Adversarial Similarity Network for Evaluating Image Alignment in Deep Learning Based Registration

Jingfan Fan[1], Xiaohuan Cao[1,2], Zhong Xue[3], Pew-Thian Yap[1], and Dinggang Shen[1(✉)]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
dgshen@med.unc.edu
[2] School of Automation, Northwestern Polytechnical University, Xi'an, China
[3] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Abstract. This paper introduces an unsupervised adversarial similarity network for image registration. Unlike existing deep learning registration frameworks, our approach does not require 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 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. Experiments on four brain MRI datasets indicate that our method yields registration performance that is promising in both accuracy and efficiency 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 based on intensity-based similarity metrics. However, these methods often involve computationally expensive high-dimensional optimization and task-dependent parameter tuning. Deep learning methods, such as convolutional neural networks (CNN), have been shown recently to be capable of addressing the limitations of conventional registration methods.

In *supervised learning* methods, the registration network is trained with ground-truth deformations. Sokooti et al. [1] proposed RegNet to estimate the displacement vector field for a pair of chest CT images. Yang et al. [2] predicted the momenta in LDDMM. Rohe et al. [3] built reference deformations for training by registering manually delineated regions of interests (ROIs). While effective, these methods are however limited by the availability of ground-truth deformations.

In *unsupervised learning* methods [4, 5], the deformable transformations are learned without ground-truth deformations by maximizing the similarity between a pair

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11070, pp. 739–746, 2018.
https://doi.org/10.1007/978-3-030-00928-1_83

of images, such as the sum of squared difference (SSD) and cross-correlation (CC). However, these similarity metrics are closely related to the nature of the images and might not be suitable when dealing with diverse datasets.

In this paper, we propose an *adversarial similarity network* to automatically learn the similarity metric for training a deformable registration network. The network is unsupervised and is inspired by generative adversarial network (GAN) [6]. More specifically, the generator is a *registration network* that predicts the deformations. The discriminator is a *discrimination network* that judges whether images are well aligned and feeds misalignment information to the registration network during training. The registration and discrimination networks are learned via *adversarial training,* learning a metric for accurate registration. 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, without the need for parameter tuning.
- Compared with supervised learning registration methods, the proposed network does not need ground-truth deformations. 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.

2 Method

Image registration aims to determine a deformation field ϕ that warps a subject image $S \in \mathbb{R}^3$ to a template image $T \in \mathbb{R}^3$, so that the warped image $S \circ \phi$ is similar to T. Deformation ϕ is typically determined by minimizing energy functional

$$\phi = \underset{\phi}{argmin}\, M(S \circ \phi, T) + Reg(\phi), \tag{1}$$

where $M(S \circ \phi, T)$ quantifies the dissimilarity between the template image T and the warped subject image $S \circ \phi$. $Reg(\phi)$ is the regularization to preserve the smoothness of the deformation field ϕ.

In this paper, we design a *registration network R*, to learn the deformation field ϕ for subject and template images (S, T). The mapping can be written as $R : (S, T) \Rightarrow \phi$. **First**, the *registration network R* is trained under the guidance of image similarity, therefore no ground-truth deformation field is needed. Instead of specifying a similarity metric, the similarity guidance is derived from the *discrimination network D*, which can automatically 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 as possible to convince the *discrimination network D.* **Second**, in order to preserve the smoothness of the predicted deformation field ϕ, a regularization is incorporated in the training of the *registration network R.*

As shown in Fig. 1, a deformable transformation layer connects the output of the *registration network R* (i.e., the deformation field ϕ) and input of the *discrimination network D* (i.e., a pair of registered images). The input of R is $64 \times 64 \times 64$ image patches and the output is the corresponding deformation field of size $24 \times 24 \times 24$. Here, the output size is smaller than the input, in order to adapt to the displacement range in the deformable deformation. In testing stage, the deformation field is predicted by the trained R.

