SAMPLE ALIGNMENT FOR IMAGE-TO-IMAGE TRANSLATION BASED MEDICAL DOMAIN ADAPTATION

Heng Li[1,*], *Haofeng Liu*[1], *Xiaoxuan Wang*[1], *Chenlang Yi*[1], *Hao Chen*[3], *Yan Hu*[1,*], *Jiang Liu*[1,2]

[1] School of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[2] Cixi Institute of Biomedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo 315201, China

[3] Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, China

ABSTRACT

Image-to-image (I2I) translation is a popular paradigm in domain adaptation (DA), and has been frequently used to address the lack of labeled data. However, as a result of the sample bias in medical data caused by the attributes of imaging modality or pathology, the I2I translation based DA always suffers from synthesis artifacts. For boosting the DA in medical scenarios, a sample alignment algorithm is proposed to correct the sample bias in medical data. Specifically, diffeomorphic transformation and symmetric resampling are employed to implement the sample alignment. The topological structure in medical samples is first aligned using diffeomorphic transformation. Then paired image data are collected from the aligned samples by symmetric resampling to train the I2I translation models. In the experiment, the proposed algorithm was applied to boost the DA of cross-modality data and pathological ones. Our algorithm not only improved the quality of synthesized images, but also promoted the DA of diagnosis models learned from synthesized data.

Index Terms— Domain adaptation, image-to-image translation, sample alignment, sample bias

1. INTRODUCTION

Enough labeled data (target data) for training is a fundamental requirement in many computer vision applications. However, collecting labeled data is usually troublesome in medical scenarios. Whenever target data is either not available, or it is expensive to collect and/or label it, the typical approach is to use available datasets (source data), representative of a closely related task [1]. Since this practice is leading to suboptimal performance, DA has been developed to address the issue.

I2I translation, which learns the mapping from a given image to a specific target image, can be employed to align the source and target domain. Therefore, I2I translation has been developed into a paradigm technique for implementing DA [2, 3, 4]. Especially the invention of CycleGAN, which uses unpaired images rather than paired ones to learn the mapping between domains, has remarkably promoted the application of DA in medical scenarios [5, 6]. The unpaired I2I translation has been successfully used to adapt segmentation and diagnosis models to various image conditions, such as noisy fundus images [7], high-resolution MRI [8].

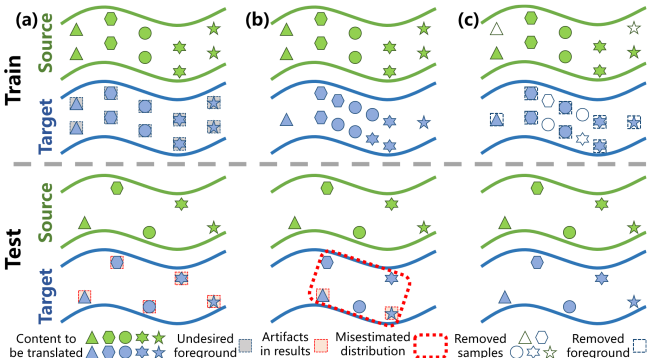

Fig. 1. Sample bias in the training of I2I translation models leads to undesired artifacts in DA. (a) the different foreground distributions in paired samples are misunderstood as domain-specific features; (b) the non-uniform sampling of the target domain misleads the estimating of the marginal probability distribution; (c) the proposed algorithm corrects the sample bias to properly learn the translation mapping. Details refer to Sec. 2.1.

