



A stability-enhanced CycleGAN for effective domain transformation of unpaired ultrasound images

Lihong Huang^a, Zixia Zhou^a, Yi Guo^{a,b,*}, Yuanyuan Wang^{a,b,*}

^a Department of Electronic Engineering, Fudan University, Shanghai 200433, China

^b Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention of Shanghai, Shanghai 200032, China

ARTICLE INFO

Keywords:

Ultrasound imaging
Domain transformation
Stability enhancement
Spectral normalization

ABSTRACT

As a cost-effective, non-invasive and radiation-free medical imaging modality, ultrasonic imaging is widely used in clinical diagnosis. However, database bias is commonplace among different medical centers. Deep learning based ultrasound image analysis algorithms are usually data driven, rendering high requirements on the uniformity of image datasets. In this paper, we propose a stability-enhanced cycle-consistent generative adversarial network (CycleGAN) method with well detail preservation for the domain transformation to normalize ultrasound images from various medical centers. To stabilize the training process of CycleGAN model, we adopt spectral normalization for 1-Lipschitz continuity to reduce model oscillation. Besides, two time-scale update rule and label smoothing strategy are also utilized to maintain the balance between generators and discriminators for further stability enhancement. Moreover, our method applies skip connections to preserve ultrasound image details and prevent resolution loss during the domain transformation process. Experiments were conducted on clinical thyroid and carotid image datasets acquired from several medical centers. Massive results demonstrate that our proposed model is easier to reach a steady state when training, outstanding 50% from the basic CycleGAN model. Compared with representative algorithms, our proposed method reaches the state-of-the-art performance, with a 11.3% decrease in the mean absolute error and a 9.8% increase in the structural similarity. Hence, our proposed algorithm has a strong capacity of the domain transformation in ultrasound images to reduce the database bias for uniformly distributed datasets. We believe that our method can contribute to the development of the ultrasound image analysis and computer aided clinical diagnosis.

1. Introduction

Ultrasonic imaging is not only inexpensive and fast, but also noninvasive and accurate, becoming one of popular medical imaging methods for clinical diagnosis [1–3]. Many studies have been conducted to reveal the underlying diagnosis and treatment information from ultrasound images [4–7]. With the rapid development of computing power and memory resources, many advanced techniques such as deep learning have been generally applied in medical ultrasound image analysis, involving tissue and lesion detection, segmentation and classification [8], which help to improve the efficiency and accuracy of medical diagnosis and treatment process greatly.

Most of researches and algorithms based on deep learning for ultrasound image analysis are data-driven, which render higher requirements on the uniformity of image datasets. However, as shown in Fig. 1, variation including gray distribution, contrast and so on is

commonplace among ultrasound images acquired from different medical centers. Such variation can be great due to different imaging settings. It is difficult to remain the performance and robustness of a network once deployed at a medical center. Newly introduced protocols or scanners bring a degradation in the performance of an optimized or tuned algorithm for a specific center, because the impact on image feature differences caused by imaging settings may exceed that induced by the pathological characteristic itself [9]. Maybe we can resolve the problem by retraining the model, but it is burdensome and time-consuming. Therefore, algorithms should deal with aforementioned image dissimilarities before being applied in the workflow of ultrasound image analysis [10]. Besides, different from natural images, extensive data especially paired data is difficult to obtain in the medical ultrasound domain [11]. Many problems, such as the mode collapse and overfitting, would be caused by training models with limited and not uniformly distributed datasets. Hence, we should perform the domain

* Corresponding authors at: Fudan University, No. 220, Handan Road, Yangpu District, Shanghai 200433, China.

E-mail addresses: guoyi@fudan.edu.cn (Y. Guo), yywang@fudan.edu.cn (Y. Wang).

<https://doi.org/10.1016/j.bspc.2022.103831>

Received 25 March 2022; Received in revised form 6 May 2022; Accepted 16 May 2022

Available online 21 May 2022

1746-8094/© 2022 Elsevier Ltd. All rights reserved.

transformation to reduce discrepancies between ultrasound images from various medical centers. This can contribute to data-driven studies and reliability enhancement of leaning based algorithms.

During last decades, domain transformation researches have been conducted in some medical imaging domains (e.g., magnetic resource images (MRIs) and histopathological images). Conventional approaches including global histogram-matching methods [12] and joint histogram registration methods [13] are not applicable to unpaired data. Instead, some creative learning-based approaches extracting powerful high-level features have superior performance to traditional modeling methods [14]. Adversarial learning [15] is usually applied for style transfer tasks. Among achievements in image-to-image translation, Pix2Pix, adapted from the conditional GANs [16], requires paired data [17]. For a remission of demand for paired data, cycle-consistent GANs (CycleGANs) [18] appeared as a milestone. Thomas et al. [10] applied residual learning from input to output to force CycleGAN to learn the residual for stain transformation in histopathology. Liu et al. [19] added the pathology consistency constraint into CycleGAN model for stain

transformation. Gao et al. [9] utilized a forward GAN path and multiple backward GAN paths to perform many-to-one transformation in MRIs. Based on the theory of “loss-correction”, Kong et al. [20] proposed a new unsupervised mode called RegGAN, which applies an additional registration network in GAN model. Inspired by the profound ability of deep learning and previous work in domain transformation, we explore the potential of CycleGANs to perform domain transformation in ultrasound images to normalize images from multi-centers.

Although deep learning has gained a lot of traction in MRIs and histopathology, there still exist challenges of the application in the domain transformation of unpaired ultrasound images. The first challenge is how to stabilize the training of the CycleGAN model. Except for the inherent difficulty of training unsupervised learning model, the low quality of ultrasound images and great variation among ultrasound images from different medical centers make it harder for generators to extract image features and generate images similar to target domain to “fool” discriminators. It leads to model oscillation or unbalance between generators and discriminators resulting weak stability problem when

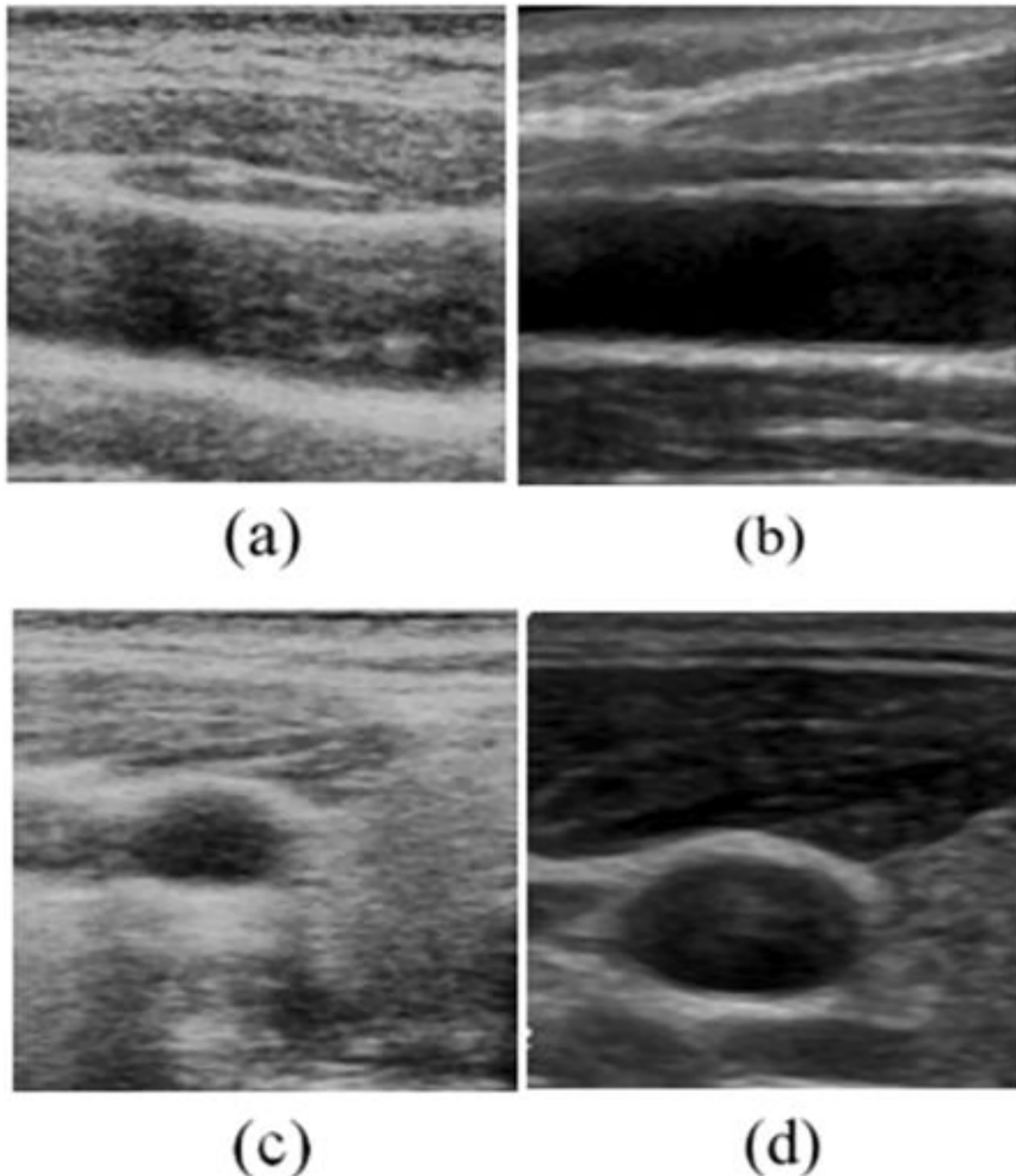


Fig. 1. Comparison between ultrasound images obtained from various medical centers. (a), (b) display ultrasound carotid images, and (c), (d) display ultrasound thyroid images. Images shown in the first column were scanned by mSonics MU1, and those shown in the second column were scanned by Toshiba Aplio 500.

training with unpaired ultrasound images. The stability of training the network can be influenced by many factors such as the performance of discriminators, the balance between generators and discriminators. An unbalance phenomenon that discriminators have been optimized to converge while generators not, usually appears when training GAN models. The appearance causes overfitting or model collapse, leading to the weak stability and low effectiveness of the CycleGAN model. The second challenge is that the image detailed information loss and resolution loss during domain transformation process. The cycle-consistency constraint does well in the style transfer but seems to just achieve the low-level style mapping between two domains. It shows insufficient constraint on preserving ultrasound image details like tissue structures, organ boundary and textures, which are of critical significance for ultrasound diagnosis and treatment. To overcome aforementioned challenges, some strategies including the spectral normalization, the two time-scale update rule and label smoothing strategy and addition of skip connections are adopted to optimize the original CycleGAN. We propose a stability-enhanced CycleGAN with better detailed information preservation for the domain transformation of ultrasound images to obtain uniformly distributed datasets. Summarizing, major contributions made by this paper are as follows:

(1) Based on the analysis of the weak stability of training CycleGAN model, we develop the basic CycleGAN by incorporating spectral normalization layers in discriminators for 1-Lipschitz continuity to stabilize the training process. Besides, training strategies including two time-scale update rule and label smoothing strategy are adopted to maintain the balance between generators and discriminators.