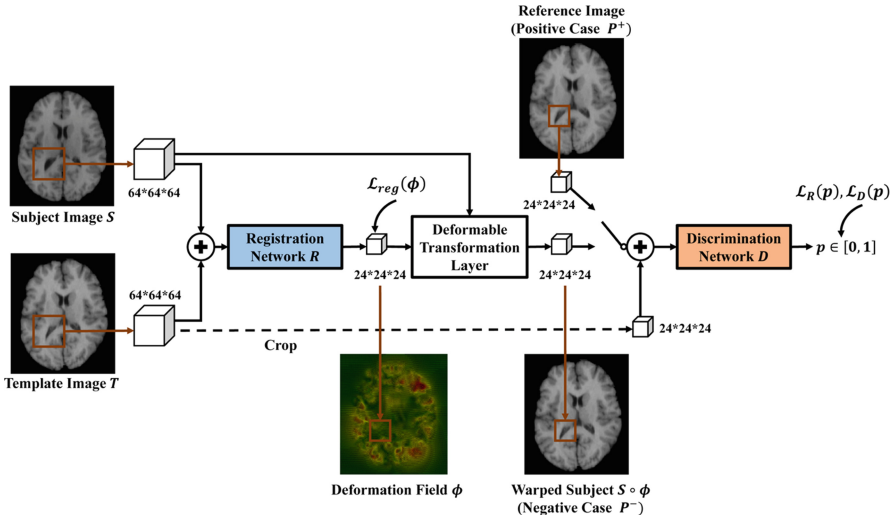

Fig. 1. The proposed adversarial similarity network for deformable image registration. The input image pair is already linearly aligned.

2.1 Adversarial Training

The adversarial training strategy, described below, is used to train *registration network R* and the *discrimination network D* is inspired by GAN [6].

(1) **Training the *discrimination network D*.**

The discrimination network D aims to determine whether the input image pair is similar (i.e., well registered). Two cases are fed into the network alternatively: (1) the *positive* case (P^+) where the images are well registered, and (2) the *negative* case (P^-) where the images are not well registered. The loss function of D can be defined as

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

where, p is the output of the *discrimination network D* that indicates the similarity probability. During training, the *positive* case is derived from the predefined aligned images and the output of D is expected to be 1, indicating the image pair is similar. The

negative case is derived from the registration network R, which means the image pairs are under registration and currently not well registered. Thus the output of D is expected to yield 0, indicating the image pair is dissimilar. The discrimination network can be optimized by minimizing the loss function in Eq. (2).

The ideal *positive* case is when the two images are exactly same. However, this cannot happen in real-world registration tasks. We therefore add some disturbance in the positive image pair. Specifically, for each image pair, the template image T is fixed. The perturbed subject image is created from the original subject image S as $\alpha \cdot S + (1-\alpha) \cdot T$ with $0 < \alpha < 1$. We set $\alpha = 0.2$ in the initial training stage to weaken the similarity requirement and $\alpha = 0.1$ in later stage for greater accuracy.

(2) **Training the registration network *R*.**

The registration network R is supervised by the image similarity based on the discrimination network D. As mentioned, the image pair that registered by the registration network R is regarded as the negative case (P^-.) for the discrimination network D. However, the registration network aims to make the registered images as similar as possible, i.e., the output similarity probability p of discrimination network D should approximate to 1. Therefore, the loss function of registration network R can be defined as

$$\mathcal{L}_R(p) = \log(1-p), p \in P^-. \tag{3}$$

In addition to the similarity guidance, the smoothness of the predicted deformation field ϕ is also enforced with loss

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

where v represents the voxel location. By jointly considering Eqs. (3) and (4), the total loss function for the registration network R is:

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

where λ is the weight of the smoothness term, which we set it to 1.

The overall network is trained by alternating between optimizing the registration network R and the discrimination network D. Convergence occurs when the discrimination network cannot distinguish the *positive* cases and the *negative* cases.

2.2 Network Details

Registration Network *R*. The registration network follows the same architecture in [7], which is a hierarchical U-Net regression model [8]. The network takes 3D patches from the subject and template images as input and produces the deformation fields associated with the patches as output.

Discrimination Network *D*. The network architecture of D is shown in Fig. 2. Basically, the input is the image pair and the output is the similarity probability $p \in [0, 1]$, with 1 indicating similarity and 0 indicating dissimilarity. Each convolution layer

is followed by ReLU activations, and 0-padding is applied in each convolution layer. The fully connected (FC) layer is used to gather information from the entire image into one value.

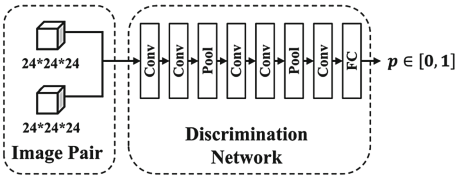

Fig. 2. The discrimination network.

Deformable Transformation Layer. A deformable transformation layer is used to warp the subject image using the deformation field ϕ. Each voxel in the warped subject image is calculated by interpolating in the corresponding location, as given by the displacement vector, in the subject image. The gradient is back propagated from the discrimination network D to train the registration network R.

