishnan et al., 2018; Fan et al., 2018; Lei et al., 2020). However, the existing image registration approaches do not explicitly enforce the criterion to guarantee topology preservation, which often result in inaccurate registration with the loss of structural information.

To overcome the potential degeneracy problem of registration fields, here we present a novel deformable image registration method called CycleMorph, which uses cycle consistency to force the deformed image to return to the original image (Zhu et al., 2017). In contrast to the existing approaches that enforce the inverse-consistency to the deformation vector fields generated from additional inverse networks (Zhang, 2018), one of the most important contributions of this work is the demonstra-

☆ Part of this work was presented at the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2019 conference (Kim et al., 2019).

* Corresponding author.
E-mail address: jong.ye@kaist.ac.kr (J.C. Ye).

https://doi.org/10.1016/j.media.2021.102036
1361-8415/

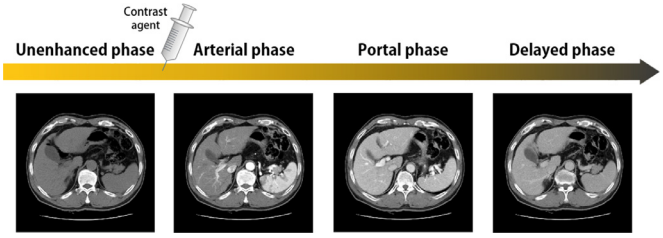

Fig. 1. Example of 2D slices taken from 3D liver CT volumes before and after injection of contrast agent. Images at different phases show various contrasts and shapes of liver and other organs.

tion of the topological preservation by imposing the cycle consistency simply on the images. This cycle consistency on image domain allows to avoid discretization error from the real-world implementation of inverse deformation fields. Moreover, as our approach is not based on the generative adversarial network (GAN) (Goodfellow et al., 2014) used in Zhu et al. (2017), but based on convolutional neural network (CNN) with a spatial transform layer (STL) (Jaderberg et al., 2015), our method does not lead to false alteration in image content.

More specifically, we train two convolutional neural networks (CNN), G_X and G_Y, that generate forward and reverse directional deformation vector fields, respectively. When a moving source image is deformed to the other fixed image by the deformation fields from G_X, then the deformed image can be reversed to the original image using the deformation fields from G_Y, by applying the cycle consistency to the reversed image and the original image. It turns out that this inverse path with the cyclic constraint is a direct way of providing high performance topology preservation with less folding problem during the deformation.

Another important innovation of this work is the extension to multi-scale implementation to deal with the large volume image registration. Specifically, due to the GPU memory limitation, training with the whole 3D volume for image registration may not be possible. To deal with this, we propose a coarse 3D registration using the subsampled volume for large deformation, followed by local deformation estimation to improve the registration accuracy.

In order to verify the performance of the proposed method, we apply our algorithm to various applications from different domains with varying memory requirement, including 2D face registration, 3D brain MR registration, and multiphase 3D abdominal contrast enhanced CT (CECT) volume registration for liver cancer evaluation. Qualitative and quantitative evaluation of the experimental results demonstrate the robustness of the proposed method and confirm the efficacy of the cycle consistency for topology preservation.

The paper is organized as follows. Section 2 reviews the related works. Section 3 describes our theory and proposed method. Section 4 describes experiment datasets we used, implementation details, and evaluation methods. Section 5 presents experimental results on registration of face expression images, MRI, and CT datasets. Section 6 discusses the proposed method with the results, and we conclude in Section 7.

2. Related works

2.1. Diffeomorphic image registration

In classical variational image registration approaches, an energy function is typically composed of two terms:

$$\mathcal{L}(X, Y, \phi) = \mathcal{L}_{sim}(\mathcal{T}(X, \phi), Y) + \mathcal{L}_{reg}(\phi), \quad (1)$$

where X and Y denote the moving image and fixed image, respectively; ϕ represents the displacement vector fields, and $\mathcal{T}$ is the transformation function which warps X to Y using the deformation vector fields ϕ. In (1), the first term is a similarity function which evaluates the shape differences between deformed images and reference images, whereas the second term is a regularization function to penalize deformation fields.

In particular, diffeomorphic image registration methods impose the constraint on the vector fields ϕ such that the resulting deformable mapping becomes a diffeomorphism. The diffeomorphic deformation ensures certain desirable properties between two image volumes like continuous, differentiable, and preserving topology (Beg et al., 2005; Avants et al., 2008; Vercauteren et al., 2009). The popular examples of these algorithmic extensions to large deformation are Large Deformation Diffeomorphic Metric Matching (LDDMM) (Beg et al., 2005; Zhang et al., 2017; Cao et al., 2005; Ceritoglu et al., 2009) and Symmetric image Normalization method (SyN) (Avants et al., 2008).

Unfortunately, these algorithms are usually computationally expensive, which prohibits its routine use in clinical workflow.

2.2. Deep-learning-based image registration

On the other hand, the learning-based registration algorithms are inductive in the sense that once a neural network is trained, it can instantaneously predict deformation vector fields for a new data. Therefore, it is ideally suitable for clinical environment. Specifically, many of learning based registration methods have been developed based on the supervised or unsupervised setting. We provide more details in the following.

2.2.1. Supervised learning methods

In supervised learning approaches, the ground-truth deformation vector fields are required, which are typically generated by the classical registration methods (Cao et al., 2017; Yang et al., 2017; Cao et al., 2018; Rohé et al., 2017; Sokooti et al., 2017). Yang et al. (2017) proposed an encoder-decoder network for patch-wise prediction of the deformation fields, and a correction network to improve deformation prediction. Cao et al. (2018) developed a non-rigid inter-modality image registration network that estimates registration fields of two-modal images. However, since the registration performance of these approaches depends on quality of the ground-truths, these works often require high quality ground-truth deformation fields and complicated pre-processing, both of which are often difficult to obtain in practice.

2.2.2. Unsupervised learning methods

To overcome the limitation of supervised learning approaches, unsupervised learning methods have been recently developed, which learn the image registration by minimizing the loss between the deformed image and fixed target image. Krebs et al. (2018) proposed an unsupervised learning approach for image registration via a low-dimensional stochastic parameterization of the deformation, in which the KL divergence between the encoding feature distribution from moving and fixed images and a multivariate unit Gaussian distribution is used as one of the loss terms. Balakrishnan et al. (2018, 2019) presented a pairwise 3D medical image registration algorithm using a CNN with a spatial transform layer (STL), which parameters are learned by the local normalized cross-correlation function. For large volume image registration, de Vos et al. (2019) proposed affine and non-rigid image registration framework, and Lei et al. (2020) presented a multiscale unsupervised learning method called MS-DIRNet through global and local registration networks.

However, most of these methods do not usually impose the constraint on the consistency, so that they can cause a folding problem from degeneracy of the mapping. To address this issue, Dalca et al. (2018) introduced a diffeomorphic integration layer and

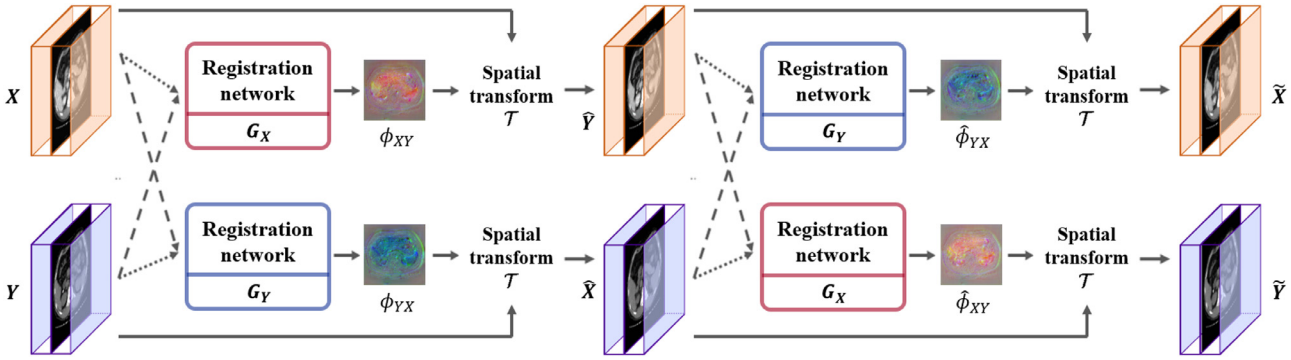

Fig. 2. The overall framework of the proposed cycle consistent deep learning model, CycleMorph, for deformable image registration. Two registration networks (G_X, G_Y) are used to take inputs by switching their orders. Each network takes two volumes (X, Y) and computes displacement vector fields. Short and long dashed lines denote the moving images and fixed images, respectively. The spatial transform function deforms the moving image according to the vector fields to match a shape of the fixed image. These transformed images ($\hat{X}, \hat{Y}$) are taken to the networks followed by transform function to ensure that the deformed images can be returned to original state.

Krebs et al. (2018) proposed a differentiable exponentiation layer. But these diffeomorphic layers should be also applied at the test phase, which might incur additional computational cost.

2.3. Consistent image registration

Although the classical diffeomorphic deformable registration algorithms have been proposed to ensure the one-to-one correspondence, deformations are generally represented discretely with a finite number of parameters, so there may be some small violations. Thus, the estimated deformation F from X to Y is not equal to the inverse of the estimated deformation R from Y to X. In consistent image registration approaches (Christensen and Johnson, 2001; Ashburner, 2007; Leow et al., 2005), this problem is alleviated by imposing additional inverse consistency:

$$R \simeq F^{-1}. \tag{2}$$

In particular, the forward and inverse mappings F and R are only defined through the corresponding deformation fields ϕ_{XY} and ϕ_{YX}, so the corresponding inverse-consistency is usually enforced as a regularization term to the deformation vector fields.

Recently, Zhang (2018) proposed an inverse-consistency enforced deep learning model that simultaneously trains both forward and inverse neural networks. The forward network estimates the deformation fields that can map a source to the target, whereas the inverse network generates the inverse flow under the inverse consistency condition of the deformation fields. On the other hand, Mahapatra et al. (2018) proposed a GAN-based image registration method by exploiting cycle consistency (Zhu et al., 2017) on the deformed images, and showed the effectiveness of their approach using perfectly aligned images.

3. Theory

The overall learning framework of the proposed CycleMorph is illustrated in Fig. 2. Specifically, for the moving source and fixed target images, X and Y, which may come from different subjects (i.e. different anatomical shapes) regardless of the contrast of each image (i.e. both inter/intramodal registration), we define two registration networks as $G_X : (X, Y) \rightarrow \phi_{XY}$ and $G_Y : (Y, X) \rightarrow \phi_{YX}$, where ϕ_{XY} (resp. ϕ_{YX}) denotes the deformation fields from X to Y (resp. Y to X). We use a spatial transformation layer $\mathcal{T}$ in the networks to warp the moving image by the estimated deformation fields, so that the registration networks can be trained by minimizing the loss function on the deformed image and fixed image. Accordingly, when a pair of images are given to the registration networks, the moving image is deformed to align with the fixed image.

