Structure and Intensity Unbiased Translation for 2D Medical Image Segmentation

Tianyang Zhang, Shaoming Zheng, Jun Cheng, *Senior Member, IEEE*, Xi Jia, Joseph Bartlett, Xinxing Cheng, Zhaowen Qiu, Huazhu Fu, *Senior Member, IEEE*, Jiang Liu, *Senior Member, IEEE*, Aleš Leonardis, *Member, IEEE*, and Jinming Duan

***Abstract*—Data distribution gaps often pose significant challenges to the use of deep segmentation models. However, retraining models for each distribution is expensive and time-consuming. In clinical contexts, device-embedded algorithms and networks, typically unretrainable and unaccessable post-manufacture, exacerbate this issue. Generative translation methods offer a solution to mitigate the gap by transferring data across domains. However, existing methods mainly focus on intensity distributions while ignoring the gaps due to structure disparities. In this paper, we formulate a new image-to-image translation task to reduce structural gaps. We propose a simple, yet powerful Structure-Unbiased Adversarial (SUA) network which accounts for both intensity and structural differences between the training and test sets for segmentation. It consists of a spatial transformation block followed by an intensity distribution rendering module. The spatial transformation block is proposed to reduce the structural gaps between the two images. The intensity distribution rendering module then renders the deformed structure to an image with the target intensity distribution. Experimental results show that the proposed SUA method has the capability to transfer both intensity distribution and structural content between multiple pairs of datasets and is superior to prior arts in closing the gaps for improving segmentation.**

***Index Terms*—Cardiovascular imaging (CMR), diffeomorphic image registration, generative adversarial network, medical image segmentation, medical image translation, optical coherence tomography (OCT).**

Manuscript received 3 December 2023; revised 27 April 2024; accepted 18 July 2024. Date of publication 29 July 2024; date of current version 5 November 2024. This work was supported in part by the EPSRC and UKRI through Advanced Research Computing at the University of Birmingham under Grant EP/T022221/1 and Grant EP/W032244/1, in part by BHF Accelerator Award under Grant AA/18/2/34218, in part by Korea Cardiovascular Bioresearch Foundation under Grant CHORUS Seoul 2022, in part by UK Biobank Resource under Application 40119, in part by Huazhu Fu's A*STAR Central Research Fund, in part by AI Singapore Programme under Grant AISG2-TC-2021-003, and in part by Heilongjiang Province Key Research and Development Program under Grant 2023ZX02C10. Recommended for acceptance by S. Gao. *(Corresponding authors: Jun Cheng; Zhaowen Qiu; Jinming Duan.)*

Tianyang Zhang, Xi Jia, Xinxing Cheng, and Aleš Leonardis are with the University of Birmingham, B15 2TT Birmingham, U.K.

Shaoming Zheng is with Imperial College London, SW7 2AZ London, U.K.

Jun Cheng is with the Institute for Infocomm Research, A*STAR, Singapore 138632 (e-mail: sam.j.cheng@gmail.com).

Joseph Bartlett is with the University of Birmingham, B15 2TT Birmingham, U.K., and also with the University of Melbourne, Victoria, VIC 3052, Australia.

Zhaowen Qiu is with NorthEast Forestry University, Harbin 150040, China (e-mail: qiuzw@nefu.edu.cn).

Huazhu Fu is with the Institute of High Performance Computing, A*STAR, Singapore 138632.

Jiang Liu is with the Southern University of Science and Technology, Shenzhen 518055, China.

Jinming Duan is with University of Birmingham, B15 2TT Birmingham, U.K., also with the Alan Turing Institute, NW1 2DB London, U.K., and also with the University of Manchester, M13 9PL Manchester, U.K. (e-mail: jinming.duan@manchester.ac.uk).

Digital Object Identifier 10.1109/TPAMI.2024.3434435

I. Introduction

MEDICAL imagesegmentation [1], [2] has been a hot topic in the last few decades, in particular, deep learning based techniques [3], [4] have drawn lots of attention. These methods often assume that the training and test data follow the same distribution. However, domain gaps often exist between data from different sources, e.g., the data from different hospitals or clinic centers is often captured by different machines with different settings. In addition, as imaging technology continues to progress, old machines become outdated and are therefore replaced by their modern counterparts. Therefore, even the new data and the accumulated data from the same place have different distributions. When it comes to performing inference using deep learning techniques, such differences between the training and inference distributions can degrade performance significantly. However, it is clear that recollecting and labelling the required training data for each distribution is hugely expensive and time-consuming. Therefore, the effective use of labelled data from previous devices or settings is vital.

To solve this issue, transfer learning methods [5], [6], [7], such as unsupervised domain adaptation (UDA) [2], [8], [9], [10], are possible ways to map the data into a different space such that the domain gaps between the source domain and target domain are minimized. A limitation of such approaches is that they often require the model to be retrained or its latent features to be accessed. However, networks and labels compiled to software can not be touched in clinical applications. Recently, generative adversarial models have been proposed to tackle this problem by transferring the intensity distributions from source to target domain and reducing the domain gaps. In this task, the input image is transferred (adapted) and then tested with a model trained on the original labelled data. Specifically, this task has been defined in [11] and is considered to be different from UDA or other similar problems.

Previously, Chen et al. [12] proposed a MUNIT based model to transfer the intensity distribution of the source dataset to that of the target dataset by maintaining the content in the latent space. Zhang et al. [13] proposed to generate new images by maintaining the main edges in the images. However, these methods overlooked the differences in structure statistics. In fact,

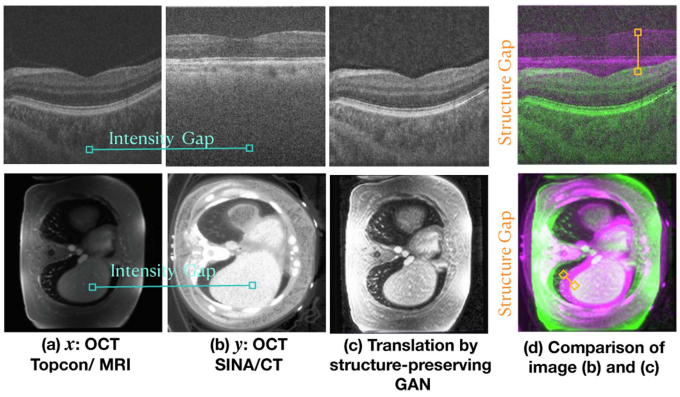


Fig. 1. Images illustrate the domain shift problem, with green and orange lines showing differences in intensity distribution and structure compared to the target image. The comparison images on the right side highlight that structural gaps, e.g., curve, position and anatomy, are challenging to mitigate through translation methods that keep the structure unchanged. Specifically, $\boldsymbol{x}$ represents a sample from the source dataset, while $\boldsymbol{y}$ represents a sample from the target dataset. The purple and green regions indicate areas where the representation in (b) differs from (c), and where the representation in (c) differs from (b), respectively.

the structures captured using various imaging techniques can be different due to the nature of imaging principles. For instance, the curvatures of the objects of interest in images are largely affected by scales. Fig. 1 shows two OCT images (a) and (b) from different manufacturers. As shown in sub-image (a), not only structures but also intensities are different from the target (b). Therefore, we have a task with contradictive objectives: on the one hand, we hope to maintain the contents/edges in the generated data for the subsequent analysis. It leads to the propagation of the structural gap. On the other hand, we hope to eliminate the structural gap as the generated and real data are supposed to be indistinguishable to the discriminator. Overall, we present the translation problem to closing the existing domain gaps of structure and intensity distribution and improve the segmentation performance. For this segmentation problem under consideration, a model is meticulously trained using target domain images alongside their associated masks. The trained model is subsequently tested on the samples translated from source domain to target domain.

In this paper, we propose a simple yet important structure unbiased adversarial approach (SUA) to overcome the above issue. It extracts the main structures from the image and translates them separately such that both structures and intensities gaps are reduced. With this translation, it is expected that the performance of models trained on the training data (target) will improve when tested on data (source) from a different distribution. Finally, we also compute an inverse deformation field to warp the segmentation result back to its original domain, which is an essential step to obtain the segmentation for the original images. Note that our method is different from a spatial deformation followed by a structure-preserving GAN. To better understand the motivation of the proposed method, we use T-SNE [14] to visualize OCT images after different processing in Fig. 2. As shown, there is a large gap between the data from the two domains indicated by the yellow and blue dots. Although a deformation could reduce the gap, the gap due to intensity distributions remain, as shown in (b). Similarly, (c) illustrates that after applying a structure preserving GAN to the images they can still be distinguished. Whilst applying a deformation followed by a structure preserving GAN is able to reduce the gap, the deformation causes significant texture distortions. Such distortions cannot be removed by the structure-preserving translation GAN, as is described in Section IV-A3. Additionally, a brief illustration of the proposed method compared with other translation methods is shown in Fig. 3.

The main contributions of the paper are summarized as follows:

- We find that the domain gaps exist in not only the intensity distributions but also in the image structures. Then we identify an important image-to-image translation problem which helps to reduce domain gaps associated with both image structure and intensity distributions. By taking this advantage, we demonstrate the effectiveness of our method for medical image segmentation.
- We propose a novel SUA network to tackle domain gaps in segmentation problem by a novel image-to-image translation strategy. Specifically, the proposed method has the capability to translate images by reducing the structural and intensity gaps with a spatial transformation block and a structure rendering mapping respectively.
- Extensive experiments on two retinal OCT datasets, a chest CT & MRI paired dataset and two cardiac datasets show that the proposed method is able to transfer both structure shapes and intensity distributions effectively with improved subsequent segmentation results.

II. Related Works

A. Generative Adversarial Networks (GANs)

GANs [15], [16] are originally proposed to generate images from random inputs in an unconditional manner. They contain a generator and a discriminator; the generator aims to output samples to be indistinguishable from the training samples while the discriminator tries to differentiate them. Recently, many conditions have been proposed and integrated into GANs for various applications such as image segmentation [8], [17], [18], image synthesis [19], super-resolution [20], image reconstruction [21], [22], [23], medical image translation [24], [25], [26], [27] and text-to-images synthesis [28]. Such conditional GANs are generally divided into paired image-to-image translation and unpaired image-to-image translation.

B. Paired Image-to-Image Translation

Paired image-to-image translation methods are trained on a paired dataset to obtain a mapping which converts an image from one domain to another [17], [29]. Earlier, many paired image-to-image translation algorithms have been proposed for various tasks, e.g, super resolution [20], image segmentation [30], [31], [32], [33], spatial transformation, denoising, etc. Isola et al. [17] proposed pix2pix which can be applied to general translation tasks. However, it is often difficult to obtain paired training data to train models of this type.