(2) Considering the weakness of the cycle-consistency constraint, we utilize the addition of skip connections in generators to enhance image detailed information preservation and reduce the resolution loss during

cycle transformation process.

(3) Massive qualitative and quantitative results of the domain transformation on clinical thyroid and carotid ultrasound image datasets are provided to demonstrate the state-of-the-art performance of our proposed method.

2. Method

2.1. Model overview

Fig. 2 shows the end-to-end framework of our proposed method at a component level, which consists of two GANs for two transformations, i. e., $X \rightarrow Y$ and $Y \rightarrow X$, respectively. In the training stage, the proposed model is trained with a mass of unpaired training data, aiming to obtain end-to-end mapping between two domains. In the testing stage, transformed ultrasound images are directly generated by the trained model. The data preprocessing step, including random cropping, rotation and image scaling, is for the data augmentation. The registration preprocessing, including a B-spline-based nonrigid registration and a landmark-based nonrigid registration, is adopted to obtain paired testing data for evaluating the domain transformation performance.

2.2. Model description

2.2.1. Architecture of the proposed model

The architecture of generators can be viewed in Fig. 3. Every generative network is composed of an encoder-decoder architecture inspired by the U-net [21], which is a widely used feature extractor, and a residual convolution path. The encoder is responsible for extracting the principal feature and morphological content from the source

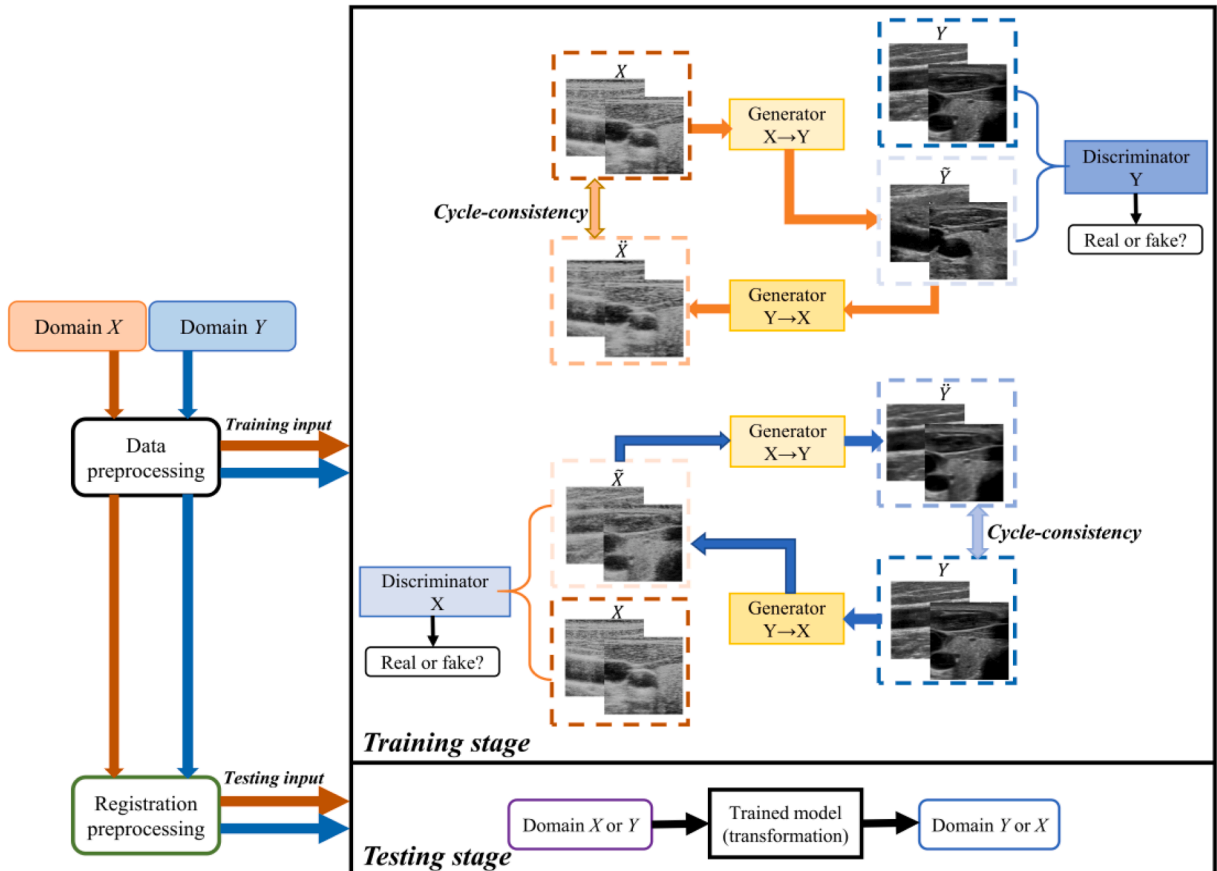


Fig. 2. Overview of the proposed method (X, Y denote ultrasound images from two domains respectively. $\tilde{X}$ and $\tilde{Y}$ represent transformed images. $\bar{X}$ and $\bar{Y}$ denote reconstructed images).

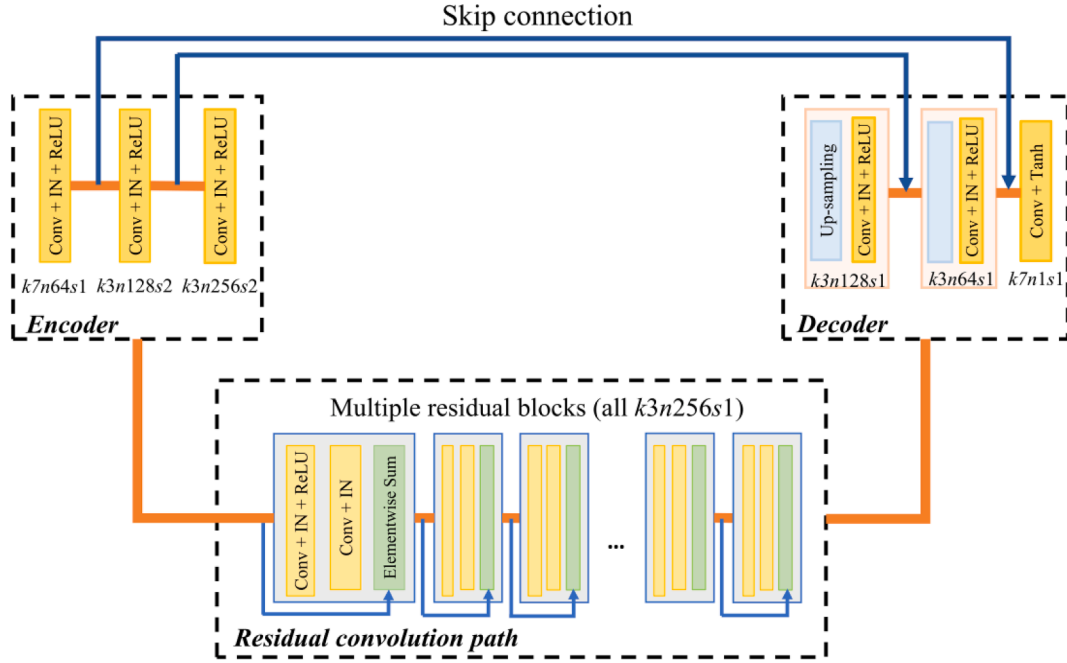


Fig. 3. The generator of the proposed model (k , n and s denote the kernel size, the number of feature maps and stride, respectively).

ultrasound image. It starts by a 7×7 convolution. Subsequently, to maintain the spatial continuity and reserve more structural details, we adopt two 3×3 convolutions instead of the pooling operation. Each of the convolutional layers is followed by a rectified linear unit (ReLU) [22] to increase nonlinear characteristics of the model and avoid the slow convergence caused by the disappearance of gradient. Instance normalization and reflection padding are also used with every convolution. The following module is a residual convolution path, a high-performance feature extractor containing nine successive residual convolution blocks [23]. It can integrate different-level image features efficiently and deepen the depth of the network yet not increase the training difficulty. Each residual convolution block contains two 3×3 convolutions with fixed-number filters. For the decoder, we utilize two

2.2.2. Loss function

We used the L1-loss to minimize the cycle-consistency error. The least-square loss [27] is adopted as the adversarial loss to reduce the chance of the disappearance of the gradient [28]. In addition, we add an identity loss term to ensure the identity mapping in generators. It can alleviate the style over-transferring problem that may reduce the quality of transformed ultrasound images and lead to the instability of the CycleGAN model. The identity loss is computed in the L1 norm:

$$\mathcal{L}_{identity}(G, F) = \mathbb{E}_{y \sim p_{data}(y)} [\|G(y) - y\|_1] + \mathbb{E}_{x \sim p_{data}(x)} [\|F(x) - x\|_1] \quad (1)$$

The overall objective function of our model can be expressed as following:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{LSGAN}(G, D_Y) + \mathcal{L}_{LSGAN}(F, D_X) + \lambda \mathcal{L}_{cycle}(G, F) + \gamma \mathcal{L}_{identity}(G, F) \quad (2)$$

nearest-neighbor up-sampling instead of transposed convolutions to reduce the extent of checkerboard artefacts [24], as checkerboard artefacts will bring great errors and deformations in the transformed results or lead to the failure of training model. Finally, the output image is obtained through a 7×7 convolution and a ‘Tanh’ activation function.

