IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 20, 2020, accepted June 3, 2020, date of publication June 8, 2020, date of current version June 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3000666

Interactive Translation in Echocardiography Training System With Enhanced Cycle-GAN

LONG TENG[1,2], **ZHONGLIANG FU**[1], **AND YU YAO**[1]
[1]Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610041, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Long Teng (long_teng@ieee.org)

This work was supported by the Sichuan Province's New Generation of Artificial Intelligence Major Special Project under Grant 2018GZDZX0036.

ABSTRACT Interactive translation in echocardiography training system refers to the pixel-wise translation between ultrasound cardiac and theoretical sketch images in the course of hand-on operation. It is capable of efficiently gaining more insights into clinical ultrasound anatomy. However, major studies on the synthesis of ultrasound cardiac image primarily discuss the physical model simulation, while studies on cardiac image segmentation place an emphasis on image processing. Thus, they cannot be easily integrated into one pipeline for interactive translation. This paper presents an enhanced Cycle-GAN for interactive translation. Perceptual loss is introduced to enhance the quality of synthetic ultrasound texture, while Cycle-GAN translates between two modalities. The proposed method is trained on 300 pair images and tested on 68 pair images. As revealed from the experiment results, the proposed method is feasible in interactive translation, and it is superior over Cycle-GAN for ultrasound image synthesis.

INDEX TERMS Echocardiography, generative adversarial network (GAN), interactive translation, computer vision, medical image analysis.

I. INTRODUCTION

The echocardiography training system refers to a computer-based novel training machine demonstrating ultrasound cardiac features interactively. In the past two decades, it has extensively acted as an efficient method to enhance professional skills in clinical assessment [1]. Echocardiography training system can help the trainee master left ventricular structure and function, right ventricular structure and function, valve function, etc. [2]. For this reason, numerous corresponding training courses have also been formulated to delve into the significant enhancement of trainee's clinical capability in different hands-on operations [2], [3]. A simple introduction of the training system here and corresponding courses is presented through the link: http://www.iechoonline.com/product/tee/.

The interactive translation acts as the underlying function of the echocardiography training system, illustrating different images during hand-on operation. Theoretically, it covers two types of translation, namely, ultrasound image to sketch image(U2S) and sketch image to ultrasound image(S2U).

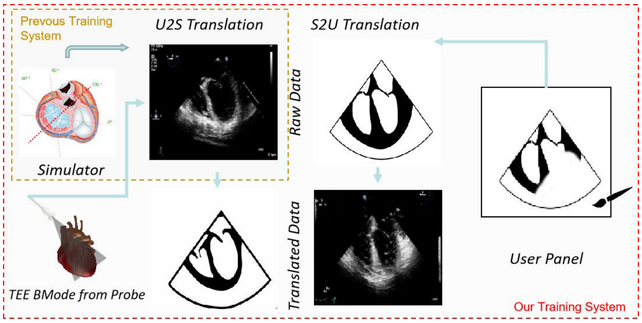

FIGURE 1. Example of interactive translation: U2S translates ultrasound images into sketch images. This procedure helps the trainee to comprehend cardiac structure. S2U translates sketch images into ultrasound images. It helps to reinforce the understanding of ultrasound characteristics.

However, existing studies on U2S and S2U are different in principle. Previous S2U is determined by either built-in data or built-in simulation model [4]. Existing U2S relies on image segmentation (e.g., level-set [5] and active contour [6]). Such fundamental difference prevents those two modalities from interactive translation, and even makes it difficult for real-time interaction.

The associate editor coordinating the review of this manuscript and approving it for publication was Kin Fong Lei.

Previous S2U consists of three important methods, namely, interpolative method, generative image-based method, and generative model-based methods, each of which has its advantage and drawback.

Interpolative method aims to simulate 2-dimensional (2D) ultrasound images by interpolation from build-in 3-dimensional (3D) data [7], [8]. Since it is an interpolation from real data, the ultrasound texture can be highly realistic, and the sketch image can be build-in for pixel-wise correspondence as well. Note that if the user changes views, no view-dependent effect will be simulated. Moreover, preparing the build-in corresponding sketch data is also time-consuming.

The generative image-based method is the simulation from other images (e.g., CT and MR). It simulates wave propagation on 3D volumes of CT or MR [9]. Bürger *et al.* manually assigned ultrasound tissue texture to segmented CT and MR images [10]. Besides, Zhu *et al.* manually assigned labels to segmented CT voxels and then textured the 2D slice exhibiting real ultrasound texture [11]. The mentioned methods solve the view-dependent simulation, whereas they are time-consuming for data preparation and propagation calculation.

The generative model-based method aims to simulate the ultrasound images according to a physical anatomy model. Sun and McKenzie built a heart model, sliced a 2D image from the heart model, and then assigned the 2D image exhibiting ultrasound texture [12]. Köhn *et al.* developed a mathematical heart model based on an MR image to calculate the ultrasound cardiac images [13]. The mentioned methods provide more details of cardiac motion. However, their model requires more verification experiments, and they cannot build a connection to U2S as well.

Previous U2S primarily focuses on segmentation of specific region of interest (ROI): endocardial segmentation, myocardium segmentation, and valve segmentation [14].

The endocardial border exhibits the property that contrasts around the left ventricle (LV) chamber changes under the relative orientation between the border and the transducer direction. Thus, intensity gradient-based methods exhibit limited performance on endocardial segmentation. Accordingly, alternative methods of active shape model (ASM), active contours model (ACM), and level set are adopted for such segmentation. Nikos Paragios *et al.* employed ASM for LV segmentation [15]. They considered the time-consistent ASM to achieve precise segmentation on fifty patients. G Hamarneh and T Gustavsson built an active shape model (ASM) and ACM based method to achieve LV endocardial segmentation [16]. Their method exploits ACM to achieve smooth and connected boundaries while employing ASM to achieve shapes similar to the given training set. Ning Lin *et al.* developed a multi-scale level-set framework for endocardial segmentation [17]. They assumed that Gaussian is capable of approximately modeling the intensity distribution of an ultrasound image at a certain coarse scale. Subsequently, they adopted region homogeneity and edge features in a level set approach to extract boundaries at this coarse scale. The mentioned methods highly apply to endocardial segmentation, whereas they may not able to directly transfer to other ROI segmentations.