In particular, to guarantee the topology preservation between the deformed and fixed images, we employ the cycle consistency constraint between the original moving image and its re-deformed image. That is, the two deformed images are given as an input to the networks again by switching their order to impose cycle consistency on a pixel level of images. This constraint stems from a mathematical observation that a homeomorphic mapping between two topological spaces preserves all topological properties, but requires much relaxed constraints compared to a diffeomorphism that requires an additional condition of a differentiable mapping in terms of network implementation. Since the diffeomorphism is a subset of the homeomorphic deformation, imposing the cycle consistency to the networks by ensuring that the shape of deformed images successively return to the original shape allows the networks to offer homeomorphic mappings capable of preserving topology.

3.1. Loss function

We train the proposed cycle consistent learning model by solving the following optimization problem:

$$\min_{G_X, G_Y} \mathcal{L}(X, Y, G_X, G_Y), \tag{3}$$

where

$$\begin{aligned}\mathcal{L}(X, Y, G_X, G_Y) =& \mathcal{L}_{regist}(X, Y, G_X) \\ &+ \mathcal{L}_{regist}(Y, X, G_Y) \\ &+ \alpha \mathcal{L}_{cycle}(X, Y, G_X, G_Y) \\ &+ \beta \mathcal{L}_{identity}(X, Y, G_X, G_Y),\end{aligned} \tag{4}$$

where $\mathcal{L}_{regist}$, $\mathcal{L}_{cycle}$, and $\mathcal{L}_{identity}$ are registration loss, cycle loss, and identity loss, respectively, and α and β are hyper-parameters. As shown in Fig. 3, our method is trained in an unsupervised manner without ground-truth deformation fields. More detailed description of each loss functions is as following.

3.1.1. Registration loss

The registration loss function is based on the energy function of traditional variational image registration (1) that has similarity and smoothness penalized terms. We employ the local cross-correlation for the similarity function to be less sensitive to the contrast variations (Avants et al., 2008), and the l_2-loss for the regularization function. Accordingly, our registration loss function can be written as:

$$\mathcal{L}_{regist}(X, Y, G_X) = -(\mathcal{T}(X, \phi_{XY}) \otimes Y) + \lambda \sum ||\nabla \phi_{XY}||^2, \tag{5}$$

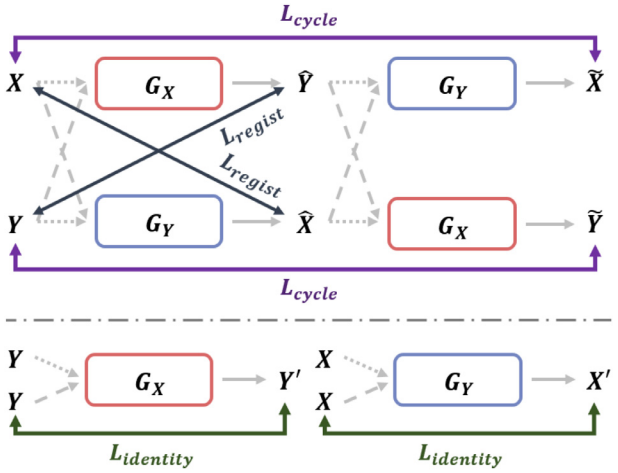

Fig. 3. The diagram of loss function structure in our proposed method. The registration loss function, $\mathcal{L}_{regist}$, computes dissimilarity in shape of the deformed and fixed images. The cycle loss function, $\mathcal{L}_{cycle}$, allows the displacement fields to preserve topology between the moving and deformed images. The identity loss function, $\mathcal{L}_{identity}$, prevents the network from generating deformation fields that deform stationary regions.

where λ is a hyper-parameter, ϕ_{XY} is deformation vector fields from G_X with the input X and Y, and $\otimes$ denotes the local normalized cross-correlation (Balakrishnan et al., 2019), which is computed by:

$$A \otimes B = \sum_{\mathbf{v}\in\Omega} \frac{\left(\sum_{\mathbf{v}_i}(A(\mathbf{v}_i)-\bar{A}(\mathbf{v}))(B(\mathbf{v}_i)-\bar{B}(\mathbf{v}))\right)^2}{\left(\sum_{\mathbf{v}_i}(A(\mathbf{v}_i)-\bar{A}(\mathbf{v}))^2\right)\left(\sum_{\mathbf{v}_i}(B(\mathbf{v}_i)-\bar{B}(\mathbf{v}))^2\right)}, \tag{6}$$

where Ω denotes the whole 3D volume, and $\bar{A}(\mathbf{v})$ and $\bar{B}(\mathbf{v})$ denote the local mean value of volume $A(\mathbf{v})$ and $B(\mathbf{v})$, respectively. Here, $\mathbf{v}_i$ iterates over a $w \times w \times w$ pixels around $\mathbf{v}$ (or $w \times w$ for the 2-D registration case), with $w = 9$ in our study.

3.1.2. Cycle loss

To retain the topology between the moving and deformed images, we design the cycle consistency on a pixel level of images as shown in Fig. 3. Specifically, an image X is first deformed to an image $\hat{Y}$, after which the deformed image is registered again by another network to generate image $\tilde{X}$ in the proposed framework. Then, the cycle consistency is applied between the re-deformed image $\tilde{X}$ and its original image X to impose $X \simeq \tilde{X}$. Similarly, an image Y should be successively deformed by the two networks to generate image $\tilde{Y}$, and the cycle consistency allows to impose $Y \simeq \tilde{Y}$.

Here, one of the important parts of cycle loss for our registration framework is that the network receives two inputs: moving and fixed images. Thus, the correct implementation of the cycle consistency should be given by as the vector-form of the cycle consistency condition:

$$(X, Y) \simeq \left(\mathcal{T}(\hat{Y}, \hat{\phi}_{YX}), \mathcal{T}(\hat{X}, \hat{\phi}_{XY})\right), \tag{7}$$

where

$$(\hat{Y}, \hat{X}) := (\mathcal{T}(X, \phi_{XY}), \mathcal{T}(Y, \phi_{YX})). \tag{8}$$

Therefore, the cycle loss is computed by:

$$\mathcal{L}_{cycle}(X, Y, G_X, G_Y) = \|\mathcal{T}(\hat{Y}, \hat{\phi}_{YX}) - X\|_1 + \|\mathcal{T}(\hat{X}, \hat{\phi}_{XY}) - Y\|_1, \tag{9}$$

where $||\cdot||_1$ denotes the l_1-norm.

3.1.3. Identity loss

When deforming images by displacement vector fields, the stationary regions of images should not be changed as fixed points. To consider this and improve the registration accuracy, as shown in Fig. 3, we design the identity constraint by imposing that the input image should not be deformed when the identical images are used as both a moving and fixed images. We implement the identity loss as following:

$$\begin{aligned}\mathcal{L}_{identity}(X, Y, G_X, G_Y) \\ = -(\mathcal{T}(Y, G_X(Y, Y)) \otimes Y) - (\mathcal{T}(X, G_Y(X, X)) \otimes X),\end{aligned} \tag{10}$$

where $\otimes$ denotes the local cross correlation defined in (6). Since minimizing the negative of cross correlation loss allows the similarity between deformed image and fixed images to be maximized, the maximum for the identical inputs can be achieved by not performing deformation (or trivial identity deformation). Thus, this identity loss prevents unnecessary deformation, increasing the stability of the deformation vector fields estimation in stationary regions.

3.2. Spatial transformation layer

In order to deform a moving image X with the displacement vector fields ϕ from the network, we add the spatial transform layer $\mathcal{T}$ proposed in (Jaderberg et al., 2015) to the network. Specifically, for 3D image registration in our experiments, we adopt the 3D transformation function with trilinear interpolation, which can be defined as:

$$\mathcal{T}(X, \phi) = \sum\nolimits_{q\in\mathcal{N}(p+\phi(p))} X(q) \prod\nolimits_{d\in\{i,j,k\}} (1 - |p_d + \phi(p_d) - p_d|), \tag{11}$$

where p indicates the pixel index, $\mathcal{N}(p+\phi(p))$ denotes the 8-pixel cubic neighborhood around $p+\phi(p)$, and d is three directions in 3D image space. Similarly, in case of 2D image registration, we deform the image by applying bilinear interpolation in the spatial transform layer. Since this grid sampling via spatial transformer network is differentiable, our deep learning model can be trained by backpropagating errors during optimization.

3.3. Multiscale image registration

Although the proposed CycleMorph provides powerful deformation on various image domains, deep neural networks should be trained using GPU, whose bottleneck is the limited memory. Especially, this is a problem for 3D image registration, such as contrast enhanced CT registration of a liver at multiple time points.

Since CycleMorph can be applied not only to full-sized images but also to downsampled images and local patches, the issue of memory limitation can be resolved by multiscale image registration method, i.e. global registration followed by local registration. Fig. 4 shows the schematic flow diagram of training and test stages in the proposed multiscale registration method. More details are as follows.

3.3.1. Training stage

In the training phase, the global and local registration networks are trained separately one after the other using the proposed cycle-consistent model. Specially, the global image registration model is trained on given image pairs that can be full-resolution images or sub-sampled images. In the latter case, the full-resolution deformed images are obtained by up-sampled deformation fields. Then, the local image registration model is trained on patches extracted from the deformed images from the global registration and the original fixed images.