For discriminators, an architecture inspired by the classic ‘PatchGAN’ setup [25] is adopted, which is composed of four 4×4 down-sampling convolutions followed by a leaky rectified linear unit (LeakyReLU) [26] to retain the information brought by the negative axis and reduce the inactive neurons number. Four convolutions are with 64, 128, 256 and 512 filters, respectively. Each of them except the first one is followed by a spectral normalization layer. The final layer is a 3×3 convolution to reduce the output to a single filter map. Finally, the discriminator produces a 30×30 discriminative feature map instead of a single value to judge whether the generated image is real or not. Based on the architecture, the discriminator has a reduced receptive field, which restricts the network to smaller parts of input images for more attention to high-frequency changes in images [10].

where G and F are generators for the transformation $X \rightarrow Y$ and $Y \rightarrow X$ respectively, D_X and D_Y are corresponding discriminators; λ and γ are constants. The process of training the proposed model is to obtain the optimal solution to the following optimization problem:

$$G^*, F^* = \underset{G, F, D_X, D_Y}{\operatorname{argminmax}} \mathcal{L}(G, F, D_X, D_Y) \quad (3)$$

2.3. Enhancing the stability of the model

As an unsupervised learning method, the CycleGAN model is not easy to be trained to reach the Nash equilibrium. In order to stabilize the training of the CycleGAN model, we adopt other strategies apart from the identity loss to obtain an improved model with more steady performance. According to the reference [29], the influence brought by loss functions on training GANs is not so strong, i.e., no one loss function is absolutely superior to others. Therefore, we don’t make much adjustment on the loss function and pay more attention on changing the composition of networks and training strategies. Optimizations are further detailed below.

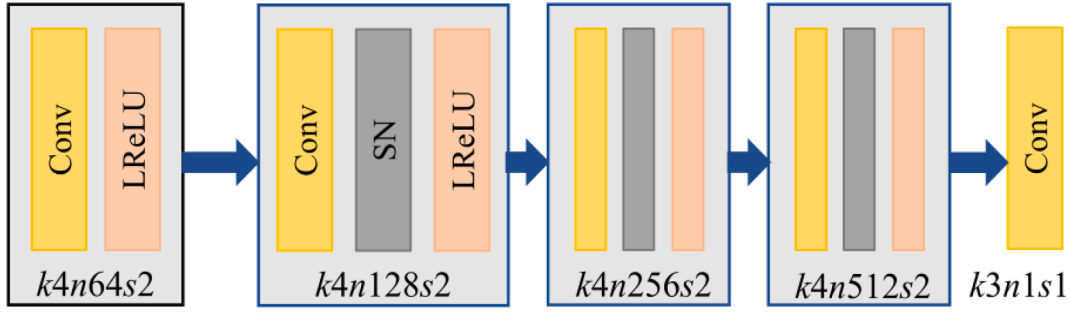


Fig. 4. The discriminator of the proposed model (k , n and s denote the kernel size, the number of feature maps and stride, respectively).

2.3.1. Spectral normalization

Although the least-square loss applied in the discriminator loss function helps to alleviate the gradient disappearance problem, it results in weaker stability of the discriminative network with a high learning rate [30]. Spectral normalization was firstly proposed by Miyato et al. [31], and since then this useful tool has been adopted for GAN training in many studies. It has been proved that the spectral normalization is a better practical choice than other regularization and normalization methods [32]. As illustrated in Fig. 4, we add a spectral normalization layer after each convolution of the discriminative network to normalize parameters of the convolutional layer to meet the 1-Lipschitz continuity. According to the stability theorem of GANs, the control capacity of the discriminative network can be enhanced when the input and output of the network satisfy the 1-Lipschitz continuity [33], which efficiently reduces the model oscillation and speeds up the convergence. Hence, to a great extent, the spectral normalization stabilizes the gradient and maintains the balance between the generator and discriminator so as to stabilize the training of the CycleGAN model. The principle of the spectral normalization is to estimate the spectral norm of the parameter matrix of the convolutional layer. The formulation to compute the spectral norm is defined as:

$$\sigma(W) = \sup_{x \neq 0} \frac{\|Wx\|_2}{\|x\|_2} = \sup_{\|x\|_2 \leq 1} \|Wx\|_2 \quad (4)$$

where x is the input, W is the weight matrix of the convolution. By enforcing $\sup_{\|x\|_2 \leq 1} \|Wx\|_2 = 1$, the parameter-normalized weight matrix can be expressed as:

$$W_{SN} = \frac{W}{\sigma(W)} \quad (5)$$

2.3.2. Two time-scale update rule

The two time-scale update rule was firstly proposed by Heusel et al. [34] and has been applied to many other GANs such as DCGAN [35] and WGAN-GP [36] to solve the unbalance problem between generators and discriminators for the stability enhancement of GAN models. In simple terms, the training strategy is to choose different learning rates instead of the same value for generators and discriminators. In practice, we use the Adam optimizer with a lower learning rate for training generators and a higher one for discriminators. The strategy is aimed to make discriminators produce feedback (i.e. discriminative results) for generators in time and then update generators with a smaller step. Thus, the learning process of generators is more scientific so that weights of the generative network can be updated more accurately. Generators are enforced to generate images with more features close to reference ones. Discriminators will be fooled and hard to distinguish between generated images and real ones. The adversarial loss can be reduced and the model will be trained towards the stability. Therefore, the two time-scale update rule enhances the balance between generators and discriminators during the adversarial learning process and brings stability enhancement of CycleGAN model.

2.3.3. Label smooth strategy

Assuming that discriminators rely on few features to judge images generated from generators, generators can generate only those features in response to discriminators' detection. The optimization of generative networks will become greedy without long-term benefit. In order to alleviate such problem, we utilize the label smooth strategy. We set the target label value at 0.9 or lower when the prediction probability given by discriminators to judge the image whether belonged to the target domain reaches 0.9. The operation is adopted to appropriately increase the difficulty of training discriminators to maintain the balance between generators and discriminators ensuring the stability of the CycleGAN model.

2.4. Enhancing detailed information preservation

In clinical ultrasound diagnosis, tissue structures, details and speckle of ultrasound images play a critical role in extracting image features to provide diagnostic information. Consequently, in the domain transformation of ultrasound images, the detailed information preservation is of significant importance. Although the cycle-consistency constraint enforces the transformation back to source domain and contributes to better performance in style transfer tasks, it seems to show insufficient constraint on reserving details. As the reference [37] indicates that first few layers of convolutional neural networks usually contain more transferable feature information of input images, two skip connection paths are established between layers in the encoder and decoder that mirror each other as shown in Fig. 3. The addition of skip connections feeds the low-level information from input to output, which provides shortcuts for reserving and reconstructing details from source images. We hypothesize that these skip connections are also especially important in the transformation to reduce the loss of high-resolution information in deeper layers of the network.

3. Experiments

3.1. Materials and preprocessing

Experiments were carried on two clinical ultrasound datasets acquired using ultrasound imaging devices mSonics MU1 (Ultimedical Technology, China) and Toshiba Aplio 500 (Toshiba Medical Systems Corporation, Japan), respectively. Transducer central frequencies are 6 MHz and 7.5 MHz respectively. We collected 120 pieces of carotid and thyroid ultrasound images with mSonics MU1 and the same quantity of images with Toshiba Aplio 500. Carotid and thyroid images were scanned for 47 healthy volunteers by an experienced doctor. To ensure the robustness of our method, the training set and testing set are both a mixture of carotid and thyroid images.

To expand the training set, augmentation techniques like the random cropping, rotation and image scaling, were carried out as data pre-processing. After augmentations, our training set contains 1656 unpaired pieces of images with the size of 400×400 from two aforementioned medical centers. For the testing set, 78 pairs carotid and

thyroid images with the size of 400×400 were sampled, which contains 26 pairs of carotid ultrasound images and 52 pairs of thyroid ultrasound images. In order to evaluate our method properly, the 78 pairs of ultrasound images were registered to each other using a B-spline-based nonrigid registration method [38] and a landmark-based nonrigid registration method [39]. Specifically speaking, volunteers held their breath for about 10 s when being scanned to reduce the deformation. In the scanning process, landmark points were recorded on volunteers' scanning positions for the landmark-based nonrigid registration to obtain paired testing data pairs for appropriate evaluation of our method. After registration, 78 ultrasound image pairs can be regarded as the source domain and reference domain respectively to test and compare the domain transformation performance of different models. Some samples from the testing set are shown in Fig. 5. It is worth to note that the intersection of training set and testing set is empty.

3.2. Experimental setting

In this paper, the commonly used unpaired image-to-image

translation method 'Cycle-GAN' is regarded as the baseline method. We conducted a series experiments to demonstrate the model evolution process and the superiority of our proposed method.

3.2.1. Ablation study

We first explored the effectiveness of the optimizations including the spectral normalization, two time-scale update rule and label smooth strategy to stabilize the training of the proposed model. The stability performance comparisons between the training process with/without such optimizations were conducted. Besides, through the study of skip connections on the image detailed information reservation, we can demonstrate why it is employed.