In myocardium segmentation, the epicardial features are represented by low-intensity differences rather than endocardial features for the acoustic density difference between tissue to tissue and blood to tissue [18]. Boukerroui *et al.* employed image enhancement to down-regulate the effect of attenuation and enhance features before segmentation [19]. Vivek Walimbe *et al.* introduced a deformable model for myocardium segmentation [18]. Sarah Leclerc *et al.* established an open access large-scale 2D dataset to delve into this segmentation with deep learning-based methods [20].

It is challenging to segment the valve since it is relatively small with rapid deformation. Ivana Mikic *et al.* presented a segmentation with additional information on optical flow [21]. ML Siqueira *et al.* attempted to segment all the three ROIs together by k-means based algorithm. However, they achieved limited performance for either of the regions [22].

This paper proposes an enhanced Cycle-GAN for interactive translation in the echocardiography training system. For U2S, myocardium and valve segmentations are simultaneously achieved, and the sector boundary is derived. For S2U, the sketch image undergoes pixel-wise translation into an ultrasound image, and additional constraint is given to ensure the ultrasound texture fidelity. The major contributions of this research include,

- Cycle-GAN is adopted to fuse ultrasound to sketch (U2S) and sketch to ultrasound (S2U) together for interactive translation in one pipeline.
- The Cycle-GAN method is enhanced for S2U by introducing perceptual loss besides Cycle-GAN loss.
- U2S is adopted to achieve myocardium segmentation, valve segmentation, and derive the sector boundary simultaneously. No existing studies have delved into this task.

The rest sections here are organized as follows. Section II gives a brief overview of related works. Section III presents our proposed method. Section IV presents the results of the proposed method and some discussion toward the results. Section V discusses how to apply the proposed method to interactive translation. Section VI draws the conclusion of the proposed method and further work in the further.

II. RELATED WORK

As suggested in the introduction, the GAN framework is adopted as a guideline to achieve echocardiography interactive translation. Though rare works have been published on echocardiography interactive translation, there are some related works on GAN based echocardiography enhancement, as well as the translation of GAN based cardiac cross-modalities. Those works are related to ours as part of the GAN based echocardiography studies.

A. GAN BASED ECHOCARDIOGRAPHY ENHANCEMENT

The quality of acquired ultrasound images varies noticeably depending on the ultrasound equipment and the operator, while ultrasound image quality significantly affects the diagnosis. In this regard, Zhibin Liao *et al.* proposed a quality transfer StarGAN [23]. They prepared four groups of echocardiography data. A different group of data exhibits different image quality, ranging from poor to excellent. Subsequently, the StarGAN framework is adopted to transfer the quality level between different data groups. After training, a given ultrasound image can be transferred to an arbitrary image quality level. However, the collection of different quality level data is delicate and affects the final result profoundly.

Likewise, Deepak Mishra *et al.* proposed an ultrasound image enhancement GAN [24]. It can input low-resolution ultrasound images and output high-resolution ultrasound images. Besides, adversarial loss and structural loss are combined to train the proposed network and achieve state-of-the-art performance over existing methods. However, their image translation is irreversible. The high-resolution images cannot be translated into low-resolution images, as well as other modalities.

Mohammad H. Jafari *et al.* proposed a GAN framework to translate low quality ultrasound images into high quality by introducing structure regularization to CycleGAN [25]. Subsequently, the translated high-quality ultrasound image is adopted for segmentation. Their experiment suggests the enhancement of segmentation accuracy using ultrasound image quality.

AH Abdi *et al.* presented a GAN based generation model to generate high-quality ultrasound images [26]. Their network inputs an ultrasound image, and a segmentation mask outputs a novel ultrasound image complying with the structure of the given segmentation mask. In the training course, the input ultrasound image acts as a fixed frame. Thus, all of their generated images are similar to each other. In other word, the ultrasound texture is generated from a fixed ultrasound image.

B. GAN BASED CARDIAC CROSS-MODALITIES TRANSLATE

Cardiac radiological images are acquired by different imaging modalities, covering ultrasound imaging, computed tomography (CT), as well as magnetic resonance imaging (MRI). Those modalities exhibit significantly different data distribution. In such case, the sketch image can also be considered a particular modality acquired manually. Recently, some cardiac cross-modality researches have been explored.

Qi Dou *et al.* proposed an unsupervised cross-modalities translation between CT images and MRI images [27]. Via the translation from MRI images to CT images, the segmentation label of MRI is also translated into that of CT images. Thus, given MRI segmentation label can be exploited for CT images. Their work presents a novel cross-modality image segmentation method.

Ziqi Zhou *et al.* employed the cross-modality correlated information to enhance segmentation [28]. First, GAN is adopted to translate MRI images into CT images and CT images into MRI images. Subsequently, an attention-based auto-encoder is applied for segmenting both modalities. The final segmentation result of each modality outperforms that of baseline segmentation methods.

Libao Guo *et al.* developed a dual network GAN for pediatric echocardiography segmentation [29]. A fully connected network and a U-Net [30] are combined in the generator for chamber segmentation. The experiment is performed on four-chamber view echocardiography. Their segmentation reduces the sonographer's work intensity in manual segmentation and enhances the reliability of segmentation. All of the mentioned works solve a specific issue of echocardiography translation, whereas they remain far from a useful interactive translation for echocardiography training.

Meanwhile, the related cross-modality image generation and reconstruction methods also provides ways to solve the image translation issue. Zhen Zhu *et al.* proposed a new generative adversarial network for pose image transfer on condition of target pose key points [31]. Hongfeng You *et al.* proposed a feature extraction method that to reflect the contextual relationships between the pixels [32]. Weiwei Cai *et al.* proposed a GAN base network to recover the realistic image content [33]. They are consistent with the starting point of this article, all to find comprehensive pixel-level correspondence between modalities.

III. METHODS

In the present section, the enhanced Cycle-GAN is introduced for echocardiography interactive translation. First, the perceptual loss is explained for S2U. Subsequently, the adversarial and overall loss function is given for training. Lastly, we illustrate the network structure detail for U2S and S2U in this paper.

A. PERCEPTUAL LOSS

Theoretically, the ultrasound texture is acquired by convolving point-like scatterers in the tissue with the ultrasonic impulse response, which is termed as the point-spread function (PSF) [34]. Subsequently, the PSF is determined by factors (e.g., probe aperture and ultrasound frequency). It is therefore suggested that the statistical intensity of ultrasound texture can be only approximated, rather than an accurate reconstruction. Given this, either the L1 or L2 loss of Cycle-GAN is not sufficient for S2U. We need to seek for a visually realistic constraint for S2U.