3.3.2. Test stage

Although the global and local registration networks are trained separately, the successive deformation of a moving image with

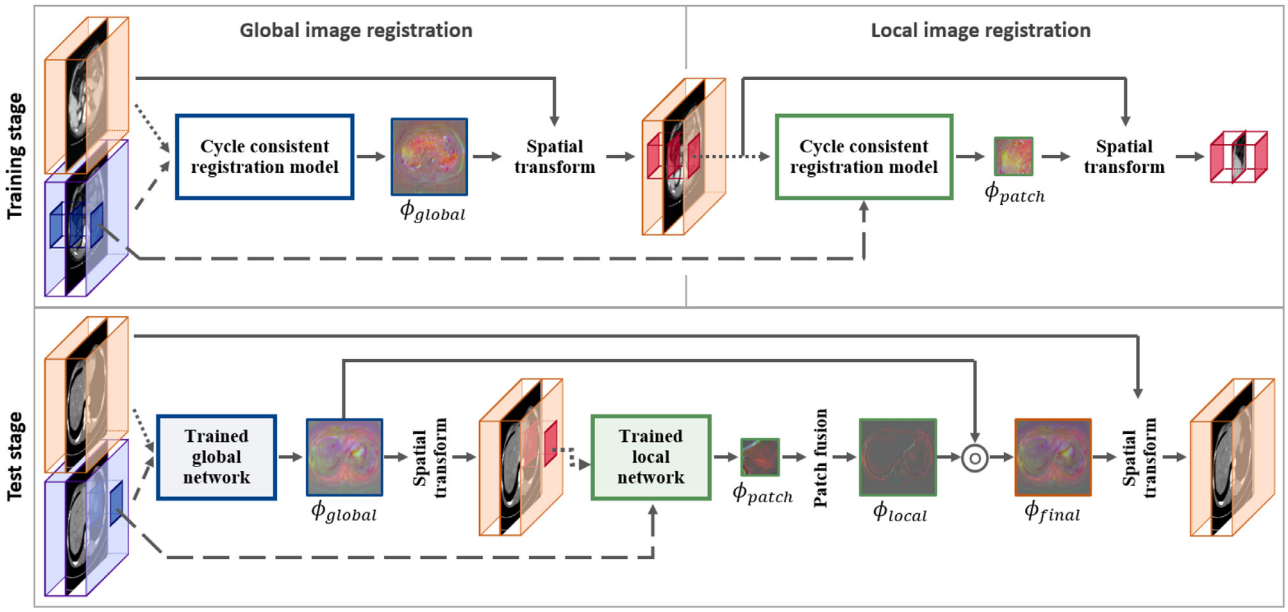

Fig. 4. Flow diagram of the multiscale CycleMorph registration method for large volume images. The upper part illustrates the flow of training stage for global and local image registration. The lower part shows the flow of test stage using the trained global and local registration networks for a given moving and fixed images. The short- and long-dashed lines indicate moving image and fixed image, respectively.

two registration networks potentially reduces the registration accuracy due to the accumulation of interpolation errors at each stage. Therefore, rather than deforming a moving image twice, the trained global and local networks are applied successively to estimate the deformation field at each scale, and the final deformation of the moving image is performed only once using the refined deformation field (see Fig. 4).

Specifically, given a new pair of input composed of moving and fixed images, the trained global registration network generates an intermediate deformed image and the corresponding deformation vector field ϕ_{global}. Then, the local registration network takes an input of patches extracted from the deformed image and the fixed image, so that it can generate the deformation field ϕ_{patch} for each patches. And then, to fuse the deformation fields and obtain a local deformation field ϕ_{local} at the fine scale, all patches of deformation fields generated by the local network are stitched and overlapped with regular intervals. Here, in order to get displacements that form plausible deformation at the boundaries of each patches, we set the overlap size to be large. Then, a final deformation vector field ϕ_{final} is estimated by composition of global and local deformation fields, $\phi_{global} \circ \phi_{local}$. This composition method can be implemented by warping ϕ_{global} with ϕ_{local} and adding the result with ϕ_{local} (Vercauteren et al., 2009). In Fig. 5, the Jacobian determinants map of global and fused fine deformation shows plausible local deformation of our approach. Finally, with this ϕ_{final} and the spatial transformer, the moving image is deformed once to align with the fixed target image.

Accordingly, the final deformed image can have a resolution similar to that of the original moving image without accumulating an interpolation error.

4. Method

To demonstrate the flexibility and improved performance of the proposed method, we conduct experiments using images from various application domains. First, we apply our method to face expression images to show the registration performance on 2D images. Second, we apply our method for 3D brain MR registration benchmark dataset, in which individual brain images are registered to a common atlas. Finally, we verify our approach using a very challenging registration problem with liver CECT data set, where extensive deformation from large 3D volumes should be estimated for multiphase contrast enhancement pattern analysis.

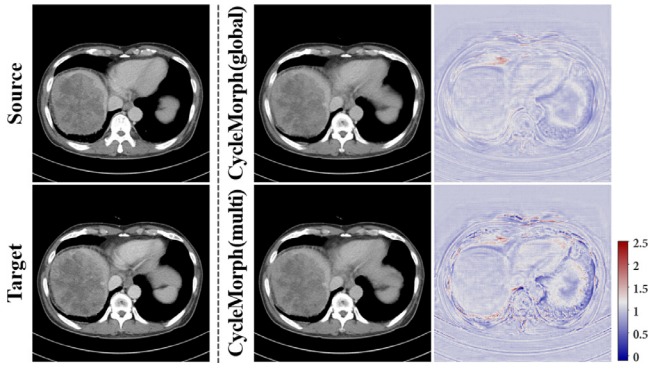

Fig. 5. Examples of multiscale image registration. The first column shows source and target CT images, and the second and third columns show deformed images and Jacobian determinants maps of global and multiscale deformation fields from the proposed CycleMorph, respectively. The intensity range of CT images is [-150, 250]HU.

4.1. Datasets

4.1.1. Facial expression image

The 2D face expression images are obtained from Radboud Faces Database (RaFD) (Langner et al., 2010). This provides eight different facial expression images for each 67 subjects; neutral, angry, contemptuous, disgusted, fearful, happy, sad, and surprised. This dataset also provides three different gaze directions for all facial expressed images so that there are total 1608 images. We divided the dataset by 53, 7, and, 7 participants for training, validation, and test images, respectively, and used all pairs of face images gazing not only the same direction but also different directions. We cropped all images to 640×640 and resized them into 128×128.

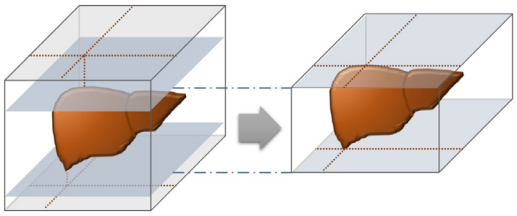

Fig. 6. Process to extract slices including liver in abdominal CT images.

4.1.2. Brain MRI

For brain MR image registration task, we used OASIS-3 (LaMontagne et al., 2018) dataset. This provides 1249 T1-weighted 3D brain MR images and corresponding volumetric segmentation results produced through FreeSurfer (Fischl, 2012). Specifically, we first preprocessed the data using standard preprocessing steps: resampling all scans to $256 \times 256 \times 256$ grid with 1mm^3 isotropic voxels, affine spatial normalization, and brain extraction. Then, we cropped the images to $160 \times 192 \times 224$, and divided by 255. We used 1027 scans for network training, 93 scans for validation, and 129 scans for test data.

4.1.3. Multiphase liver CT

The multiphase liver CT scans are provided by Asan Medical Center, Seoul, South Korea. Each scans was acquired from the patients with risk factors for HCC in the liver. The scan is 4D liver CT in that 3D volumes in four-phase (unenhanced, arterial, portal and 180-s delayed phases) before and after the contrast agent injection. The data have a resolution of $512 \times 512 \times depth$, where $depth$ is the number of slices for each CT images, and the slice thickness is 5mm. We trained the networks for image registration using 555 scans and evaluated our method on 50 test scans.

Here, since $depth$ of multiphase images may be all different due to their different scanning time, image coverage, and field-of-view of images, we extracted slices including liver using a segmentation network trained by an improved U-Net rather than resampling data into same image size. Then, to stack moving and fixed images along the channel direction as a network input, we performed zero-padding to the above and below volumes to make the number of slices same as shown in Fig. 6, which allows the input images to have same characteristics with the original images without any information loss in liver region. We normalized the images with the maximum value of each volume.

4.2. Implementation details

The proposed deformable registration method was implemented in Python using PyTorch library. The specific implementation details for face and medical image registration tasks are as follows. The code is available at https://github.com/jongcye/MEDIA_CycleMorph.

4.2.1. 2D Face expression image registration

For the face image registration, we employed 2D U-Net (Ronneberger et al., 2015) that takes 2D images as an input of moving and fixed images and generates a deformation field in width and height directions. In training of the network, we used the Adam optimization algorithm with learning rate 2×10^{-5} and batch size 1. We set the hyper-parameters as $\alpha = 1$, $\beta = 1$, and $\lambda = 1$. We augmented the data by horizontal flipping and trained the model for 133,560 iterations using a single GPU, NVIDIA GeForce GTX 1080 Ti. For the input, we converted the RGB images to gray-scale, but to obtain deformed images with RGB channels, we applied the same deformation fields of gray-scale images to each RGB channels at the test stage.

4.2.2. 3D Medical image registration

In order to evaluate the proposed model with 3D medical image registration task, we adopted 3D CNN that takes 3D volumes and generates displacement vector fields in width-, height-, and depth directions. We used VoxelMorph-1 (Balakrishnan et al., 2018) as a baseline network, so that our deep learning model without both the cycle and identity loss is equivalent to VoxelMorph-1. This network architecture consists of encoder, decoder and their connections similar to U-Net (Ronneberger et al., 2015). Here, because of the high memory usage for training the 3D CNN, we set the batch size to 1. For data augmentation, we adopted horizontal or vertical flipping and rotation with 90 for each training volume pair to improve registration performance without over-fitting.

For brain MRI registration, we set the hyper-parameters as $\alpha = 0.1$, $\beta = 0.5$, and $\lambda = 1$. To train the networks, we applied Adam with a momentum optimization algorithm with the learning rate of 2×10^{-4}. Using a single GPU, NVIDIA Titan RTX, we trained the model for 30 epochs. Here, even though the brain registration task fits in the GPU memory, we also tested our multiscale registration method to compare with the existing registration approach. For training of local registration model in the multiscale approach, we extracted patches from the globally deformed images and fixed images with size of $p \times p \times p$, where $p = 64$ in our experiment. We set the learning rate as 1×10^{-4} and trained the model for 70 epochs. At the inference phase, we got the local registration fields by overlapping the patches by $\frac{3}{4}p \times \frac{3}{4}p \times \frac{7}{8}p$.

For multiphase liver CT image registration, we adopted the multiscale registration method to address GPU memory limitation. For the global registration model, we sub-sampled the pair of input images from $512 \times 512 \times depth$ to $128 \times 128 \times depth$ to fit in the GPU memory size, but at the inference phase, we obtained full-resolution deformation fields by upsampling. Also, the training method of local registration model in multiscale approach was same with the brain registration method mentioned above. Using a single GPU, NVIDIA GeForce GTX 1080 Ti, we trained the global and local registration networks for 50 and 60 epochs, respectively, by Adam optimization with learning rate 10^{-4}. We set the hyper-parameters as $\alpha = 0.1$, $\beta = 1$, and $\lambda = 1$.

4.3. Evaluation

To verify the proposed method quantitatively, we evaluated the registration accuracy between the deformed and fixed images. First, we use the common evaluation criterion by measuring the regularity of the deformation fields ϕ. This can be done by computing the percentage of non-positive values in determinant of Jacobian matrix on ϕ, which can be defined by:

$$|J_\phi(\mathbf{v})| = |\nabla\phi(\mathbf{v})| \leq 0. \tag{12}$$

where $\mathbf{v}$ denotes the voxel location and $|\cdot|$ is the determinant of a matrix. Since the diffeomorphism is a one-to-one smooth and continuous mapping that has nonzero Jacobian determinant (Ashburner, 2007), the percentage of negative Jacobian determinants indicates how much the registration is different from diffeomorphic registration.