However, I2I translation is still an ideal scenario, where existing translation methods, no matter paired or unpaired ones, require training samples delicately reflecting the domain-independent and domain-specific features. Unfortunately, some medical scenarios may not meet this assumption. Caused by the attributes of certain modalities and pathology, I2I translation encounters challenges. As shown in Fig. 1, the sample bias between the source and target domain con-

∗ Correspondence: lih3, huy3@sustech.edu.cn

fuses the domain-independent and domain-specific features and mislead the training of translation mapping. In cross-modality translation, weakly paired images with unidentical foreground distributions may be paired collected from different modalities (Fig. 1 (a)). For synthesizing pathologic images, non-uniformly sampled images may be collected by the sampling strategy distorted by specific lesions (Fig. 1 (b)). Thus a bias appears between the sample distributions of the source and target domain, leading to artifacts in the domain adapted data. Although strategies, such as updating the network only with samples having a small loss [9], selecting robust loss functions [10], have been proposed to reduce the effect of noise samples in DA. The sampling bias remains a challenge in the I2I translation based DA for medical scenarios.

To boost the I2I translation based DA, this paper proposes a sample alignment (SA) algorithm to fundamentally correct the sampling bias (as shown in Fig. 1 (c)). For implementing SA, diffeomorphic transformation is first introduced to align the structures between the original samples. Subsequently, symmetric resampling is performed on the aligned samples to collect paired data for I2I translation. Thus the DA is boosted by improving the training of translation models with the paired data. The main contributions of this paper are summarized as follows:

- For boosting the I2I translation based DA in medical scenarios, a sample alignment algorithm is proposed to correct the sample bias in medical data.
- The I2I translation mapping for DA is superiorly trained with the proposed algorithm, which not only corrects the sample bias, but also constructs paired training data.
- Cross-modality translation and pathological image synthesis are performed to validate the algorithm. The I2I translation and automatic diagnosis are both boosted in the experiments.

2. METHODOLOGY

2.1. Problem Formulation

A domain $\mathbb{D}$ consists of two components: a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$, where $X = \{x_1, ..., x_n\} \in \mathcal{X}$. X is the sample set of the ith term feature vector x_i. Define the source and target domain as $\mathbb{D}^S = \{\mathcal{X}^S, P^S(X^S)\}$ and $\mathbb{D}^T = \{\mathcal{X}^T, P^T(X^T)\}$. The paradigm of I2I translation achieves DA by developing a translation mapping $\mathcal{M}^{S\to T}$ to align the source and target domains, i.e. $\mathcal{M}^{S\to T}(\mathbb{D}^S) \sim \mathbb{D}^T$.

In practice, the feature space $\mathcal{X}$ and marginal probability distribution $P(X)$ are estimated using the acquired sample set X. However, the various imaging modalities and pathological conditions may distort the estimating of $\mathcal{X}$ and $P(X)$. As the examples presented in Fig. 2 (a), for synthesizing En face OCT from color fundus photography, due to the different imaging fields, these two modalities have various foreground distributions in the image samples. This distribution variation will distort the mapping $\mathcal{M}^{S\to T}$ (as the challenge shown in Fig. 1 (a)), since the foreground distribution in X^T is treated as a characteristic to be mapped between the feature spaces. On the other hand, in Fig. 2 (b) the simulation of pneumonia CT scans, the scans without pneumonia are uniformly sampled from the whole lung scans. While the sampled infected scans are concentrated in the scans with larger lung areas, where are more likely to contain lesions. Therefore, the non-uniform sampling of infected data imports distortion to the estimating of $P(X^T)$, and further misleads the translation mapping $\mathcal{M}^{S\to T}$ (as the issue shown in Fig. 1 (b)).

To circumvent the distortions resulting from the mentioned scenarios, as demonstrated in Fig. 1 (c) sample alignment is performed by aligning the samples used to learn $\mathcal{M}^{S\to T}$. Consequently, the proper translation mapping can be learned from the aligned samples.

2.2. Sample alignment

SA is proposed to correct the sample bias in I2I translation. As shown in Fig. 2, the anatomical structures in original samples are first aligned using diffeomorphic transformation, and then symmetric resampling is performed to eliminate the bias.

2.2.1. Structure alignment by diffeomorphic transformation

Among subjects and imaging modalities, certain topologies are invariant in the anatomical structures of human organs or tissue. Accordingly, atlas-based algorithms [11] are developed to convert the segmentation tasks to image registration tasks by generating a pixel-wise alignment mapping for segmentation masks. Inspired by atlas, structure alignment is conducted between the source and target samples by registering the anatomical topologies.