Existing super-resolution researches have delved into the semantic feature-based loss function for maintaining image details [35]. In the S2U scenario here, the VGG16 based semantic feature is explored for ultrasound texture synthesis. As shown in figure 2, the overall network adopts a VGG16 structure [36]. The pre-trained VGG16 is loaded, and the middle channels of the 8th, 15th, 22th, 28th and 31st layers are presented. The low-level feature contains

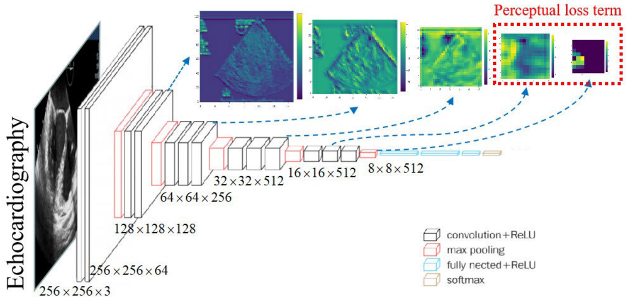

FIGURE 2. Perceptual loss: VGG16 as a guideline to explore the visually realistic constraint for S2U. As is shown, low-level feature trends to represent structural information, while high-level feature trends to extract semantic information.

structural information. Thus, the echocardiography contour in the 8th, 15th, 22nd layer's feature map can be recognized. As the network goes deeper, the feature map tends to represent semantic information. The 28th and 31st layers' feature maps contain more semantic information, so the echocardiography contour cannot be recognized in those high-level feature maps. Hence, this paper uses the 28th, 31st layer's feature map as the perceptual loss term. The formula is expressed as follows,

$$\mathcal{L}_{perceptual} = \|\phi(I_u) - \phi(G_u(I_s)) + \|\phi(I_u) - \phi(G_u(G_s(I_u))) \tag{1}$$

In Eq. 1, ϕ represents the 28th and 31st layer's output of pre-trained VGG16. I_u denotes the ground truth ultrasound image. I_s is the input sketch image. G_u, G_s are the S2U and U2S networks, as introduced in the subsections below. The perceptual loss function $\mathcal{L}_{perceptual}$ is to minimize the semantic difference between translated ultrasound image and ground truth image and then make the translated ultrasound image more realistic.

B. FULL OBJECTIVE

The enhanced Cycle-GAN here reinforces the S2U translation with perceptual loss, suggesting that the U2S and S2U translations are trained under the Cycle-GAN framework. Each translation is constraint with GAN loss,

$$\begin{aligned}\mathcal{L}_{GAN}^{u2s} &= \mathbb{E}_{I_s\sim p_{data}(I_s)}[logD_s(I_s)] \\ &\quad + \mathbb{E}_{I_u\sim p_{data}(I_u)}[log(1 - D_s(G_s(I_u)))] \end{aligned}\tag{2}$$

$$\begin{aligned}\mathcal{L}_{GAN}^{s2u} &= \mathbb{E}_{I_u\sim p_{data}(I_u)}[logD_u(I_u)] \\ &\quad + \mathbb{E}_{I_s\sim p_{data}(I_s)}[log(1 - D_u(G_u(I_s)))] \end{aligned}\tag{3}$$

where the D_s and D_u are the discriminators of U2S and S2U networks. The Eq. 5 forces the translated sketch image to satisfy the real sketch image's distribution. Moreover, Eq. 3 forces the translated ultrasound image to comply with a real ultrasound image's distribution. The complementary cycle consistency loss is as follows,

$$\begin{aligned}\mathcal{L}_{cyc} &= \mathbb{E}_{I_u\sim p_{data}(I_u)}\{\|G_u(G_s(I_u)) - I_u\|_1\} \\ &\quad + \mathbb{E}_{I_s\sim p_{data}(I_s)}\{\|G_s(G_u(I_s)) - I_s\|_1\} \end{aligned}\tag{4}$$

Eq. 4 is the constraint that the reconstructed ultrasound and sketch images are pixel-wise the same as ground truth. For U2S, this refers to the segmentation accurate between reconstructed sketch and input sketch. For S2U, this presents a supplement of similarity between reconstructed ultrasound image and an input ultrasound image. Hence, the overall objective loss of enhanced Cycle-GAN is expressed as:

$$\mathcal{L} = \mathcal{L}_{GAN}^{u2s} + \mathcal{L}_{GAN}^{s2u} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{per}\mathcal{L}_{perceptual} \tag{5}$$

C. NETWORK ARCHITECTURES

Figure 3 illustrates the framework of our enhance Cycle-GAN. It covers two parts. Part 1 indicates the overall framework and loss function. Part 2 is the network details in part 1.

The U2S and S2U translation exploit the UNet architecture, as shown at the top part of part 2. It inputs a three-channel image and outputs a three-channel image. The input image size reaches 256 × 256. The first 8 gray blocks are convolution blocks. Each block covers a convolution layer, batch normalization layer, as well as a relu layer. The convolution layer's kernel size reaches 4, and the stride is 2. Thus, the bottleneck feature is 1/256 times of the input size.

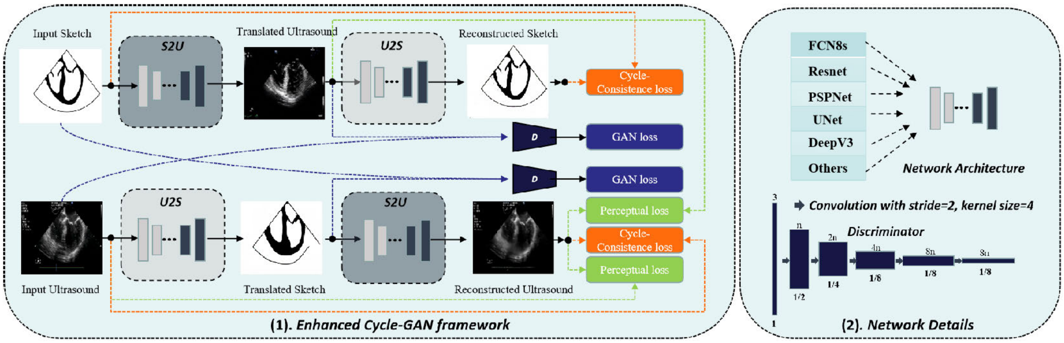

FIGURE 3. Overview of our proposed Networks: Part 1 is the overall framework, Part 2 is the detail of individual network blocks. The blue dash line, dash orange line, and dash green line respectively represents the loss function of GAN, Cycle-consistence, and perceptual constraint. The U2S and S2U blocks in part 1 use 6 kinds of network architecture, which shows in part 2. The discriminator block, which is marked as D, is also shown with detail layers in part 2.