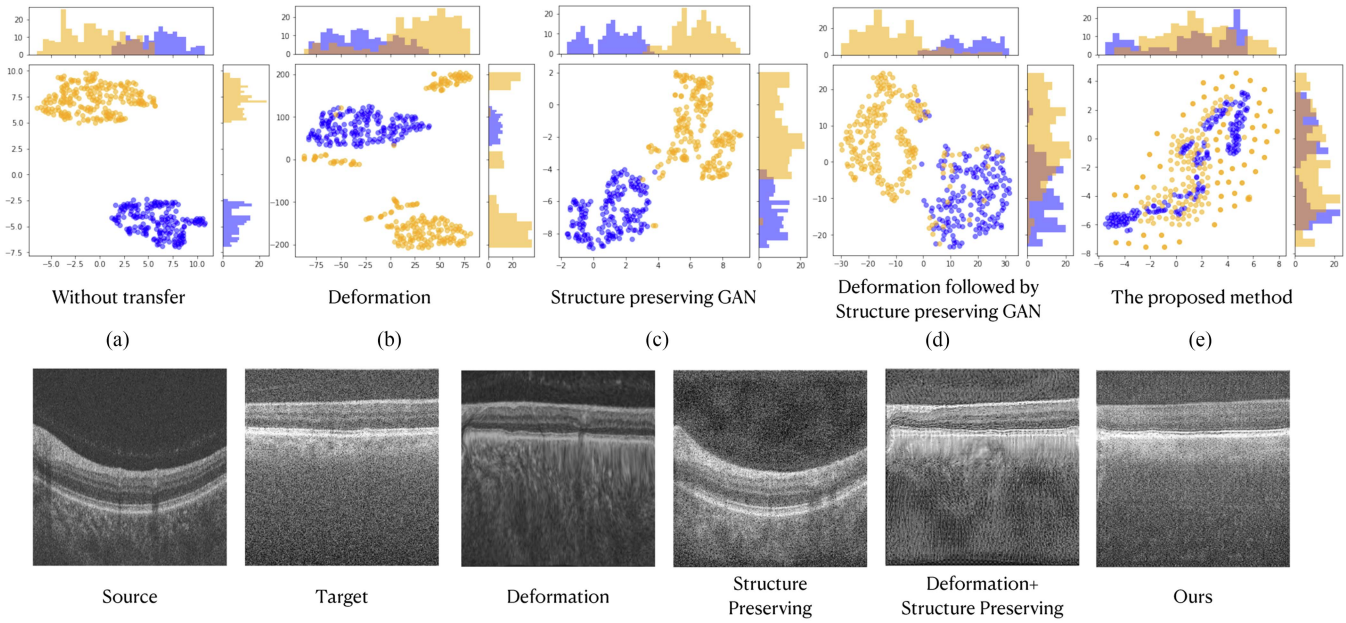


Fig. 2. Illustrations of domain shift issues using T-SNE. Yellow and blue dots indicate images from two different domains. From left to right, figures are plotted based on: (a) original data points; (b) data points transformed by spatial deformation; (c) data points transferred by structure-preserving GAN; (d) data points transformed first by spatial deformation followed by structure-preserving GAN; and (e) data points transferred by our proposed SUA method. The second row shows samples from the source domain, target domain, and results transformed by each method: Spatial deformation, structure-preserving GAN, spatial deformation followed by structure-preserving GAN, and our method.

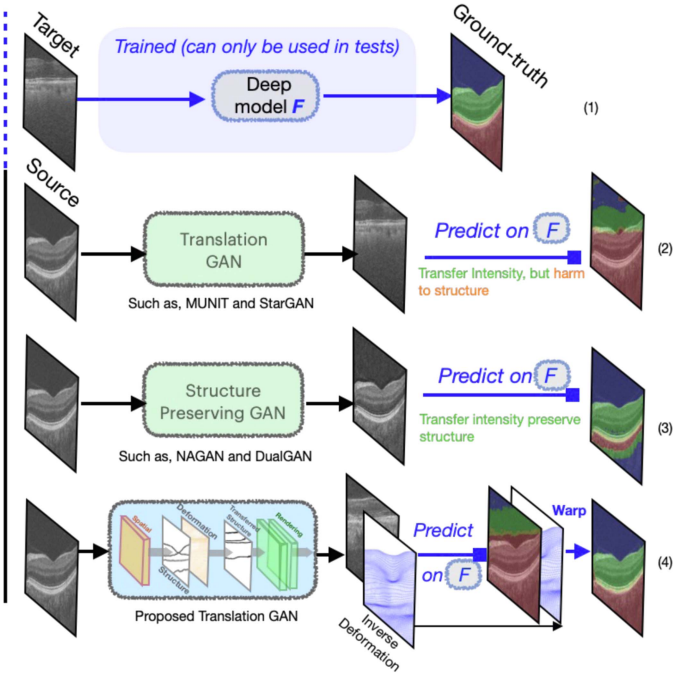


Fig. 3. Illustration of pipelines of translation methods for segmentation: (1) Trained segmentation model without domain adaptation; (2) Translation GAN; (3) Structure preserving GANs, and (4) our GAN with learnable deformation (more details in Fig. 4).

C. Unpaired Image-to-Image Translation

Unpaired methods are trained on unpaired image datasets, i.e., for each sample in one dataset, there are no corresponding samples in the other dataset. Donahue et al. [34] established an unsupervised generative adversarial method named BiGAN, which uses feature learning and representation to translate the data. Zhu et al. [35] proposed a cycle-consistent adversarial network (Cycle-GAN), which consists of forward and backward cycle models for unsupervised image-to-image translation. Kim et al. [36] invented DiscoGAN to seek the relationship between source and target domains. Huang et al. [37] proposed MUNIT, which separates content and intensity distributions in the latent space and achieves translation by switching intensity distributions. Yan et al. [38] presented an improvement of CycleGAN, which reduces the domain gaps between cardiac images obtained on devices from different vendors. Zhang et al. [13] proposed an approach to generate new images while preserving the main edges in the original images. Guo et al. [39] proposed SCCGAN which has a similar ability to preserve structures of source images during translation through a structure consistency constraint. Fourier-transform-based style translation methods [40], [41], [42] also demonstrate notable proficiency in efficiently translating images while preserving essential structural details. These techniques function by decomposing images into low-frequency and high-frequency components, with the latter capturing object structures resembling the identity. Choi et al. [43] proposed an image-to-image translation network named StarGAN v2 based on StarGAN which can translate the image with richer textures than CycleGAN. Many other implementations of StarGAN in medical image translation have since been established. For example, Abu-Srhan et al. [44] proposed a TarGAN based on StarGAN architecture and cycle-consistency loss for multi-modal image translation, and Bashyam et al. [45] utilized StarGAN v2 on Brain MRI domain translations. Diffusion-based

method DDIB [46], which transfers images by connecting source and target noising distributions through Schrödinger bridge. Since unpaired methods do not require paired samples, these methods are more convenient to practically apply and attract more attention for reducing domain gaps. These unpaired methods transfer the intensity distributions between domains and reduce the domain gaps to a certain extent. However, existing methods mainly consider the gap between intensity distributions while failing to capture changes in structure, which is vital for medical image segmentation.

D. Spatial Deformation

Spatial deformation estimates a mapping between a source image and a target image, and the resultant deformation can be used to warp the source image such that it has similar structure to that in the target domain. Among different approaches, diffeomorphism is an important feature which describes an invertible function that maps one differentiable deformation to another such that both the function and its inverse are differentiable [47]. Conventional iterative diffeomorphic methods such as LDDMM [47], Dartel [48], ANTs [49], and Demons [50] are accurate but suffer from high computational costs. Very recently, Thorley et al. [51] proposed a fast iterative method based on the Nesterov accelerated ADMM for diffeomorphic routine, but its performance is limited for cross-domain problems. Building on the spatial transformer network [52], the last few years have seen a boom in image spatial transformer methods based on deep learning. These include VoxelMorph [53], SYMNet [54], B-spline Network [55] and VR-Net [56], just to name a few. To achieve diffeomorphisms, most deep learning methods use multiple squaring and scaling as a neural layer [53], [54], [55]. However, these methods are designed without considering differences in intensity distributions.

E. Geometry Accuracy in Synthetic Data

The assessment of geometry accuracy in synthetic data serves as a crucial evaluation metric, as highlighted by [57]. For instance, the Dice similarity coefficient is commonly computed to evaluate the accuracy of representing specific tissue classes or structures such as bones, fat, muscle, air, and the overall body. It has been observed that registration errors in synthetic data can introduce blurring artifacts in high-contrast regions, leading to inaccuracies in treatment localization, as noted by [58]. In certain applications, such as subsequent segmentation tasks, achieving structure correspondence becomes essential. Jiang et al. [59] investigated the use of MRI-to-CT translation to enhance the robustness of segmentation, while Kieselmann et al. [60] generated synthetic MRI data from CT scans to train segmentation networks for reliable auto-segmentation algorithms of organs-at-risk and radiation targets. Additionally, Zhang et al. [13] explored the adaptation segmentation task for medical images, emphasizing the importance of geometry accuracy in medical image translation when applying a trained segmentation model to fit images from a different domain by translating them to the source domain. In Section IV of our work, we have demonstrated the efficiency of our proposed method in this segmentation task.

III. Method

In this paper, we propose an image translation approach such that an input image can be converted to an output image while the domain gaps in both intensity distribution and structure can be minimized. We define the test (source) data to be translated as X_S and the set of training (target) data with labelled segmentation ground truth as X_T. Our goal is to learn a mapping from X_S to X_T, such that the element $\boldsymbol{x}_S \sim P_{X_S}$ will be translated to $\boldsymbol{x}_T \sim P_{X_T}$ which has the same underlying structure and intensity distribution as in X_T. Since the objective of the intensity distribution translation is to overcome the domain gap issue for subsequent analysis tasks such as segmentation, the content of the images or the underlying clean part of the images is what really matters and shall be kept unchanged. However, arbitrarily maintaining the edges would lead to the propagation of the structural gap. In this paper, we first obtain the main structure $\boldsymbol{u}$ via a preprocessing step. Then, a spatial transformation is used to get the forward and inverse deformation fields (ϕ and ϕ^{-1}) between the input and target images. After that, we utilize ϕ to warp structure masks which are formed by filling the structural images to get $\boldsymbol{u}(\phi)$ which we re-obtain the edges from to get the warped structure and feed into the structure-to-image rendering generator G. Finally, the outputs $G(\boldsymbol{u}(\phi), \boldsymbol{x}_S(\phi))$ of G are expected to have similar structures and intensities to images in the target domain. The overall process is shown in Fig. 4. Since the structures are deformed in the spatial transformation block, we use the inverse deformation field ϕ^{-1} to warp the segmentation outputs back to original shape as the final segmentation output. An illustration of the intermediate results is given in Fig. 5 for a better understanding of the proposed method.

A. Pixel Clustering Map

We first compute a basic pixel clustering map $\boldsymbol{u}^*$ for input image $\boldsymbol{x}$ using the joint image reconstruction and pixel clustering Potts model [61]. Then its edge sketch $\boldsymbol{u}$ is obtained. Meanwhile, we combine the clustering map into binary masks. Then these masks are given a Gaussian gradient and multiplied by their corresponding source images $\boldsymbol{x}_S$ and $\boldsymbol{x}_T$ to get the composed structure images $\boldsymbol{I}_S$ and $\boldsymbol{I}_T$, which is shown in Fig. 6.

B. Spatial Transformation Block

In this section, we establish a spatial transformation block which obtains a deformation to reduce the domain shift between dominant structures in the datasets. We note that most related methods only include a forward deformation estimation strategy. However, an inverse deformation is crucial for our application in image segmentation. If we perform an inverse computing by switching source and target, the obtained inverse deformation will not match the forward deformation well, thus, we propose a spatial transformation block which learns the forward and inverse deformation fields simultaneously.