3.2.2. Comparison methods

To further confirm the superiority of our proposed method, we compared the proposed method with RegGAN [20], Residual CycleGAN [10], SRGAN [40]. We reimplemented the RegGAN method in the "C (cycle-consistency) + R(registration)" mode. We reimplemented Residual CycleGAN method with the same function of the adversarial loss,

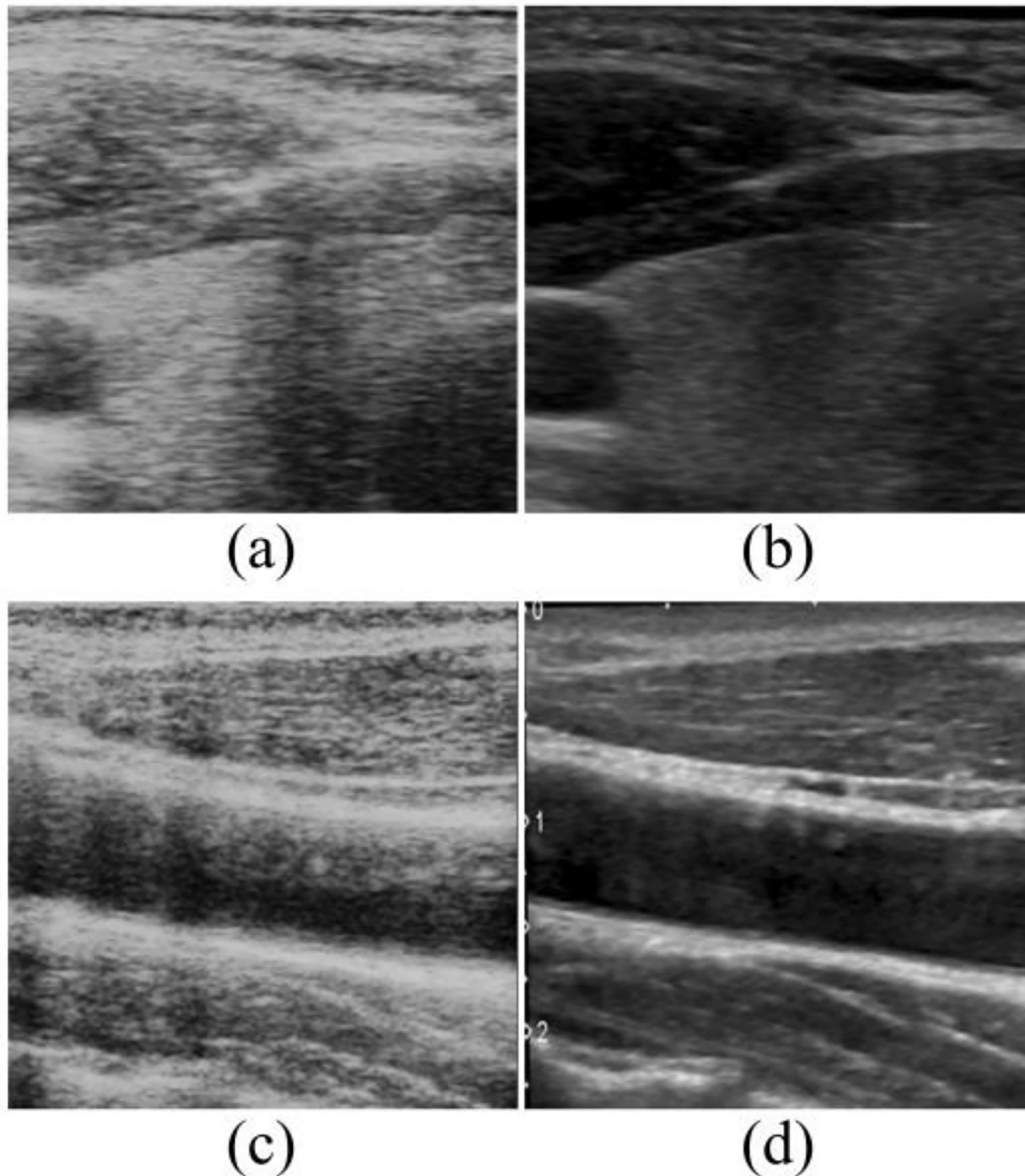


Fig. 5. Some samples in the testing set. (a), (b) present ultrasound thyroid image pairs of the same location, and (c), (d) display ultrasound carotid image pairs of the same location. Images scanned by mSonics MU1 are shown in the first column, and images scanned by Toshiba Aplio 500 are shown in the second column.

cycle-consistency loss and extra identity loss as that of our method. As there are few previous works conducted in the domain transformation of ultrasound images, we also reimplemented the SRGAN to perform the transformation. The training process of different methods converged after 300 epochs. In different experiments, both the visual observation and quantitative comparison results were demonstrated.

3.3. Evaluation measures

For the ultrasound image domain transfer task of $X \rightarrow Y$, there are mainly four types of images, i.e., the source image X , the transformed image $\tilde{Y}$, the reconstructed image $\tilde{X}$ and the target image Y and vice versa. Therefore, six types of images, i.e., X , Y , $\tilde{X}$, $\tilde{Y}$, $\tilde{X}$, and $\tilde{Y}$ can be used to evaluate the transformation performance of two directions $X \rightarrow Y$ and $Y \rightarrow X$.

3.3.1. For $X \leftrightarrow \tilde{X}/Y \leftrightarrow \tilde{Y}$ or $X \leftrightarrow \tilde{X}/Y \leftrightarrow \tilde{Y}$

To comprehensive evaluate the uniformity of transformed ultrasound images ($\tilde{X}/\tilde{Y}$) and corresponding reference ones, we use several indices including the mean absolute error (MAE) to estimate their gray distribution, the multi-scale Structural Similarity Index Measure (MS-SSIM) to measure their multi-scale structural similarity, the Pearson correlation coefficient (Pearson-R) to indicate their correlation and the mutual information (MI) to evaluate the mutual dependence. The greater similarity signifies more uniformly distributed datasets can be obtained with the transformation model and better cycle-consistency of the transformation model. Evaluation indicators are detailed below:

1) MAE. The MAE metric is used to evaluate the average pixel gray difference between different images. The lower value indicates the smaller difference so that transformed images or reconstructed ones are more similar to the corresponding domain.

2) SSIM/MS-SSIM. The similarity/multi-scale similarity between different images can be evaluated by the SSIM/MS-SSIM, which are calculated to assess the performance in preserving tissue structures and texture information [41].

3) Pearson-R. In order to evaluate the relationship of different images, the Pearson-R is used to evaluate the correlation. A higher value demonstrates the stronger relationship between them.

4) PSNR. The PSNR metric is adopted to evaluate the quality of transformed images and reconstructed ones [42]. A higher PSNR value represents the transformed images and reconstructed ones with the high quality, containing more clear details.

5) MI. The MI metric is adopted to evaluate the mutual dependence between the transformed images and reference ones [43], which can be calculated by:

$$MI(x, y) = \sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \quad (6)$$

where $P(x)$, $P(y)$ and $P(x, y)$ denote the marginal probability distribution functions and joint probability function of the transformed image and corresponding reference one, respectively [44].

The indexes are applicable for the performance assessment of both the transformation direction $X \rightarrow Y$ and $Y \rightarrow X$ to comprehensively evaluate the effectiveness of the proposed transformation model.

3.3.2. For $X \leftrightarrow \tilde{Y}/Y \leftrightarrow \tilde{X}$

Considering the inherent distance between two different ultrasound image domains, the Contrast-Structure Similarity (CSS) [19] is adopted to evaluate how much the structural information is preserved from the source image. The CSS is a variant of the SSIM and can be defined as following:

$$CSS(x, y) = \frac{2\sigma_{xy} + c}{\sigma_x^2 + \sigma_y^2 + c} \quad (7)$$

where σ_x , σ_y are the standard derivations, σ_{xy} is the covariance. c is a stabilizing factor variable to stabilize the division with weak denominator.

4. Results and discussion

4.1. Performance improvement brought by the optimizations on stabilizing the training process

Based on the analysis of the weak stability of training CycleGAN model, we utilize the addition of the spectral normalization, two time-scale update rule and label smooth strategies to stabilize the training process of the model. In order to study the stability improvement of training, we trained the CycleGAN model with/without optimizations with the same training set.

Fig. 6 shows loss value curves of generators and discriminators during the training process. It can be seen from Fig. 6(a) and (b) that the loss of generators is hard to converge while the loss of discriminators drops rapidly, which reveals the unbalance and weak stability of training the basic CycleGAN model. By contrast, Fig. 6(c) and (d) show that the loss of generators and discriminators converges to a lower value and keeps steady as the training continues when training the stability-enhanced CycleGAN, which intuitively demonstrates that it is effective to stabilize the training process of the CycleGAN model with the optimizations described in Section 2.3.

To further confirm the performance improvement, more quantitative and qualitative results were obtained. Some samples in the domain transformation process using the original and stability-enhanced CycleGAN model are given in Fig. 7, which reveals that the quality of images produced by the basic CycleGAN is much lower than those generated by the stability-enhanced CycleGAN. As labeled with the red boxes in Fig. 7, the basic CycleGAN may generate images with many deformations and artifacts in transformation results. Besides, comparing the first row of (c) to (f), it is easy to find that transformed images generated by the basic CycleGAN model vary from epoch to epoch, such as the greatly different contrast, resolution, detailed textures. Huge discrepancies show the unsteady training process of the basic CycleGAN model. In contrast to the first row of (c) to (f), transformed images produced by the stability-enhanced model (i.e., the second row of (c) to (f)) have more clear and accurate textures while those in the first row suffer from over-smoothing or the over-contrast problem. Thus, transformed ultrasound images produced by the stability-enhanced CycleGAN model with the same training iteration are much more similar to reference ones. The visual effect demonstrates the great improvement of the domain transformation performance brought by optimizations to against the weak stability when training the CycleGAN model.