The channel size of each layer's output is illustrated on the block as well.

The last eight layers of UNet architecture are the de-convolution layer blocks. Each block covers a de-convolution layer, a batch normalization layer, as well as a relu layer. The de-convolution layer's kernel size is 4, and the stride is 2. In each of the blocks, the input data is first concatenated with a corresponding output from the first eight layers. Subsequently, a convolution layer is applied to down-sample its channel into half size. Lastly, a de-convolution block up-sampling the image size while down-sampling its channel size. The up-sampling and down-sampling factors are all 2.

The discriminator network shows a cascade structure of convolution blocks. The 2nd to the fourth block consists of a convolution layer, a batch normalization layer, as well as a relu layer. Kernel size is 4 × 4 in each convolution layer. The detailed architecture of discriminator is listed in table 1.

TABLE 1. The detail architecture of discriminator.

Discriminator Network D	
Input: Image I	
[Layer1]:	Conv2d: (4, 4, 64), stride=2, padding=1; $ReLU$;
[Layer2]:	Conv2d: (4, 4, 128), stride=2, padding=1; Batchnorm; $ReLU$;
[Layer3]:	Conv2d: (4, 4, 256), stride=2, padding=1; Batchnorm; $ReLU$;
[Layer4]:	Conv2d: (4, 4, 512), stride=1, padding=1; Batchnorm; $ReLU$;
[Layer5]:	Conv2d: (4, 4, 1), stride=1, padding=1; $Sigmoid$;
Output: Real or Fake (Probability)	

IV. EXPERIMENTS

This section presents the experiment result of the proposed method. We first draw the comparison between ground truth, the result here, and Cycle-GAN' result in S2U translation. Subsequently, we analyze the U2S performance of the proposed method. The proposed method is achieved with Pytorch platform. All experiments are performed with GTX1080 GPU.

A. DATASET

1) OUR DATASET

Our dataset is collected in the hospital as guided by doctors. The annotations are made by the teamwork of doctors and art teachers. The dataset consists of exams from 100 patients. It contains 736 (368 pairs) B-mode transesophageal echocardiography (TEE). 600 images (300 pairs) are employed for training, while the left 136 images (68 pairs) are adopted for testing. The raw image height and width are 600 × 800 pixels. During the training and testing processes, they are random cropped and resized into 256 × 256 pixels.

2) CAMUS DATASET

The CAMUS dataset is an open access echocardiography dataset which consists of clinical exams from 500 patients [20]. The training data contains 450 patients, 1800 images. The testing data contains 50 patients, 200 images. The dataset involves a wide variability of acquisition settings. During the training and testing processes, they are random cropped and resized into 256 × 256 pixels. To verify our proposed interactive translation, the annotation of myocardium is revised into sketch image with sector boundaries.

3) DATA AUGMENTATION

The data augmentation is employed before training. The data augmentation covers random crops, resize and random flip operation in the horizon and vertical directions, thereby making our model robust for image scales and local details. The data is first resized into 280 × 280, and then is random cropped into 256 × 256 pixels. Taking into account the factor of random flip. It provides 24 × 24 × 4 = 2304 times data scale than the original dataset.

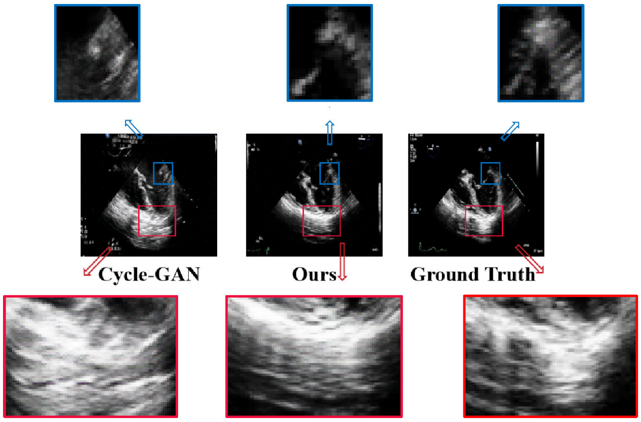

FIGURE 4. Amplified S2U: The second row illustrates the S2U result of Cycle-GAN, S2U result of ours(the baseline network is UNet), and the ground truth. The blue block represents the valve part of the ultrasound image. The red block marks a region of the myocardium. The blue block marked region is amplified in the first row, while the red block marked region is amplified in the third row. Our reconstructed texture is more realistically approximate to the ground truth.

B. SKETCH TO ULTRASOUND TRANSLATION

The experiment here is performed on the introduced dataset. For S2U translation, the S2U network inputs the sketch image and outputs the translated ultrasound images. Figure 4 illustrates the qualitative comparison between our result and the advanced method. Representatively, two ROIs are amplified in the cardiac valve and myocardium.

In the region of the cardiac valve, more texture information is reconstructed with the proposed method. The benefit of the texture information results in the translation of different tiny structures. The Cycle-GAN based translation ignores the cardiac valve, while the proposed method maintains the valve.

In the region of the myocardium, the speckle from Cycle-GAN trends to blur, while the result here maintains more realistic textures. Figure 5 presents two groups of S2U translation. The first column is the input sketch, and the

