Medical Image Analysis 58 (2019) 101545

Contents lists available at ScienceDirect

Medical Image Analysis

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

Adversarial learning for mono- or multi-modal registration

Jingfan Fan[a], Xiaohuan Cao[a], Qian Wang[b], Pew-Thian Yap[a,*], Dinggang Shen[a,c,*]

[a] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[b] Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[c] Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

ARTICLE INFO

Article history:
Received 2 February 2019
Revised 16 June 2019
Accepted 19 August 2019
Available online 24 August 2019

Keywords:
Deformable image registration
Fully convolutional neural network
Generative adversarial network

ABSTRACT

This paper introduces an unsupervised adversarial similarity network for image registration. Unlike existing deep learning registration methods, our approach can train a deformable registration network without the need of ground-truth deformations and specific similarity metrics. We connect a registration network and a discrimination network with a deformable transformation layer. The registration network is trained with the feedback from the discrimination network, which is designed to judge whether a pair of registered images are sufficiently similar. Using adversarial training, the registration network is trained to predict deformations that are accurate enough to fool the discrimination network. The proposed method is thus a general registration framework, which can be applied for both mono-modal and multi-modal image registration. Experiments on four brain MRI datasets and a multi-modal pelvic image dataset indicate that our method yields promising registration performance in accuracy, efficiency and generalizability compared with state-of-the-art registration methods, including those based on deep learning.

1. Introduction

Deformable registration establishes anatomical correspondences between a pair of images. Traditional registration methods seek to estimate smooth deformation fields by optimizing the cost functions in association with similarity metrics (*i.e.*, volumetric-based, landmark-based, *etc.*). However, these methods typically involve high-dimensional numerical optimization and are often computationally expensive.

Deep learning methods have been shown recently to be capable of addressing the limitations of conventional registration methods. Other than using convolutional neural networks (CNN) (LeCun et al., 2015) to predict rigid transformation parameters (Miao et al., 2016; Salehi et al., 2018) or central displacement (Cao et al., 2017) from a local patch, more of the current works focus on predicting the voxel-to-voxel mapping (*i.e.*, dense deformation fields) from images by using U-net (Ronneberger et al., 2015) like fully convolutional neural networks (FCNN) (Long et al., 2015). By performing end-to-end learning, the registration network can be trained with ground-truth deformations, which cannot be manually delineated and are usually obtained by a traditional registration algorithm under the help of tissue segmentation maps (Cao et al., 2017; Yang et al., 2017; Rohé et al., 2017). While effective, these methods are however limited by the availability of the ground-truth deformations. In contrast, unsupervised learning methods (de Vos et al., 2017; Balakrishnan et al., 2018; Li and Fan, 2018) aim to learn the deformable transformations without ground-truth deformations by maximizing the similarity between a pair of images instead, such as the sum of squared difference (SSD) and cross-correlation (CC). However, it is often difficult to determine the most effective similarity metrics for different registration purposes. When dealing with multi-modal registration, this issue may become even more challenging, since most current similarity metrics cannot work well for deformable multi-modal registration.

In this paper, we propose an adversarial similarity network to automatically judge the image similarity through a network rather than using any arbitrary similarity metrics. The network is unsupervised. It is inspired by generative adversarial network (GAN) (Goodfellow et al., 2014). More specifically, we implement the framework by connecting a U-net based generator with a CNN based discriminator. The generator is a *registration network* that takes two input image volumes and outputs the same size predicted deformations. The discriminator is a *discrimination network* that judges whether images are well aligned and feeds misalignment information to the registration network during the training. The registration and discrimination networks are learned via adversarial training, *i.e.*, and the registration network is trained by the guidance provided by the discrimination network. The

* Corresponding authors.
E-mail addresses: ptyap@med.unc.edu (P.-T. Yap), dgshen@med.unc.edu (D. Shen).

https://doi.org/10.1016/j.media.2019.101545
1361-8415/

discrimination network is trained by using the output of the registration network. The main contributions of this work are summarized as follows:

- Compared with the traditional registration methods, a robust and fast end-to-end registration network is developed for predicting the deformation in one-pass.
- Compared with supervised learning methods for registration, the proposed network does not need ground-truth deformations in training. The network is trained in an adversarial and unsupervised manner.
- The proposed adversarial similarity network learns a meaningful metric for effective training of the registration network, without any prior assumption on the image pairs.
- The proposed method can be generalized for both mono-modal and multi-modal registration problem.

This paper extends a preliminary version of the work presented at the 2018 International Conference on Medical Image Computing and Computer-Assisted Intervention (Fan et al., 2018). This paper extends the design of the discrimination network for multi-modal images. We implement both whole-image-based and patch-based networks for the registration model.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 details our adversarial training strategy for medical image registration. Section 4 presents experimental results on both brain MRI data and CT-MR pelvic data. We discuss insights of the results and conclude in Section 6.

2. Related work

Image registration is a crucial and fundamental procedure in medical image analysis (Maintz and Viergever, 1998; Holden, 2008; Sotiras et al., 2013; Viergever et al., 2016; Zhou et al., 2019; 2018; Tang et al., 2019; 2018; Xue et al., 2006a; Xue et al., 2006b). The aim of registration algorithm is to obtain a spatial transformation that can align a subject (moving or source) image to a template (fixed or target) image. The spatial transformation includes linear transformation (Fan et al., 2016a; 2016b; 2017) such as translation, rotation, scaling, and shearing; and non-linear transformation determining voxel-to-voxel correspondences. Image registration involves determining the transformation ϕ^* that minimizes the image dissimilarity and keeps deformation smooth:

$$\phi^* = \underset{\phi}{\operatorname{argmin}}\, \operatorname{dissim}(I_M \circ \phi, I_F) + \lambda \operatorname{reg}(\phi), \tag{1}$$

where $I_M \in \mathbb{R}^3$ and $I_F \in \mathbb{R}^3$ denote the moving image and the fixed image, respectively. $I_M \circ \phi$ is the warped moving image using deformation ϕ. Image dissimilarity $\operatorname{dissim}(I_M \circ \phi, I_F)$ can be defined as the intensity sum of squared distance (SSD) (Rueckert et al., 1999), (normalized) cross-correlation (CC/NCC) (Wu et al., 2012; Wang et al., 2016), (normalized) mutual information (MI/NMI) (Studholme et al., 1999; Viola and Wells III, 1997), *etc.* $\operatorname{reg}(\phi)$ is a regularization term for ensuring smoothness of the estimated deformation field ϕ. λ is a regularization parameter that balances the similarity and smoothness. Regularity of the deformation field can be achieved by Gaussian smoothing (Viola and Wells III, 1997; Woods et al., 1998), utilizing a spline (Wu et al., 2012; Hellier et al., 2002) or diffeomorphic (Vercauteren et al., 2009; Avants et al., 2008) deformation model.

Numerous algorithms are proposed for medical image registration by optimizing (1). Demons (Thirion, 1998), HAMMER (Shen and Davatzikos, 2002), and Elastix (Klein et al., 2010) are standard methods for image registration. There are methods that strive to keep the deformation field smooth, topology-preserving, and diffeomorphic, such as diffeomorphic demons (Vercauteren et al., 2009), log-demons (Vercauteren et al., 2008), LCC-demons (Lorenzi et al., 2013), large deformation diffeomorphic metric mapping (LDDMM) (Cao et al., 2005), symmetric normalization (SyN) (Avants et al., 2008), and DARTEL (Ashburner, 2007). However, these methods often involve computationally expensive high-dimensional optimization.