After conducting the step mentioned in Section III-A, a pair of the composed structure images $\boldsymbol{I}_S$ and $\boldsymbol{I}_T$ (the pair with the maximum SSIM score) are obtained. Here, we propose an

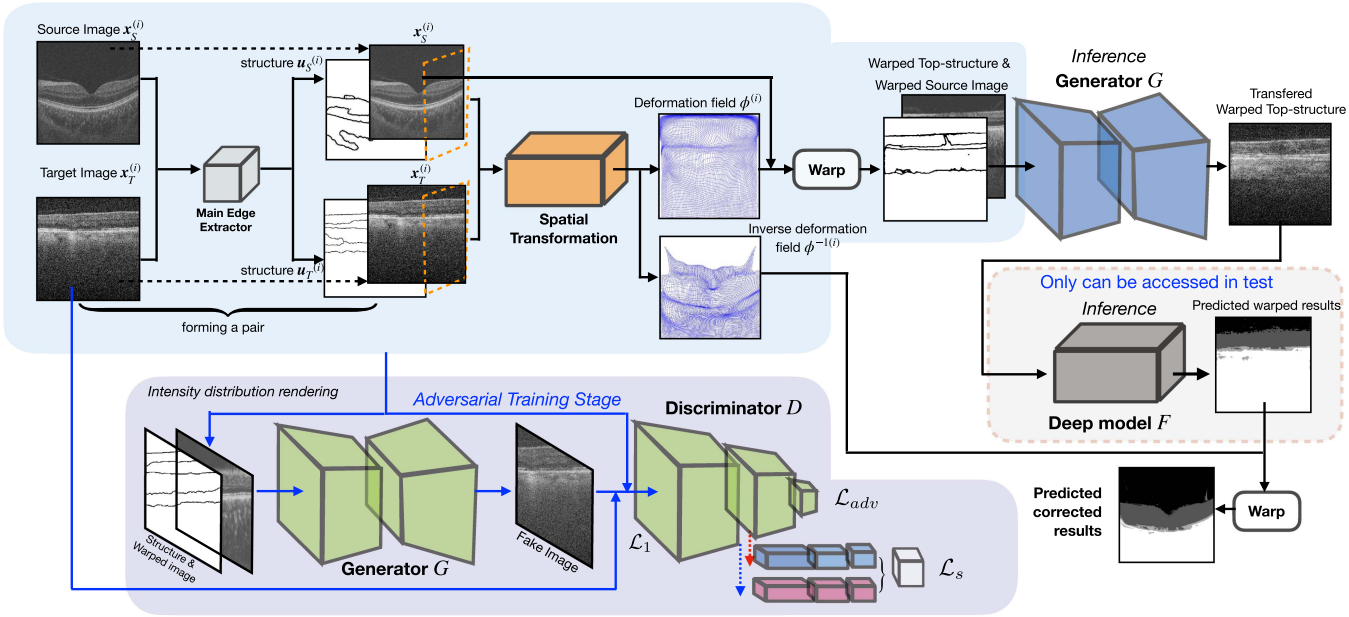


Fig. 4. An illustration of the SUA: First, the source and target images ($\boldsymbol{x}_S^{(i)}$ and $\boldsymbol{x}_T^{(i)}$) are processed to compute the dominant structure of the input image. Then, the obtained dominant structures $\boldsymbol{u}_S{}^{(i)}$ and $\boldsymbol{u}_T{}^{(i)}$ are used to compute the deformation field $\boldsymbol{\phi}^{(i)}$ and its inverse $\boldsymbol{\phi}^{-1(i)}$. The deformation field $\boldsymbol{\phi}^{(i)}$ is used to warp the $\boldsymbol{u}_S{}^{(i)}$, which is further processed by the generator G. The resultant image is fed to the trained segmentation model, whose output is warped back by the inverse deformation field $\boldsymbol{\phi}^{-1(i)}$ to get the final segmentation result.

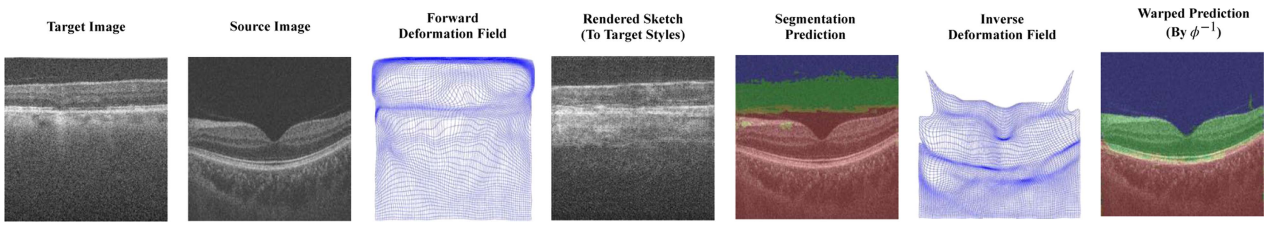


Fig. 5. Intermediate results produced by the pipeline. From left to right are target image $\boldsymbol{x}_T$, source image $\boldsymbol{x}_S$, the forward deformation field ϕ, the warped top-structure, the rendered top-structure, the prediction results, the inverse deformation field ϕ^{-1}, and the prediction results warped back to the original structure.

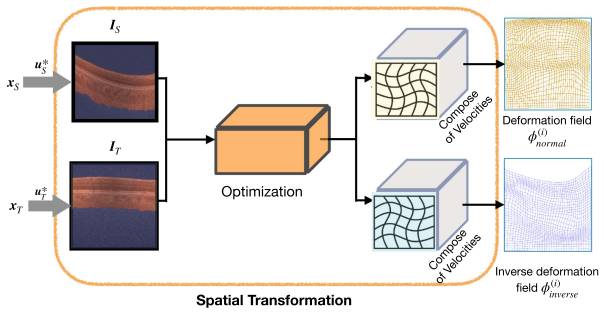


Fig. 6. Illustration of the spatial transformation procedure.

invertible spatial transformation to align them. Since we do so by multiplying Gaussian masks with the original images, this can be considered performing a transformation on the masks rather than a cross-domain transformation. Computing a diffeomorphic deformation can be treated as modelling a dynamical system [47], given by an ordinary differential equation (ODE): $\partial\phi/\partial t = \mathbf{v}_t(\phi_t)$, where $\phi_0 = \mathrm{Id}$ is the identity transformation and $\mathbf{v}_t$ indicates the velocity field at time t ($\in [0, 1]$). To solve the ODE, we use Euler integration, in which the forward deformation field ϕ is calculated as the compositions of a series of small deformations, defined as

$$\phi = \left(\mathrm{Id} + \frac{\mathbf{v}_{t_{N-1}}}{N}\right) \circ \cdots \circ \left(\mathrm{Id} + \frac{\mathbf{v}_{t_1}}{N}\right) \circ \left(\mathrm{Id} + \frac{\mathbf{v}_0}{N}\right). \quad (1)$$

The backward deformation can be computed reversely as

$$\phi^{-1} = \left(\mathrm{Id} - \frac{\mathbf{v}_0}{N}\right) \circ \left(\mathrm{Id} - \frac{\mathbf{v}_{t_1}}{N}\right) \circ \cdots \circ \left(\mathrm{Id} - \frac{\mathbf{v}_{t_{N-1}}}{N}\right). \quad (2)$$

In the above equations, if the velocity fields $\mathbf{v}_{t_i}, \forall i \in \{0, \ldots, N-1\}$ are sufficiently small whilst satisfying some smoothness constraints, the resulting composition is a diffeomorphic deformation. In addition, note that the composition between ϕ and ϕ^{-1} will give an approximate identity grid and one can use ϕ^{-1} to warp an image back.

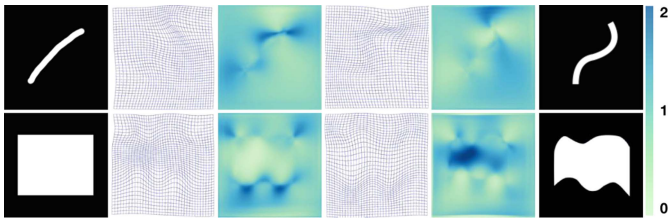


Fig. 7. Visualisations of diffeomorphic forward and backward (inverse) deformations. The first and the last columns show the source and target image, respectively. The second and the fourth columns show the forward and backward deformation fields, respectively. Finally, the third and the fifth columns show the Jacobian determinants of corresponding forward and backward deformations, respectively.

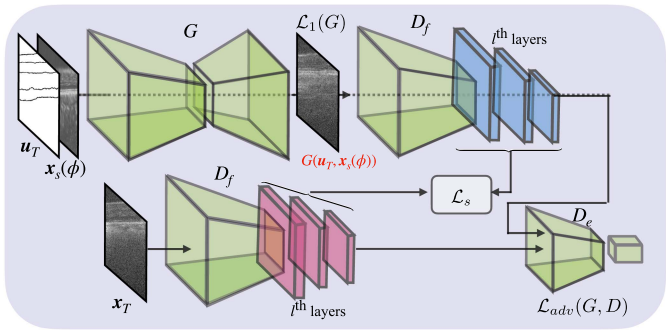


Fig. 8. Architecture of the intensity distribution rendering network, where D_f and D_e denotes the start and the end parts of the discriminator D.

To compute the velocity fields whilst satisfying these diffeomorphic constraints, we use the following model

$$\min_{\mathbf{v}} \left\{ \frac{1}{2}\|\rho(\mathbf{v})\|^2 + \frac{\lambda}{2}\|\nabla^n \mathbf{v}\|^2 \right\}, \tag{3}$$

where $\rho(\mathbf{v}) = \langle \nabla \boldsymbol{I}_S, \mathbf{v} \rangle + \boldsymbol{I}_S - \boldsymbol{I}_T$. ∇^n denotes the n^{th} order gradient, and here we use $n = 3$. For a scalar-valued function $f(x, y, z)$ in the continuous setting, $\nabla^n f = \left[\binom{n}{k_1, k_2, k_3} \frac{\partial^2 f}{\partial x^{k_1} \partial y^{k_2} \partial z^{k_3}} \right]^{\mathrm{T}}$, where $k_i = 0, \ldots, n, i \in \{1, 2, 3\}$ and $k_1 + k_2 + k_3 = n$. Moreover, $\binom{n}{k_1, k_2, k_3}$ are known as the multinomial coefficient which is computed by $\frac{n!}{k_1! k_2! k_3!}$. To solve the model effectively, we use a multiscale ADMM algorithm developed in [51]. An illustration of forward and inverse deformations is shown in Fig. 7.