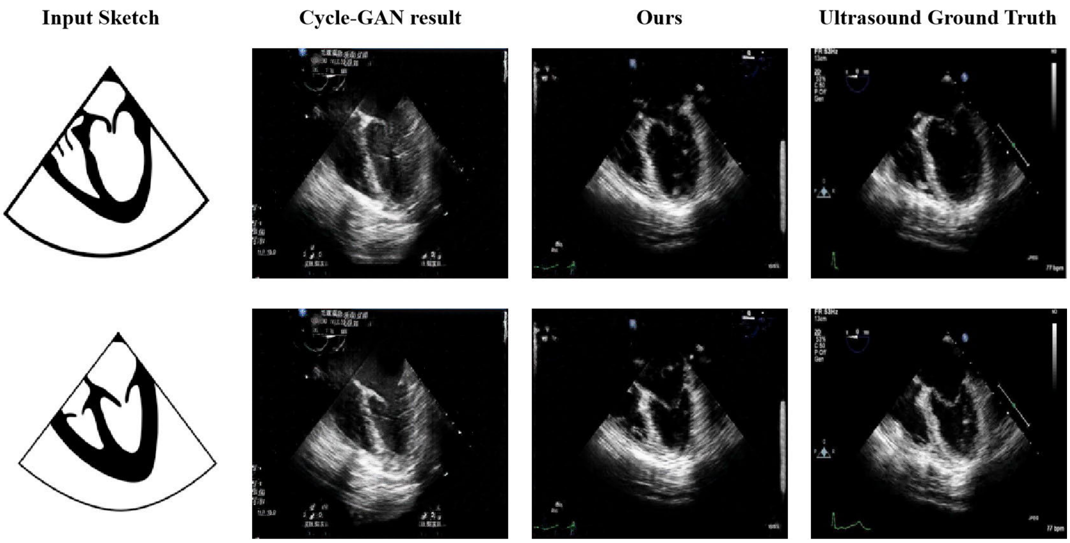

FIGURE 5. Qualitative results of S2U(the baseline network is UNet): The first column is the input sketch image. The second to fourth columns represent translated ultrasound images from Cycle-GAN, our method, and the ground truth. Our result visually more close to the ground truth.

rest columns represent the results and ground truth. Since our perceptual loss enriches the texture information of S2U, the translated myocardium is visually more realistic.

For quantitative analysis of translated ultrasound image quality, the peak signal to noise ratio (PSNR) and structural similarity index (SSIM) index are adopted for translated images. The PSNR acts as a quality measurement between the translated and ground truth images. The higher the PSNR, the better the quality of the translated image will be. The SSIM can measure how similar the translated image and ground truth are. A higher SSIM indicates the translated image is more similar to the ground truth.

Table 2 draws a quantitative comparison between our result and the baseline method's result. In the testing dataset, the result here exhibits better performance on both PSNR and SSIM, and these qualitative and quantitative results indicate consistent performance. It is therefore verified that the proposed perceptual loss enhances the performance of S2U.

TABLE 2. S2U translation evaluation with index of PSNR and SSIM.

Our Dataset				CAMUS Dataset	
S2U		PSNR	SSIM	PSNR	SSIM
FCN8s	CycleGAN	14.71	0.47	15.59	0.47
	Ours	**14.92**	**0.49**	**16.12**	**0.49**
ResNet	CycleGAN	14.25	0.44	15.32	0.47
	Ours	**14.44**	**0.46**	**15.64**	**0.48**
PSPNet	CycleGAN	14.04	0.43	14.32	0.38
	Ours	**14.41**	**0.46**	**14.66**	**0.39**
UNet	CycleGAN	13.72	0.41	14.24	0.38
	Ours	**14.38**	**0.45**	**14.50**	**0.38**
DeepV3	CycleGAN	13.21	0.37	13.42	0.36
	Ours	**13.92**	**0.39**	**13.74**	**0.36**

The U2S is vital for the trainee to master cardiac structure. The experiment here is performed on the testing data. One of the results is illustrated in figure 6. The first image is an input ultrasound image, and the second image is our result based on the comparison of ground truth (the red color overlay represents the ground truth), and the third image is the ground truth.

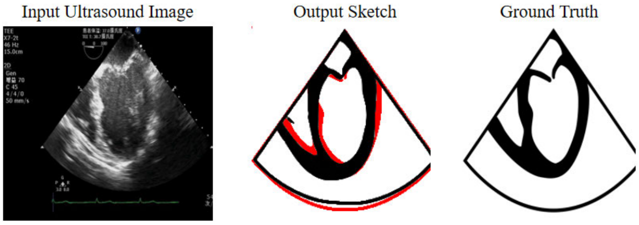

FIGURE 6. U2S translation: The middle image is our result. The red background represents the ground truth of the third image.

C. ULTRASOUND TO SKETCH TRANSLATION

The contour of the result here approximates to the ground truth. For the sector boundary, the raw data is blurry with a low contrast ratio, since there exists no myocardium close to the boundary. Compared with the ground truth, our inference is contracted to an inner place. Nevertheless, it remains an identifiable shape.

For the cardiac regions, the chamber achieves similar segmentation to the ground truth. However, the myocardium and cardiac valve region remain not as accurate as truth ground truth. Since the high cost and time-consuming property of ground truth labels, this remains a future research topic.

The segmentation index of the Dice coefficient and Intersection over Union (IOU) is adopted to delve into the overall translation accuracy. Table 3 lists the statistical results of U2S. Dice is a statistic adopted to gauge the similarity of

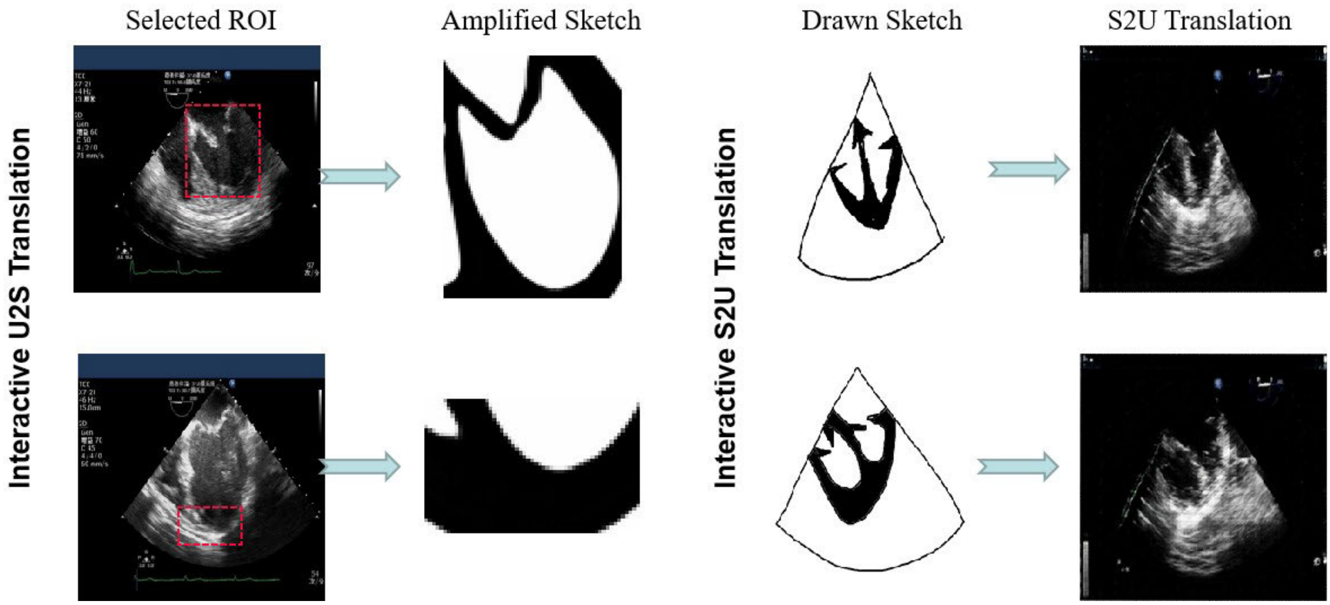

FIGURE 7. Interactive translation in echocardiography training system: The first two columns illustrate interactive U2S operation, and the last two columns represent the interactive S2U translation.