Additional quantitative evaluation criterion for each datasets depends on each application: the facial expression image dataset has ground-truth labels of deformed images; the brain MR dataset has segmentation map for several brain structures; and the liver CT dataset has anatomical landmark points. Therefore, we adopted different evaluation methods for each datasets. The details are as follows.

4.3.1. Analysis of face expression image registration

In the face expression image registration tasks, we deform different facial expression images of the same person looking in the

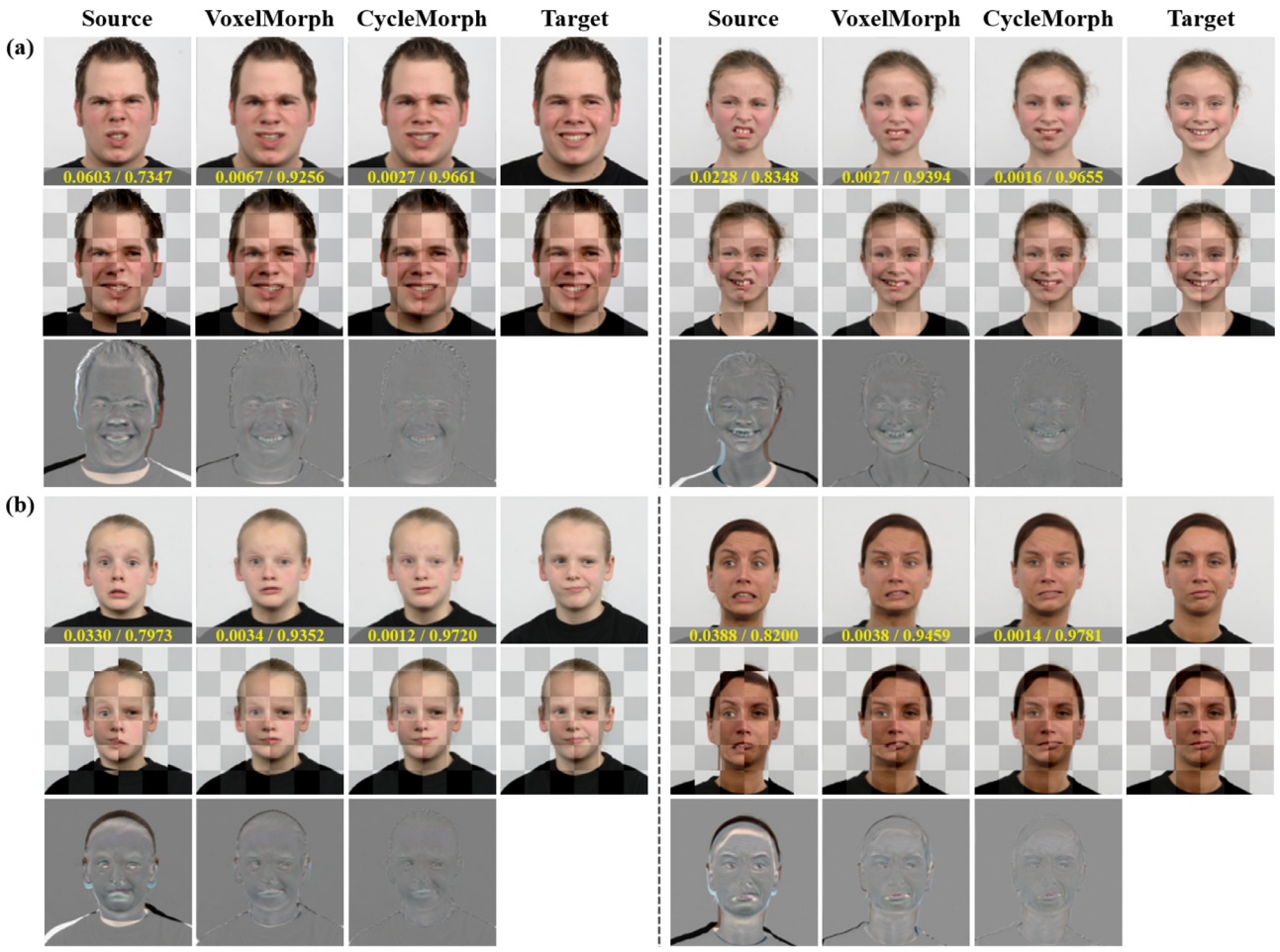

Fig. 7. Qualitative comparison results of face image registration for different gaze and expression images. Left panel: Results of face images gazing from front to right directions. Right panel: Results of face images gazing from left to front directions. (a) Results from disgusted to happy face registration. (b) Results from fearful to contemptuous face registration. In each (a) and (b), the first row shows source, deformed results, and target images with the average NMSE / SSIM values, the second row shows checkerboard visualization of image registration (bright: source/results/target, dark: target), and the last row shows difference images between the source/results and target images.

same or different direction. Accordingly, there are ground-truth labels for all deformed images, so we evaluated the results of the face image registration by the normalized mean square error (NMSE) and structural similarity (SSIM) between deformed images and fixed target images. For all pairs of face expression images, we averaged the scores for quantitative analysis.

4.3.2. Analysis of brain MRI registration

Since the brain MR dataset we used has segmentation labels for anatomical structures of brain, we evaluated the registration performance using the Dice score between the deformed segmentation map and fixed atlas segmentation label, which can be computed as:

$$Dice(A,B) = \frac{2TP}{2TP + FP + FN}, \tag{13}$$

where TP, FP, and FN are the number of pixels of true positive, false positive, and false negative regions. Among the segmented anatomical structures, we extracted 30 structures that are typically composed of over 100 pixels in a volume. To get segmentation maps for the registered images, we deformed the original segmentation map of moving image with the deformation fields computed from the registration networks between the original image and the atlas.

4.3.3. Analysis of liver CT registration

For the quantitative evaluation of liver CT registration, we computed the target registration error (TRE) on the 20 anatomical and pathological points in the liver and adjacent organs on the axial portal-phase images of the 50 test CT scans, which are marked by radiologists. The TRE can be computed by the average Euclidean distance as following:

$$TRE(A,B) = \frac{1}{N}\sum_{i=1}^{N} \|a_i - b_i\|, \tag{14}$$

where N is the number of landmark points, a_i and b_i is the i-th landmark coordinate vectors in the moving image A and fixed image B, respectively. Also, we measured differences of liver cancer size with major and minor lengths of cancer region to verify the performance in the view point of tumor diagnosis. The specific information of the marking points is described in Appendix A.

4.3.4. Comparative methods

In order to verify the improved performance of the proposed method, we employed several comparative methods that show the state-of-the-art performance in the image registration: Elastix (Klein et al., 2009), SyN (Avants et al., 2008) by Advanced Normalization Tools (ANTs) (Avants et al., 2011), VoxelMorph (Balakrishnan et al., 2018), VoxelMorph-diff (Dalca et al., 2018), ICNet (Zhang, 2018), and MS-DIRNet (Lei et al., 2020). For

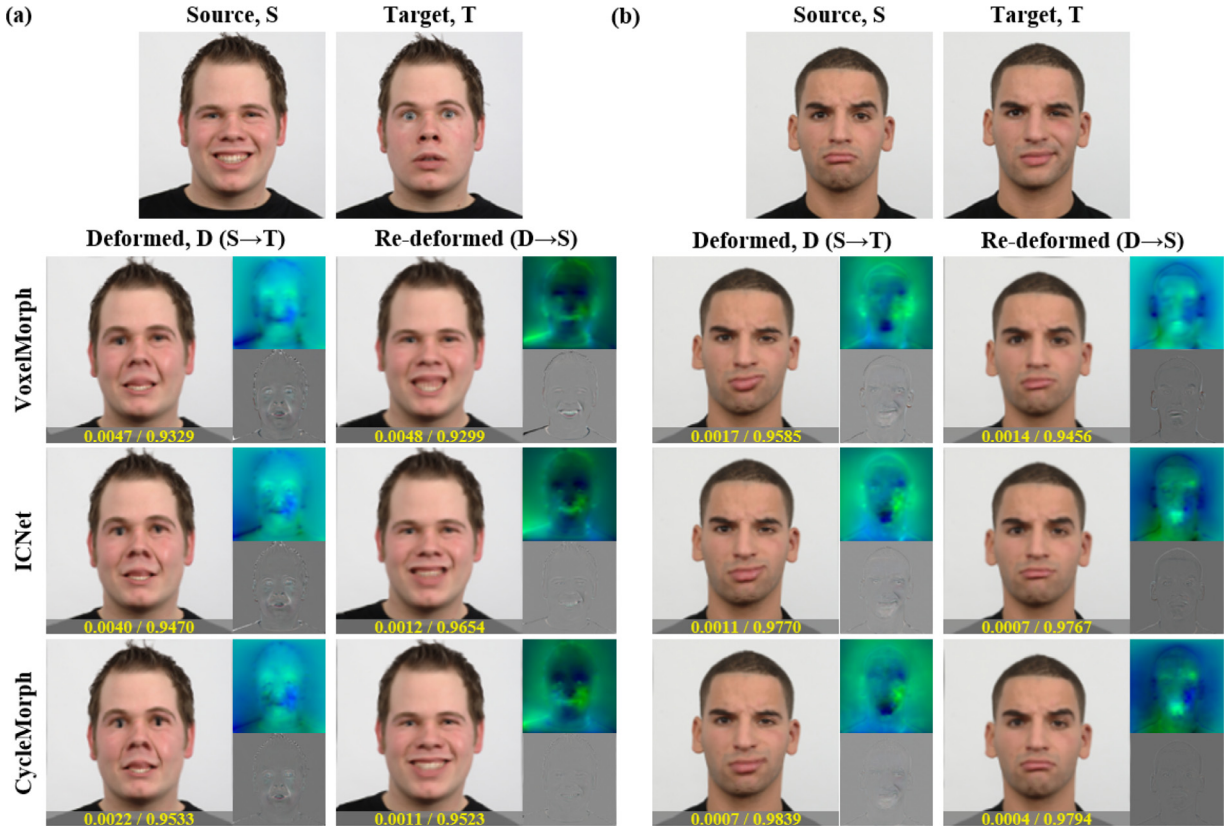

Fig. 8. Face expression image registration performance with various qualitative results. (a) Results from happy to fearful face registration. (b) Results from sad to contemptuous face registration. For each (a) and (b), the first row shows source and target images. From the second row, the first column shows deformed images of the source into the target, the third column shows re-deformed images from the deformed images into the original source images, and both the second and the fourth columns show deformation fields (top) and difference images between the results and each fixed images (bottom). The average values of NMSE / SSIM are displayed on each results.

the deep learning methods, we used VoxelMorph-1 proposed by Balakrishnan et al. (2018) as a baseline network, and employed same parameters for fair comparison. For VoxelMorph-diff, we set $\sigma = 0.05$ and $\lambda = 50,000$ for better results. Since MS-DIRNet (Lei et al., 2020) is one of the representative multi-scale approaches to address GPU memory issues, we used MS-DIRNet as a baseline method to compare our multiscale implementation of CycleMorph. The specific choices of comparative algorithms are determined by applications.