C. Intensity Distribution Rendering

An intensity distribution rendering network is used to render the warped structures with the targeted intensity distribution. As shown in Fig. 8, it is a paired image-to-image translation network, including a "U-Net-like" [62] structure for translation mapping, a discriminator for adversarial training, and a feature alignment mechanism for loss computation. Mathematically, the losses $\mathcal{L}_{adv}$, $\mathcal{L}_1$ and $\mathcal{L}_s$ corresponding to the three components are denoted as:

$$\mathcal{L}_{adv} = \mathbb{E}_{\boldsymbol{u}_T \sim U_T}[\log(1 - D(G(\boldsymbol{u}_T, \boldsymbol{x}_S(\phi)), \boldsymbol{u}_T))] + \mathbb{E}_{(\boldsymbol{x}_T, \boldsymbol{u}_T) \sim (X_T, U_T)}[\log D(\boldsymbol{x}_T, \boldsymbol{u}_T)], \tag{4}$$

where $\boldsymbol{u}_T \sim U_T$ represents the targeted structure, $(\boldsymbol{x}_T, \boldsymbol{u}_T) \sim (X_T, U_T)$ represent the targeted image and structure pair and D represents the discriminator of the mapping G, and ϕ represents the deformation between $\boldsymbol{x}_T$ and its closest $\boldsymbol{x}_S$ (maximum SSIM).

$$\mathcal{L}_1 = \mathbb{E}_{(\boldsymbol{x}_T, \boldsymbol{u}_T) \sim (X_T, U_T)} \|G(\boldsymbol{u}_T, \boldsymbol{x}_S(\phi)) - \boldsymbol{x}_T\|_1, \tag{5}$$

where $\|\cdot\|_1$ denotes the ℓ_1-norm. Then feature correlations are given by Gram matrix Gr, where $Gr(\boldsymbol{x}_T)_{ij}^{(l)}$ correspond to the Gram matrix of lth layer of D with input $\boldsymbol{x}_T$ and $\boldsymbol{u}_T$. It is computed as follows:

$$Gr(\boldsymbol{x}_T)_{ij}^{(l)} = \mathrm{vec}\left[\mathcal{F}(\boldsymbol{x}_T)_i^{(l)}\right]^T \mathrm{vec}\left[\mathcal{F}(\boldsymbol{x}_T)_j^{(l)}\right], \tag{6}$$

where $\mathcal{F}(\boldsymbol{x}_T)_i^{(l)}$ represents the lth layer of D's feature maps, the subscripts i and j denote the ith and jth channel, and $\mathrm{vec}(\cdot)$ denotes the vectorization operation.

$$\mathcal{L}_s = \mathbb{E}_{(\boldsymbol{x}_T, \boldsymbol{u}_T) \sim (X_T, U_T)} \sum_{l=1}^{3} \|Gr(G(\boldsymbol{u}_T, \boldsymbol{x}_S(\phi)))^{(l)} - Gr(\boldsymbol{x}_T)^{(l)}\|_F \tag{7}$$

The full objective of the intensity distribution rendering network is described as the following equation.

$$G^*, D^* = \arg\min_G \max_D [\mathcal{L}_{adv} + \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_s], \tag{8}$$

where λ_1 and λ_2 are hyper-parameters that balance different losses. The optimization procedures of the whole model are summarized in Algorithm 1.

Algorithm 1: Optimization Procedures.

Input: $\boldsymbol{X}_S$ and $\boldsymbol{X}_T$.
Require: G and D. Two Adam optimizers A_1 and A_2.
for $epoch < N$ **do**
  if $epoch = N - 1$ **then**
    for $\boldsymbol{x}_S^{(j)} \in \boldsymbol{S}$ **do**
      Obtain $\boldsymbol{u}_S$ and $\boldsymbol{u}_S^*$;
      Calculate the SSIM of $\boldsymbol{I}_S^{(i)}$ and $\boldsymbol{I}_T^{(j)}$;
    end
    Solve deformations ϕ and ϕ^{-1} for indices of the pair that has the maximum SSIM via (3);
  end
  for $\boldsymbol{x}_S^{(i)} \subset \boldsymbol{X}_S$ **do**
    for $\boldsymbol{x}_T^{(j)} \subset \boldsymbol{X}_T$ **do**
      Obtain $\boldsymbol{u}_T^{(i)}$; Calculate $\mathcal{L}_{adv}(G, D)$, $\mathcal{L}_s(G)$ and $\mathcal{L}_1(G)$;
      $D \leftarrow A_1(D, [\boldsymbol{x}_T^{(j)}, \boldsymbol{u}_T^{(i)}], \mathcal{L}_{adv})$; $G \leftarrow A_2(G, D, [\boldsymbol{x}_T^{(j)}, \boldsymbol{u}_T^{(i)}], [\mathcal{L}_{adv}, \mathcal{L}_1, \mathcal{L}_s])$;
    end
  end
end
Output: G, ϕ, ϕ^{-1}.

D. Implementation

Network Architecture: The generator in our networks includes 15 residual 2D convolution blocks. Specifically, the first 8 blocks are encoder blocks, where each block contains a 4×4 convolution layer sequentially followed by a ReLU activation function, an instance normalization layer and a dropout layer; the rest are decoder blocks, where each block contains a 4×4 transposed convolution layer sequentially followed by a ReLU function, an instance normalization layer and a dropout layer, respectively. The discriminator in the proposed model contains four convolution blocks and a fully connected layer, where these convolution blocks include a 4×4 convolution layer (padding equals to 1), an instance normalization layer and a Leaky-ReLU function. The core code will be released after the acceptance of our work.

Training Details: In intensity distribution rendering, we set the balancing hyper-parameter $\lambda_1 = 1$ and $\lambda_2 = 100$ in (8) for all the experiments. We apply the Adam [63] optimizer with a learning rate of 0.0002, which decays to zero following a linear principle from the 100th epoch to the 200th epoch in all experiments. Moreover, the padding pixel number is 8, the tolerance ratio is 0.001, the balance hyper-parameter is 5 and the max-iteration is 50 in the spatial transformer of all experiments. Additionally, we set the $\gamma = 0.35$ for the OCT and the cardiac experiments, and $\gamma = 0.55$ for the MRI and CT experiments in the Potts model.

IV. Experimental Results

We conduct a comprehensive evaluation of the Structure-Unbiased Adversarial (SUA) network across multiple domain adaptation tasks in medical image segmentation. In retinal Optical Coherence Tomography (OCT), we utilize the SINA and ATLANTIS datasets to address structural and intensity distribution differences. Similarly, in MRI-to-CT domain transfer, we leverage publicly available datasets to demonstrate the method's applicability to multi-modal imaging. Additionally, we assess the adaptability of SUA to pathological conditions by transferring images from diseased (ACDC) to healthy (UKBB) subjects in cardiac MRI. Through extensive experiments, we evaluate both structure and intensity translation, employing various evaluation metrics and comparisons with state-of-the-art methods. These experiments collectively showcase the effectiveness and versatility of our SUA in addressing domain gaps in medical image segmentation tasks.

A. Domain Translation on Retinal OCT

1) Datasets: We apply SUA on retinal OCT to transfer the shape characteristics and intensity distribution of one OCT dataset to another. The SINA[1] and ATLANTIS datasets are used. The SINA dataset contains 220 B-scans from 20 volumes of eyes with drusen and geographic atrophy, collected using a spectral domain-OCT imaging system from Bioptigen. Three boundaries have been manually annotated, including boundary 1: internal limiting membrane (ILM); boundary 2: between the outer segments and the retinal pigment epithelium (OS/RPE); and boundary 3: between Bruchs membrane and the choroid (BM/Choroid). We use this dataset as the target T. The ATLANTIS dataset is a local dataset, containing 176 B-scans, collected using a swept-source OCT machine. The same three boundaries as SINA are annotated and used in this study. Since the two datasets are collected from different machines under different protocols, there exist gaps in both structure and intensity distributions, which makes the models trained from one dataset perform poorly on the other.

2) Comparison With Prior Arts: The performance of the image translation is evaluated in terms of both structure and intensity distribution.

Evaluation on Structure Translation: The performance evaluation of the structure transfer is challenging. Ideally, it would be evaluated by comparing the translated image with the corresponding image from the other machine. However, it is not practical in image to identify the exact same region of the object using different machines. Since the main objective of the translation is to improve subsequent analysis, we evaluate its performance using the segmentation results indirectly.

A segmentation network based on the U-Net architecture is first trained using the SINA dataset (training/target dataset). The proposed generative model is trained to translate ATLANTIS images such that the structure characteristics and intensity distributions of the translated images are similar to those in SINA. The trained U-Net segmentation model is then applied on the translated data (test/source dataset) to detect the three boundaries, which are subsequently used to evaluate the performance of the image-to-image translation model.

The mean intersection-over-union ($mIoU$) [65], the mean S ϕ rensen–Dice coefficient ($Dice$), the accuracy (Acc), the sensitivity (Sen), the specificity (Spe), and the false discovery rate (FDR) are used as evaluation metrics.

To evaluate the effectiveness of the proposed SUA network, we compare it with state-of-the-art translation methods including CycleGAN [35], MUNIT [37], DualGAN [64], StarGAN v2 [43] NAGAN [13], SCCGAN [39] and DDIB [46] as well as the spatial transformation methods VoxelMorph [53], VoxelMorph with mutual information loss (VoxelMorph+MI) and VR-Net [56]. For all SOTA methods and the test tool U-net, we use the same hyper-parameters described in the open-sourced codes or the papers from the original authors. In the experiments, we train these GANs in a similar way to learn the translation from ATLANTIS to SINA. The transferred ATLANTIS is then fed into the U-Net segmentation network to segment the three boundaries for comparison with the manual ground truth. For the spatial transformation methods, the deformation is computed to warp the image for segmentation. The output of the segmentation is warped back by the inverse deformation for comparison. Table I shows the comparison between the proposed method and other methods, where values are shown in the form of mean(std). As shown from the results, the proposed SUA outperforms the state-of-the-art performance on all metrics.

Fig. 9 shows some sample results for visual comparison. As can be observed, Cycle-GAN indiscriminately transfers both

[1] https://people.duke.edu/\,sf59/Chiu_IOVS_2011_dataset.htm

TABLE I
DISTRIBUTION AND SEGMENTATION EVALUATION OF RETINAL OCT

Methods	Distribution		Segmentation					
	$D_{Bhat}\downarrow$	$Corr\uparrow$	$Acc\uparrow$	$Dice\uparrow$	$mIoU\uparrow$	$Sen\uparrow$	$Spe\uparrow$	$FDR\downarrow$
Without Translation	0.391	0.587	0.854 (0.028)	0.538 (0.060)	0.520 (0.057)	0.586 (0.056)	0.891 (0.023)	0.384 (0.027)
VoxelMorph [53]	0.340	0.369	0.760 (0.013)	0.324 (0.016)	0.235 (0.015)	0.500 (0.015)	0.806 (0.006)	0.533 (0.115)
VoxelMorph+MI [53]	0.367	0.586	0.785 (0.008)	0.501 (0.026)	0.281 (0.018)	0.561 (0.036)	0.833 (0.005)	0.558 (0.022)
VR-Net [56]	0.237	0.677	0.761 (0.015)	0.333 (0.036)	0.240 (0.027)	0.515 (0.031)	0.812 (0.009)	0.551 (0.058)
CycleGAN [35]	0.139	0.810	0.875 (0.024)	0.624 (0.058)	0.597 (0.059)	0.656 (0.065)	0.909 (0.017)	0.363 (0.052)
MUNIT [37]	**0.071**	**0.959**	0.811 (0.021)	0.522 (0.033)	0.323 (0.032)	0.539 (0.026)	0.867 (0.014)	0.529 (0.021)
DualGAN [64]	0.313	0.597	0.839 (0.014)	0.500 (0.024)	0.390 (0.021)	0.574 (0.018)	0.881 (0.010)	0.394 (0.012)
NAGAN [13]	0.177	0.750	0.955 (0.025)	0.748 (0.032)	0.681 (0.053)	0.771 (0.034)	0.969 (0.021)	0.250 (0.021)
StarGAN v2 [43]	0.244	0.724	0.754 (0.034)	0.312 (0.044)	0.231 (0.044)	0.335 (0.044)	0.823 (0.027)	0.639 (0.050)
SCCGAN [39]	0.354	0.647	0.945 (0.007)	0.749 (0.045)	0.621 (0.048)	**0.847** (0.052)	0.955 (0.008)	0.298 (0.032)
DDIB [46]	0.355	0.598	0.881 (0.052)	0.560 (0.098)	0.569 (0.104)	0.567 (0.090)	0.918 (0.036)	0.527 (0.085)
Ours	0.102	0.874	**0.980** (0.008)	**0.816** (0.007)	**0.763** (0.087)	0.804 (0.072)	**0.985** (0.006)	**0.165** (0.067)