TABLE 3. U2S translation evaluation with index of DICE and IOU.

		Our Dataset			CAMUS Dataset		
U2S		Average	Minimal	Maximal	Average	Minimal	Maximal
FCN	DICE	0.947	0.931	0.959	0.935	0.886	0.958
	IOU	0.900	0.871	0.921	0.878	0.796	0.920
ResNet	DICE	0.944	0.928	0.960	0.896	0.865	0.926
	IOU	0.894	0.866	0.922	0.812	0.762	0.863
PSPNet	DICE	0.937	0.912	0.960	0.923	0.894	0.953
	IOU	0.882	0.938	0.924	0.858	0.809	0.911
UNet	DICE	0.927	0.905	0.944	0.935	0.872	0.965
	IOU	0.864	0.827	0.894	0.879	0.772	0.933
DeepV3	DICE	0.876	0.847	0.912	0.927	0.890	0.955
	IOU	0.780	0.734	0.839	0.864	0.802	0.914

two samples, while IOU gives the ratio between the intersection and union of two sets. The proposed method presents acceptable performance results on both Dice and IOU index.

D. ABLATION STUDIES OF BASELINE NETWORKS

To investigate the advantage and limitation of proposed method on echocardiography translation application, we compare the S2U translation results with different baseline networks.

As is shown in figure 8 and figure 9. We choose 5 different baseline networks, FCN8s [37], Resnet(9 blocks) [38], PSPNet [39], UNet [30], and DeepV3 [40]. The S2U translation performance varies with different baseline networks. Statistically, the FCN8s baseline network achieves the best result, the DeepV3 baseline network has the worst performance. The ultrasound texture is also statistically getting more realistic with an appropriate baseline network and perceptual loss function.

In figure 8 and figure 9, the synthetic ultrasound texture is not exactly the same as the real texture. It is because the synthetic ultrasound texture is statistically getting from the training dataset. This means, if we want to synthetic ultrasound texture with cardiac disease, the training data should contain corresponding samples.

V. DISCUSSION

In this paper, the enhanced Cycle-GAN method is discussed for interactive translation between ultrasound image and sketch image, whereas no application in the echocardiography training system is mentioned above. In the present section, how we apply the proposed method to interactive translation is briefly discussed in the following.

Currently, we have developed two functions based on the proposed method. The figure 7 illustrates the interactive U2S and S2U translations. The first two columns represent the U2S translation during hands-on operation, and the last two columns denote the S2U translation for offline training.

During hands-on operation, automatic U2S can help trainees accelerate their comprehension. Moreover, one can

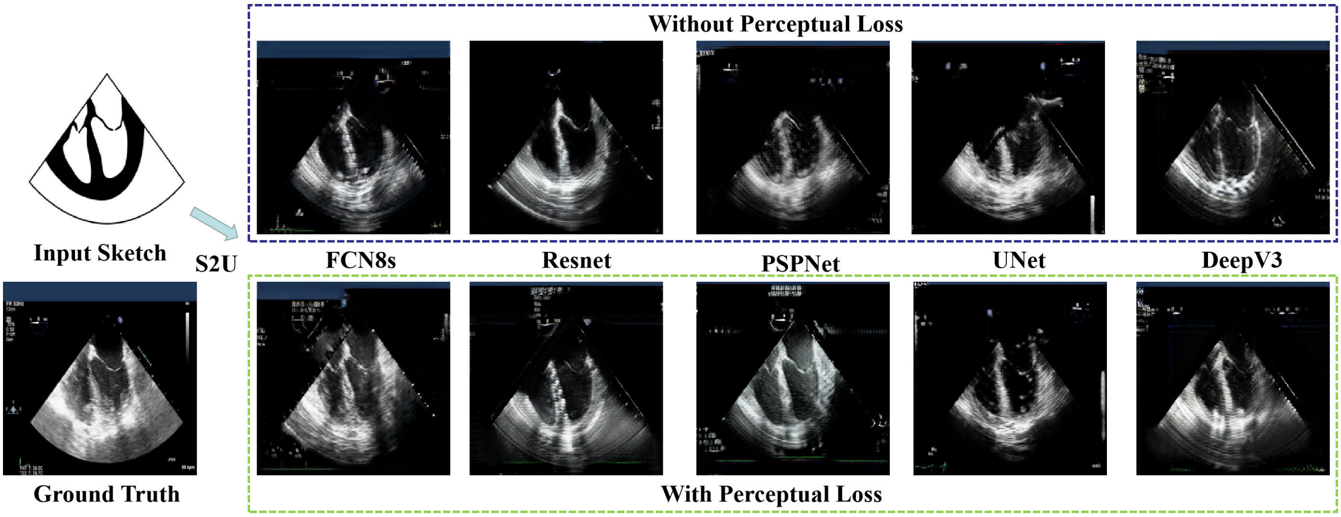

FIGURE 8. Result of different baseline model on our dataset: The first row is the S2U result without perceptual loss. The second row is the S2U result with perceptual loss.