Recently, deep learning methods have been shown to be promising in addressing the limitations of conventional registration methods. Deep learning methods can learn 1) transformation parameters and 2) deformation field as detailed below.

2.1. Learning transformation parameters

Image registration is traditionally a high dimensional optimization task as the deformation field consists of dense and smooth displacement vectors. To simplify the complicated optimization procedure, some studies focus on learning preliminary transformation parameters, such as rigid transformation parameters (Miao et al., 2016; Salehi et al., 2018), displacements of key points (Cao et al., 2017; 2018b) and registration momentum (Yang et al., 2017), by using patch-based CNN architecture.

In order to assess the pose and location of an implanted object during surgery, Miao *et al.* (Miao et al., 2016) used a CNN regression approach to achieve real-time 2D/3D registration. The transformation parameter space was partitioned into different zones and the CNN model was trained in each zone separately. Then, the transformation parameters were decomposed in a hierarchical manner. Salehi et al. (2018) also proposed a deep CNN regression model for 3D rigid registration. They estimated rigid transformation based on sectional 3D volumetric, and the bi-invariant geodesic distance was used as the loss function. These CNN models were trained using simulated data generated by manually adapting the transformation parameters.

Predicting the registration parameters for deformable registration is more challenging than rigid registration. Cao et al. (2017) used an equalized active-points sampling strategy to build a similarity-steered CNN model for predicting the displacements associated with the active points, and then the dense deformation field can be obtained by interpolation. This strategy significantly enhanced the accuracy when estimating the deformation field and did not require further refinement by traditional registration methods. Based on this framework, a cue-aware deep regression network was further proposed to more effectively train a registration network (Cao et al., 2018b). In these two methods, the ground-truth displacements were obtained with the help of tissue segmentations by using existing registration tools. In another study, Yang et al. (2017) predicted the momenta of the deformation in LDDMM setting (Cao et al., 2005). LDDMM model takes an initial momentum value for each voxel (which is often computationally expensive) as input to calculate the final deformation field. The authors circumvented this by training a U-Net-like architecture to predict the dense momentum map from the input images. They trained the prediction network using training images and the ground-truth initial momentum obtained by numerical optimization of LDDMM.

2.2. Learning deformation field

The deformation field can be learned directly using deep learning. Recently, FCNN (Long et al., 2015) and U-Net (Ronneberger et al., 2015) have shown effective for end-to-end learning of voxel-to-voxel prediction. Considering the success of these deep networks in image segmentation tasks, researchers now show keen interest in using these networks for predicting the dense deformation field in the image registration task.

Training of the registration network can be supervised with ground-truth deformation fields. Uzunova et al. (2017) focused

on synthesizing a huge amount of realistic ground-truth training data for deep learning based medical image registration. Basically, they learned a statistical appearance model from the available training images and applied this model to synthesize an arbitrary number of new images with varying object shapes and appearance. Rohé et al. (2017) proposed a U-Net-like architecture, namely SVF-Net, to perform image registration. In order to obtain an accurate ground-truth mapping between the image pair, they built reference deformations for training by registering the segmented regions-of-interest (ROIs) instead of registering the intensity images. Sokooti et al. (2017) proposed RegNet to estimate the displacement field from a pair of chest CT images. The training process was conducted on a variety of simulated deformations acting as the ground truth, while the testing stage used the trained model for aligning the baseline and follow-up CT images of a same patient.

Additional guidance, such as image similarity metric and segmented labels, can be employed to refine training. Fan et al. (2019) proposed a hierarchical dual-supervised U-Net-like network. The deformation field achieved by a conventional registration method was used as the coarse guidance to pre-train the network, then the similarity between the reference image and the warped floating image was used as fine guidance to further refine training. Hu et al. (2018b,c) used labeled corresponding structures for training but without labeling for registration during testing. These approaches improve the accuracy of deformation learning when the ground-truth deformations are not accurate.

Unsupervised learning allows the deformation fields to be learned directly from the to-be-registered image pair (Balakrishnan et al., 2018; Li and Fan, 2018; de Vos et al., 2017; Yan et al., 2018). Unsupervised learning models estimate voxel-to-voxel deformable transformation by maximizing image similarity. Some standard similarity metrics, which are differentiable, such as SSD and NCC, can be employed to define the loss function for training registration networks. In addition, regularization loss (Vishnevskiy et al., 2016) can also be used to constrain the smoothness of the predicted deformation field. Unsupervised learning makes it possible to train the registration network using large-scale unlabeled images.

3. Method

3.1. Registartion network

In this paper, we propose a general adversarial learning framework for 3D image registration of mono- or multi-modality images. In the rest of the paper, the mono-modal images are exemplified by brain MR images and the multi-modal images are exemplified by pelvic MR images and CT images.

We suppose all the moving images $\{I_M \in \mathbb{R}^3\}$ and fixed images $\{I_F \in \mathbb{R}^3\}$ are linearly aligned. The adversarial learning framework aims to predict the dense voxel-to-voxel correspondences, in the form of a deformation field, from the moving image to the fixed image. We design a *registration network R*, to learn the deformation field ϕ: $R : (I_M, I_F) \rightarrow \phi$. The registration network R is trained to maximize image similarity and does not require ground-truth deformation fields. Instead of a similarity metric, image similarity is determined based on a *discrimination network D*, which can be trained adversarially to judge whether the two images are well aligned with probability $p \in [0, 1]$. The registration network R is trained to register the images as accurate to fool the discrimination network D so that the registered images do not differ. The *registration network R* is employed to ensure smooth deformation.

As illustrated in Fig. 1, in the training stage, the registration network R and the discrimination network D are connected by a spatial transformation layer, which connects the output of the R (*i.e.*,