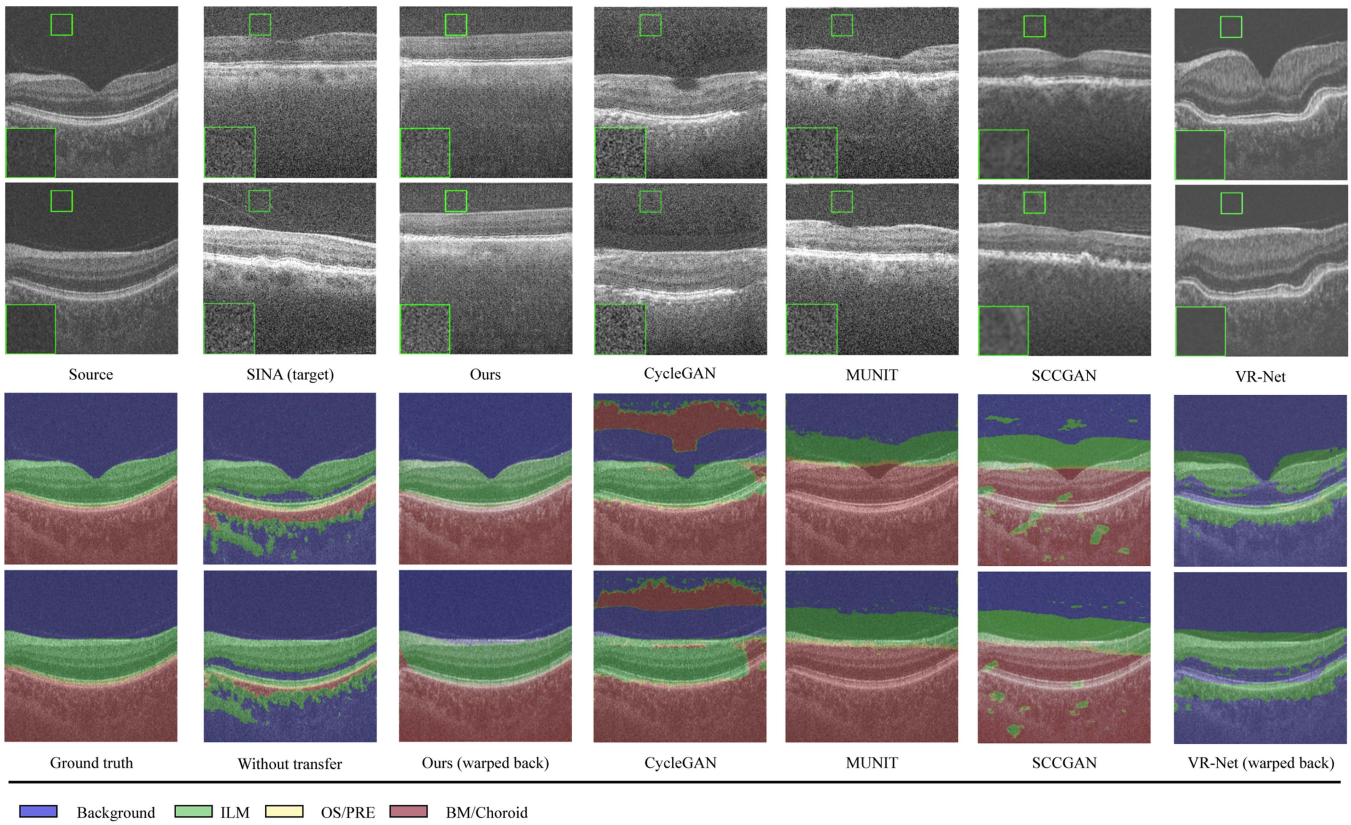


Fig. 9. Results of translated images from ATLANTIS to SINA. The first and second rows show translated images of two input images by different methods. The third and fourth rows show the corresponding segmentation results.

intensity distribution and structure, resulting in a SINA-shaped segmentation input and prediction. Since the prediction is in a different domain than the input, we cannot forcibly apply SINA-to-ATLANTIS of Cycle-GAN and there is no displacement information as we obtained in our method. Therefore, the raw SINA-shaped prediction shows very low performance. This phenomenon is more obvious on MUNIT and StarGAN v2 which transfer the structure even more effectively. Respectively, our method achieves a gain of 0.251 on $mIoU$ and 0.189 on $Dice$ compared to Cycle-GAN, and a gain of 0.525 on $mIoU$ and 0.391 on $Dice$ compared to MUNIT. Similarly to the NAGAN, SCCGAN has the ability to preserve structures through a structure consistency constraint based on squared-loss Mutual Information. There are two reasons why DualGAN does not achieve

desired segmentation performance. First, they cannot differentiate the difference between structure and intensity distribution. Second, they suffer from the domain gap in shape which we tackled using the spatial transformation block. This happens to NAGAN and SCCGAN as well, and our method achieves a gain of 0.358 and 0.067 on $mIoU$, 0.313 and 0.065 on $Dice$ compared to DualGAN and NAGAN respectively. The superior performance of our proposed method over MUNIT stems from a particular drawback in MUNIT. Specifically, MUNIT translates the style by switching the style features in the latent space and doesn't have spatial movement predictions, which struggles to track changes in structure and shape during the translation process. As a result, the altered structure and shape do not align well with the original image, leading to diminished segmentation performance. Similar to MUNIT, DDIB struggles to track changes in structure and shape during the translation process, which decrease the segmentation results by lacking correspondence. In comparison with NAGAN, the SCCGAN is designed to keep the structure and shape unchanged during the translation, which maintains the structure gap to the target dataset. Therefore, it does face challenges in reducing the domain gaps related to the structure. These remaining structural domain gaps subsequently lead to reduced segmentation performance. Moreover, spatial transformation methods, i.e., Voxelmorph and VR-Net, do not achieve acceptable results because they ignore the difference in texture and intensity distribution.

Evaluation on Intensity Distribution Translation: We also evaluate the performance of intensity distribution translation from the SINA to ATLANTIS. Two metrics Bhattacharyya distance D_{Bhat} [66] and correlation $Corr$ [67] are computed to evaluate the performance of intensity distribution transfer referring to [13]. To calculate the two metrics, we calculate the intensity histograms of both transferred images and target images:

$$Corr(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H_1})(H_2(I) - \bar{H_2})}{\sqrt{\sum_I (H_1(I) - \bar{H_1})^2 \sum_I (H_2(I) - \bar{H_2})^2}}, \tag{9}$$

$$D_{Bhat}(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{\bar{H_1}\bar{H_2}N^2}} \sum_I \sqrt{H_1(I)H_2(I)}}, \tag{10}$$

where H_1 and H_2 denote the normalized histograms. I denotes background regions, $\bar{H_k} = \frac{1}{N}\sum_J H_k(J)$ represents the mean value of histogram H_k, $k = 1, 2$, and N denotes the total number of histogram bins.

As shown in Fig. 9 and Table I, most GAN based methods could achieve good results in translating the intensity distributions. However, MUNIT cannot preserve the content information as it tends to change the structure to fit the target distribution, which seriously reduces the segmentation performance. CycleGAN achieves balanced results in shifting the intensities and keeping structural information, however, some unnatural shapes and structures appear. Although spatial transformation methods reduce the shape gap between source and target data, the structure differences still persist to a certain extent and the intensity difference cannot be conquered.

TABLE II
RESULTS OF ABLATION STUDY

Method	$mIoU \uparrow$	$Dice \uparrow$	$FDR \downarrow$
w/o Translation	0.528	0.547	0.381
DiffR	0.367	0.504	0.391
DiffR + u^*	0.316	0.545	0.519
DiffR + GAN	0.305	0.508	0.530
DiffR + GAN + u^*	0.326	0.538	0.525
Ours w/o SSIM & $\boldsymbol{x}_s(\phi)$	0.613	0.702	0.251
Ours w/o $\boldsymbol{x}_s(\phi)$	0.748	0.813	0.202
Ours + Fourier	0.644	0.732	0.259
Ours	**0.763**	**0.816**	**0.165**

3) Ablation Study and Discussion: In order to justify each component of the proposed method, we conduct the following ablation studies. The following methods are compared: (1) The baseline approach is the direct use of the model trained on SINA without any other processing, denoted as w/o Translation. (2) The proposed diffeomorphic spatial transformation method to warp ATLANTIS images to SINA, denoted as DiffR. (3) The proposed DiffR method with clustering map u^*, denoted as DiffR + u^*. (4) A combination of spatial transformation and structure-preserving GAN to the DiffR spatial transformation results, denoted as Transformation + GAN. (5) The combination in (4) with u^*, denoted as Transformation + GAN + u^*. (6) Our prpoposed method without the maximum SSIM mechanism. (7) The intensity translation module of the proposed method is replaced by the Fourier-transform-based method [40]. (8) Our proposed method. The results show that all the components are effective, as shown in Table II.

Here we find out the reason that a straightforward combination of proposed spatial transformation method and the structure-preserving GAN cannot work well. It is because that the texture distortion caused by the spatial transformation is hard to be removed by the structure-preserving translation GAN. Fig. 10 shows an example for visual comparison. As shown, if we only employ a spatial transformation method, we are unable to obtain an accurate deformation. However, there are still many distortions in the translated results, such as the line-shape noise patterns. Such distortions could lower the segmentation performance and can hardly be eased by intensity distribution translation methods. For example, source image warped by DiffR Attention and transferred intensity distributions still remains distorted. Because it is hard to separate these distortions from structure, the network might consider them as structures rather than intensity distribution. Additionally, Fig. 11 shows that the Fourier-transform-based method can not be effectively applied to replace the intensity translation module. Although Fourier-transform-based methods demonstrate notable proficiency in efficiently translating images, these methods are susceptible to introducing distortions in the translated results due to deformations. This susceptibility arises from the sensitivity of high-frequency components to deformations. Particularly, during the

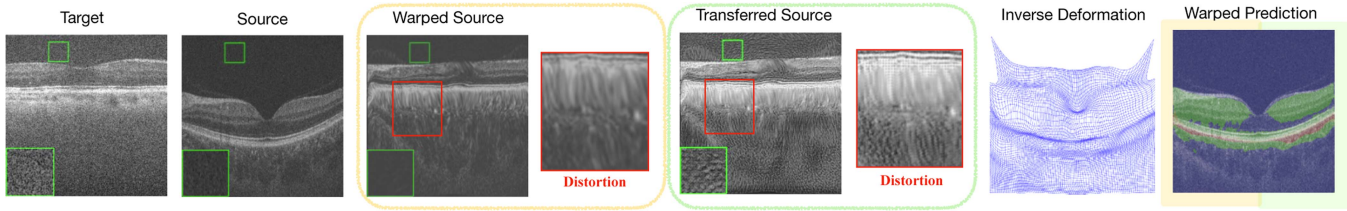


Fig. 10. Illustrations of results and deformations in the ablation study. It shows the reasons (distortions caused by deformations) why a straight forward combination of spatial transformation model and structure-preserving GAN will not work well.

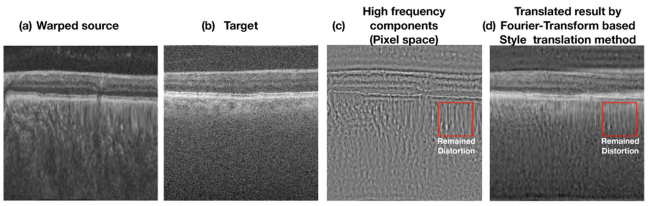


Fig. 11. Illustrating distortion artifacts caused by the Fourier-transform-based style translation. (a) shows a warped source image as input to the intensity translation module; (b) is the target style we want (a) to transfer to; (c) demonstrates high-frequency sensitivity to deformations, whereas (d) showcases the retention of distortions, as highlighted in the red boxes.

intensity translation phase, where inputs comprise deformed source images and basic structure contours, the high-frequency components of these deformed source images often contain distortions resulting from the deformations.