5. Experimental results

5.1. Face expression image registration

In 2D face image registration, we adopted the VoxelMorph and ICNet to compare with the proposed method and evaluated the results.

5.1.1. Qualitative evaluation

Fig. 7 shows visual comparisons of the results on various face expression photos of men, women, and children. We deformed the source image into the target image. We can observe that the proposed method deforms source image to be more similar to the target image compared to VoxelMorph, especially on the region of eyes and mouth. ICNet provides similar artifacts as VoxelMorph (not shown in the figure). In most of the data set, we found that the proposed CycleMorph provides better quality results of image registration compared to VoxelMorph and ICNet.

To explicitly analyze the effect of cycle consistency to preserve diffeomorphism, we additionally performed the study on deformation fields whether the deformed images preserve topology and can be returned to their original images. In order to register the deformed images into the original images reversely, we set the forward deformed images as new source images and the original source images as new target images, and applied the same registration networks. The figures in the right column of Fig. 8(a)(b) illustrate the visual comparison results of backward image registration. It shows that the proposed method provides deformed images that can be reversed to the original images, while deformed images from VoxelMorph and ICNet reversed worse.

Table 1
Quantitative evaluation results on the face expression image registration. NMSE, SSIM, and the percentage of non-positive values in determinant of Jacobian matrix of deformation fields are evaluated on the all test pairs of face expression images (Parentheses: standard deviations across data). Two asterisks (**) denote $p \ll 0.05$ from the statistical significance test using t-test between CycleMorph and the second best method, ICNet.

Method	NMSE $\times 10^{-1}$	SSIM	$\lvert J_\phi \rvert \leq 0$ (%)
Initial	0.363 (0.268)	0.823 (0.066)	0
VoxelMorph	0.047 (0.057)	0.936 (0.024)	0.050 (0.106)
ICNet	0.029 (0.041)	0.959 (0.018)	0.022 (0.066)
CycleMorph	0.019 (0.023) **	0.968 (0.012) **	0.018 (0.059) **

5.1.2. Quantitative evaluation

Table 1 includes the quantitative evaluation results of the comparative methods and our proposed method. To compare the performance of image registration effectively, we computed evaluation scores on the source and target images before the registration. By comparison, we found that the proposed CycleMorph decreases NMSE by 0.034 and increases SSIM by 0.145 compared to the initial. Also, our method outperforms VoxelMorph and ICNet by

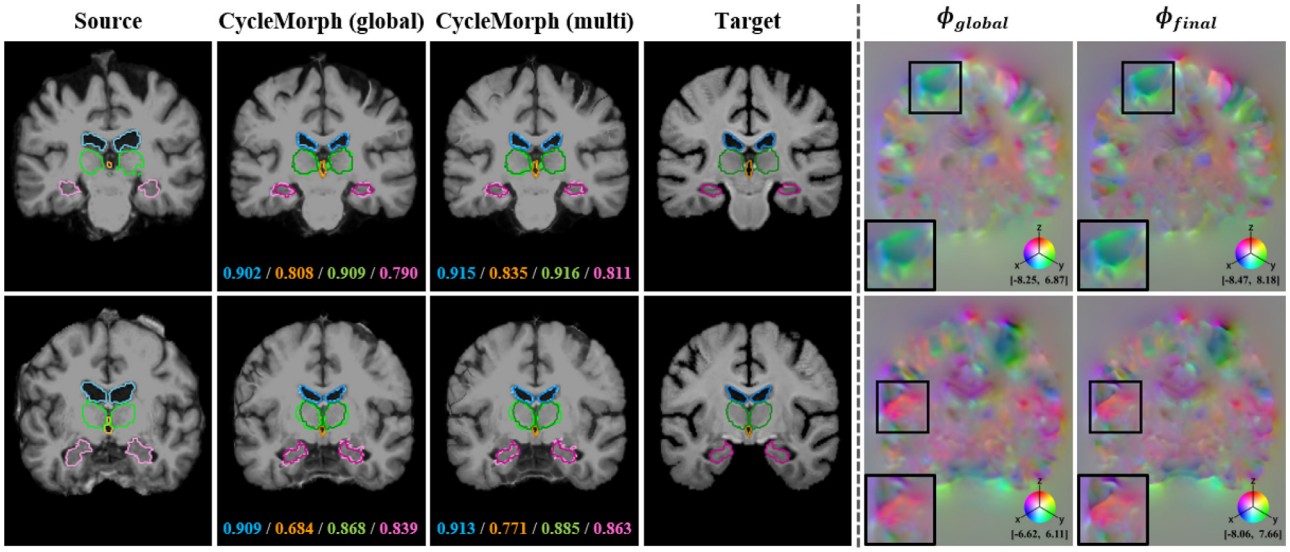

Fig. 9. Qualitative results of atlas-based brain MR image registration of the proposed method. We overlaid boundaries of several anatomical structures with dark color for target contours and light color for source/deformed contours (blue: ventricles, orange: third ventricle, green: thalami, pink: hippocampi). The Dice scores for each structures are displayed with the corresponding colors on each results. The moving source images are in the first column, deformed images from the proposed CycleMorph are in the second column (global) and the third column (multiscale), and the fixed target images are in the fourth column. The last two columns show the corresponding deformation fields for global and multiscale image registration, where the RGB color map represents the 3D fields in x-(blue), y-(green), and z-(red) directions, and $[p, q]$ in colorbars denotes the magnitude range of the fields. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison of consistency of reversed images in the face expression image registration (Parentheses: standard deviations across data). Two asterisks (**) denote $p \ll 0.05$ from the statistical significance test using t-test between CycleMorph and the second best method, ICNet.

Method	NMSE $\times 10^{-1}$	SSIM
VoxelMorph	0.025 (0.031)	0.945 (0.016)
ICNet	0.013 (0.016)	0.968 (0.012)
CycleMorph	0.007 (0.004) **	0.971 (0.009) **

Table 3
Quantitative evaluation results on the brain MR image registration. The Dice score, the percentage of non-positive values in determinant of Jacobian matrix of deformation fields, and the runtime (min) are computed for all test scans (Parentheses: standard deviations across data). Two asterisks (**) denote $p \ll 0.05$ from the statistical significance test. We performed t-test on global CycleMorph with VoxelMorph and multi CycleMorph with MS-DIRNet.

Method	Dice	$\lvert J_\phi \rvert \leq 0$ (%)	Time
Initial	0.616 (0.171)	0	0
ANTs SyN	0.752 (0.140)	0.400 (0.100)	122 (CPU)
VoxelMorph	0.749 (0.145)	0.553 (0.075)	0.01 (GPU)
VoxelMorph-diff	0.731 (0.139)	0.631 (0.073)	0.01 (GPU)
ICNet	0.743 (0.146)	0.488 (0.083)	0.01 (GPU)
MS-DIRNet	0.751 (0.142)	0.804 (0.089)	2.06 (GPU)
CycleMorph (global)	0.750 (0.144) **	0.510 (0.087) **	0.01 (GPU)
CycleMorph (multi)	0.758 (0.141) **	0.450 (0.081) **	2.19 (GPU)

0.032% and 0.004% additional reduction, respectively, in the metric on Jacobian determinant. When we performed hypothesis testing with paired t-test, p-values on all metrics for our method over the second best method were $p \ll 0.05$, which indicates that our CycleMorph outperforms the comparative methods.

Additionally, Table 2 shows the NMSE and SSIM between the re-deformed images (the figures in the right column of Fig. 8(a)(b)) and their original moving images. The reversed images from our method are very similar to the original images with lower NMSE and higher SSIM compared to VoxelMorph and ICNet. Therefore, we can confirm that the cycle constraint in our proposed method plays an important role in producing deformation fields that improve topology preservation of input images with less folding problem.

5.2. Brain MR image registration

To evaluate the proposed method on atlas-based brain MR image registration, we compared the method with several comparative methods: ANTs SyN for traditional method, VoxelMorph, VoxelMorph-diff, and ICNet for deep-learning-based global registration approaches, and MS-DIRNet for learning-based multiscale registration method.

5.2.1. Qualitative evaluation

The results of atlas-based brain MR image registration are shown in Fig. 9. The proposed CycleMorph method deforms images accurately for each pairs of the moving source and fixed target images, which can be specifically verified with the segmentation boundaries of several brain structures. In addition, thanks to the cycle constraint, we can observe that the image registration is performed by the smooth deformation fields. We displayed the qualitative results of all methods in Fig. B.1 in Appendix B, which also shows the high quality of image deformation from CycleMorph compared to the comparative learning based methods.

5.2.2. Quantitative evaluation

Fig. 10 represents Dice scores for the evaluated anatomical structures across test scans. The scores of left and right brain structures are averaged into one score. Our CycleMorph models achieve higher scores on most of structures than the comparative methods of global registration and MS-DIRNet of multiscale registration. In particular, on some structures such as brain stem, forth ventricle, and amygdala, our global and multiscale CycleMorph models perform better than the other comparative methods. We showed the score values in Table B.1 in Appendix B.

Table 3 shows the quantitative evaluation results with average Dice scores across all structures and scans, the percentage of non-positive values in Jacobian determinant, and runtime. For the global registration, the proposed CycleMorph shows higher Dice

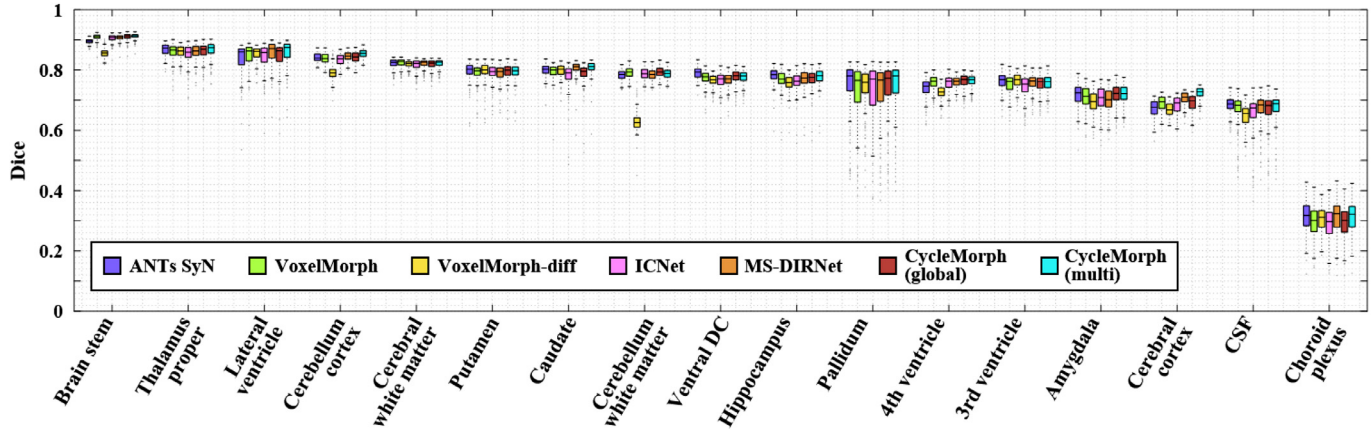

Fig. 10. Dice scores on the deformed segmentation maps of brain anatomical structures for quantitative comparisons of atlas-based brain MR image registration.