Diffeomorphic transformation [12] is employed to implement the registration. Denote $x^S(\theta) \in X^S$ and $x^T(\theta) \in X^T$ as the samples separately from the source and target domain, $\theta \in \mathbb{R}^d$ represents points in the coordinates and d is the dimension of the image space. The transformation map $\hat{\tau}$ from $x^T(\theta)$ to $x^S(\theta)$ is given by:

$$\hat{\tau} = \arg\min_{\tau} \mathcal{C}(x^S(\theta), x^T(\tau(\theta))). \tag{1}$$

Diffeomorphic transformation defines a time-dependent, smooth velocity field $v(\tau(\theta, t), t) = \frac{\mathrm{d}\tau(\theta,t)}{\mathrm{d}t}$. $\hat{\tau}$ is computered by intergrating $\hat{v}$, where

$$\hat{v} = \underset{v}{\operatorname{argmin}} \left[\int_0^1 \|v\|^2 \mathrm{d}t + \lambda \int_\Omega \|x^T(\tau(\theta, 1)) - x^S(\theta)\|^2 \mathrm{d}\Omega \right], \tag{2}$$

Ω is the image space.

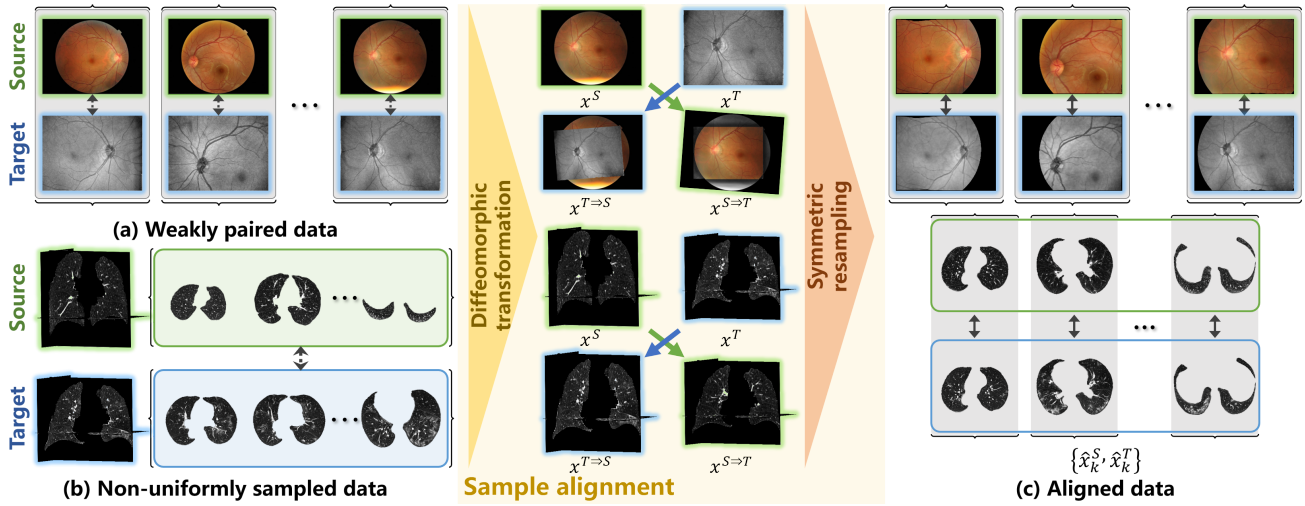

Fig. 2. To address the samples bias in (a) weakly paired data and (b) non-uniformly sampled data, sample alignment is implemented to acquire (c) aligned data $\{\hat{x}_k^S, \hat{x}_k^T\}$. Using diffeomorphic transformation and symmetric resampling, the aligned data corrects the discrepant foreground distributions in (a) and distribution bias in (b) to boost the training of I2I translation models.