B. Domain Translation From MRI to CT

1) Datasets: To further justify the generalization of the proposed method, we also apply SUA on multi-modal data. In the second set of experiments, we apply it on domain transfer from MRI to CT. Two publicly available datasets[2] [68], [69] with manually labelled segmentation ground truth are used. The first dataset is a MRI dataset that includes three regions in the chest, i.e., cardiac, lung, and liver. The second dataset is a CT dataset, taken from the same patients as the MRI dataset and containing the same organs. We use the MRI dataset as the source, X_S, and the CT dataset as target, X_T. In this experiment, we only use the common areas of 2D images to evaluate the effectiveness of the proposed method for its generalization.

Cross modality translation often faces more comprehensive gaps in structure and intensity distribution. In some situations, the boundaries in the images can be fundamentally different because of the difference in imaging principles. For example, we observe gradually changing brightness from the center to the top/bottom in MRI, whereas, a close to uniform distribution is observed in CT. In addition, shape is more flat in MRI than CT.

2) Comparison With Prior Arts: Similar to retinal OCT experiments, we train the segmentation network on CT and use it to infer transferred MRI images. Similarly, the segmentation outputs are warped back and evaluated against the original MRI ground truth. We compare the proposed method with the same set of methods as in Section IV-A.

Evaluation of Structure Transfer: We first evaluate the structure transfer from MRI to CT. As shown in the Table III and Fig. 12, our proposed SUA method outperforms the state-of-the-art methods on $mIoU$, $Dice$, Acc, Spe and FDR. Structures of cardiac, lung and liver are transferred successfully compared with the VR-Net, Voxelmorph and DualGAN methods. While our inverse deformation can maintain the original shapes compare with diffusion based translation model DDIB and translation GANs such StarGAN v2, MUNIT, NAGAN, SCCGAN and CycleGAN.

Evaluation of Intensity Distribution Transfer: Table III compares the proposed method with other methods in transferring intensity distribution. Fig. 12 gives some examples for visual comparison. As shown in the results, translation GANs could also obtain good results on translating the intensity distributions similar to the OCT experiment. Unsurprisingly, MUNIT achieves the best distribution results but obtains bad results in segmentation. This is reasonable since MUNIT changes the semantic information and leads to misaligned masks. Furthermore, we can see that the translated images by our method not only reach good results in distribution alignment but also capture the structure information by obtaining the inverse deformations. Although spatial transformation methods reduce the shape gap between source and target datasets, the structure difference still remains to a certain extent and the intensity difference cannot be conquered.

C. Translation Between Data From Healthy and Unhealthy Subjects

1) Datasets: Next, we examine how our method performs when translating between data from healthy subjects and that from unhealthy subjects. For this we use MRI images from both the ACDC [70] and the Biobank (UKBB) [71] datasets. For ACDC, it is composed of 100 patients with three types of pathologies: infarction, dilated cardiomyopathy, and hypertrophic cardiomyopathy. For UKBB, we randomly select 100 healthy subjects. The aim is to transfer the style of images in ACDC to that in UKBB, so that the segmentation model trained from healthy patients also works for pathological cases. For each subject in both datasets, we select images at the end-diastolic (ED) frames for experiments.

2) Translation From Unhealthy Subjects Data to Healthy Subjects Data: Similar to the retinal OCT experiments and the

[2]https://learn2reg.grand-challenge.org

TABLE III
DISTRIBUTION AND SEGMENTATION EVALUATION OF CHEST MRI TO CT

Methods	Distribution		Segmentation					
	$D_{Bhat}\downarrow$	$Corr\uparrow$	$Acc\uparrow$	$Dice\uparrow$	$mIoU\uparrow$	$Sen\uparrow$	$Spe\uparrow$	$FDR\downarrow$
Without Translation	0.592	0.043	0.832 (0.003)	0.337 (0.008)	0.253 (0.006)	0.325 (0.012)	0.796 (0.004)	0.635 (0.012)
VoxelMorph [53]	0.371	0.125	0.859 (0.003)	0.506 (0.009)	0.314 (0.005)	0.520 (0.012)	0.834 (0.003)	0.582 (0.013)
VoxelMorph+MI [53]	0.328	0.132	0.827 (0.006)	0.344 (0.016)	0.257 (0.011)	0.348 (0.025)	0.803 (0.007)	0.645 (0.018)
VR-Net [56]	0.346	0.135	0.870 (0.004)	0.541 (0.015)	0.336 (0.012)	0.507 (0.015)	0.849 (0.005)	0.557 (0.017)
CycleGAN [35]	0.300	0.314	0.930 (0.014)	0.683 (0.076)	0.549 (0.073)	0.781 (0.092)	0.938 (0.024)	0.343 (0.046)
MUNIT [37]	**0.286**	**0.507**	0.924 (0.015)	0.651 (0.086)	0.517 (0.080)	0.744 (0.104)	0.929 (0.026)	0.364 (0.055)
DualGAN [64]	0.386	0.102	0.818 (0.004)	0.348 (0.007)	0.260 (0.010)	0.358 (0.009)	0.816 (0.012)	0.635 (0.006)
NAGAN [13]	0.320	0.059	0.841 (0.037)	0.557 (0.066)	0.532 (0.067)	0.608 (0.046)	0.885 (0.027)	0.389 (0.035)
StarGAN v2 [43]	0.329	0.188	0.907 (0.005)	0.552 (0.041)	0.521 (0.036)	0.589 (0.051)	0.876 (0.009)	0.502 (0.034)
SCCGAN [39]	0.321	0.280	0.842 (0.024)	0.592 (0.043)	0.392 (0.041)	0.598 (0.043)	0.892 (0.017)	0.589 (0.035)
DDIB [46]	0.339	0.205	0.889 (0.006)	0.565 (0.027)	0.527 (0.024)	0.682 (0.026)	0.915 (0.009)	0.574 (0.021)
Ours	0.304	0.296	**0.946** (0.012)	**0.750** (0.042)	**0.622** (0.048)	**0.837** (0.038)	**0.959** (0.008)	**0.280** (0.039)

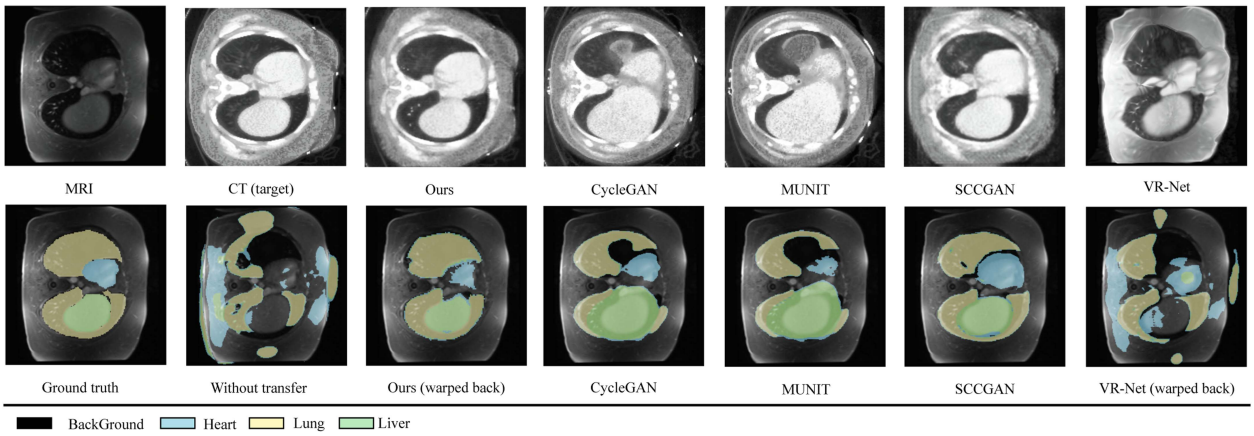


Fig. 12. Results of translated images in the MRI to CT experiment.

MRI-to-CT experiments, we train a segmentation model on normal images in UKBB which is then used to segment the diseased images in ACDC dataset after image translation. We then evaluate the performance by comparing warped segmentation outputs and original ACDC ground truth. Similarly, the proposed method is compared with the same set of methods as in Section IV-A.

Evaluation of Structure Translation: Table IV shows the performance of different models measured by $mIoU$, $Dice$, Acc, Spe and FDR. Compared with the VR-Net, Voxelmorph, and DualGAN methods, our SUA model obtains better results. Additionally, Fig. 13 shows that the transferred structures are on par with that by VoxelMorph, VoxelMorph(MI), and VR-net. Furthermore, the inverse deformation obtained by our method can maintain the shape information of the source dataset, thus outperforming diffusion based translation model DDIB and translation GANs, such as CycleGAN, NAGAN, SCCGAN and StarGAN v2.

Evaluation of Intensity Distribution Transfer: Similarly, we evaluate translation on intensity distributions. Table IV shows the comparison between the proposed method and other methods. Specifically, translation GANs have advantages in the translation of intensity distribution, which has been observed similarly in previous experiments. Not unexpectedly, the same issues occur in MUNIT: though it gives the best distribution results visually, it loses the semantic information of the input source samples. In contrast, the translated images by our method not only produce good results in aligning distributions, but also capture the structure information by using the inverse deformations. Spatial transformation methods, such as VoxelMorph with mutual information loss, reduce the structure gaps to a certain extent, but the intensity difference still could not be decreased.

3) Translation From Healthy Subjects Data to Unhealthy Subjects Data: To further understand the relationship between normal and abnormal image characteristics, we conducted an additional experiment in the reverse direction. This involved translating images representative of a healthy state into the distribution typically associated with disease images. The results, as depicted in the subsequent Table V and Fig. 14, the gaps between normal cases and abnormal cases are reduced by

TABLE IV
DISTRIBUTION AND SEGMENTATION EVALUATION OF CARDIAC IMAGES

Methods	Distribution		Segmentation					
	$D_{Bhat}\downarrow$	$Corr\uparrow$	$Acc\uparrow$	$Dice\uparrow$	$mIoU\uparrow$	$Sen\uparrow$	$Spe\uparrow$	$FDR\downarrow$
Without Translation	0.204	0.528	0.956 (0.012)	0.253 (0.036)	0.237 (0.027)	0.259 (0.027)	0.755 (0.011)	0.599 (0.251)
VoxelMorph [53]	0.153	0.748	0.957 (0.012)	0.298 (0.097)	0.267 (0.067)	0.290 (0.072)	0.767 (0.029)	0.511 (0.281)
VoxelMorph+MI [53]	0.143	0.755	**0.959** (0.012)	0.341 (0.142)	0.299 (0.105)	0.335 (0.128)	0.786 (0.050)	**0.504** (0.213)
VR-Net [56]	0.144	0.776	**0.959** (0.012)	0.335 (0.118)	0.292 (0.084)	0.323 (0.098)	0.781 (0.037)	0.581 (0.185)
CycleGAN [35]	0.109	0.881	0.958 (0.011)	0.589 (0.127)	0.399 (0.097)	0.577 (0.221)	**0.979** (0.010)	0.599 (0.136)
MUNIT [37]	**0.059**	**0.960**	0.946 (0.010)	0.510 (0.067)	0.333 (0.046)	0.508 (0.193)	0.969 (0.005)	0.593 (0.049)
DualGAN [64]	0.259	0.825	0.954 (0.012)	0.275 (0.051)	0.249 (0.034)	0.272 (0.035)	0.761 (0.016)	0.561 (0.181)
NAGAN [13]	0.122	0.902	0.956 (0.012)	0.271 (0.057)	0.248 (0.037)	0.270 (0.038)	0.759 (0.016)	0.538 (0.228)
StarGAN v2 [43]	0.185	0.735	0.944 (0.008)	0.539 (0.066)	0.353 (0.050)	0.521 (0.150)	0.966 (0.006)	0.578 (0.053)
SCCGAN [39]	0.117	0.940	0.950 (0.010)	0.553 (0.096)	0.367 (0.074)	0.605 (0.137)	0.939 (0.045)	0.553 (0.083)
DDIB [46]	0.114	0.940	0.946 (0.012)	0.236 (0.003)	0.223 (0.006)	0.244 (0.004)	0.745 (0.003)	0.739 (0.113)
Ours	0.116	0.948	0.953 (0.025)	**0.541** (0.221)	**0.564** (0.240)	**0.633** (0.209)	0.903 (0.073)	0.556 (0.215)