Implementation Details. The network is implemented using 3D Caffe using Adam optimization. The learning rate is initially set to 1e–3, with 0.5 weight decay after every 50 K iterations. During testing, the registration network is used without the discrimination network to predict the deformation field.

3 Experiments and Results

In this section, we compare the proposed method with different training strategies and several state-of-the-art deformable registration algorithms. Four public datasets [9], including LPBA40, IBSR18, CUMC12, and MGH10, are used to validate the proposed method. After affine registration, all the images are resampled to the same size ($220 \times 220 \times 184$) and resolution ($1 \times 1 \times 1$ mm^3). Two state-of-the-art registration methods, i.e., diffeomorphic demons (D. Demons) [10] and SyN [11], are used as the comparison methods. We also compare our method with other deep learning registration strategies, including (1) supervised training (i.e., ground-truth deformations obtained by SyN), (2) unsupervised training with similarity metrics SSD [4] and CC [5].

The training images are derived from LPBA40. Among the 40 subjects, 30 images are selected as the training data, in which 30×29 image pairs can be drawn. The remaining 10 images are used as the testing data. Specifically, 300 patch pairs are extracted from each training image pair, giving a total of 26,000 training samples.

3.1 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 (names

defined in [9]) is shown in Fig. 3. The proposed algorithm achieves the best performance for 42 out of the 54 ROIs, while the performance of the remaining 12 ROIs are comparable, compared with other deep learning registration algorithms. The average DSC value in Table 1 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.

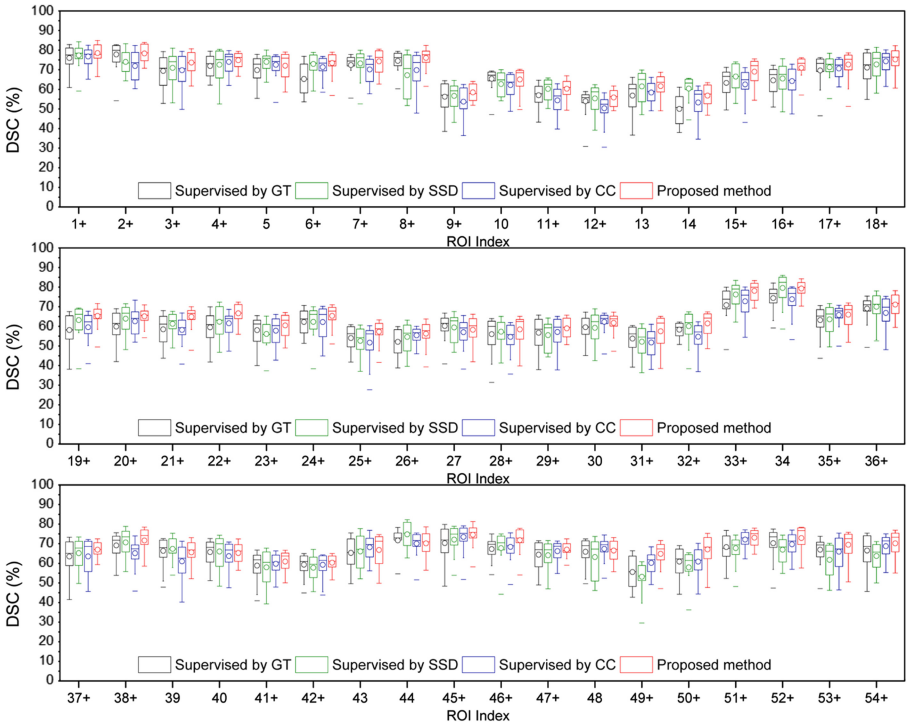

Fig. 3. Boxplot of DSC (%) in 54 ROIs for the 10 testing subjects from LPBA40 dataset, after performing registration under different training strategies: (1) supervised learning, (2) similarity metrics SSD and CC, and (3) the proposed adversarial similarity network. " + " marks improvements given by the proposed method over the three other methods.

3.2 Evaluation on IBSR18, CUMC12, MGH10

To further evaluate generalizability of the proposed method, we apply the network trained on LPBA40 dataset on a total of 40 brain images from three different datasets (i.e., IBSR18, CUMC12, and MGH10). We register each image pair in the same dataset. Figure 4 shows a typical set of results from MGH10. The results for Diffeomorphic Demons and SyN are obtained via careful parameter tuning.

Table 1 provides the DSC for all methods. The average DSC is calculated based on all the ROIs for each individual dataset. The results indicate that, when applied directly

Table 1. Mean DSC (%) on LPBA40, IBSR18, CUMC12, and MGH10 datasets.