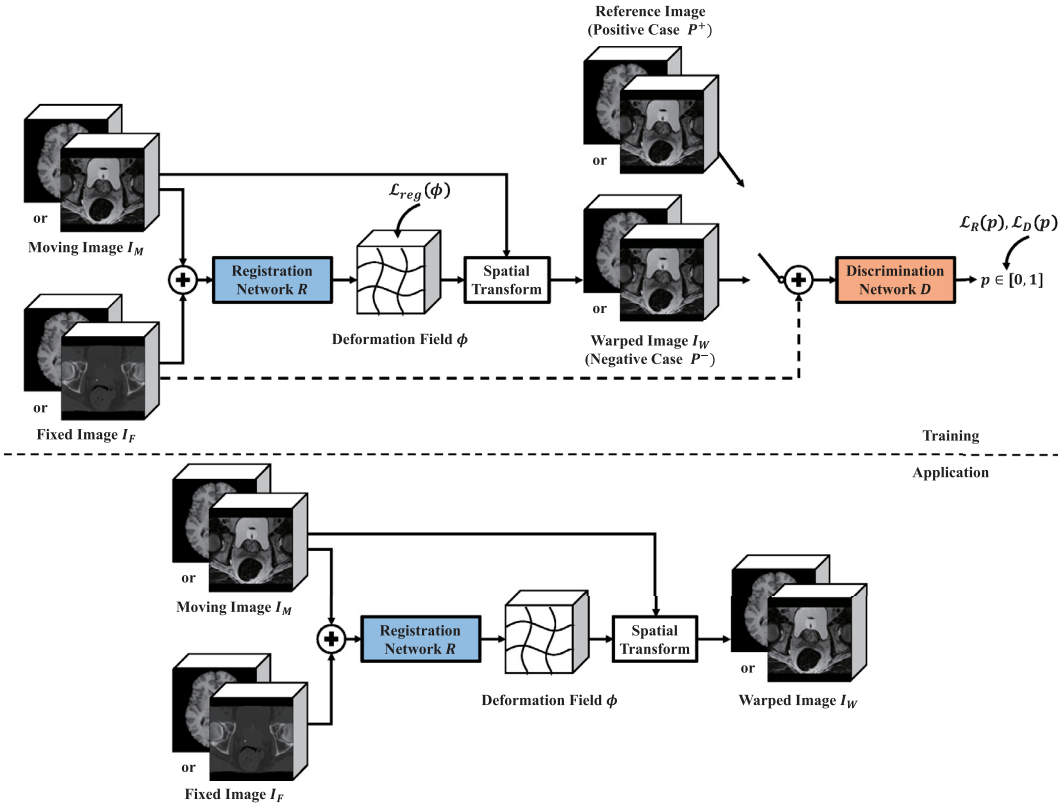

Fig. 1. The proposed adversarial similarity network for deformable image registration. The input image pair is already linearly aligned.

the deformation field ϕ) to the input of D (the warped moving image). In the application stage, the deformation field is predicted by R.

3.2. Adversarial training

Networks R and D are trained alternatively in an adversarial fashion. The training involves 1) feeding to D a reference image (an image similar to the fixed image, defined in Section 3.5) with the fixed image to learn how a pair of well-registered images look like; 2) feeding to D the moving image warped using the deformation field estimated by the registration network and the fixed image to learn how a misaligned pair of images look like; and 3) feeding to R the moving image and the fixed image to learn a deformation field with high score given by D. These steps are iterated to train the D (Section 3.3) and R (Section 3.4).

3.3. Training the discrimination network

The discrimination network D determines whether the input image pair is similar (i.e., being well-registered). Two cases are fed into the network alternatively: 1) well-registered image pairs, called *positive case* (P^+) and 2) the misaligned image pairs, called *negative case* (P^-). The loss function of D is defined as

$$\mathcal{L}_D(p) = \begin{cases} -\log(p), & c \in P^+ \\ -\log(1-p), & c \in P^- \end{cases}, \tag{2}$$

where, p is the output of the D indicating the probability of similarity, and c indicates the input case. During training, each positive case is expected to give a value close to 1 and each negative case a value close to 0. The *discrimination network D* can be optimized by minimizing the loss function (2).

3.4. Training the registration network

The registration network R aims to make the registered images as similar as possible, giving a high p-value from D. The similarity loss function is defined as

$$\mathcal{L}_R(p) = -\log(p), c \in P^-. \tag{3}$$

The smoothness of deformation field ϕ is enforced with loss

$$\mathcal{L}_{\text{reg}}(\phi) = \sum_{v \in \mathbb{R}^3} \nabla\phi^2(v), \tag{4}$$

where v represents the voxel location. In addition, the anti-folding penalization (Zhang, 2018) is utilized to enhance the regularization term, by penalizing large foldings (*i.e.*, $\nabla\phi(v) + 1 < 0$). By jointly considering (3) and (4), the total loss function for R is

$$\mathcal{L} = \mathcal{L}_R(p) + \lambda \cdot \mathcal{L}_{\text{reg}}(\phi), \tag{5}$$

where λ balances the two losses and is set to 1000. More discussion about this parameter is provided in Section 5.

Training is carried out iteratively by alternatively optimizing R and D. Convergence occurs when D cannot distinguish between positive and negative cases.

3.5. Definition of a positive case

Each positive case is a well-registered image pair consisting of a reference image and the fixed image. The reference image is defined differently for mono-modal registration and multi-modal registration. For mono-modal registration, the ideal positive case is when the image pair is exactly identical. However, this is an over-strict requirement and is not practical since anatomical structures vary across subjects. To avoid this, we generate the reference image based on the original moving image I_M and fixed image I_F (see Fig. 2) using

$$I_R = \alpha \cdot I_M + (1-\alpha) \cdot I_F, 0 < \alpha < 1, \tag{6}$$

where we set $\alpha = 0.2$ in the initial training stage (*i.e.*, the first 5 epochs) to weaken the similarity requirement and $\alpha = 0.1$ in later stage for greater accuracy.

For multi-modal registration, the reference image and the moving image are from the same modality but the fixed image is from a different modality. We use a small number of paired MR and CT images (Cao et al., 2018a) from the same subjects as reference. The discrimination network is trained using unpaired MR and CT images.

3.6. Network details

3.6.1. Registration network

A number of registration networks (Yang et al., 2017; Fan et al., 2019; Cao et al., 2018b; Balakrishnan et al., 2018) can be used in the proposed adversarial framework. We chose U-Net (Ronneberger et al., 2015) for its capability in localized pixel-wise learning, owing to its contracting encoder and expansive decoder paths (see Fig. 3). In the encoder path, the multi-channel inputs first go through a convolutional layer with $3 \times 3 \times 3$ kernels, followed by rectified linear unit (ReLU) activation (He et al., 2015). A

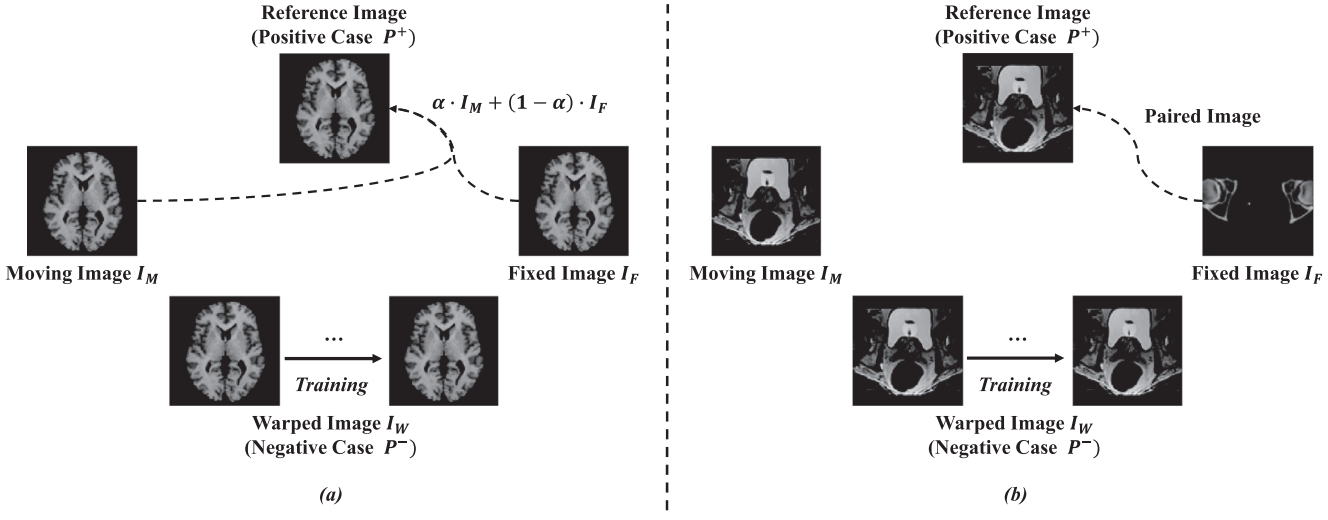

Fig. 2. The definitions of the positive and negative cases for the discrimination network. (a) definition in mono-modal images, (b) definition in multi-modal images.