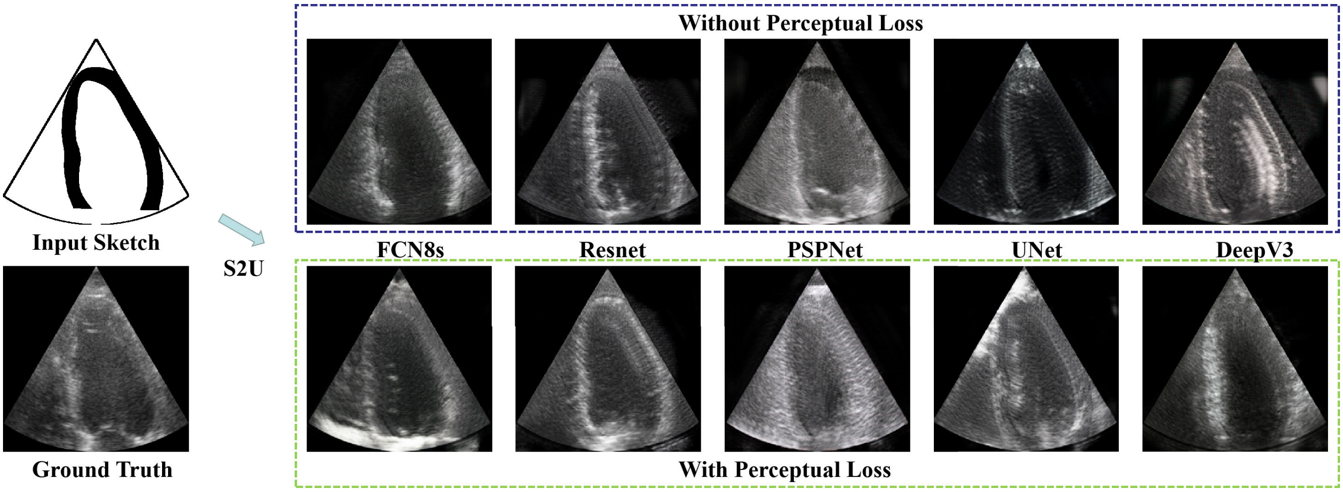

FIGURE 9. Result of different baseline model on CAMUS dataset: The first row is the S2U result without perceptual loss. The second row is the S2U result with perceptual loss.

look at the amplified sketch of arbitrary select ROI. During offline training, the trainee can draw a sketch to present its corresponding echocardiography.

Our proposed method could only statistically approximate the ground truth in S2U translation. It could not be able to generate the same ultrasound image as ground truth. The training dataset determines its synthetic texture. Despite such restrictions, it enables a useful interactive tool in echocardiography training applications with considerable performance.

VI. CONCLUSION

In this paper, an enhanced Cycle-GAN is proposed for interactive translation in the echocardiography training system. The S2U directly reconstructs the ultrasound texture without estimating the parameters of physical models. The U2S translates the ultrasound image into a sketch image with additional inferring of the sector boundary. The PSNR, SSIM, DICE, IOU are quantitatively analyzed, while visualized interactive translation is qualitatively analyzed for U2S and S2U. The presented networks achieve excellent performance as well as higher PSNR and SSIM than Cycle-GAN.

Moreover, we present additional experiments of U2S on arbitrary selected ROI and S2U on flexible hand drawing. The results reveal that the proposed interactive translation method is promising in the echocardiography training system.

Our subsequent work will be divided into two parts, namely, how to obtain more accurate segmentation with the existing dataset, as well as how to enhance the interactive experience.

REFERENCES

[1] Y. Singh, C. C. Roehr, C. Tissot, S. Rogerson, S. Gupta, K. Bohlin, M. Breindahl, A. El-Khuffash, and W. P. de Boode, ''Education, training, and accreditation of neonatologist performed echocardiography in Europe—framework for practice,'' *Pediatric Res.*, vol. 84, no. S1, pp. 13–17, Jul. 2018.

[2] P. K. Sarkar, M. Boivin, and P. H. Mayo, ''Effectiveness of an advanced critical care echocardiography course,'' *J. Intensive Care Med.*, p. 0885066619867678, Aug. 2019.

[3] A. E. Jones, V. S. Tayal, and J. A. Kline, ''Focused training of emergency medicine residents in goal-directed echocardiography: A prospective study,'' *Academic Emergency Med.*, vol. 10, no. 10, pp. 1054–1058, Oct. 2003.

[4] T. Blum, A. Rieger, N. Navab, H. Friess, and M. Martignoni, ''A review of computer-based simulators for ultrasound training,'' *Simul. Healthcare, J. Soc. Simul. Healthcare*, vol. 8, no. 2, pp. 98–108, Apr. 2013.

[5] R. Malladi, J. A. Sethian, and B. C. Vemuri, ''Shape modeling with front propagation: A level set approach,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 2, pp. 158–175, Feb. 1995.

[6] M. Kass, A. Witkin, and D. Terzopoulos, ''Snakes: Active contour models,'' *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988.

[7] W. Arkhurst, A. Pommert, E. Richter, H. Frederking, S.-I. Kim, R. Schubert, and K. H. Höhne, ''A virtual reality training system for pediatric sonography,'' *Int. Congr. Ser.*, vol. 1230, pp. 483–487, Jun. 2001.

[8] M. Weidenbach, H. Drachsler, F. Wild, S. Kreutter, V. Razek, G. Grunst, J. Ender, T. Berlage, and J. Janousek, ''EchoComTEE—A simulator for transoesophageal echocardiography,'' *Anaesthesia*, vol. 62, no. 4, pp. 347–353, Mar. 2007.

[9] J. A. Jensen, ''Field: A program for simulating ultrasound systems,'' in *Proc. 10th Nordicbaltic Conf. Biomed. Imag.*, vol. 4, 1996, pp. 351–353.

[10] B. Burger, C. Abkai, and J. Hesser, ''Simulation of dynamic ultrasound based on ct models for medical education,'' *Stud. Health Technol. Informat.*, vol. 132, p. 56, Jan. 2008.

[11] D. Magee, Y. Zhu, R. Ratnalingam, P. Gardner, and D. Kessel, ''An augmented reality simulator for ultrasound guided needle placement training,'' *Med. Biol. Eng. Comput.*, vol. 45, no. 10, pp. 957–967, Sep. 2007.

[12] B. Sun and F. D. McKenzie, ''Medical student evaluation using virtual pathology echocardiography (VPE) for augmented standardized patients,'' *Stud. Health Technol. Informat.*, vol. 132, pp. 508–510, Jan. 2008.

[13] G. Reis, B. Lappé, S. Köhn, C. Weber, M. Bertram, and A. H. Hagen, ''Towards a virtual echocardiographic tutoring system,'' in *Visualization in Medicine and Life Sciences*. Springer, 2008, pp. 99–119.