Expect for visual comparisons, some quantitative evaluation results on transforming X to Y are showed in Table 1, where the evaluation of "original images" shows the huge distance between two different domains. As shown in Table 1, the MAE decreases from 28.43 ± 1.86 to 14.27 ± 1.88 , which demonstrates the capability of the CycleGAN model with stronger stability to predict more similar ultrasound images with corresponding reference ones. The improvement in the SSIM/MS-SSIM, which increase from $0.33 \pm 0.08/0.49 \pm 0.08$ to $0.41 \pm 0.07/0.58 \pm 0.08$, proves the competence of the stability-enhanced model in maintaining ultrasound image details. The PSNR increases from 14.38 ± 2.36 to 16.53 ± 1.58 , which indicates the ability of the improved model to generate images with higher quality. The CSS is increased from 0.39 ± 0.06 to 0.50 ± 0.05 , which means that more information is preserved from source images. The MI is improved from 0.58 ± 0.18 to 0.87 ± 0.11 , which proves the higher mutual dependence between the transformed images and reference ones and stronger capability of our modified model to predict similar images with corresponding target domain. Comparing with original images, the distance between transformed images and target ones is greatly shortened by the stability-

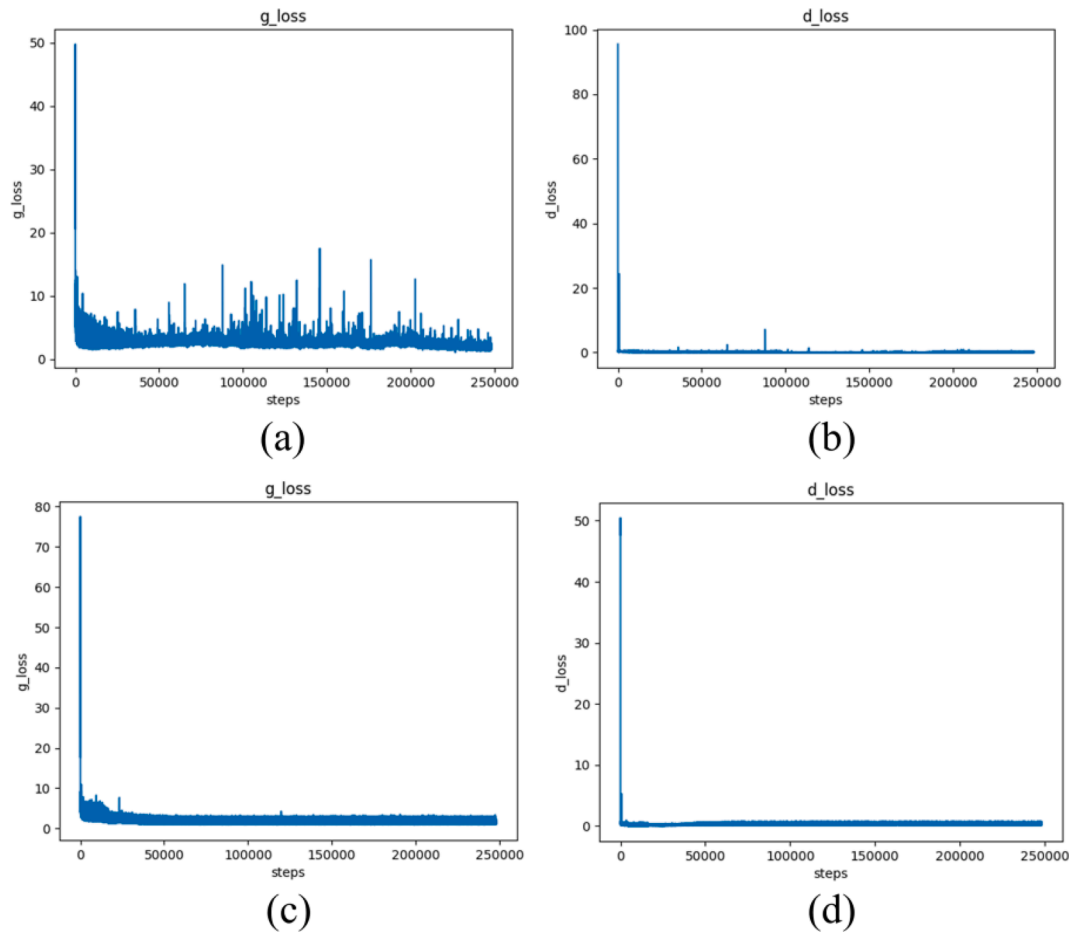


Fig. 6. The loss value curves of generators and discriminators during the training process. (a), (b) are loss curves of generators/discriminators when training the basic CycleGAN model. (c), (d) are loss curves of generators/discriminators when training the stability-enhanced CycleGAN model.

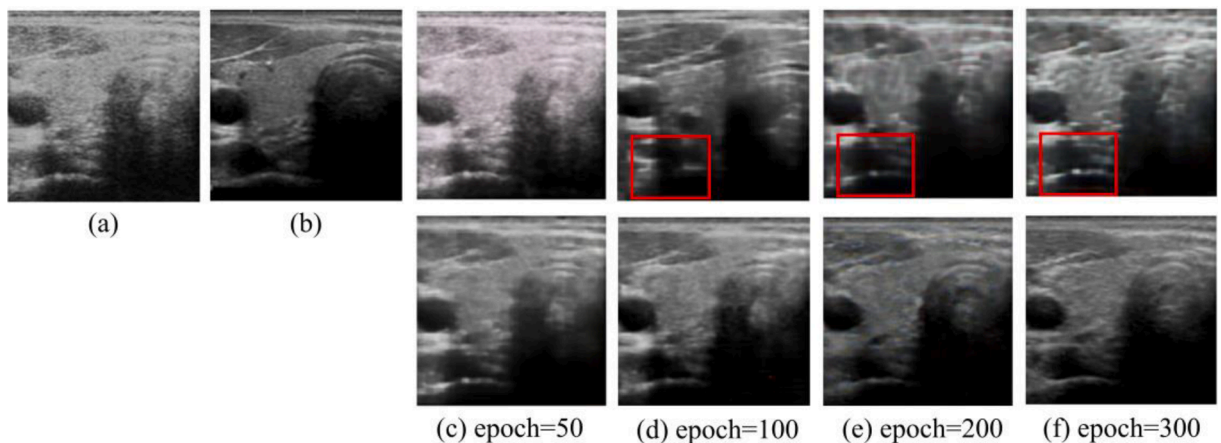


Fig. 7. Some testing samples produced by the original and stability-enhanced CycleGAN model with different training iterations. (a), (b) present a pair samples from the testing set of source images from different domains. (a), (b) are scanned using mSonic MU1 and Toshiba Aplio 500 respectively. (a) denotes the source image and (b) is the corresponding reference one. (c) to (f) present the corresponding transformed images produced by the original CycleGAN and the stability-enhanced CycleGAN model with different iterations. The first row of (c) to (f) present transformed images produced by the original CycleGAN and those in the second row are generated by the stability-enhanced CycleGAN model. Some red ovals are labeled to highlight the deformation in images.

enhanced CycleGAN model, which demonstrates the strong ability of the model for the domain transformation in ultrasound images. All of the above comparisons on the visual effect and evaluation indices reveal the importance and effectiveness of the optimizations on stabilizing the training process of the CycleGAN model.

4.2. Importance of skip connections

Based on the analysis of the insufficient constraint of the cycle-consistency loss to preserve the image detailed information, the stability-enhanced CycleGAN model is incorporated with skip connec-

Table 1

Quantitative evaluation results for stabilizing the training process of the model.

Model	$Y \leftrightarrow \tilde{Y}$		MS-SSIM	PSNR/ dB	MI	$X \leftrightarrow \tilde{Y}$ CSS
	MAE	SSIM				
Original images	51.32	0.22	0.36	12.58	0.37	0.34
CycleGAN	± 1.52	± 0.05	± 0.07	± 1.53	± 0.18	± 0.51
	28.43	0.33	0.49	14.38	0.58	0.39
CycleGAN + SN + TTU + LS (Stability- enhanced CycleGAN)	± 1.86	± 0.08	± 0.08	± 2.36	± 0.18	± 0.06
	14.27	0.41	0.58	16.53	0.87	0.50
	± 1.88	± 0.07	± 0.08	± 1.58	± 0.11	± 0.05

NOTE: 'SN' means the spectral normalization; 'TTU' means the two time-scale update rule; 'LS' means the label smooth; 'Y' means images like Fig. 7 (b); $\tilde{Y}$ means 'fake_Y' (transformed images) like Fig. 7 (c) to (f).

tions. In order to study the importance of the addition of skip connections, examples of visual results and overall evaluation indices were obtained as shown in Fig. 8 and Table 2. In Fig. 8, (a) to (f) show the transformation process $X \rightarrow \tilde{Y} \rightarrow \tilde{X}$, and (a), (b), (g), (h), (i), (j) display the transformation process $Y \rightarrow \tilde{X} \rightarrow \tilde{Y}$. As labeled with red boxes and blue ovals in Fig. 8, some parts of sampled images are highlighted and magnified to better show the detailed difference in the structural information. Comparing Fig. 8(c) and Fig. 8(d), which are $\tilde{Y}$ produced by the stability-enhanced CycleGAN without/with skip connections, detailed textures of Fig. 8(c) labeled with red boxes are blurry while those of Fig. 8(d) are more clear. Besides, the organ boundary labeled with the blue oval in Fig. 8(c) are over-smooth in contrast to that of Fig. 8(d). Misty structural details and over-smoothing problem lead to the deformation and error in transformation results, which will provide inaccurate or mistaken image feature information and even worse rise risk of incorrect medical diagnosis. In the same way, comparing Fig. 8(g) with Fig. 8(h), it can be demonstrated again that the model with skip connections achieve a better structural information preservation effect.