Table 4
Results of study on local patch size. Dice and the percentage of non-positive values in determinant of Jacobian matrix of deformation fields are computed on the all test scans. "p#" denotes patch size # × # × # used in local image registration (Parentheses: standard deviations across data).

Method (CycleMorph)	Dice	$\lvert J_\phi \rvert \leq 0$ (%)
global	0.7502 (0.144)	0.5101 (0.087)
global + local(p64)	0.7580 (0.141)	0.4497 (0.081)
global + local(p80)	0.7572 (0.141)	0.4472 (0.082)
global + local(p96)	0.7562 (0.141)	0.4564 (0.082)

measures with less percentage of non-positive Jacobian determinant compared to VoxelMorph and VoxelMorph-diff. These results are similarly shown in the comparison of multiscale registration methods between MS-DIRNet and our method. In ICNet method, the average Dice score was lower than the CycleMorph (global), although Jacobian determinant index was slightly better. On the other hand, among the learning-based methods, our multiscale CycleMorph method was the best by reducing the non-positive Jacobian determinants and increasing the accuracy of 3D image registration in terms of Dice score.

5.2.3. Study on local patch size

Since the patch size in local image registration of our method can vary, we also studied on the effect of local patch size in the proposed model. As shown in Table 4, we set the global registration results as a baseline, and conducted the experiment of local registration with different patch sizes.

When we compared the results with Dice scores and Jacobian determinant, all multiscale methods improve the global registration results with higher Dice score while generating deformation fields with less non-positive values in Jacobian determinant. Also, when the patch size is smaller, we can observe that the result on Dice score of anatomical structures tends to be higher, and the deformation regularity tends to be better with less folding problem. From these results, we used the patch size of $64 \times 64 \times 64$ in our experiments.

5.3. Multiphase liver CT image registration

For the evaluation of the proposed method, we adopted several comparative methods: Elastix and ANTs SyN for classical methods, VoxelMorph for the comparison of learning-based global registration method, MS-DIRNet for the comparison of learning-based multiscale registration method.

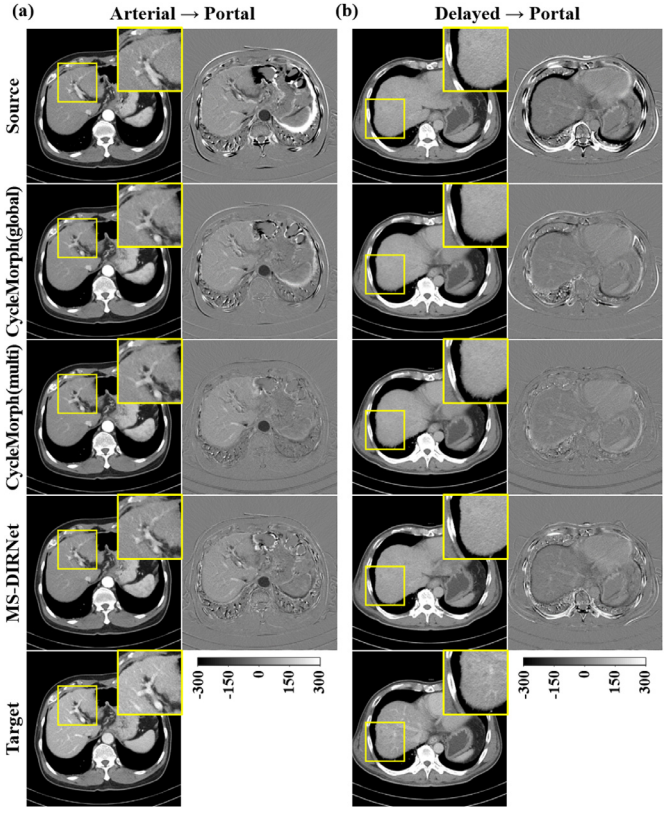

Fig. 11. Global and multiscale (global followed by local) registration results of the proposed CycleMorph on the multiphase liver CT dataset. (a) Results from the images in arterial to portal phases. (b) Results from the images in delayed to portal phases. For each (a) and (b), the yellow box shows the remarkable parts, and the second column shows difference images between the source/results and target images. Intensity range of the CT images is [-150, 250]HU, and that of the difference images is [-300,300]HU. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.3.1. Qualitative evaluation

Fig. 11 and Fig. 12 illustrate the multiscale registration results by CycleMorph. Specifically, Fig. 11 shows that the multiscale registration performance is improved over the global registration and MS-DIRNet, which is well visualized in the difference images between the deformed images and target images. From this result, we can confirm that the global registration tends to deform whole shape of the source images to fit into the fixed target images, while the local registration provides the local region deformation. In ad-

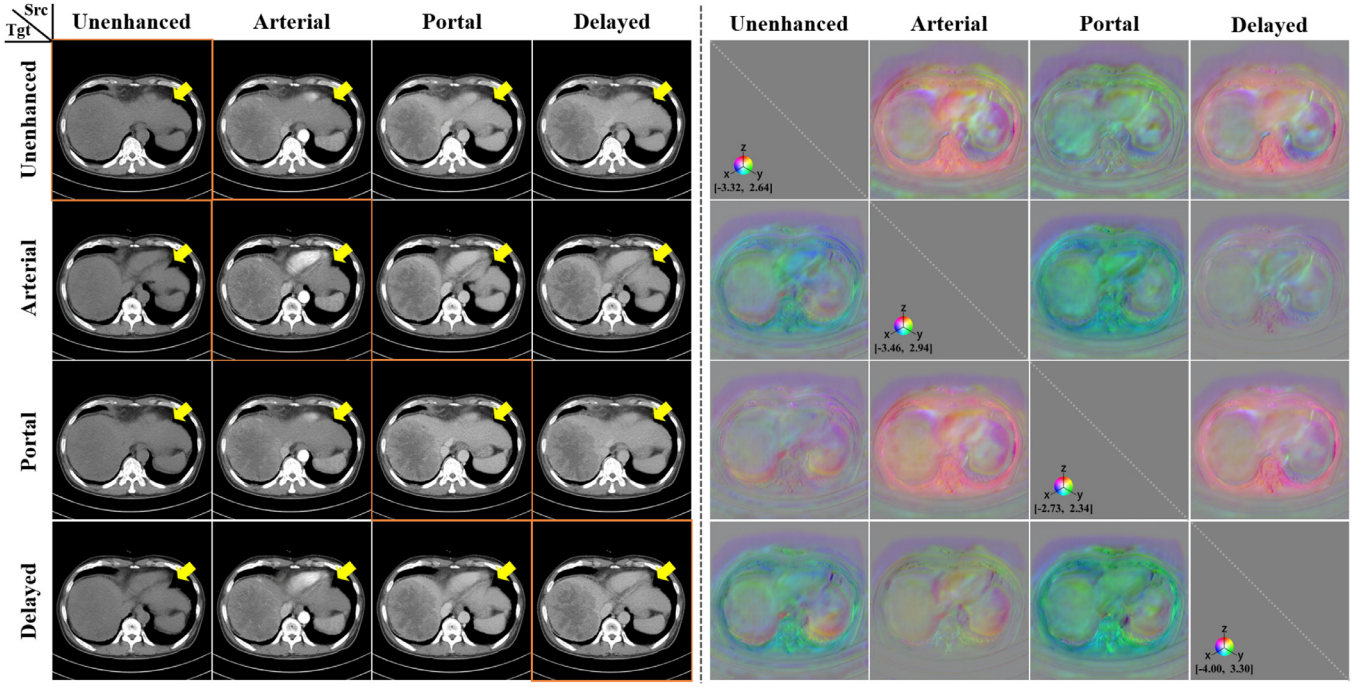

Fig. 12. Qualitative results of multiphase liver CT registration with a single trained network. Left: deformed images from the source (src) to target (tgt) images with intensity range of [-150, 250]HU. The yellow arrows with the same position that indicate the remarkable parts of the results. The diagonal images with orange box are original images, and they are deformed to other phase images as indicated by each row. The $(i, j), i \neq j$, element of the figure represents the deformed image to the i-th phase from the j-th phase original image. Right: deformation vector fields for the left deformed results, where the RGB color map is to show the 3D fields in x-(blue), y-(green), and z-(red) directions. For each deformation fields from one source image to the others three target images, we apply the same magnitude ranges $[p, q]$ that are indicated in colorbars on the diagonal images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Quantitative evaluation results on the multiphase liver CT image registration. TRE (*mm*), tumor size differences, and the average test time (min) are evaluated on the deformed arterial/delayed images into the fixed portal image (Parentheses: standard deviations across subjects). Asterisks denote statistically significant difference of global CycleMorph over VoxelMorph and multi CycleMorph over MS-DIRNet. We denoted $p \ll 0.05$ with two asterisks ($**$) and $p \leq 0.05$ with one asterisk ($*$).

Method	Arterial → Portal					Delayed → Portal				
	TRE	Tumor size diff		Time		TRE	Tumor size diff		Time	
		Major	Minor	GPU	CPU		Major	Minor	GPU	CPU
Elastix	3.261 (1.143)	0.981	0.610	-	19.64	2.963 (0.913)	0.910	0.577	-	19.64
ANTs SyN	6.269 (3.726)	0.631	0.613	-	158	5.173 (2.167)	0.508	0.591	-	158
VoxelMorph	6.674 (4.217)	0.789	1.638	0.11	0.80	5.351 (1.892)	0.610	0.868	0.11	0.80
MS-DIRNet	5.021 (4.175)	0.186	0.178	0.69	14.22	4.042 (1.938)	0.136	0.102	0.69	14.10
CycleMorph (global)	4.722 (3.294) $**$	0.631	0.563	0.06	0.86	3.902 (1.694) $**$	0.275	0.209	0.06	0.86
CycleMorph (multi)	4.715 (3.407) $*$	0.493	0.611	0.74	14.32	3.915 (1.674)	0.282	0.220	0.74	14.26

dition, Fig. 12 shows that the proposed multiscale method provides accurate registration results with smooth deformation vector fields on the all pairs of multiphase 3D images with different contrast.