Thus the registered image sample $x^{T\Rightarrow S}(\theta)$ from the target to the source domain and the inverse one $x^{S\Rightarrow T}(\theta)$ are obtained by:

$$x^{T\Rightarrow S}(\theta) = x^T\left(\hat{\tau}\left(\theta\right)\right), \tag{3}$$

$$x^{S\Rightarrow T}(\theta) = x^S\left(\hat{\tau}^{-1}\left(\theta\right)\right), \tag{4}$$

where $\hat{\tau}^{-1}$ the inverse map of $\hat{\tau}$. Two sample pairs with aligned structures are constructed by $\{x^S(\theta), x^{T\Rightarrow S}(\theta)\}$ and $\{x^{S\Rightarrow T}(\theta), x^T(\theta)\}$. To simplify the manuscript, only the first pair is used in the following processing, as the two pairs are equally processed.

2.2.2. *Symmetric resampling*

Images without bias are symmetrically resampled from the structure aligned samples to learn the mapping $\mathcal{M}^{S\to T}$.

For weakly paired 2D data, the sample is collected from the intersection of the foreground in the aligned source and target images. Specifically, denote the foreground mask as $M(\cdot)$, the symmetric samples are given by:

$$\begin{aligned}\text{if } \theta \in \mathbb{R}^2, \{\hat{x}_k^S, \hat{x}_k^T\} =& \{x_k^S(\theta), x_k^{T\Rightarrow S}(\theta)\} \\ &\odot [M(x_k^S(\theta)) \cap M(x_k^{T\Rightarrow S}(\theta))],\end{aligned} \tag{5}$$

where k is the index of samples and $\odot$ represents the represents element-wise multiplication.

For unpaired 3D data, the training samples for I2I translation are collected from the aligned 3D images, such as CT, MRI. The 3D samples with the overlapped foreground are defined as:

$$\begin{aligned}\text{if } \theta \in \mathbb{R}^3, \{\widetilde{x}_{ij}^S, \widetilde{x}_{ij}^T\} =& \{x_i^S(\theta), x_{ij}^{T\Rightarrow S}(\theta)\} \\ &\odot [M(x_i^S(\theta)) \cap M(x_{ij}^{T\Rightarrow S}(\theta))].\end{aligned} \tag{6}$$

where i and j are the indexes of source and target samples.

Subsequently, the scan sets are collected, i.e. $\widetilde{X}^S = \{\widetilde{x}_{11}^S, ..., \widetilde{x}_{1J}^S, ..., \widetilde{x}_{IJ}^S\}$ and $\widetilde{X}^T = \{\widetilde{x}_{11}^T, ..., \widetilde{x}_{1J}^T, ..., \widetilde{x}_{IJ}^T\}$. And the aligned training samples $\{\hat{x}_k^S, \hat{x}_k^T\}$ are symmetrically resampled from $\widetilde{X}^S$ and $\widetilde{X}^T$ under the condition $\hat{x}_k^S \in \mathcal{X}^S$ and $\hat{x}_k^T \in \mathcal{X}^T$.

Through diffeomorphic transformation and symmetric resampling, SA not only corrects the bias between the source and target domain, but also constructs paired images $\{\hat{x}_k^S, \hat{x}_k^T\}$ to boost the training of I2I translation mapping $\mathcal{M}^{S\to T}$.

3. EXPERIMENTS AND EVALUATION

3.1. Compliance with Ethical Standards

To validate the effectiveness of the proposed sample alignment, it has been implemented in two DA scenarios of medical image translation, including the cross-modality translation between the weakly paired color fundus photography and En face OCT, and the pathological image synthesis of CT scans with COVID-19. The cross-modality translation is performed on 50 image pairs, and 40 CT volumes collected from CORONACASES.ORG [13] and LUNA16 [14] to synthesize the CT scans infected COVID-19 [15]. The experimental protocol is in accordance with the Declaration of Helsinki and has been approved by the local Ethics Committee.