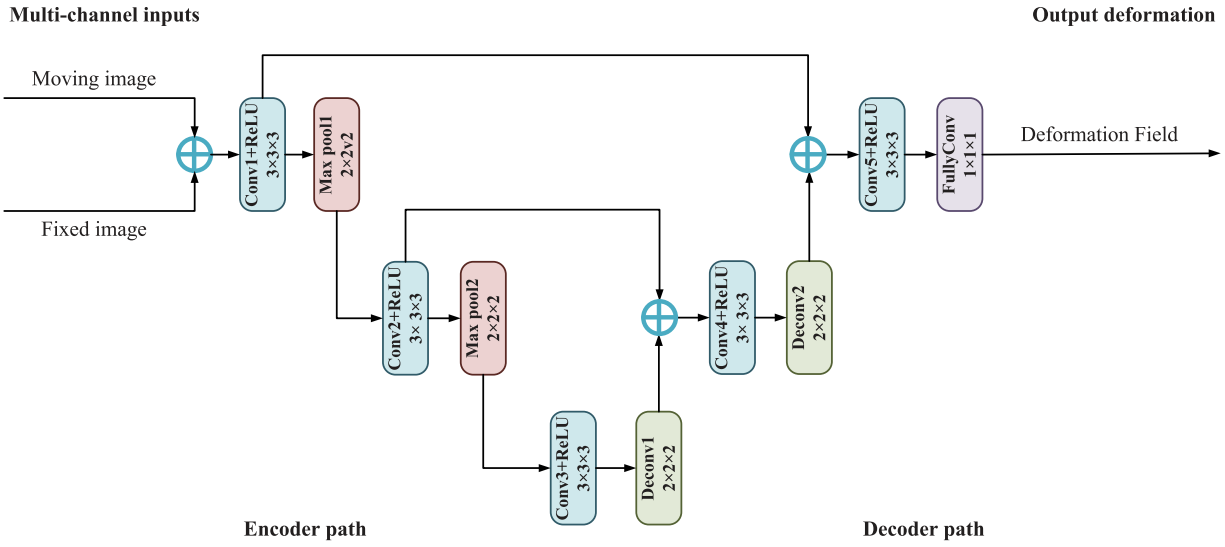

Fig. 3. Architecture of registration network.

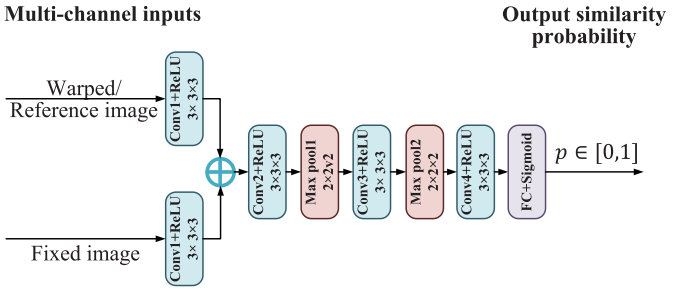

Fig. 4. Architecture of discrimination network.

$2 \times 2 \times 2$ max pooling layer is then used to downsample the feature maps. These layers are included twice to capture information from multiple resolutions.

Symmetric to the encoder path, the decoder path increases the resolution level twice by using $2 \times 2 \times 2$ deconvolutional layers. A convolutional layer with ReLU activation is performed in each resolution level. A concatenation operator is included to fuse the upsampled feature maps with the corresponding feature maps from the encoder path. The final deformation field is obtained by a $1 \times 1 \times 1$ fully convolutional layer without any activation function. The weights of the network can be learned using the constraint from the discrimination network.

3.6.2. Discrimination network

The architecture of discrimination network is shown in Fig. 4. Basically, the input is an image pair and the output is the similarity probability with 1 indicating similarity and 0 indicating dissimilarity. Each convolution layer is zero-padded and is followed by ReLU activations. After applying max pooling for two times, the fully connected (FC) layer with sigmoid activation function is used to gather information from the entire image to give a single value. The images in the different channels, which can come from different modalities, are passed through a convolution layer independently before concatenation.

3.6.3. Spatial transformation layer

A spatial transformation layer (Jaderberg et al., 2015) is used to warp the moving image using the deformation field ϕ. Each voxel x in the warped subject image is calculated by tri-linear interpolation in the corresponding location, as given by the displacement vector, in the subject image:

$$I_W(x) = I_M(x + \phi(x)) \tag{7}$$

$$\approx \Sigma_{y \in \mathcal{N}(x+\phi(x))} I_M(y) \cdot \prod_{d \in \{0,1,2\}} (1 - |x_d + \phi_d(x) - y_d|), \tag{8}$$

where I_W is warped version of I_M based on deformation ϕ, x is a voxel location in I_W, $y \in \mathcal{N}(x + \phi(x))$ is an 8-voxel cubic neighborhood of location $x + \phi(x)$, and d is the dimension index in the 3D image space. There are no trainable parameters in this layer. The gradients of this layer are back propagated from the discrimination network D to train the registration network R.

3.6.4. Implementation details

In this paper, we have implemented both patch-based (Fan et al., 2019) and full-image-based (Balakrishnan et al., 2018) registration networks. The detailed architectures are described in Table 1.

The patch-based implementation takes as input overlapping $68 \times 68 \times 68$ images patches extracted from the moving and fixed images and produce as output a $28 \times 28 \times 28$ patch of displacement vectors. The patch size is reduced because zero-padding is not performed during convolution. We deliberately predict only central deformations of the patches, because the deformable prediction is highly related to the local information of the images, and the boundary regions of the patches may lose their real correspondences. That means the maximum offset of displacement vector in each direction is 20, which is sufficient for measuring the local deformations. When dealing with larger scale deformed images, the range of receptive field needs to be enlarged by adding more convolutional and pooling layers. When training or deploying the patch-based network for the full images, we extract overlapping patches by the step size of 28, *i.e.*, the output patch size. Thus, all the non-overlapping output patches can form the entire deformation field. The more detailed discussion about patch-based registration network can be found in our previous work (Fan et al., 2019).

The full-image-based implementation is more straightforward with the same input and output sizes without zero-padding during convolution. But limited by GPU memory, the number of feature channels of the full-image-based implementation is much smaller than the patch-based implementation.

Table 1
The implementation details of the network architecture. (Findicates the size of full image).