[14] S. Mazaheri, P. S. B. Sulaiman, R. Wirza, F. Khalid, S. Kadiman, M. Z. Dimon, and R. M. Tayebi, ''Echocardiography image segmentation: A survey,'' in *Proc. Int. Conf. Adv. Comput. Sci. Appl. Technol.*, Dec. 2013, pp. 327–332.

[15] N. Paragios, M.-P. Jolly, M. Taron, and R. Ramaraj, ''Active shape models and segmentation of the left ventricle in echocardiography,'' in *Proc. Int. Conf. Scale-Space Theories Comput. Vis.*, Springer, 2005, pp. 131–142.

[16] G. Hamarneh and T. Gustavsson, ''Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images,'' in *Proc. Comput. Cardiol.*, vol. 27, Sep. 2000, pp. 115–118.

[17] N. Lin, W. Yu, and J. S. Duncan, ''Combinative multi-scale level set framework for echocardiographic image segmentation,'' *Med. Image Anal.*, vol. 7, no. 4, pp. 529–537, Dec. 2003.

[18] V. Walimbe, V. Zagrodsky, and R. Shekhar, ''Fully automatic segmentation of left ventricular myocardium in real-time three-dimensional echocardiography,'' in *Proc. SPIE*, vol. 6144, Mar. 2006, Art. no. 61444H.

[19] D. Boukerroui, J. A. Noble, M. C. Robini, and M. Brady, ''Enhancement of contrast regions in suboptimal ultrasound images with application to echocardiography,'' *Ultrasound Med. Biol.*, vol. 27, no. 12, pp. 1583–1594, Dec. 2001.

[20] S. Leclerc, E. Smistad, J. Pedrosa, A. Ostvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, C. Lartizien, J. Dhooge, L. Lovstakken, and O. Bernard, ''Deep learning for segmentation using an open large-scale dataset in 2D echocardiography,'' *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019.

[21] I. Mikic, S. Krucinski, and J. D. Thomas, ''Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates,'' *IEEE Trans. Med. Imag.*, vol. 17, no. 2, pp. 274–284, Apr. 1998.

[22] M. L. Siqueira, J. Scharcanski, and P. O. Navaux, ''Echocardiographic image sequence segmentation and analysis using self-organizing maps,'' *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, vol. 32, nos. 1–2, pp. 135–145, 2002.

[23] Z. Liao, M. H. Jafari, H. Girgis, K. Gin, R. Rohling, P. Abolmaesumi, and T. Tsang, ''Echocardiography view classification using quality transfer star generative adversarial networks,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Springer, 2019, pp. 687–695.

[24] D. Mishra, S. Chaudhury, M. Sarkar, and A. S. Soin, ''Ultrasound image enhancement using structure oriented adversarial network,'' *IEEE Signal Process. Lett.*, vol. 25, no. 9, pp. 1349–1353, Sep. 2018.

[25] M. H. Jafari, Z. Liao, H. Girgis, M. Pesteie, R. Rohling, K. Gin, T. Tsang, and P. Abolmaesumi, ''Echocardiography segmentation by quality translation using anatomically constrained cyclegan,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Springer, 2019, pp. 655–663.

[26] A. H. Abdi, T. Tsang, and P. Abolmaesumi, ''GAN-enhanced conditional echocardiogram generation,'' 2019, *arXiv:1911.02121*. [Online]. Available: http://arxiv.org/abs/1911.02121

[27] 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,'' 2018, *arXiv:1804.10916*. [Online]. Available: http://arxiv.org/abs/1804.10916

[28] Z. Zhou, X. Guo, W. Yang, Y. Shi, L. Zhou, L. Wang, and M. Yang, ''Cross-modal attention-guided convolutional network for multi-modal cardiac segmentation,'' in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Springer, 2019, pp. 601–610.

[29] Q. Wang, A. Gomez, J. Hutter, K. McLeod, V. Zimmer, O. Zettinig, R. Licandro, E. Robinson, D. Christiaens, E. A. Turk, and A. Melbourne, ''Smart ultrasound imaging and perinatal, preterm and paediatric image analysis,'' in *Proc. Int. Workshop Preterm, Perinatal Paediatric Image Anal.*, 2019.

[30] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Springer, 2015, pp. 234–241.

[31] Z. Zhu, T. Huang, B. Shi, M. Yu, B. Wang, and X. Bai, ''Progressive pose attention transfer for person image generation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2347–2356.

[32] H. You, S. Tian, L. Yu, and Y. Lv, ''Pixel-level remote sensing image recognition based on bidirectional word vectors,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1281–1293, Feb. 2020.

[33] W. Cai and Z. Wei, ''PiiGAN: Generative adversarial networks for pluralistic image inpainting,'' *IEEE Access*, vol. 8, pp. 48451–48463, 2020.

[34] O. Mattausch, E. Ren, M. Bajka, K. Vanhoey, and O. Goksel, ''Comparison of texture synthesis methods for content generation in ultrasound simulation for training,'' *Proc. SPIE*, vol. 10135, Mar. 2017, Art. no. 1013523.

[35] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, ''Photo-realistic single image super-resolution using a generative adversarial network,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[36] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[37] J. Long, E. Shelhamer, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[38] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[39] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, ''Pyramid scene parsing network,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[40] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, ''Rethinking atrous convolution for semantic image segmentation,'' 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

LONG TENG was born in 1988. He received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 2006 and 2010, respectively. He is currently pursuing the Ph.D. degree with the University of Chinese Academy of Sciences. From 2019 to 2020, he was a Visiting Student in bioengineering with Yale University. His research interests include medical image processing, deep learning, and machine learning.

ZHONGLIANG FU was born in 1967. He is currently a Professor. His research interests include machine learning and image processing.

YU YAO was born in 1980. He is currently pursuing the Ph.D. degree in computer software and theory. He is also a Professor with the University of Chinese Academy of Sciences. He is also a Researcher with the Chengdu Institute of Computer Application, Chinese Academy of Sciences. He is also a Ph.D. Supervisor. He is also a Director of the Research and Development Center, Chengdu Information Technology Company Ltd. He is also the Talent of Western Light with the Chinese Academy of Sciences. He has long been engaged in academic research and industrial transformation in medical image processing, machine learning, and pattern recognition. He is also the Director of the Sichuan Computer Society, an Executive Director of the Intelligent Medical Branch of the Sichuan Computer Society, the Secretary-General, and a member of the Artificial Intelligence and Big Data Special Committee of the Sichuan Oncology Society.