5.3.2. Quantitative evaluation

We performed quantitative evaluation of the registration results on the deformed images in arterial/delayed phases into portal phase that is often used as a standard in the clinical practice. Here, since the images in unenhanced phase are difficult to obtain the landmark points, we did not compute the evaluation metrics on the unenhanced phase images. Table 5 shows the results of average TRE, tumor size differences, and average runtime for a 3D image registration with various comparative methods.

Specifically, we can observe that the proposed global CycleMorph achieves significant improvement of registration performance compared to the existing deep learning methods of VoxelMorph with paired t-test p-values of quite less than 0.05. Also, in arterial to portal phase image registration, the performance improvement by our multiscale registration with CycleMorph was statistically significant from MS-DIRNet with p-values of 0.048 and with lower TRE values. In addition, the tumor size differences between the deformed images and the portal phase images from our method are smaller than Elastix, ANTs SyN, and VoxelMorph. These results show that the proposed method provides comparable deformation over other methods on 3D CT images. Although MS-DIRNet shows the smallest tumor size differences among the comparisons, we can confirm that the registration quality of the multiscale registration with our CycleMorph is much better than MS-DIRNet as shown in Fig. 11.

Furthermore, we calculated the runtime of image registration using a CPU and GPU for the comparative methods. When computing the registration on the CPU, ANTs SyN takes more than two hours, while Elastix requires approximately 20 minutes. The computational time of global CycleMorph was within 1 minutes. This verifies that the runtime of our learning-based global registration method is much faster on the CPU compared to the classical meth-

Table 6
Results of ablation study on loss function. TRE (mm) and the percentage of non-positive values in determinant of Jacobian matrix of deformation fields are computed on the deformed arterial/delayed images into the fixed portal image (Parentheses: standard deviations across subjects). Asterisks denote statistically significant difference of the ablated methods over the proposed CycleMorph. We denoted $p \ll 0.05$ with two asterisks ($**$) and $p \leq 0.05$ with one asterisk ($*$).

Method	Arterial → Portal		Delayed → Portal	
	TRE	$\lvert J_\phi \rvert \leq 0$ (%)	TRE	$\lvert J_\phi \rvert \leq 0$ (%)
Proposed w/o $\mathcal{L}_{cycle} + \mathcal{L}_{identity}$	5.377 (3.888) $**$	0.058 (0.170)	4.415 (1.831) $**$	0.039 (0.064)
Proposed w/o $\mathcal{L}_{cycle}$	5.241 (4.017) $**$	0.085 (0.217)	4.210 (1.737) $*$	0.083 (0.176)
Proposed w/o $\mathcal{L}_{identity}$	5.006 (3.864) $*$	0.049 (0.131)	4.212 (1.931) $*$	0.049 (0.117)
Proposed (CycleMorph)	4.722 (3.294)	0.032 (0.099)	3.902 (1.694)	0.029 (0.084)

ods, ANTs SyN and Elastix. Also, the multiscale registration with CycleMorph on the CPU takes about 14 minutes, which even shows faster runtime than Elastix. On the other hand, when we tested the registration on the GPU, the deep learning based models takes less than 1 minutes. Here, the global registration of the proposed method only takes about 4 seconds, and the total runtime of multiscale registration is 41 seconds.

6. Discussion

6.1. Ablation study on loss function

To verify the effect of cycle constraint in our designed loss function, we performed an ablation study on liver CT data by excluding the cycle loss and/or identity loss. For this study, we analyzed results of the global image registration with the same training and test procedure for fair comparison. Table 6 shows that the percentage of the number of non-positive values in the determinant of Jacobian matrix on deformation fields as well as TRE are dependent upon the loss functions. Specifically, the network only trained by the registration loss, i.e. without $\mathcal{L}_{cycle}$ and $\mathcal{L}_{identity}$, deforms images with the largest errors among the methods. And both of the cycle and identity loss functions increase the accuracy of the registration. Here, there is a statistically significant difference between CycleMorph and the proposed without $\mathcal{L}_{cycle} + \mathcal{L}_{identity}$ ($p \ll 0.05$). In the registration from arterial to portal phase images, CycleMorph was statistically significantly better than the proposed method without $\mathcal{L}_{cycle}$ or $\mathcal{L}_{identity}$ with p-values of 0.0006 and 0.029, respectively. This difference was similarly observed in the task of delayed to portal phase image registration with p-values of 0.001 and 0.002.

Additionally, the evaluation metric of Jacobian matrix emphasizes the effect of cycle consistency. Specifically, the proposed method without the cycle loss produces deformation fields with more non-positive voxels of determinant of Jacobian matrix than the proposed method with the cycle constraint. Here, the reason that the method only with the registration loss has smaller percentage of non-positive values of Jacobian determinants than the other ablated methods is because the network provides registration fields that hardly deform on the large-scale images, which can be confirmed with TRE values. In contrast, thanks to the cycle loss, the proposed method is less prone to folding problem and enhances topological preservation on 3D image registration.

6.2. Convergence

In the proposed CycleMorph, the network sees two times the same amount of data than other methods at each iterations. However, as shown in Fig. 13, when we draw convergence curves of the registration loss over epochs in training stage, the rate of convergence and convergence value are very similar across the other methods even though CycleMorph is seen to have better generalization properties. This suggests that the performance improvement of CycleMorph does not come from the number of seen data, but the effect of the cycle constraint in network training.

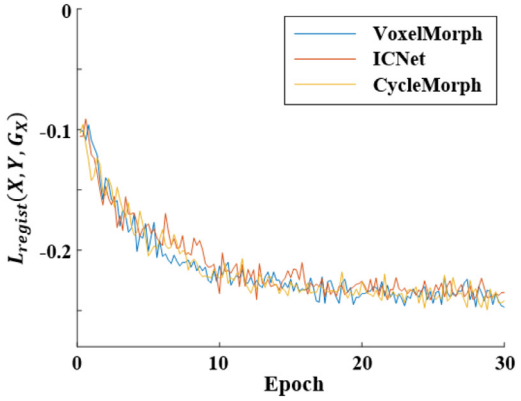

Fig. 13. Convergence curves of the registration loss $L_{regist}(X, Y, G_X)$ for the global registration methods in the 3D brain registration experiment, where G_X deforms a moving image X into the fixed atlas image Y.

6.3. 2D Vs. 3D image registration

We showed the image registration performance using 2D natural images as well as 3D medical images. In the experiment of 2D face expression image registration, CycleMorph achieved much larger gain over the comparative learning based methods, compared to the 3D medical registration results.

The difference in the amount of gain may come from the difference of image dimension in that two-dimensional natural images can be easier deformed than three-dimensional images. Also, the difference on the number of learning parameters causes the performance gap. In the 3D medical image registration experiments, we could not use the same number of channels and depth used in 2D image registration due to the limitation of GPU memory. Although we set the 3D network as deep as possible, the 3D CNN has 200 times fewer learning parameters than the 2D CNN we used, which makes the 3D network less expressive than that for 2D image registration. We expect that the performance improvement of CycleMorph in 3D medical image registration would be higher when the networks deepen and learning parameters increase with a larger GPU memory.

6.4. Deformation fields composition in multiscale registration

To combine two global and local deformation fields as one deformation fields, CycleMorph used the composition method that warps the global and local fields and adds the result with the local fields (Vercauteren et al., 2009). This method improves the registration performance compared to the additive method that adds

Table 7
Comparison of composite and additive methods for combining global and local deformation fields in multiscale image registration (Parenthesis: standard deviation across data). The metrics for brain MR and multiphase liver CT datasets are the average Dice and TRE, respectively, and we computed the average runtime (min) on a GPU.

Method	Metric	$\lvert J_\phi \rvert \leq 0$ (%)	Time (GPU)
Brain MRI			
addition	0.756 (0.141)	0.788 (0.100)	2.18
composition	0.758 (0.141)	0.450 (0.082)	2.19
Multiphase liver CT			
addition	4.323 (2.637)	0.178 (0.414)	0.69
composition	4.315 (2.701)	0.074 (0.177)	0.74

the two fields. Specifically, as shown in Table 7, while the employed composition method takes slightly longer time than the simple addition method because of additional computational cost, we obtained higher Dice score on brain MR image registration with 0.2% gain and less TRE on multiphase liver CT image registration. The Jacobian determinant index was also better in the composition method, since the composition method can compensate for the degraded fields at the boundaries of patches from local image registration.

7. Conclusion

In this paper, we presented a CycleMorph, a novel cycle consistent deep learning model for unsupervised deformable image registration method. CycleMorph imposes cycle consistency between a pair of images. In addition, to deal with the memory issues for 3D registration, multiscale registration is possible by combining the global and local registration. Once the networks are trained in a certain image domain, we found that a single network can provide accurate image registration on unseen data in the same image domain.

Throughout the experiments using the various image datasets, we demonstrated that CycleMorph can provide topology-preserved image deformation between moving and deformed images. Also, CycleMorph achieves performance improvement over the several comparative learning based methods. Accordingly, CycleMorph may be useful to deform images with high accuracy and fast runtime in various image domains.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

CRediT authorship contribution statement

Boah Kim: Conceptualization, Methodology, Software, Writing - original draft, Visualization. **Dong Hwan Kim:** Investigation. **Seong Ho Park:** Investigation, Resources. **Jieun Kim:** Software, Investigation. **June-Goo Lee:** Validation, Investigation, Data curation. **Jong Chul Ye:** Conceptualization, Writing - review & editing, Supervision.

Acknowledgements

This work was supported in part by the Industrial Strategic technology development program (10072064, Development of Novel Artificial Intelligence Technologies To Assist Imaging Diagnosis of Pulmonary, Hepatic, and Cardiac Diseases and Their Integration into Commercial Clinical PACS Platforms) funded by the Ministry of Trade Industry and Energy (MI, Korea), in part by KAIST R&D Program (KI Meta-Convergence Program) 2020 through Korea Advanced Institute of Science and Technology (KAIST), and also in part by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2020-0-01461) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation). We would like to thank Junyoung Kim for the discussion on the face expression dataset.

Appendix A. Quantitative evaluation of multiphase liver CT registration

For the quantitative evaluation of registration performance on multiphase liver CT data, the marking of anatomical and pathological points and tumor size measurement were performed by an expert. Specifically, an abdominal radiologist (D.H.K., with 8-year experience with liver CT) marked the following anatomical and pathological points in the liver and adjacent organs on the axial portal-phase images of the 50 CT datasets using a dedicated software (Medical Imaging Interaction Toolkit Workbench): (1) the uppermost point of the liver (i.e., right hepatic dome); (2) the left end of the left lateral hepatic section; (3) the inferior tip of the right hepatic lobe; (4) the innermost border of the caudate lobe of the liver; (5) the most caudal part of the gallbladder fundus; (6) gallbladder stones if present; (7) the points where right, middle, left, and right inferior (if present) hepatic veins meet the inferior vena cava; (8) suprahepatic inferior vena cava at the level that shows its maximum width; (9) the caudal boundary of the splenic veins entry into the main portal vein; (10) the caudal boundary of the branching-off of the left portal vein from the main portal; (11) the origins of P2, P3, and P4 portal branches from the left portal vein; (12) the point of right portal veins branching into anterior and posterior segmental portal veins; (13) the points of right anterior and right posterior portal veins branching into segmental portal veins; (14) fissure for ligamentum teres or recanalized umbilical vein; (15) fissure for ligamentum venosum (at the level that shows the umbilical segment of the left portal vein); and (16) hepatic cysts or calcifications if present. In addition, the long- and short-diameters of hepatic HCCs (in the larges lesion if multiple nodules were present) were measured.