Network	Layer	Patch-based implementation					Full-image-based implementation				
		Input size	Output size	Number of channels	Repeat times	0-padding	Input size	Output size	Number of channels	Repeat times	0-padding
Registration Network	Conv1	68^3	64^3	64	2	No	1 • F	1 • F	16	1	Yes
	Pool1	64^3	32^3	64	1	-	1 • F	1/2 • F	16	1	-
	Conv2	32^3	28^3	128	2	No	1/2 • F	1/2 • F	32	1	Yes
	Pool2	28^3	14^3	128	1	-	1/2 • F	1/4 • F	32	1	-
	Conv3	14^3	10^3	256	2	No	1/4 • F	1/4 • F	32	1	Yes
		10^3	20^3	128	1	-	1/4 • F	1/2 • F	32	1	-
	Deconv1										
	Conv4	20^3	16^3	128	2	No	1/2 • F	1/2 • F	32	1	Yes
		16^3	32^3	64	1	-	1/2 • F	1 • F	32	1	-
	Deconv2										
	Conv5	32^3	28^3	64	2	No	1 • F	1 • F	8	2	Yes
		28^3	28^3	3	1	-	1 • F	1 • F	3	1	-
	FullyConv										
	Conv1	28^3	28^3	8	1	Yes	1 • F	1 • F	8	1	Yes
	Conv2	28^3	28^3	8	1	Yes	1 • F	1 • F	8	1	Yes
	Pool1	28^3	14^3	8	1	-	1/2 • F	1/2 • F	8	1	-
	Conv3	14^3	14^3	16	1	Yes	1/2 • F	1/2 • F	16	1	Yes
	Pool2	7^3	7^3	16	1	-	1/4 • F	1/4 • F	16	1	-
	Conv4	7^3	7^3	32	1	Yes	1/4 • F	1/4 • F	32	1	Yes
	FC	7^3	1	1	1	-	1/4 • F	1	1	1	-

The same network architecture is used for mono-modal and multi-model registration. The network is implemented using 3D Keras (Chollet et al., 2015; Abadi et al., 2016) and trained on a single Nvidia TitanX GPU. We use the Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 1e-4 and 0.5 wt decay after every 50K iterations. Training a patch-based network typically takes 10 epochs and a full-image-based network takes 20 epochs.

4. Experimental results

In this section we evaluate the performance of the proposed method in both mono-modal and multi-modal image registration.

4.1. Competing methods

We compared the proposed method with two state-of-the-art registration methods, i.e., LCC-demons (Lorenzi et al., 2013) and SyN (Avants et al., 2008). Comparison was also performed with respect to deep learning registration methods that use different forms of guidance, including ground-truth deformations and intensity SSD and CC.

(1) **LCC-demons** (Lorenzi et al., 2013): A fast and robust registration framework based on the log-Demons diffeomorphic registration algorithm. The transformation is parameterized by stationary velocity fields. The similarity metric is the symmetric local correlation coefficient (LCC).
(2) **SyN** (Avants et al., 2008): A symmetric image normalization method (SyN) for maximizing the cross-correlation within the space of diffeomorphic maps.
(3) **DL_GT** (Fan et al., 2019): Supervised deep learning registration using ground-truth deformations produced by SyN.
(4) **DL_SSD** (de Vos et al., 2017): Unsupervised deep learning registration using SSD similarity metric.
(5) **DL_CC** (Balakrishnan et al., 2018): Unsupervised deep learning registration using CC similarity metric.
(6) **DL_ASN** (Proposed): Unsupervised deep learning registration with adversarial similarity network.

Unless otherwise specified, all the deep learning methods are based on patch-based implementation. Comparison results of the patch-based and full-image-based implementations are provided in Section 4.2.4

4.2. Mono-Modal registration

4.2.1. Datasets and settings

For evaluation, we utilize four public datasets (Klein et al., 2009): LPBA40, IBSR18, CUMC12, and MGH10. In preprocessing, all the subjects are linear registered to the same space by using FLIRT (Jenkinson and Smith, 2001). After affine registration, all the images are resampled to the same size ($224 \times 224 \times 160$) and resolution ($1 \times 1 \times 1\,\text{mm}^3$). The training images are derived from LPBA40. Among 40 subjects, 30 subjects are selected for training. 30×29 image pairs can be drawn. Specifically, 300 patch pairs were extracted from each training image pair by a 28-voxel sliding step, yielding a total of 26,000 training samples. The remaining 10 images (10×9 image pairs) are used for validation. IBSR18, CUMC12, and MGH10 are used for testing.

4.2.2. Evaluation on LPBA40

For the 10 testing subjects in the LPBA40 dataset, we perform deformable registration on each image pair. The Dice Similarity Coefficient (DSC) of 54 brain ROIs (with names defined in Klein et al. (2009)) is shown in Fig. 5. The proposed algorithm achieves the best performance for 42 out of the 54 ROIs. The performance for the remaining 12 ROIs is comparable with other deep learning registration algorithms. The improvements for 35 ROIs are statistically significantly ($p - value < 0.05$). The average DSC value in Table 2 also shows the best accuracy of the proposed method, which indicates that the proposed adversarial similarity guidance is effective to train an accurate registration network in an unsupervised manner.

4.2.3. Evaluation on IBSR18, CUMC12, MGH10

To evaluate the generalizability of the proposed method, we apply the network trained on only 30 images of LPBA40 to three different datasets (i.e., IBSR18, CUMC12, and MGH10). We register each image pair in the same dataset. Fig. 6 shows a typical set of results from MGH10, demonstrating that the proposed method yields better structural alignment. The results for LCC-demons and SyN are obtained after careful parameter tuning. Fig. 7 shows the corresponding cortical surfaces. We observe that our proposed method achieves the most accurate aligned sulci and gyri, indicated by yellow curves.