The state-of-the-art I2I translation algorithms, including pix2pix [16], CycleGAN [17], and CUT [18], are employed to conduct the experiments. The synthesis quality is qualitatively and quantitatively evaluated, and the synthesis is also verified in the automatic diagnosis of COVID-19.

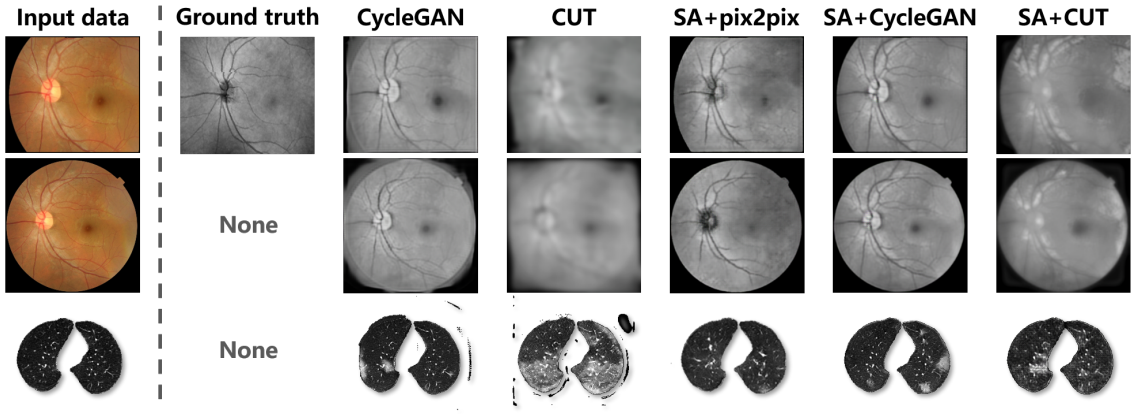

Fig. 3. Visual comparison of the synthesis results. The first two row present the cross-modality translation of fundus images using the aligned and original data, and the third row is the synthesis of COVID-19 infected CT scan.

3.2. Synthesis quality evaluation

To demonstrate the effectiveness of SA, the translation models are respectively trained with and without SA then test on aligned and original data.

Qualitative evaluation is exhibited in Fig. 3. Consistent with Section 2.1, due to the discrepancy in foreground distribution, the OCT images synthesized without SA attempt to fill the background of fundus images. And as a result of the non-uniform sampling, the translation models make efforts on extending the foreground of the lung area.

For the training of I2I translation models, SA not only eliminates the bias caused by the distorted sampling, but also constructs paired samples from the weakly paired or unpaired data. Accordingly, SA boosts the training of translation models. Thus remarkable progress is observed in the image synthesis, and the model of pix2pix is enabled to perform. Moreover, the first two rows of Fig. 3 demonstrate that the data aligned by SA endows the proper training of translation models such that superior performances are achieved in both aligned and original data.

Table 1. Quantitative evaluation of the synthesis quality.

Methods	2D fundus		3D lung	
	PA	mIoU	PA	mIoU
pix2pix	—-	—-	—-	—-
CycleGAN	0.896	0.644	0.964	0.910
CUT	0.847	0.478	0.918	0.835
SA+pix2pix	0.983	0.927	0.993	0.977
SA+CycleGAN	0.988	0.957	0.995	0.985
SA+CUT	0.905	0.704	0.986	0.962

A quantitative evaluation of the synthesis quality is summarized in Table 1. As the pixel-wise paired ground truth is unavailable in this experiment, the synthesis artifacts are quantified using the segmentation evaluation metrics of pixel accuracy (PA) and mean intersection over union(MIoU) to evaluate the synthesis quality. Table 1 shows that the synthesis artifacts have been effectively suppressed by SA.

3.3. Boosting of clinical diagnosis

As I2I translation has been extensively used as a paradigm of DA to boost computer-aided diagnosis, DA on the automatic diagnosis of COVID-19 is performed to verify SA. CT scans are augmented by I2I translation with and without SA, then VGG16 and ResNet18 are trained as the classifier to diagnose CT scans infected COVID-19. The diagnosis accuracy is summarized in Table 2 .