Dataset	D.Demons	SyN	Supervised by GT	Supervised by SSD	Supervised by CC	Proposed method
LPBA40	68.7 ± 2.4	71.3 ± 1.8	70.7 ± 2.3	70.4 ± 2.2	71.2 ± 2.8	**71.8 ± 2.3**
IBSR18	54.6 ± 2.2	57.4 ± 2.4	52.4 ± 3.1	53.1 ± 1.8	54.2 ± 3.4	**57.8 ± 2.7**
CUMC12	53.1 ± 3.4	54.1 ± 2.8	52.7 ± 3.1	51.6 ± 2.3	51.8 ± 4.1	**54.4 ± 2.9**
MGH10	60.4 ± 2.5	**62.4 ± 2.4**	59.7 ± 2.5	58.2 ± 1.6	59.6 ± 2.9	61.7 ± 2.1

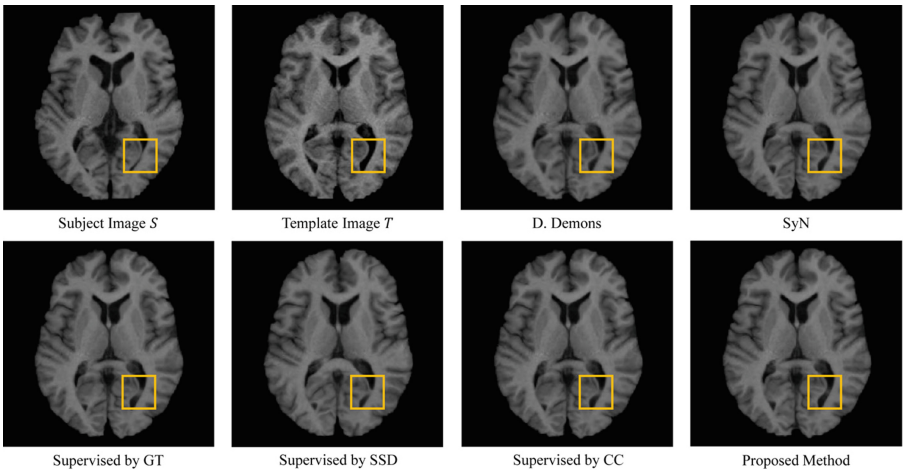

Fig. 4. Typical registration results from MGH10. The boxes mark significant improvements.

to unseen datasets, other learning strategies do not work well. Our method gives the best overall performance. Compared with the fine-tuned Diffeomorphic Demons and SyN, the proposed method exhibit better performance.

The proposed algorithm is implemented based on a single Nvidia TitanX (Pascal) GPU. The average computation time for registering a pair of 3D brain images ($220 \times 220 \times 184$) is 18.3 s, which is considered efficient for deformable registration.

4 Conclusions

In this paper, an adversarial training strategy is designed for unsupervised registration. Our network does not need ground-truth deformations or predefined similarity metrics. Instead, the similarity metric is learned automatically based on the discrimination network. The experimental results indicate that the proposed method exhibits higher registration accuracy compared with state-of-the-art registration methods.

Acknowledgment. This work was supported in part by NIH grants (EB006733, EB008374, MH100217, AG041721, AG053867).

References

1. Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B.P.F., Išgum, I., Staring, M.: 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.) MICCAI 2017. LNCS, vol. 10433, pp. 232–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_27
2. Yang, X., et al.: Quicksilver fast predictive image registration–a deep learning approach. NeuroImage **158**, 378–396 (2017)
3. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: 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.) MICCAI 2017. LNCS, vol. 10433, pp. 266–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_31
4. Li, H., Fan, Y.: Non-Rigid Image Registration Using Self-Supervised Fully Convolutional Networks without Training Data. arXiv preprint arXiv:1801.04012 (2018)
5. Balakrishnan, G., et al.: An Unsupervised Learning Model for Deformable Medical Image Registration. arXiv preprint arXiv:1802.02604 (2018)
6. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (2014)
7. Fan, J., et al.: BIRNet: Brain Image Registration Using Dual-Supervised Fully Convolutional Networks. arXiv preprint arXiv:1802.04692 (2018)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Klein, A., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage **46**(3), 786–802 (2009)
10. Vercauteren, T., et al.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage **45**(1), S61–S72 (2009)
11. Avants, B.B., et al.: Symmetric diffeomorphic image registration with cross-correlation. Med. Image Anal. **12**(1), 26–41 (2008)