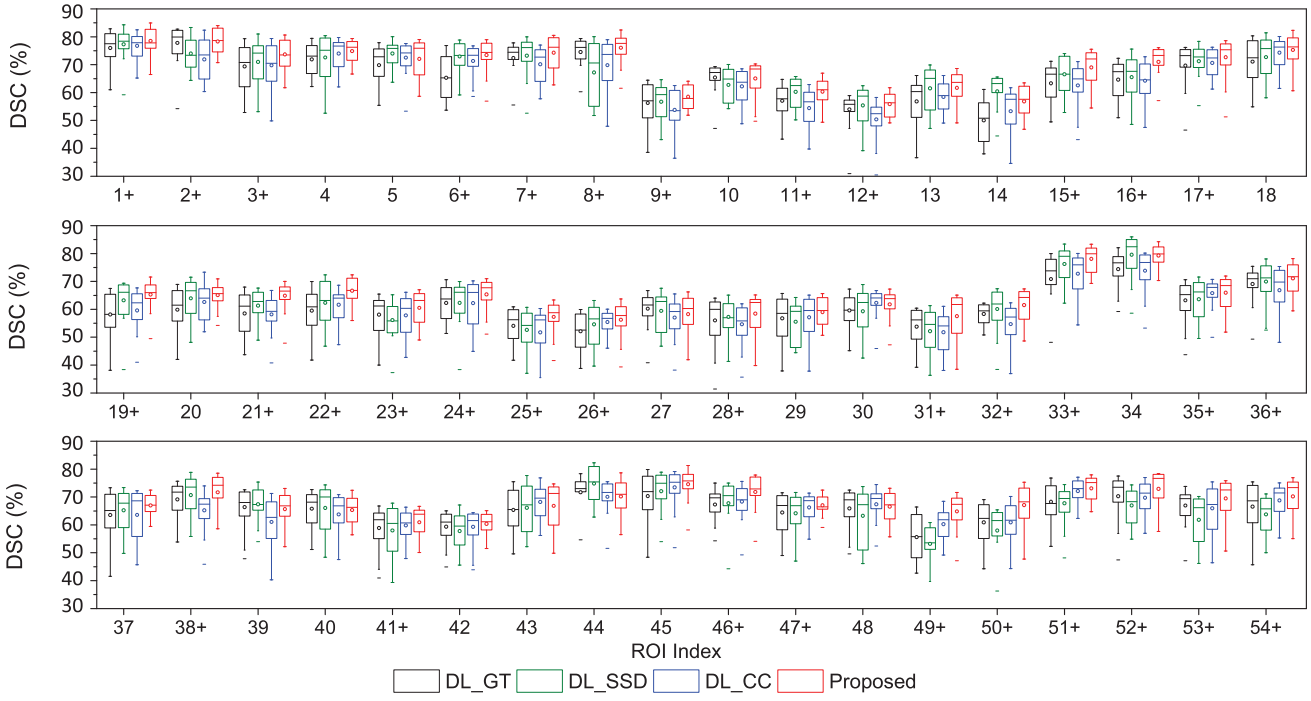

Fig. 5. Boxplots of DSC values (%) of 54 ROIs for the 10 testing subjects from the LPBA40 dataset, for registration using 1) supervised learning, 2) similarity metrics SSD and CC, and 3) the proposed adversarial similarity network. "+" marks statistically significant improvements ($p - value < 0.05$) given by the proposed method over the other three methods.

Table 2
Quantitative results for LPBA40, IBSR18, CUMC12, and MGH10 datasets.

	Dataset	Affine	LCC-demons	SyN	DL_GT	DL_SSD	DL_CC	Proposed
DSC (%)	LPBA40	60.4 ± 2.6	70.7 ± 2.2	71.3 ± 1.8	70.7 ± 2.3	70.4 ± 2.2	71.2 ± 2.8	71.8 ± 2.3
	IBSR18	39.8 ± 3.2	56.8 ± 2.0	57.4 ± 2.4	52.4 ± 3.1	53.1 ± 1.8	54.2 ± 3.4	57.8 ± 2.7
	CUMC12	40.2 ± 3.3	53.9 ± 2.7	54.1 ± 2.8	52.7 ± 3.1	51.6 ± 2.3	51.8 ± 4.1	54.4 ± 2.9
	MGH10	46.3 ± 3.8	61.4 ± 2.3	62.4 ± 2.4	59.7 ± 2.5	58.2 ± 1.6	59.6 ± 2.9	61.7 ± 2.1
ASD (mm)	LPBA40	1.17 ± 0.14	0.53 ± 0.04	0.49 ± 0.03	0.59 ± 0.06	0.62 ± 0.05	0.54 ± 0.04	0.47 ± 0.03
	IBSR18	1.43 ± 0.20	0.71 ± 0.04	0.70 ± 0.04	0.78 ± 0.07	0.82 ± 0.08	0.75 ± 0.05	0.68 ± 0.05
	CUMC12	1.53 ± 0.23	0.77 ± 0.05	0.72 ± 0.04	0.82 ± 0.06	0.85 ± 0.06	0.79 ± 0.04	0.70 ± 0.04
	MGH10	1.24 ± 0.19	0.60 ± 0.04	0.57 ± 0.04	0.70 ± 0.06	0.71 ± 0.06	0.65 ± 0.04	0.58 ± 0.04

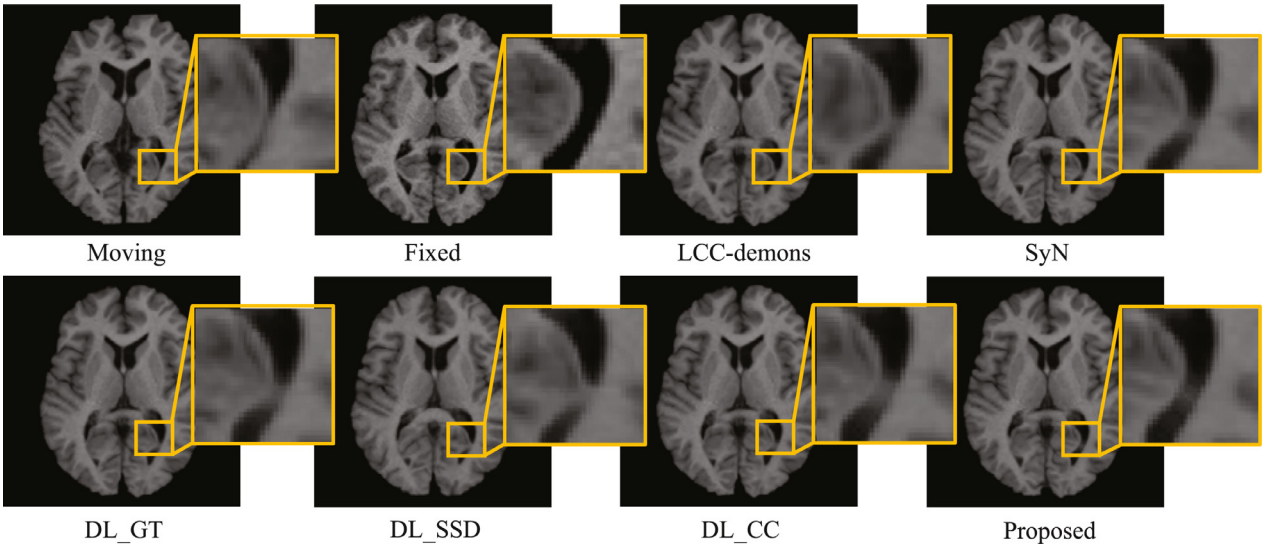

Fig. 6. Typical registration results from MGH10.

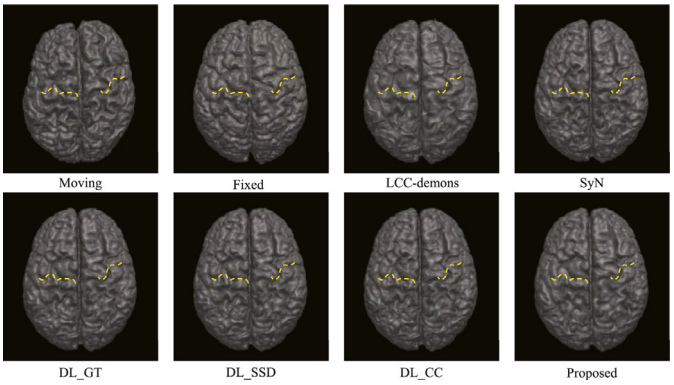

Fig. 7. Visualized registration results. Yellow curves draw the alignment of sulcus and gyrus.

Table 3
Comparison of patch-based and full-image-based implementations in terms of training time per epoch for 30×29 image pairs in LPBA40 and testing time per image pair with size $224 \times 224 \times 160$.

Methods	Training time (h)	Testing time (s)	DSC (%)
Patch-based	45	18.8	71.8 ± 2.3
Full-image-based	1.8	1.3	71.2 ± 2.7

Table 2 provides the Dice Similarity Coefficient (DSC) and Average Surface Distance (ASD) for all competing methods. The average DSC is calculated based on all the ROIs for each individual dataset, and ASD is calculated based on the surfaces representing the boundaries between white matter and gray matter. The LPBA40 results are obtained from the validation set, *i.e.*, the 10 images not used for training. The proposed method achieves the best overall performance for most of the challenging registration tasks. The deep learning methods are trained only using the LPBA40 dataset. The other three datasets are unseen datasets to training. The results indicate that, all the deep learning methods work well on LPBA40, but less so for unseen datasets except the proposed.