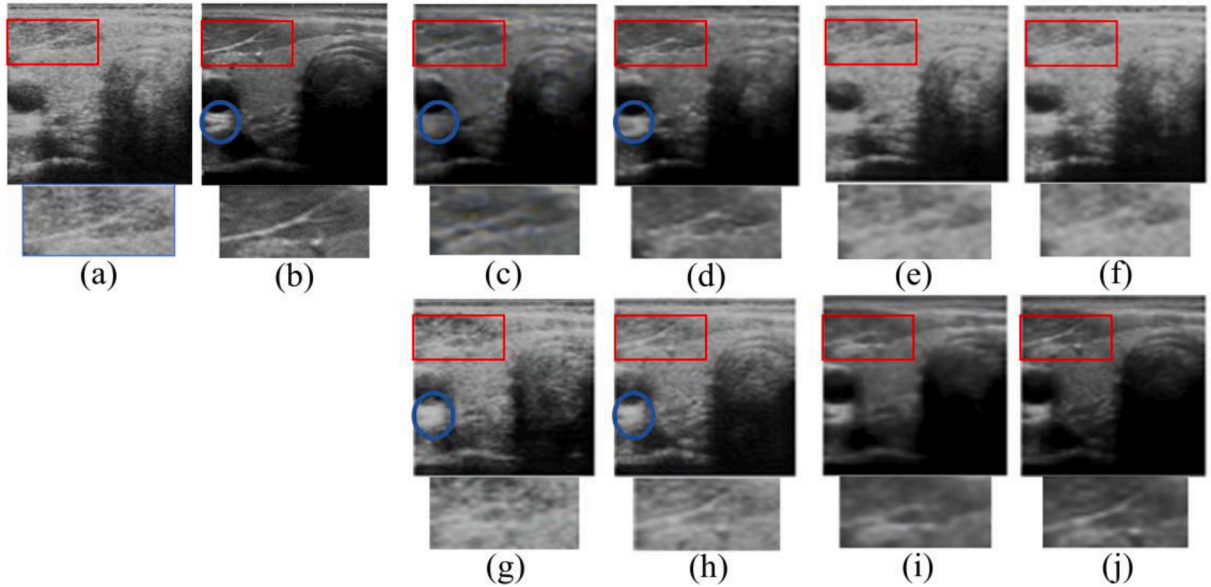


Fig. 8. Some testing results produced by the stability-enhanced CycleGAN model with/without skip connections. (a), (b) present a pair samples from the testing set of source images from different domains. (a), (b) are scanned using mSonics MU1 and Toshiba Aplio 500 respectively. (a) denotes the source image and (b) is the corresponding reference one. (c), (d) present transformed images generated by the stability-enhanced CycleGAN model without/with skip connections from (a). (e), (f) show corresponding reconstructed images. (g), (h) denote transformed images generated by the stability-enhanced CycleGAN model without/with skip connections from (b). (i), (j) display corresponding reconstructed images. Some parts of (b) to (j) are labeled with red boxes and blue ovals to better show detailed differences between them.

In addition, detailed discrepancies labeled with the red boxes between Fig. 8(i)/Fig. 8(e) and Fig. 8(j)/Fig. 8(f) indicate the better reconstruction effect of the modified model with skip connections, which reveal the higher cycle-consistency brought by the addition of skip connections.

Apart from the visual effect, objective quantitative calculations were also given. As we can see in Table 2, the mean MAE SSIM, MS-SSIM, PSNR and MI values achieved by our proposed method are 12.20 ± 1.38 , 0.45 ± 0.06 , 0.62 ± 0.08 , 19.47 ± 1.24 , and 1.04 ± 0.14 respectively, which are superior to those achieved by the model without skip connections. The CSS metric increases from 0.50 ± 0.05 to 0.75 ± 0.03 . All the results, especially the improvements of the SSIM, MS-SSIM, MI and CSS, indicate that the addition of skip connections is of significant impact on reserving image details and reducing the information loss during the domain transformation task. Moreover, the distance between generated images and reference ones is further shortened by the proposed method according to the decrease of the mean MAE.

It is obvious that the performance of our proposed method gains an improvement on each metric, which manifests that short paths offered by skip connections facilitate the broadcast of the detailed information. Hence, the significance of the addition of skip connections can be verified by intuitive results and quantitative assessments mentioned above.

4.3. Comparison with other methods

To comprehensively confirm the superiority of our proposed method, massive results were obtained to compare our method with the currently proposed RegGAN. Details of results are presented as follows.

Fig. 9 displays intuitive results. Comparing Fig. 9(c) with Fig. 9(b), the resolution and contrast of the ultrasound image are improved a little, but textures and structural details are fragmented. And the gary distribution of Fig. 9(c) is quite different from that of Fig. 9(a). Thus, the SRGAN method brings an improvement in the image resolution but performs not well in normalizing ultrasound images from different domains. From Fig. 9(d), we can find that the transformed image produced by the Residual CycleGAN suffers from the over-smoothing problem. By comparing Fig. 9(d) with Fig. 9(b), we think that the over-smoothing phenomenon is caused by the overmuch information maintained in

Table 2
Quantitative evaluation results for addition of skip connections.

Model	$Y \leftrightarrow \tilde{Y}$					$X \leftrightarrow \tilde{Y}$
	MAE	SSIM	MS-SSIM	PSNR/dB	MI	CSS
Original images	51.32	0.22	0.36	12.58	0.37	0.34
	± 1.52	± 0.05	± 0.07	± 1.53	± 0.18	± 0.51
CycleGAN	28.43	0.33	0.49	14.38	0.58	0.39
	± 1.86	± 0.08	± 0.08	± 2.36	± 0.18	± 0.06
CycleGAN + SN + TTU + LS	14.27	0.41	0.58	16.53	0.87	0.50
(Stability-enhanced CycleGAN)	± 1.88	± 0.07	± 0.08	± 1.58	± 0.11	± 0.05
CycleGAN + SN + TTU + LS + SC	12.20	0.45	0.62	19.47	1.04	0.75
(Proposed method)	± 1.38	± 0.06	± 0.08	± 1.24	± 0.14	± 0.03

NOTE: 'SC' denotes skip the connection; 'X' denotes images like Fig. 8(a); 'Y' denotes images like Fig. 8(b); 'Y' denotes transformed images like Fig. 8(c), (d).

the transformed result, which might be attributed to the forcible residual learning applied from input to output of the Residual CycleGAN. In comparison, as shown in Fig. 9(f), our proposed method applies residual blocks between the encoder and decoder of generators with supplementary skip connections, which can integrate feature information extracted in different levels appropriately and greatly alleviate the over-smoothing problem. In Fig. 9(e), the over-smoothing appearance reduces to a certain extent but still exists in contrast to Fig. 9(f). As labeled with the red ovals in Fig. 9(c)-(e), there are deformations of textures and tissue boundary in the transformation results produced by other methods, while the structural details of Fig. 9(f) are more distinct and accurate. The better detail preservation benefits from the addition of the

skip connections in the proposed method, while in the RegGAN method, misaligned target images are considered as noisy labels and the generator is trained with an additional registration network to fit the misaligned noise distribution adaptively, which might lead to fuzzy textures or deformations.

The third row of Fig. 9 illustrates distance maps of different methods between the transformed image and corresponding reference one to show the transformation effect vividly and visually. The darker the color blue, the smaller distance value between the transformed image and the reference one it has. Observing the bottom row of Fig. 9, it can be further confirmed that transformed images produced by our proposed method are closer to the reference ones. As a consequence, the proposed method has an obvious superiority of the domain transformation in ultrasound images.

Expert for qualitative contrasts, Tables 3-6 were provided to evaluate the transformation path $X \rightarrow \tilde{Y} \rightarrow \tilde{X}$ and $Y \rightarrow \tilde{X} \rightarrow \tilde{Y}$, respectively, which can reflect the overall domain transformation performance of different methods from two opposite domain transfer directions.

As shown in Tables 3 and 5, the "original image" term is calculated as a reference to show the difference between ultrasound images in two different domains. For $X \rightarrow \tilde{Y}$, comparing with the RegGAN, our proposed method achieves a decrease from 13.76 ± 1.49 to 12.20 ± 1.38 for the mean MAE value and an increase from 17.31 ± 2.04 to 19.47 ± 1.24 for the mean PSNR value. The mean SSIM, MS-SSIM, Pearson-R, MI and CSS values achieved by our proposed method are 0.53 ± 0.04 to 0.45 ± 0.06 , 0.62 ± 0.08 , 0.78 ± 0.09 , 1.04 ± 0.14 and 0.75 ± 0.03 , which are superior to those achieved by other methods. For $Y \rightarrow \tilde{X}$, the results obtained by the proposed method also surpass those gained by other algorithms in all of the above metrics.

The lower MAE and the higher PSNR, Pearson-R, and MI values obtained by our proposed method demonstrate the superior capacity of generating images similar to the target domain, which helps to reduce

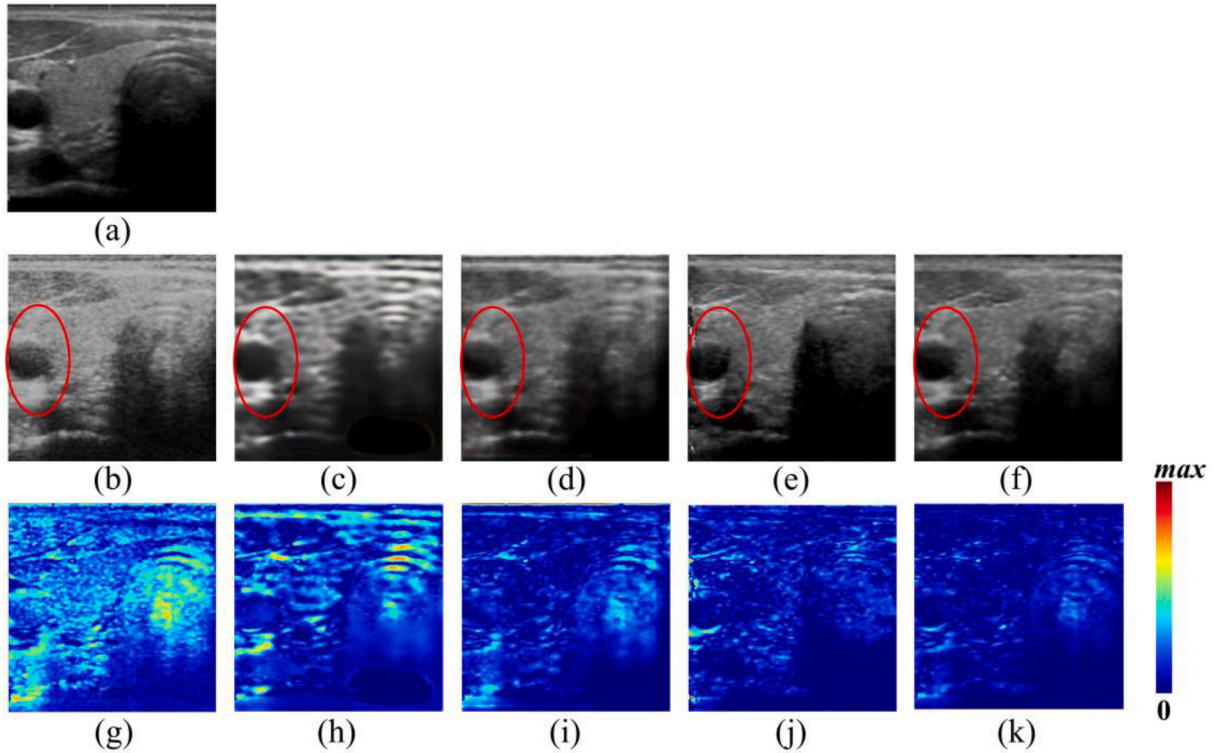


Fig. 9. Comparison between different domain transformation methods. (a) presents the reference one from the target domain scanned using Toshiba Aplio 500. (b) denotes the input from the source domain scanned using mSonics MU1. (c) to (f) display transformed images produced by SRGAN, Residual CycleGAN, RegGAN and the proposed method, respectively. (h) to (k) display corresponding distance maps of different methods between the transformed image and the reference one. (g) shows the distance map between (a) and (b) as a contrast. Red ovals are labeled to highlight the detailed difference.