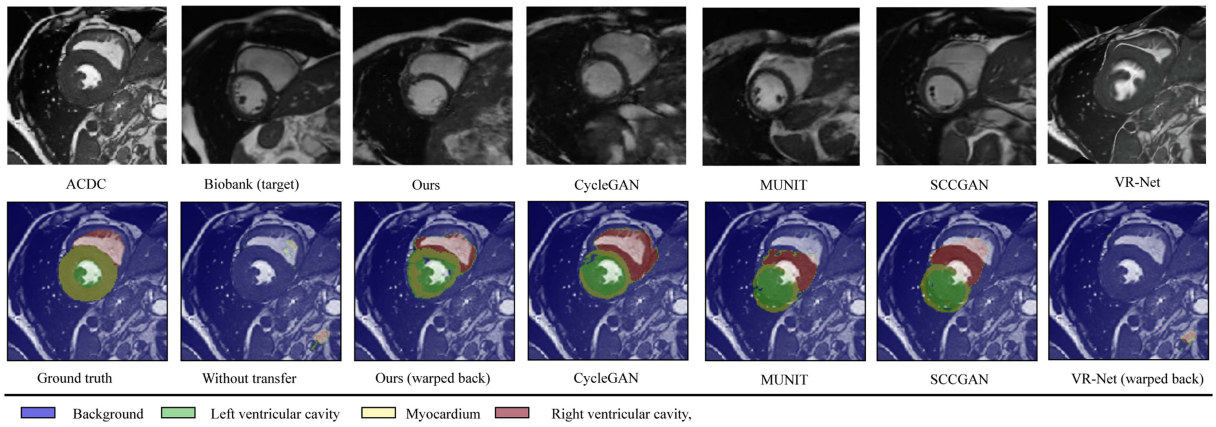


Fig. 13. Results of translated images in the ACDC to UKBB experiment.

TABLE V
TRANSLATION MEASUREMENT ON REVERSE DIRECTION

Method	$D_{Bhat}\downarrow$	$Corr\uparrow$
without Translation	0.204	0.528
Reverse direction (health to disease)	**0.107**	**0.901**

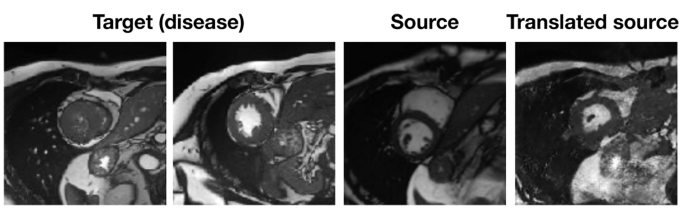


Fig. 14. Health to disease translation.

translation. Specifically, the diseases in these cardiac samples mainly manifest as morphological and structural abnormalities in images, rather than the presence of additional lesions, which makes generating images from healthy to disease not out of reach.

V. DISCUSSION

A. Quantitative Investigation

We have rigorously evaluated the proposed method across six different datasets of varied modalities, extensive experimental results have also proved the superiority of the proposed method in MRI-CT translation and pathology-normal cardiac MRI translation. In these two tasks, large structural gaps can be clearly observed. These results from the six datasets clearly prove that our method is a general approach and can effectively translate images between various domains and reduce the shifts to improve segmentation with large and small structural gaps. As quantification of the extents of structure gap could be important, we employ Mutual Information loss. This metric is adept at assessing structural differences and has been effectively used to measure structural consistency post-translation in SCCGAN [39]. A lower Mutual Information loss between two datasets indicates a greater structural divergence. Our computations reveal that the initial Mutual Information loss between OCT datasets is substantial but is significantly mitigated by our translation method, as evidenced in the translated results shown in Table VI.

TABLE VI
THE MUTUAL INFORMATION (MEDIAN) BETWEEN SOURCE AND TARGET DATASETS (SOURCE & TARGET) AND BETWEEN SOURCE AND TRANSLATED SOURCE DATASETS (SOURCE & REUSLTS)

	Retinal OCT		Chest MRI & CT		Cardiac MRI	
Structual Gap ↓	source & target	source & results	source & target	source & results	source & target	source & results
Mutual Information ↑	0.0317	0.1247	0.3072	0.9956	0.1167	0.2186

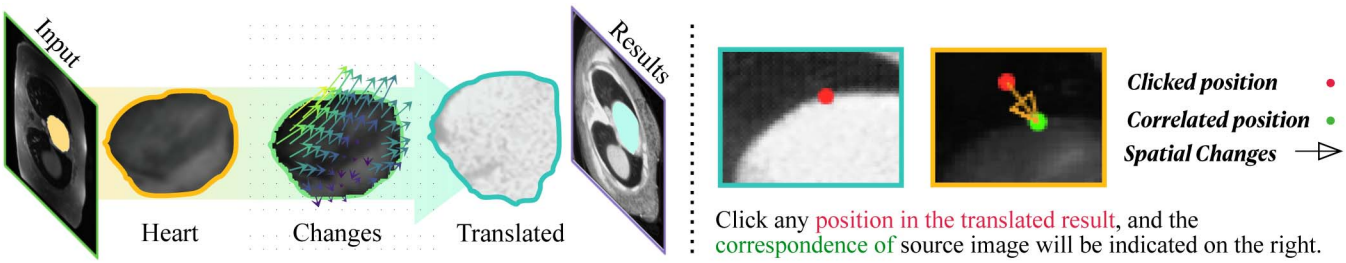


Fig. 15. Illustrating the translation traceability. Left: a semantic figure showing how the heart is deformed and translated from MRI to CT. Right: a screenshot of the online demo given at https://traceable-translation.github.io) showing the spatial connection before and after translation.

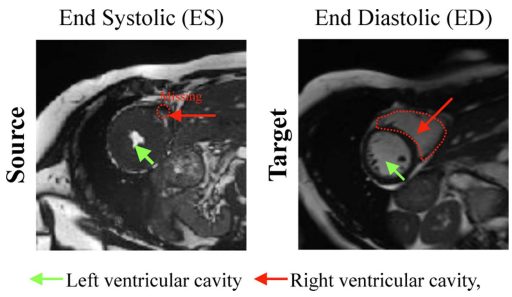


Fig. 16. The first image is one of source images, which is an abnormal image on ES state; the second image is a target image which is normal and is on ED state.

We noticed that large variances appear in the results of the third dataset, *e.g*, Ours and CycleGAN's. This is because this cardiac dataset is a very difficult one and these challenges come from data collected at different phases and translation from pathological to healthy subjects. Specifically, the source images in this dataset are from abnormal subjects at the end systolic (ES) phase, whilst the target images are from normal subjects at the end diastolic (ED) phase. As shown in Fig. 16 left, the left ventricular cavity in the source image is very small, and the right ventricular cavity even disappears. In contrast, they are completely normal in the target image. These issues make it challenging to estimate the deformation between them, which led to large variances in our final results. This high variance problem can be migrated by either using large-sized datasets (like our first OCT dataset) or datasets that have smaller structure gaps (like our second MR/CT dataset). We highlight that although our method produced higher variances in this cardiac dataset, our method still achieved the best overall results compared to other baseline methods. The superiority of our method can be confirmed by its higher means in Table IV.

B. Structure Correspondence

The significance of structure correspondence is often overlooked in existing image-to-image methods. They usually focus on whether the generated image are similar to the target distribution rather than whether the structure of the generated image is related to that of the source image [39]. Furthermore, preserving the structure without addressing the geometry gaps between source and target images is inadequate for fulfilling the translation requirements in medical image tasks. For example, the geometry difference between the imaging in MRI to CT translation would affect the treatment course [72]. It is also mentioned in [73] that establishing correspondences can help to monitor disease progression, estimate motion in radiotherapy planning. These imply that structure correspondence is essential for treatment courses.

Our proposed method possesses an ability to synthesize large quantities of paired data by employing structure unbiased image-to-image translation, while achieving pixel-level structure correspondence. This breakthrough paves a way for the development of an interactive translation method, enabling accurate translation of images with point-to-point correspondences to effectively capture structural changes. To exemplify this advancement, we presented Fig. 15 to show the spatial correspondences between source and translated images. An ***online demo*** was also developed and given at https://traceable-translation.github.io.

VI. CONCLUSION

Image to image translation is an essential task in machine learning, especially for tasks across different modalities. In this paper, we propose to reduce the domain shifts in both structure and intensity distributions. A novel SUA network which contains a structure extractor, a spatial transformation module, and an intensity-rendering module, is proposed in this paper. Our experimental results have shown that SUA is able

to transfer both the structure and intensity distributions and improve segmentation results. Although our method achieves state-of-the-art performance, it has a limitation over the domain gaps introduced by very small objects or lesions as they are often too small to be detected by the dominant structure extractor, which would be future work.

ACKNOWLEDGMENTS

The authors highly acknowledge the anonymous reviewer for his/her contribution to improving the quality of this manuscript. The computations described in this research were performed using the Baskerville Tier 2 HPC service.

REFERENCES

[1] Z. Gu et al., "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[2] C. Chen, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 865–872.

[3] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[4] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.

[5] J. Donahue et al., "DecAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.

[6] L. Duan, I. W. Tsang, and D. Xu, "Domain transfer multiple kernel learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 465–479, Mar. 2012.

[7] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 2208–2217.

[8] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss," in *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, 2018, Art. no. 691.

[9] J. Jiang et al., "PSIGAN: Joint probabilistic segmentation and image distribution matching for unpaired cross-modality adaptation-based MRI segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4071–4084, Dec. 2020.

[10] F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, and S. R. Bulo, "Multidial: Domain alignment layers for (multisource) unsupervised domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4441–4452, Dec. 2021.

[11] M. Gadermayr, L. Gupta, V. Appel, P. Boor, B. M. Klinkhammer, and D. Merhof, "Generative adversarial networks for facilitating stain-independent supervised and unsupervised segmentation: A study on kidney histology," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2293–2302, Oct. 2019.

[12] C. Chen et al., "Unsupervised multi-modal style transfer for cardiac mr segmentation," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2019, pp. 209–219.

[13] T. Zhang et al., "Noise adaptation generative adversarial network for medical image analysis," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1149–1159, Apr. 2020.

[14] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html

[15] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[16] E. L. Denton et al., "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1486–1494.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.

[18] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," in *Proc. NIPS Workshop Adversarial Training*, Barcelona, Spain, 2016. [Online]. Available: https://hal.science/hal-01398049/

[19] H. Kazemi, M. Iranmanesh, A. Dabouei, S. Soleymani, and N. M. Nasrabadi, "Facial attributes guided deep sketch-to-photo synthesis," in *Proc. IEEE Winter Appl. Comput. Vis. Workshops*, 2018, pp. 1–8.