4.2.4. *Patch*-vs. *full-Image-Based implementations*

Table 3 presents the computational efficiency (using NVIDIA TitanX GPU) and DSC accuracy on LPBA40 validation set. The patch-based implementation uses 300 patch pairs for training and the full-image-based implementation uses the image pairs as samples. The training time per epoch of patch-based implementation is much longer than that of full-image-based implementation, although the training time of a single patch is faster than a full image. The testing time of patch-based implementation is 10 times more than that of full-image-based implementation. Both implementations are much faster than conventional registration methods, such as SyN (47 minutes) and LCC-demons (23 minutes) per image pair on CPU. The GPU implementation of SyN (Luo et al., 2015) takes about 10 minutes.

4.3. *Multi-Modal registration*

The evaluation of multi-modal image registration is based on pelvic MR and CT images ($224 \times 192 \times 96$, $1 \times 1 \times 1\,\text{mm}^3$) of 22 prostate cancer patients. The prostate, bladder and rectum in both MR and CT images are manually labeled by physicians. All the CT (fixed) and MR (moving) images are linearly registered to a common space using FLIRT (Jenkinson and Smith, 2001) with mutual information (MI).

In the training stage, we randomly choose 10 image pairs of the same patients to prepare *paired data*, which is used to train the discrimination network as the well registered images for positive case, by further registering the labels of prostate, bladder and rectum (Cao et al., 2018a). Then, we randomly select 15×15 image pairs from 15 CT images and 15 MR images to form the training set, and the remaining 7 image pairs are used as the testing set. The other training settings of the networks follow that of mono-modal image registration.

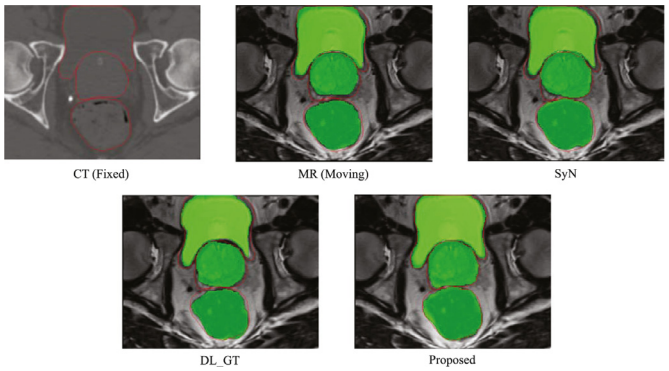

Fig. 8. Registration results of multi-modal images. The red curves outline the ROIs on the CT images, and the green regions are the corresponding ROIs on the MR images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Quantitative results for pelvic dataset.

	Organ	Affine	SyN	DL_GT	Proposed
DSC (%)	Bladder	84.7 ± 5.4	86.2 ± 4.8	86.1 ± 5.8	89.1 ± 4.3
	Prostate	80.6 ± 5.2	83.9 ± 3.7	83.4 ± 4.0	86.8 ± 3.8
	Rectum	77.4 ± 4.5	81.6 ± 4.4	80.9 ± 4.5	84.7 ± 4.2
ASD (mm)	Bladder	1.87 ± 0.63	1.59 ± 0.48	1.62 ± 0.55	1.33 ± 0.38
	Prostate	2.06 ± 0.67	1.74 ± 0.54	1.78 ± 0.60	1.57 ± 0.44
	Rectum	2.34 ± 0.79	1.94 ± 0.62	1.96 ± 0.59	1.57 ± 0.41

Since SSD and CC will not work for multi-modal registration, only MI based SyN and supervised deep learning trained using ground-truth deformations produced by SyN are used in comparison. Example qualitative (Fig. 8) and quantitative (Table 4) results indicate our method can effectively register multi-modal images.

5. Discussion

The proposed unsupervised adversarial learning strategy avoids the need for ground-truth deformations and predefined similarity metric. Experimental results have demonstrated its accuracy, generalizability and speed. However, there are some limitations that need to be addressed.

In GAN training (Goodfellow et al., 2014), the discriminator typically converges much faster than the generator, causing the generator to be under-trained. Adjusting the training proportion of the generator and the discriminator can help balance the convergence speed of these two networks. But this is still an open problem and should be further investigated and addressed with methods such as Pixel-GAN (Isola et al., 2017), Cycle-GAN (Zhu et al., 2017) and WGAN-GP (Gulrajani et al., 2017). Another way is to choose a more reasonable definition of reference data to train the discriminator. In this paper, in order to avoid pre-defined assumptions and prior knowledge, the simplest reference selection method is used to make the model learn autonomously. More attempts on reference image settings will help improve the performance of the adversarial training strategy in the registration problem.

We have evaluated the deformation smoothness for different values of λ in Eq. (5) using the range [0,2500]. As can be seen from Fig. 9, the DSC value increases with λ before 500 but decreases after that point. Table 5 presents the average number of voxels with

Table 5
Average (std) number of voxels with negative Jacobian determinant values for different deep learning registration methods.

Dataset	DL_GT	DL_SSD	DL_CC	Proposed
Brain	28745.6 (5479.1)	31475.1 (6104.5)	26541.9 (4791.2)	15713.8 (2013.5)
Pelvic	14712.8 (2514.6)	-	-	10145.2 (1204.3)

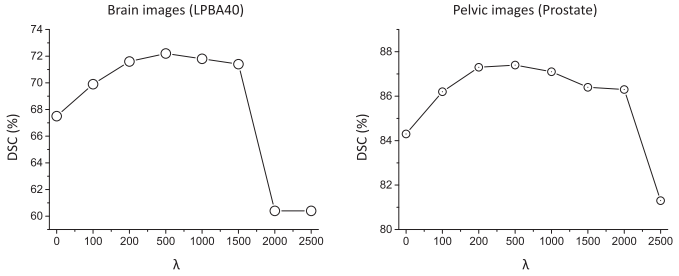

Fig. 9. Effect of varying λ for balancing $\mathcal{L}_R$ and $\mathcal{L}_{reg}$.

negative Jacobian determinant values for all the deep-learning-based registration methods. Our method yields the least number of voxels with negative Jacobian determinant values among these deep-learning-based registration methods. Since we do not use diffeomorphic constraint for regularization, the number of voxels with negative Jacobian determinant values is still larger than SyN and LCC-demons (almost all positive). For well-behaving deformation fields, diffeomorphic deformation models (Yang et al., 2017; Dalca et al., 2018) can be used. Another alternative is to use adversarial training, with ideal deformations simulated using biomechanical models (Hu et al., 2018a), to enforce regularization by discriminating predicted deformations from ideal ones.

6. Conclusion

In this paper, we have introduced an unsupervised adversarial learning strategy for mono- and multi-modal image registration. Our network does not need ground-truth deformations or predefined similarity metrics. Instead, the similarity metric is learned automatically based on a discrimination network. The experimental results indicate that the proposed method yields registration accuracy comparable to state-of-the-art methods but with significantly better generalizability.

Declaration of Competing Interest

The authors declare that they do not have any financial or non-financial conflict of interests.

Acknowledgments

This work was supported in part by NIH grants (EB008374, AG053867).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., 2016. Tensorflow: a system for large-scale machine learning.. In: OSDI, 16, pp. 265–283.