Table 3Quantitative evaluation results for the domain transformation $X \rightarrow \tilde{Y}$ using different methods.

Model	$Y \leftrightarrow \tilde{Y}$						$X \leftrightarrow \tilde{Y}$
	MAE	SSIM	MS-SSIM	PSNR/dB	Pearson-R	MI	CSS
Original images	51.32 ± 1.52	0.22 ± 0.05	0.26 ± 0.07	8.58 ± 1.53	0.64 ± 0.13	0.37 ± 0.18	0.34 ± 0.51
SRGAN	21.36 ± 1.35	0.34 ± 0.09	0.47 ± 0.08	12.42 ± 1.45	0.70 ± 0.11	0.60 ± 0.12	0.57 ± 0.10
Residual CycleGAN	18.72 ± 1.66	0.40 ± 0.07	0.52 ± 0.09	15.16 ± 1.62	0.71 ± 0.12	0.91 ± 0.17	0.62 ± 0.09
RegGAN	13.76 ± 1.49	0.41 ± 0.08	0.59 ± 0.07	17.31 ± 2.04	0.72 ± 0.13	0.96 ± 0.19	0.53 ± 0.04
Our method	12.20 ± 1.38	0.45 ± 0.06	0.62 ± 0.08	19.47 ± 1.24	0.78 ± 0.09	1.04 ± 0.14	0.75 ± 0.03

Table 4Quantitative evaluation results for the reconstruction $\tilde{Y} \rightarrow \tilde{X}$ using different methods.

Model	$X \leftrightarrow \tilde{X}$					
	MAE	SSIM	MS-SSIM	PSNR/dB	Pearson-R	MI
SRGAN	—	—	—	—	—	—
Residual CycleGAN	7.52 $\pm$ 1.41	0.79 $\pm$ 0.11	0.90 $\pm$ 0.04	20.18 $\pm$ 0.69	0.96 $\pm$ 0.01	1.17 $\pm$ 0.16
RegGAN	4.66 $\pm$ 0.89	0.77 $\pm$ 0.03	0.84 $\pm$ 0.02	23.73 $\pm$ 0.54	0.94 $\pm$ 0.02	1.33 $\pm$ 0.17
Our method	4.32 ± 1.02	0.85 ± 0.04	0.97 ± 0.01	27.94 ± 0.89	0.99 0.01	1.36 ± 0.17

NOTE: $\tilde{X}$ represents reconstructed images like Fig. 8 (e), (f).

differences between different domains so as to obtain the uniformly distributed database. The higher values of the SSIM, MS-SSIM, and CSS mean the more detailed information maintained by generated images after the domain transformation, which proves the worth and importance of skip connections again. As for the Residual CycleGAN method and SRGAN, they don't achieve desired results. Transformed ultrasound images generated by the two methods are very different from reference ones, which indicate that the Residual CycleGAN and SRGAN methods are not applicable for the domain transformation in ultrasound image domain though they do well in the stain transformation and super-resolution reconstruction respectively.

For $\tilde{Y} \rightarrow \tilde{X}$, the MAE obtained by our proposed methods reduces by 7.9% and the SSIM, MS-SSIM, PSNR, Pearson-R and MI increase by 26.9%, 15.5%, 17.7%, 5.3% and 2.3% compared to the RegGAN. For $\tilde{X} \rightarrow \tilde{Y}$, our method also reaches the best achievements in all the metrics. The decrease of the MAE and increases of the SSIM, MS-SSIM, PSNR, Pearson-R and MI metrics indicate that the proposed method achieves the better reconstruction performance during the end-to-end transformation process. In a certain sense, the less loss of content and detailed

information and better reconstruction effect during the cycle process also signify the excellent domain transformation performance of our method.

To sum up, the visual observation comparison shown in Fig. 9 reveals the better transformation results produced by the proposed method. Furthermore, the proposed method gains the superior achievement on each evaluation metric compared to other domain transformation methods whether from the transformation performance perspective or the reconstruction performance perspective. All the above comparisons demonstrate the superior capacity of our proposed method to perform the domain transformation in ultrasound images, which greatly reduces discrepancies between different domains and normalize images from different medical centers. Hence, our method is an efficient way as an image preprocessing step to obtain the uniformly distributed database for data-driven researches.

5. Conclusion and future work

In this paper, a stability-enhanced CycleGAN method with well image detailed information preservation performance is proposed for

Table 6Quantitative evaluation results for the reconstruction $\tilde{X} \rightarrow \tilde{Y}$ using different methods.

Model	$Y \leftrightarrow \tilde{Y}$					
	MAE	SSIM	MS-SSIM	PSNR/dB	Pearson-R	MI
SRGAN	—	—	—	—	—	—
Residual CycleGAN	5.81 $\pm$ 0.99	0.85 $\pm$ 0.05	0.91 $\pm$ 0.05	26.18 $\pm$ 1.94	0.97 $\pm$ 0.02	1.34 $\pm$ 0.16
RegGAN	6.54 $\pm$ 1.16	0.81 $\pm$ 0.09	0.87 $\pm$ 0.12	23.51 $\pm$ 2.76	0.90 $\pm$ 0.08	1.16 $\pm$ 0.19
Our method	3.35 ± 0.93	0.90 ± 0.01	0.95 ± 0.01	31.02 ± 1.43	0.99 0.01	1.38 ± 0.15

NOTE: $\tilde{Y}$ means reconstructed images like Fig. 8(i), (j).**Table 5**Quantitative evaluation results for the domain transformation $Y \rightarrow \tilde{X}$ using different methods.

Model	$X \leftrightarrow \tilde{X}$						$Y \leftrightarrow \tilde{X}$
	MAE	SSIM	MS-SSIM	PSNR/dB	Pearson-R	MI	CSS
Original images	51.32 ± 1.52	0.22 ± 0.05	0.26 ± 0.07	8.58 ± 1.53	0.64 ± 0.13	0.37 ± 0.18	0.34 ± 0.51
SRGAN	—	—	—	—	—	—	—
Residual CycleGAN	23.41 ± 1.62	0.30 ± 0.06	0.35 ± 0.18	12.94 ± 1.52	0.70 ± 0.14	1.00 ± 0.17	0.65 ± 0.07
RegGAN	15.69 ± 1.36	0.39 ± 0.04	0.42 ± 0.08	16.25 ± 1.60	0.81 ± 0.08	1.05 ± 0.14	0.51 ± 0.04
Our method	13.65 ± 1.45	0.42 ± 0.05	0.56 ± 0.07	17.16 ± 1.10	0.93 ± 0.03	1.10 ± 0.18	0.74 ± 0.05

the domain transformation, aiming to normalize ultrasound images from different medical centers. Based on the analysis of weakness of the basic CycleGAN, the spectral normalization, two time-scale update rule and label smooth strategy are adopted to stabilize the training of the CycleGAN model. Besides, considering the inadequate constraint of the cycle-consistency to reserve the detailed information, skip connections are added to generators to refine transformation results of the proposed method. Massive experimental results demonstrate the effectiveness of optimizations mentioned above and the feasibility of the proposed method to normalize images from different domains. Using the promising method, it is easier to reduce the database bias so as to obtain the uniformly distributed database. The algorithm can further promote the potential establishment of large ultrasound image datasets. As a faithful data processing method, it might have a contribution to the development of medical big data, which enhances the reliability and universality of computer aided diagnosis.

In the future work, more ultrasound image datasets may be obtained from more different medical centers, so that “many-to-one” transformation mode can be explored to enhance the adaptability and efficiency of our proposed method. Extended applications on segmentation or classification tasks will also be conducted to further demonstrate the

value of our study.

CRediT authorship contribution statement

Lihong Huang: Conceptualization, Methodology, Software, Writing – original draft. **Zixia Zhou:** Data curation, Visualization. **Yi Guo:** Investigation, Writing – review & editing. **Yuanyuan Wang:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant 61871135 and 81830058), the Science and Technology Commission of Shanghai Municipality (Grant 20DZ1100104).

Appendix

Implementation details

All ultrasound images were resized to 256×256 . Loss weights (λ, γ) of the loss function $\mathcal{L}_{\text{CycleGAN}}$ were tuned and chosen as (10, 1). The Adam optimizer with the momentum term $\beta_1 = 0.9$ was adopted for generators and discriminators. The learning rate linearly decay every 50 iterations with the decay factor as 0.5. Data preprocessing steps were implemented using MATLAB software. The following network training process was implemented in the TensorFlow library and trained end-to-end from scratch using a Nvidia Geforce RTX 2080 TI GPU to increase the training speed. The model is trained alternating between generators and discriminators, so that the data generated by generators is getting similar to the corresponding target one. When training the model with the training set, the training data was chosen randomly from the source domain and target domain in every training epoch.