Table 2. Diagnosis accuracy with I2I translation based DA.

Methods	VGG16	ResNet18
Original Data	0.853	0.862
pix2pix / SA+pix2pix	—- / 0.871	—- / 0.885
CycleGAN / SA+CycleGAN	0.811 / 0.903	0.830 / 0.906
CUT / SA+CUT	0.844 / 0.872	0.819 / 0.874

Compared to the diagnosis only using original data, the augmentation without SA degrades the diagnosis performance, resulting from the artifacts in synthesized images. Thanks to SA suppressing the artifacts, the training of diagnosis models is boosted by the synthesized images to outperforms the model only learned from the original data.

4. CONCLUSIONS

To boost the I2I translation based DA in the attendance of sample bias, a sample alignment algorithm is proposed in this paper. The structures in samples are first aligned by diffeomorphic transformation, and paired data are symmetrically resampled from the aligned ones. The experiments demonstrate the effectiveness of the proposed algorithm, it suppresses the artifacts in synthesized images and improved the performances of DA on diagnosis models.

Acknowledgment

This work was supported in part by Guangdong Provincial Department of Education (2020ZDZX3043), Guangdong Provincial Key Laboratory (2020B121201001), Guangdong Basic and Applied Fundamental Research Fund Committee (2020A1515110286), and Shenzhen Natural Science Fund (JCYJ20200109140820699 and 20200925174052004).

5. REFERENCES

[1] Saeid Motiian et al., "Unified deep supervised domain adaptation and generalization," in *ICCV*, 2017, pp. 5715–5725.

[2] Qing Yu et al., "Divergence optimization for noisy universal domain adaptation," in *CVPR*, 2021, pp. 2515–2524.

[3] Zak Murez et al., "Image to image translation for domain adaptation," in *CVPR*, 2018, pp. 4500–4509.

[4] Heng Li et al., "Restoration of cataract fundus images via unsupervised domain adaptation," in *ISBI*. IEEE, 2021, pp. 516–520.

[5] Tao Li et al., "Applications of deep learning in fundus images: A review," *Medical Image Analysis*, p. 101971, 2021.

[6] Heng Li et al., "An annotation-free restoration network for cataractous fundus images," *TMI*, pp. 1–1, 2022.

[7] He Zhao et al., "Supervised segmentation of unannotated retinal fundus images by synthesis," *TMI*, vol. 38, no. 1, pp. 46–56, 2018.

[8] Dong Nie et al., "Medical image synthesis with deep convolutional adversarial networks," *IEEE TBME*, vol. 65, no. 12, pp. 2720–2730, 2018.

[9] Hongxin Wei et al., "Combating noisy labels by agreement: A joint training method with co-regularization," in *CVPR*, 2020, pp. 13726–13735.

[10] Giorgio Patrini et al., "Making deep neural networks robust to label noise: A loss correction approach," in *CVPR*, 2017, pp. 1944–1952.

[11] Aristeidis Sotiras et al., "Deformable medical image registration: A survey," *IEEE TMI*, vol. 32, no. 7, pp. 1153–1190, 2013.

[12] Brian B Avants et al., "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *MIA*, vol. 12, no. 1, pp. 26–41, 2008.

[13] Joseph Paul Cohen et al., "Covid-19 image data collection," *arXiv 2003.11597*, 2020.

[14] Arnaud Arindra Adiyoso Setio et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," *MIA*, vol. 42, pp. 1–13, 2017.

[15] Heng Li et al., "Ct scan synthesis for promoting computer-aided diagnosis capacity of covid-19," in *ICIC*. Springer, 2020, pp. 413–422.

[16] Phillip Isola et al., "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134.

[17] Jun-Yan Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2223–2232.

[18] Taesung Park et al., "Contrastive learning for unpaired image-to-image translation," in *ECCV*, 2020, pp. 319–345.