Appendix B. Evaluation results on brain MR image registration

Fig. B.1 displays the qualitative results of all methods we implemented. This shows that the proposed method achieves better registration results than the comparative methods. Also, the Dice scores of segmentation maps on anatomical structures are shown in Table B.1. The segmentation classes are notated by from 1 to 17 that are in the order of brain stem, thalamus proper, lateral ventricle, cerebellum cortex, cerebral white matter, putamen, caudate, cerebellum white matter, ventral DC, hippocampus, pallidum, fourth ventricle, third ventricle, amygdala, cerebral cortex, CSF, and choroid plexus.

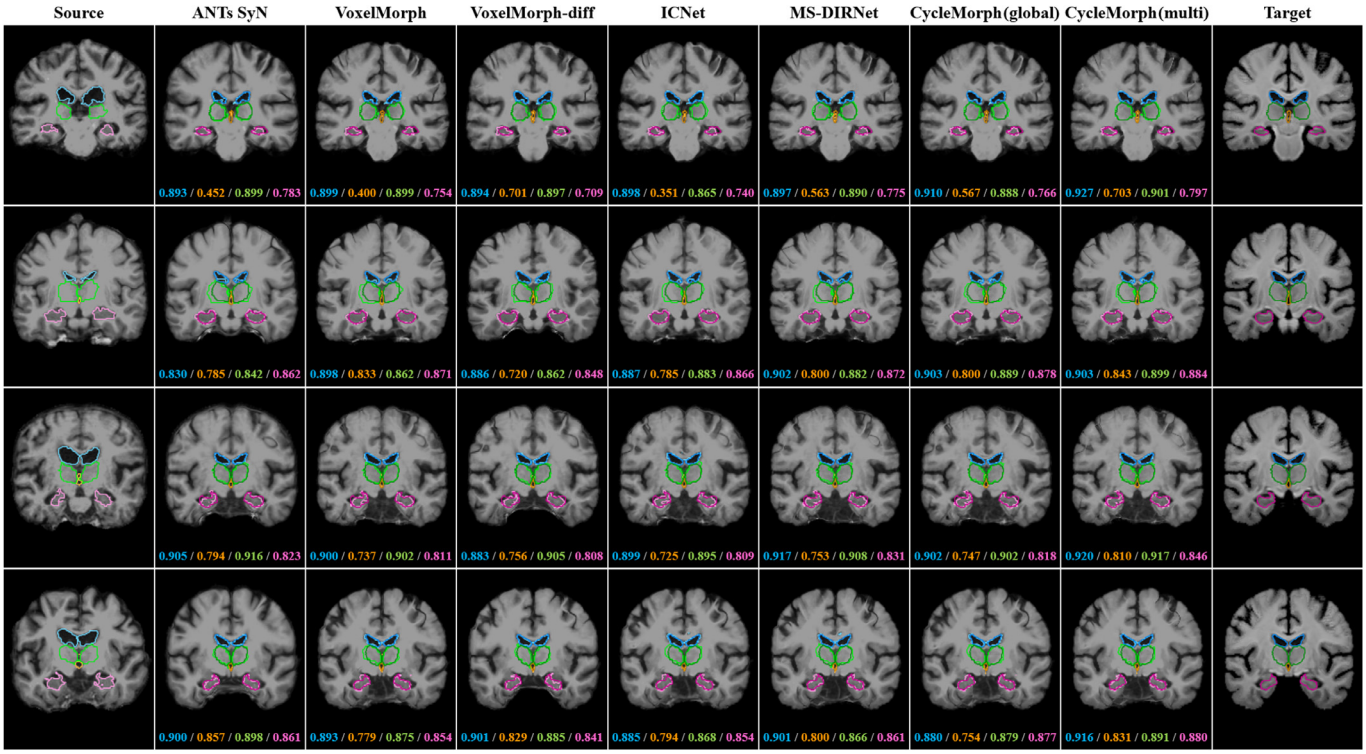

Fig. B.1. Qualitative results of atlas-based brain MR image registration from the proposed and comparative methods. We overlaid boundaries of several anatomical structures with dark color for target contours and light color for source/deformed contours (blue: ventricles, orange: third ventricle, green: thalami, pink: hippocampi). The Dice scores for each structures are displayed with the corresponding colors on each results.

Table B.1
Quantitative results with Dice scores on anatomical structures in 3D brain MR image registration experiment.

Method	Class																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ANTs SyN	0.893	0.867	0.836	0.841	0.820	0.794	0.798	0.783	0.790	0.781	0.737	0.739	0.762	0.716	0.672	0.683	0.312
VoxelMorph	0.909	0.859	0.845	0.838	0.823	0.790	0.792	0.790	0.772	0.768	0.722	0.756	0.753	0.709	0.689	0.665	0.294
VoxelMorph-diff	0.856	0.859	0.854	0.790	0.818	0.795	0.797	0.629	0.764	0.757	0.734	0.726	0.764	0.691	0.667	0.647	0.305
ICNet	0.816	0.683	0.841	0.787	0.834	0.852	0.778	0.789	0.720	0.744	0.756	0.904	0.760	0.701	0.650	0.764	0.287
MS-DIRNet	0.906	0.858	0.855	0.845	0.820	0.789	0.804	0.783	0.765	0.771	0.722	0.758	0.757	0.700	0.706	0.668	0.308
CycleMorph (global)	0.910	0.860	0.842	0.842	0.817	0.793	0.786	0.793	0.777	0.771	0.730	0.763	0.753	0.718	0.690	0.662	0.291
CycleMorph (multi)	0.911	0.865	0.857	0.854	0.820	0.792	0.806	0.787	0.775	0.777	0.737	0.762	0.757	0.718	0.723	0.669	0.309

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage 54 (3), 2033–2044.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2018. An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9252–9260.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging 38 (8), 1788–1800.

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. 61 (2), 139–157.

Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., Shen, D., 2018. Deep learning based inter-modality image registration supervised by intra-modality similarity. arXiv preprint arXiv:1804.10735.

Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D., 2017. Deformable image registration based on similarity-steered cnn regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 300–308.

Cao, Y., Miller, M.I., Winslow, R.L., Younes, L., 2005. Large deformation diffeomorphic metric mapping of vector fields. IEEE Trans. Med. Imaging 24 (9), 1216–1230.

Ceritoglu, C., Oishi, K., Li, X., Chou, M.-C., Younes, L., Albert, M., Lyketsos, C., van Zijl, P.C., Miller, M.I., Mori, S., 2009. Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. Neuroimage 47 (2), 618–627.

Christensen, G.E., Johnson, H.J., 2001. Consistent image registration. IEEE Trans. Med. Imaging 20 (7), 568–582.

Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2018. Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 729–738.

Fan, J., Cao, X., Xue, Z., Yap, P.-T., Shen, D., 2018. Adversarial similarity network for evaluating image alignment in deep learning based registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 739–746.

Fischl, B., 2012. Freesurfer. Neuroimage 62 (2), 774–781.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. In: Advances in neural information processing systems, pp. 2017–2025.

Kim, B., Kim, J., Lee, J.-G., Kim, D.H., Park, S.H., Ye, J.C., 2019. Unsupervised deformable image registration using cycle-consistent cnn. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 166–174.

Kim, K.W., Lee, J.M., Choi, B.I., 2011. Assessment of the treatment response of hcc. Abdom. Imaging 36 (3), 300–314.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2009. Elastix: a toolbox

for intensity-based medical image registration. IEEE Trans. Med. Imaging 29 (1), 196–205.
Krebs, J., Mansi, T., Mailhé, B., Ayache, N., Delingette, H., 2018. Learning structured deformations using diffeomorphic registration. arXiv preprint arXiv:1804.07172.
LaMontagne, P.J., Keefe, S., Lauren, W., Xiong, C., Grant, E.A., Moulder, K.L., Morris, J.C., Benzinger, T.L., Marcus, D.S., 2018. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimers disease. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 14 (7), P1097.
Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A., 2010. Presentation and validation of the radboud faces database. Cognition and emotion 24 (8), 1377–1388.
Lei, Y., Fu, Y., Wang, T., Liu, Y., Patel, P., Curran, W.J., Liu, T., Yang, X., 2020. 4D-ct deformable image registration using multiscale unsupervised deep learning. Physics in Medicine & Biology.
Leow, A., Huang, S.-C., Geng, A., Becker, J., Davis, S., Toga, A., Thompson, P., 2005. Inverse consistent mapping in 3d deformable image registration: its construction and statistical properties. In: Biennial International Conference on Information Processing in Medical Imaging. Springer, pp. 493–503.
Mahapatra, D., Antony, B., Sedai, S., Garnavi, R., 2018. Deformable medical image registration using generative adversarial networks. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE, pp. 1449–1453.
Onofrey, J.A., Staib, L.H., Papademetris, X., 2013. Semi-supervised learning of non-rigid deformations for image registration. In: International MICCAI Workshop on Medical Computer Vision. Springer, pp. 13–23.
Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X., 2017. Svf-net: Learning deformable image registration using shape matching. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 266–274.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.
Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B.P., Išgum, I., Staring, M., 2017. Non-rigid image registration using multi-scale 3d convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 232–239.
Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: efficient non-parametric image registration. Neuroimage 45 (1), S61–S72.
de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. Med Image Anal 52, 128–143.
Yang, X., Kwitt, R., Styner, M., Niethammer, M., 2017. Quicksilver: fast predictive image registration–a deep learning approach. Neuroimage 158, 378–396.
Zhang, J., 2018. Inverse-consistent deep networks for unsupervised deformable image registration. arXiv preprint arXiv:1809.03443.
Zhang, J., Ge, Y., Ong, S.H., Chui, C.-K., Teoh, S.-H., Yan, C.H., 2008. Rapid surface registration of 3d volumes using a neural network approach. Image Vis. Comput. 26 (2), 201–210.
Zhang, M., Liao, R., Dalca, A.V., Turk, E.A., Luo, J., Grant, P.E., Golland, P., 2017. Frequency diffeomorphisms for efficient image registration. In: International conference on information processing in medical imaging. Springer, pp. 559–570.
Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.