References

- [1] C. A. Linte, J. Moore, C. Wedlake and T. M. Peters, “Evaluation of model-enhanced ultrasound-assisted interventional guidance in a cardiac phantom,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2209–2218, Sept. 2010, <https://doi.org/10.1109/TBME.2010.2050886>.
- [2] P.M. Corry, K. Jabboury, E.P. Armour, J.S. Kong, Human cancer treatment with ultrasound, *IEEE Transactions on Sonics and Ultrasonics* 31 (5) (Sept. 1984) 444–456, <https://doi.org/10.1109/T-SU.1984.31529>.
- [3] B.G. Ewigman, Effect of Prenatal Ultrasound Screening on Perinatal Outcome, *The Nurse Practitioner* 18 (11) (1993) 5.
- [4] J. Kim, H.J. Kim, C. Kim, J.H. Lee, K.W. Kim, Y.M. Park, H.W. Kim, S.Y. Ki, Y. M. Kim, W.H. Kim, Weakly-supervised deep learning for ultrasound diagnosis of breast cancer, *Scientific Reports* 11 (1) (Dec. 2021) 24382, <https://doi.org/10.1038/s41598-021-03806-7>.
- [5] W. Guo, Q. Bian, S. Yang, J. Wang, Comparative analysis of clinical diagnosis and ultrasonic diagnosis between primary biliary cirrhosis and viral hepatitis cirrhosis, *Minerva Med* 113 (2) (2022).
- [6] M. Qi, S. Gao, S. Nie, K. Wang, L. Guo, Precise engineering of cetuximab encapsulated gadolinium nanoassemblies: *in vitro* ultrasound diagnosis and *in vivo* thyroid cancer therapy, *Drug Delivery* 28 (1) (Dec. 2021) 569–579, <https://doi.org/10.1080/10717544.2021.1889721>.
- [7] S. Kim K. Lee M. Park VP44.13: Prenatal ultrasound diagnosis of abnormal placental cord insertion and pregnancy outcome Ultrasound in Obstetrics & Gynecology 58 Oct. 2021 280–280 10.1002/UOG.24635.
- [8] S. Liu, Y.i. Wang, X. Yang, B. Lei, L.i. Liu, S.X. Li, D. Ni, T. Wang, Deep Learning in Medical Ultrasound Analysis: A Review, *Engineering* 5 (2) (2019) 261–275.
- [9] Y. Gao, Y. Liu, Y. Wang, Z. Shi and J. Yu, “A universal intensity standardization method based on a many-to-one weak-paired cycle generative adversarial network for magnetic resonance images,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2059–2069, Sept. 2019, <https://doi.org/10.1109/TMI.2019.2894692>.
- [10] T. de Bel, J.-M. Bokhorst, J. van der Laak, G. Litjens, Residual cyclegan for robust domain transformation of histopathological tissue slides, *Medical Image Analysis* 70 (2021) 102004.
- [11] N. Tajbakhsh J.Y. Shin S.R. Gurudu R. Todd Hurst C.B. Kendall M.B. Gotway J. Liang Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Transactions on Medical Imaging* 35 5 10.1109/TMI.2016.2535302. 1299 1312 May 2016 <https://doi.org/>.
- [12] L.G. Nyul J.K. Udupa Xuan Zhang New variants of a method of MRI scale standardization *IEEE Trans. Med. Imaging* 19 2 143 150.
- [13] F. Jäger, Y. Deurling-Zheng, B. Frericks, F. Wacker and J. Hornegger, “A new method for MRI intensity standardization with application to lesion detection in the brain,” in *Proceedings of the International Fall Workshop Vision Modeling and Visualization*, Aachen, Germany, November 22–24, 2006, pp. 269–276, <https://doi.org/10.1.1.638.8032>.
- [14] D. Shen, G. Wu, H. Suk, Deep learning in medical image analysis, *Annual Review of Biomedical Engineering* 19 (1) (Jun. 2017) 221–248, <https://doi.org/10.1146/annurev-bioeng-071516-044442>.
- [15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, S. Ozair, A. Courville and Y. Bengio, “Generative adversarial nets,” in *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, Montreal, QC, Canada, Dec. 2014, vol. 2, pp. 2672–2680, <https://doi.org/10.48550/arXiv.1406.2661>.
- [16] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” Nov. 2014, Preprint, <https://doi.org/10.48550/arXiv.1411.1784>.
- [17] P. Isola, J. Zhu, T. Zhou and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 21–26, 2017, pp. 5967–5976, <https://doi.org/10.1109/CVPR.2017.632>.
- [18] J. Zhu, T. Park, P. Isola and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computing Vision (ICCV)*, Venice, Italy, Oct. 22–29, 2017, pp. 2242–2251, <https://doi.org/10.1109/ICCV.2017.244>.
- [19] S. Liu, B. Zhang, Y. Liu, A. Han, H. Shi, T. Guan, Y. He, Unpaired stain transfer using pathology-consistent constrained generative adversarial networks, *IEEE Transactions on Medical Imaging* 40 (8) (Aug. 2021) 1977–1989, <https://doi.org/10.1109/TMI.2021.3069874>.
- [20] L. Kong, C. Lian, D. Huang, Z. Li, Y. Hu, Q. Zhou, Breaking the dilemma of medical image-to-image translation. in *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, Oct. 2021.
- [21] O. Ronneberger P. Fischer T. Brox U-net: convolutional networks for biomedical image segmentation 2015 Munich, Germany, May 234 241 10.1007/978-3-319-24574-4_28.
- [22] V. Nair G.E. Hinton Rectified linear units improve restricted boltzmann machines 2010 Haifa, Israel, Mar 807 814.
- [23] K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las

- Vegas, NV, USA, June 27-30, 2016, pp. 770-778, <https://doi.org/10.1109/CVPR.2016.90>.
- [24] A. Odena, V. Dumoulin, C. Olah, *Deconvolution and Checkerboard Artifacts*, *Distill* 1 (10) (2016).
- [25] C. Li M. Wand Precomputed real-time texture synthesis with markovian generative adversarial networks 2016 Amsterdam, The Netherlands 702 716.
- [26] A.L. Maas A.Y. Hannun A.Y. Ng Rectifier nonlinearities improve neural network acoustic models 2013 Atlanta, GA, USA 1 3.
- [27] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang and S. P. Smolley, "Least squares generative adversarial networks," in IEEE International Conference on Computer Vision (ICCV). Venice, Italy, Oct. 22-29, 2017, pp. 2813-2821, <https://doi.org/10.1109/ICCV.2017.304>.
- [28] X. Mao Q. Li H. Xie R.Y.K. Lau Z. Wang Multi-class generative adversarial networks with the L2 loss function 2017 Venice, Italy 2813 2821.
- [29] Y. Qin N. Mitra P. Wonka How does lipschitz regularization influence GAN training? 2020 Cham, Switzerland 310 326 [10.1007/978-3-030-58517-4_19](https://doi.org/10.1007/978-3-030-58517-4_19).
- [30] Y. Yu, J. Sun, S. GE and Q. Yang., CycleGAN-SN: Image stylization algorithm combining spectral normalization and CycleGAN, *Journal of Xi'An Jiaotong University* 54 (5) (2020) 133-141, <https://doi.org/10.7652/xjtubxb202005018>.
- [31] T. Miyato, T. Kataoka, M. Koyama and Y. Yoshida, "Spectral normalization for generative adversarial networks," in Proceedings of the International Conference on Learning Representations, Vancouver, Canada, Apr. 30 - May. 3, 2018, pp. 1-26, <https://doi.org/10.48550/arXiv.1802.05957>.
- [32] K. Kurach, M. Lucic, X. Zhai, M. Michalski and S. Gelly, "A large-scale study on regularization and normalization in GANs," in Proceedings of the International Conference on Machine Learning, Long Beach, California, Jun. 9-15, 2019, pp. 2472-2478, <https://doi.org/10.48550/arXiv.1807.04720>.
- [33] M. Arjovsky S. Chintala L. Bottou Wasserstein generative adversarial networks vol. 70 2017 Princeton, NJ, USA 214 223.
- [34] M. Heusel H. Ramsauer T. Unterthiner B. Nessler GANs trained by a two time-scale update rule converge to a nash equilibrium 2017 NY, USA 6629 6640.
- [35] A. Radford L. Metz S. Chintala Unsupervised representation learning with deep convolutional generative adversarial networks S. Juan P. Rico . Jun. in Proceedings of the International Conference on Learning Representations 2016.
- [36] I. Gulrajani F. Ahmed M. Arjovsky V. Dumoulin A. Courville Improved training of wasserstein GANs 2017 NY, USA 5769 5779.
- [37] S.J. Pan, Q. Yang, *A Survey on Transfer Learning*, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010) 1345-1359.
- [38] R. Shekhar, V. Zagrodsky, Mutual information-based rigid and nonrigid registration of ultrasound volumes, *IEEE Transactions on Medical Imaging* 21 (1) (Jan. 2002) 9-22, <https://doi.org/10.1109/42.981230>.
- [39] S. Klein, M. Staring, J.P.W. Pluim, Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines, *IEEE Transactions on Image Processing* 16 (12) (Dec. 2007) 2879-2890, <https://doi.org/10.1109/TIP.2007.909412>.
- [40] 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 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, pp. 105-114, <https://doi.org/10.1109/CVPR.2017.19>.
- [41] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in IEEE International Conference on Pattern Recognition (CVPR), Istanbul, Turkey, Aug. 23-26, 2010, pp. 2366-2369, <https://doi.org/10.1109/ICPR.2010.579>.
- [42] Q. Huynh-Thu, M. Ghanbari, Scope of validity of PSNR in image/video quality assessment, *Electronics Letters* 44 (13) (Jun. 2008) 800-801, <https://doi.org/10.1049/el:20080522>.
- [43] J. Li, X. Zhang and M. Ding, "Image quality assessment based on regional mutual information," in International Conference on Intelligent Computation and Bio-Medical Instrumentation, Wuhan, Hubei, China, Dec. 14-17, 2011, pp. 113-115, <https://doi.org/10.1109/ICBML.2011.65>.
- [44] Z. Zhou, Y.i. Guo, Y. Wang, *Ultrasound deep beamforming using a multiconstrained hybrid generative adversarial network*, *Medical Image Analysis* 71 (2021) 102086.