[20] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.

[21] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van denBerg, and I. Išgum, "Deep MR to CT synthesis using unpaired data," in *Proc. Int. Workshop Simul. Synth. Med. Imag.*, 2017, pp. 14–23.

[22] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[23] Y. Wang et al., "3D conditional generative adversarial networks for high-quality pet image estimation at low dose," *Neuroimage*, vol. 174, pp. 550–562, 2018.

[24] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Med. Image Anal.*, vol. 49, pp. 14–26, 2018.

[25] T. Iqbal and H. Ali, "Generative adversarial network for medical images (MI-GAN)," *J. Med. Syst.*, vol. 42, no. 11, 2018, Art. no. 231.

[26] K. Armanious et al., "MedGAN: Medical image translation using GANs," *Comput. Med. Imag. Graph.*, vol. 79, 2020, Art. no. 101684.

[27] K. Armanious, C. Jiang, S. Abdulatif, T. Küstner, S. Gatidis, and B. Yang, "Unsupervised medical image translation using cycle-medGAN," in *Proc. 27th Eur. Signal Process. Conf.*, 2019, pp. 1–5.

[28] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1060–1069.

[29] S. Tripathy, J. Kannala, and E. Rahtu, "Learning image-to-image translation using paired and unpaired training samples," 2018, *arXiv: 1805.03189*.

[30] B. Lei et al., "Skin lesion segmentation via generative adversarial networks with dual discriminators," *Med. Image Anal.*, vol. 64, 2020, Art. no. 101716.

[31] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SeGAN: Adversarial network with multi-scale L 1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, no. 3, pp. 383–392, 2018.

[32] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2017, pp. 408–416.

[33] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI," *Med. Image Anal.*, vol. 31, pp. 77–87, 2016.

[34] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," in *Proc. Int. Conf. Learn. Representations*, 2016. [Online]. Available: https://openreview.net/forum?id=BJtNZAFgg

[35] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.

[36] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1857–1865.

[37] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 172–189.

[38] W. Yan et al., "The domain shift problem of medical image segmentation and vendor-adaptation by UNet-GAN," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2019, pp. 623–631.

[39] J. Guo, J. Li, H. Fu, M. Gong, K. Zhang, and D. Tao, "Alleviating semantics distortion in unsupervised low-level image-to-image translation via structure consistency constraint," in *Proc. Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 18249–18259.

[40] Y. Yang and S. Soatto, "FDA: Fourier domain adaptation for semantic segmentation," in *Proc. Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4085–4095.

[41] M. Cai, H. Zhang, H. Huang, Q. Geng, Y. Li, and G. Huang, "Frequency domain image translation: More photo-realistic, better identity-preserving," in *Proc. Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 13930–13940.

[42] J. Huang et al., "Deep Fourier-based exposure correction network with spatial-frequency interaction," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 163–180.
[43] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN V2: Diverse image synthesis for multiple domains," in *Proc. Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8188–8197.
[44] A. Abu-Srhan, I. Almallahi, M. A. Abushariah, W. Mahafza, and O. S. Al-Kadi, "Paired-unpaired unsupervised attention guided GAN with transfer learning for bidirectional brain MR-CT synthesis," *Comput. Biol. Med.*, vol. 136, 2021, Art. no. 104763.
[45] V. M. Bashyam et al., "Deep generative medical image harmonization for improving cross-site generalization in deep learning predictors," *J. Magn. Reson. Imag.*, vol. 55, no. 3, pp. 908–916, 2022.
[46] X. Su, J. Song, C. Meng, and S. Ermon, "Dual diffusion implicit bridges for image-to-image translation," in *Proc. 11th Int. Conf. Learn. Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=5HLoTvVGDe
[47] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comput. Vis.*, vol. 61, no. 2, pp. 139–157, 2005.
[48] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007.
[49] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.
[50] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.
[51] A. Thorley et al., "Nesterov accelerated ADMM for fast diffeomorphic image registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2021, pp. 150–160.
[52] M. Jaderberg et al., "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.
[53] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2018, pp. 729–738.
[54] T. C. Mok and A. C. Chung, "Fast symmetric diffeomorphic image registration with convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4643–4652.
[55] H. Qiu, C. Qin, A. Schuh, K. Hammernik, and D. Rueckert, "Learning diffeomorphic and modality-invariant registration using b-splines," in *Proc. Med. Imag. Deep Learn.*, 2021, pp. 645–664.
[56] X. Jia et al., "Learning a model-driven variational network for deformable image registration," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 199–212, Jan. 2022.
[57] M. F. Spadea, M. Maspero, P. Zaffino, and J. Seco, "Deep learning based synthetic-CT generation in radiotherapy and pet: A review," *Med. Phys.*, vol. 48, no. 11, pp. 6537–6566, 2021.
[58] J. Uh, T. E. Merchant, C. Hua, Y. Li, and X. Li, "MRI-based treatment planning with pseudo CT generated through atlas registration," *Med. Phys.*, vol. 41, no. 5, 2014, Art. no. *051711*. [Online]. Available: https://www.osti.gov/biblio/22250645
[59] J. Jiang et al., "Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets," *Med. Phys.*, vol. 46, no. 10, pp. 4392–4404, 2019.
[60] J. P. Kieselmann, C. D. Fuller, O. J. Gurney-Champion, and U. Oelfke, "Cross-modality deep learning: Contouring of MRI data from annotated CT data only," *Med. Phys.*, vol. 48, no. 4, pp. 1673–1684, 2021.
[61] M. Storath, A. Weinmann, J. Frikel, and M. Unser, "Joint image reconstruction and segmentation using the potts model," *Inverse Problems*, vol. 31, no. 2, 2015, Art. no. 025003.
[62] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.
[63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[64] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2849–2857.
[65] A. Khoreva, R. Benenson, J. Hosang, M. Hein, and B. Schiele, "Simple does it: Weakly supervised instance and semantic segmentation," in *Proc. Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 876–885.
[66] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.
[67] R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (3rd Edition), Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
[68] A. D. Leow et al., "Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration," *IEEE Trans. Med. Imag.*, vol. 26, no. 6, pp. 822–832, Jun. 2007.
[69] S. Kabus, T. Klinder, K. Murphy, B. van Ginneken, C. Lorenz, and J. P. Pluim, "Evaluation of 4D-CT lung registration," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2009, pp. 747–754.
[70] O. Bernard et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.
[71] S. E. Petersen et al., "Imaging in population science: Cardiovascular magnetic resonance in 100,000 participants of U.K. biobank-rationale, challenges and approaches," *J. Cardiovasc. Magn. Reson.*, vol. 15, no. 1, 2013, Art. no. 46.
[72] E. Palmér et al., "Synthetic computed tomography data allows for accurate absorbed dose calculations in a magnetic resonance imaging only workflow for head and neck radiotherapy," *Phys. Imag. Radiat. Oncol.*, vol. 17, pp. 36–42, 2021.
[73] A. Hering et al., "Learn2Reg: Comprehensive multi-task medical image registration challenge, dataset and evaluation in the ERA of deep learning," *IEEE Trans. Med. Imag.*, vol. 42, no. 3, pp. 697–712, Mar. 2023.

Tianyang Zhang is currently working toward the PhD degree with the School of computer science, University of Birmingham, U.K. Advised by Jinming Duan and Aleš Leonardis. His research interests include, but not limited to, medical image translation, and medical image registration.

Shaoming Zheng is currently working toward the PhD degree with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. His research interests include, but not limited to, medical image analysis and synthesis.

Jun Cheng (Senior Member, IEEE) received the BE degree in electronic engineering and information science from the University of Science and Technology of China, and the PhD degree from Nanyang Technological University, Singapore. He is now a senior research scientist in the Institute for Infocomm Research, Agency for Science, technology and Research, working on AI for medical imaging, robust machine vision and perception. He is an associate editor for *IEEE Transactions on Image Processing* and *IEEE Transactions on Medical Imaging*.

Xi Jia is currently working toward the PhD degree with the University of Birmingham. His research interests include pattern recognition and medical image processing.

Joseph Bartlett is currently working toward the PhD degree with the University of Birmingham and the University of Melbourne. His research focuses on improving diffusion magnetic resonance imaging using machine learning techniques.

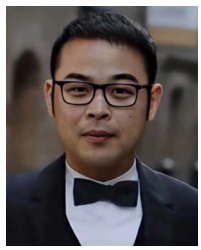
Xinxing Cheng is currently working toward the PhD degre with the School of Computer Science, the University of Birmingham, U.K. His research interests include, but not limited to, medical image analysis and synthesis.

Zhaowen Qiu is the director of China Computer Society (CCF), Outstanding member of CCF, Standing member of CCF computer application Committee, Member of Computing Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE), Artificial Intelligence Committee of China Anti Cancer Association (CACA), Digital diagnosis and treatment Committee of Geriatric Society. Director of Institute of 3D digital technology, Northeast Forestry University. Director of Heilongjiang medical image 3D visualization and 3D printing engineering technology center.

Huazhu Fu (Senior Member, IEEE) received the PhD degree from Tianjin University, in 2013. He is a senior scientist with the Institute of High Performance Computing (IHPC), A*STAR, Singapore. Previously, he was a Research Fellow (2013-2015) at NTU, Singapore, a Research Scientist (2015-2018) at I2R, A*STAR, Singapore, and a Senior Scientist (2018-2021) at Inception Institute of Artificial Intelligence, UAE. His research encompasses computer vision, AI in healthcare, and trustworthy AI. He has received a number of awards including Best Paper Award at ICME 2021, and Best Paper Award at MICCAI-OMIA 2022. He is an associate editor of *IEEE Transactions on Medical Imaging*, *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Artificial Intelligence*, and *Journal of Biomedical and Health Informatics*. He also serves on the Bio Imaging and Signal Processing Technical Committee (BISP TC) of the IEEE Signal Processing Society

Jiang Liu (Senior Member, IEEE) received the MS and PhD degrees in computer science from the National University of Singapore, in 1992 and 2004, respectively. He is a full professor in the Department of Computer Science and Engineering with the Southern University of Science and Technology. His main research interests include medical image processing and artificial intelligence.

Aleš Leonardis (Member, IEEE) is a professor with the School of Computer Science, University of Birmingham, the Chair of robotics with the University of Birmingham, and the co-director of the Centre for Computational Neuroscience and Cognitive Robotics. He is also a professor with the Faculty of Computer and Information Science, University of Ljubljana and adjunct professor with the Faculty of Computer Science, TU-Graz. His research interests include robust and adaptive methods for computer vision, object and scene recognition and categorization, and biologically motivated vision.



Jinming Duan received the PhD degree from the University of Nottingham. He is currently an associate professor within Division of Informatics, Imaging & Data Sciences, University of Manchester, a honorary associate professor with the University of Birmingham, a turing fellow with Alan Turing Institute, and a visiting researcher with Imperial College London. He is a fellow of the Higher Education Academy (FHEA) under the U.K. Professional Standards Framework for teaching and learning support in higher education. He is also an associate editor of *IEEE Transactions on Neural Networks and Learning Systems*. From 2017 to 2019, he was a research associate jointly within Department of Computing and Institute of Clinical Sciences at Imperial College London. His PhD was funded by Engineering and Physical Sciences Research Council (EPSRC). In 2016, he won "Chinese Government Award for Outstanding Self-financed Students Abroad" issued by the ambassador of Chinese Ministry of Education. Between 2019 and 2023, he and his team have developed cutting-edge machine learning algorithms that won multiple international awards, including "fastMRI", "Learn2Reg", and "RnR-ExM". He has co-authored more than 140 peer-reviewed papers.