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

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

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.

Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., Shen, D., 2018a. Deep learning based inter-modality image registration supervised by intra-modality similarity. 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: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 300–308.

Cao, X., Yang, J., Zhang, J., Wang, Q., Yap, P.-T., Shen, D., 2018b. Deformable image registration using cue-aware deep regression network. IEEE Trans. Biomed. Eng. 65 (9), 1900–1911.

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

Chollet, F., *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*, MITP-Verlags GmbH & Co. KG, 2018.

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.

Fan, J., Cao, X., Yap, P.-T., Shen, D., 2019. Birnet: brain image registration using dual-supervised fully convolutional networks. Med. Image Anal. 54, 193–206.

Fan, J., Yang, J., Ai, D., Xia, L., Zhao, Y., Gao, X., Wang, Y., 2016. Convex hull indexed gaussian mixture model (ch-gmm) for 3d point set registration. Pattern Recognit. 59, 126–141.

Fan, J., Yang, J., Lu, F., Ai, D., Zhao, Y., Wang, Y., 2016. 3-Points convex hull matching (3pchm) for fast and robust point set registration. Neurocomputing 194, 227–240.

Fan, J., Yang, J., Zhao, Y., Ai, D., Liu, Y., Wang, G., Wang, Y., 2017. Convex hull aided registration method (charm). IEEE Trans. Visual. Comput.Graph. 23 (9), 2042–2055.

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.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C., 2017. Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems, pp. 5767–5777.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034.

Hellier, P., Ashburner, J., Corouge, I., Barillot, C., Friston, K.J., 2002. Inter-subject registration of functional and anatomical data using spm. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 590–597.

Holden, M., 2008. A review of geometric transformations for nonrigid body registration. IEEE Trans. Med. Imag. 27, 111–128.

Hu, Y., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Vercauteren, T., Noble, J.A., Barratt, D.C., 2018. Adversarial deformation regularization for training image registration neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 739–746.

Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Noble, J.A., Barratt, D.C., Vercauteren, T., 2018. Label-driven weakly-supervised learning for multimodal deformarle image registration. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE, pp. 1070–1074.

Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al., 2018. Weakly-supervised convolutional neural networks for multimodal image registration. Med. image Anal. 49, 1–13.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. arXiv preprint.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. In: Advances in Neural Information Processing Systems, pp. 2017–2025.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5 (2), 143–156.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46 (3), 786–802.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2010. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imag. 29 (1), 196–205.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436.
Li, H., Fan, Y., 2018. Non-rigid image registration using self-supervised fully convolutional networks without training data. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE, pp. 1075–1078.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X., (ADNI, A.D.N.I., et al., 2013. LCC-demons: a robust and accurate symmetric diffeomorphic registration algorithm. NeuroImage 81, 470–483.
Luo, Y.-g., Liu, P., Shi, L., Luo, Y., Yi, L., Li, A., Qin, J., Heng, P.-A., Wang, D., 2015. Accelerating neuroimage registration through parallel computation of similarity metric. PloS One 10 (9), e0136718.
Maintz, J.A., Viergever, M.A., 1998. A survey of medical image registration. Med. Image Anal. 2 (1), 1–36.
Miao, S., Wang, Z.J., Liao, R., 2016. A CNN regression approach for real-time 2D/3D registration. IEEE Trans. Med. Imag. 35 (5), 1352–1363.
Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X., 2017. SVF-Net: Learning deformable image registration using shape matching. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention, 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.
Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid deformations using free form deformations: an application to breast MR images. IEEE Trans. Med. Imag. 18 (8), 712–721.
Salehi, S., Khan, S., Erdogmus, D., Gholipour, A., 2018. Real-time deep registration with geodesic loss. arXiv preprint arXiv:1803.05982.
Shen, D., Davatzikos, C., 2002. Hammer: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imag. 21 (11), 1421–1439.
Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B.P.F., Išgum, I., Staring, M., 2017. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 232–239.
Sotiras, A., Davatzikos, C., Paragios, N., 2013. Deformable medical image registration: a survey. IEEE Trans. Med. Imag. 32 (7), 1153–1190.
Studholme, C., Hill, D.L., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit. 32 (1), 71–86.
Tang, Z., Ahmad, S., Yap, P.-T., Shen, D., 2018. Multi-atlas segmentation of MR tumor brain images using low-rank based image recovery. IEEE Trans. Med. Imag. 37 (10), 2224–2235.
Tang, Z., Yap, P.-T., Shen, D., 2019. A new multi-atlas registration framework for multimodal pathological images using conventional monomodal normal atlases. IEEE Trans. Image Process. 28 (5), 2293–2304.
Thirion, J., 1998. Image matching as a diffusion process: an analogy with Maxwell's demons. Med. Image Anal. 2 (3), 243–260.
Uzunova, H., Wilms, M., Handels, H., Ehrhardt, J., 2017. Training CNNs for image registration from few samples with model-based data augmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 223–231.
Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2008. Symmetric log-domain diffeomorphic registration: ademons-based approach. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 754–761.
Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: efficient non-parametric image registration. NeuroImage 45 (1, Supplement 1), S61–S72.
Viergever, M.A., Maintz, J.A., Klein, S., Murphy, K., Staring, M., Pluim, J.P., 2016. A survey of medical image registration–under review. Med. Image Anal. 33, 140–144.
Viola, P., Wells III, W.M., 1997. Alignment by maximization of mutual information. Int. J. Comput. Vis. 24 (2), 137–154.
Vishnevskiy, V., Gass, T., Szekely, G., Tanner, C., Goksel, O., 2016. Isotropic total variation regularization of displacements in parametric image registration. IEEE Trans. Med. Imag. 36 (2), 385–395.
de Vos, B., Berendsen, F., Viergever, M., Staring, M., Išgum, I., 2017. End-to-End unsupervised deformable image registration with a convolutional neural network. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, pp. 204–212.
Wang, S., Kim, M., Wu, G., Shen, D., 2016. Scalable high performance image registration framework by unsupervised deep feature representations learning. IEEE Trans. Biomed. Eng. 63 (7), 1505–1516.
Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C., 1998. Automated image registration: I. general methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. 22 (1), 139–152.
Wu, G., Kim, M., Wang, Q., Shen, D., 2012. Hierarchical attribute-guided symmetric diffeomorphic registration for mr brain images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 90–97.
Xue, Z., Shen, D., Christos, D., 2006a. Statistical representation of high-dimensional deformation fields with application to statistically constrained 3d warping. Med. Image Anal. 10 (5), 740–751.
Xue, Z., Shen, D., Davatzikos , C., 2006b. CLASSIC: consistent longitudinal alignment and segmentation for serial image computing. NeuroImage 30 (2), 388–399.
Yan, P., Xu, S., Rastinehad, A.R., Wood, B.J., 2018. Adversarial image registration with application for MR and TRUS image fusion. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 197–204.
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.
Zhou, T., Liu, F., Bhaskar, H., Yang, J., 2018. Robust visual tracking via online discriminative and low-rank dictionary learning. IEEE Trans. Cybernet. 48 (9), 2643–2655.
Zhou, T., Thung, K.-H., Zhu, X., Shen, D., 2019. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Human Brain Mapp. 40 (3), 1001–1016.
Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint.