

Tunable Image Quality Control of 3-D Ultrasound using Switchable CycleGAN

Jaeyoung Huh^a, Shujaat Khan^a, Sungjin Choi^b, Dongkuk Shin^b, Eun Sun Lee^{*c}, Jong Chul Ye^{*a}

^aDepartment of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

^bSystem R&D Group, Samsung Medison Co., Ltd., Seoul, Korea

^cDepartment of Radiology, Chung-Ang University Hospital, 102 Heukseok-ro, Dongjak-gu, Seoul 06973, Korea

ARTICLE INFO

Keywords: 3-D Ultrasound Imaging, Deep learning, Adaptive Instance normalization (AdaIN), Obstetrics and gynecology

ABSTRACT

In contrast to 2-D ultrasound (US) for uniaxial plane imaging, a 3-D US imaging system can visualize a volume along three axial planes. This allows for a full view of the anatomy, which is useful for gynecological (GYN) and obstetrical (OB) applications. Unfortunately, the 3-D US has an inherent limitation in resolution compared to the 2-D US. In the case of 3-D US with a 3-D mechanical probe, for example, the image quality is comparable along the beam direction, but significant deterioration in image quality is often observed in the other two axial image planes. To address this, here we propose a novel unsupervised deep learning approach to improve 3-D US image quality. In particular, using *unmatched* high-quality 2-D US images as a reference, we trained a recently proposed switchable CycleGAN architecture so that every mapping plane in 3-D US can learn the image quality of 2-D US images. Thanks to the switchable architecture, our network can also provide real-time control of image enhancement level based on user preference, which is ideal for a user-centric scanner setup. Extensive experiments with clinical evaluation confirm that our method offers significantly improved image quality as well user-friendly flexibility.

© 2021

1. Introduction

In contrast to computed tomography (CT) and magnetic resonance imaging (MRI), ultrasound imaging (US) offers real-time imaging without any radiation risk. As such, US imaging is useful in many clinical areas, including gynecology (GYN) and obstetrics (OB).

Commonly used 2-dimensional (2-D) US systems capture images in a single plane along the direction of incidence of waves and deliver a high-quality image with the aid of a dense 1-D transducer array. On the other hand, 3-D US imaging systems can create 3-D images by acquiring three-axis data using mechanical scanning of 1-D transducers or 2-D array transducers (see Fig. 1(b)(c)).

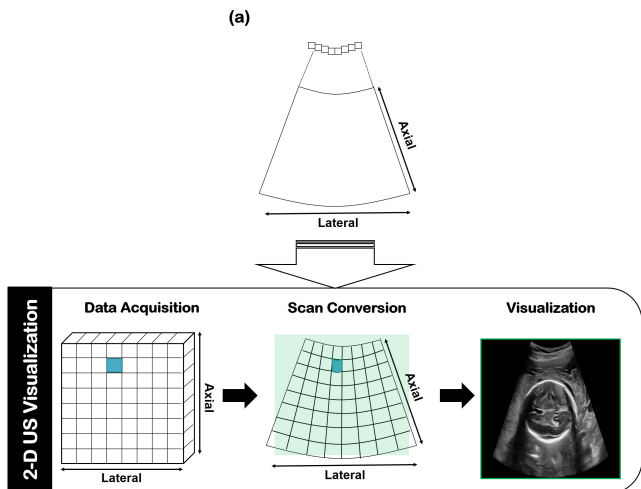
Unfortunately, when compared to a 2-D imaging system, 3-D US has fundamental limitations in terms of its resolution. For the case 3-D US using 2-D array transducer as shown in Fig. 1(c), a large number of piezo-elements are necessary to obtain high-resolution 3-D volume. However, this increases the bandwidth of the RF data considerably, so that a sparse array is often used at the expense of reduced image resolution. Al-

though 3-D US with 3-D mechanical probe using 1-D transducer scanning as shown in Fig. 1(b) may alleviate this, it also suffer from anisotropic resolution originated from the mechanical scanning. Specifically, while the axial-lateral plane, called the A-plane image, has a comparably high quality, the axial-elevation planes called B and the elevation-lateral planes called C-plane show degraded resolution. Therefore, quality improvement for 3-D US has become an important issue for ultrasound manufacturers. In 3-D US with a 3-D mechanical probe, which is the main imaging platform in our study, one of the simplest approaches to address this anisotropy in the resolution is to acquire more data along the elevation and lateral direction at the expense of the scan time. Super-resolution approaches in the 3-D volume domain may be another way to improve resolution, but they may increase the computational burden significantly, which is not appropriate for real-time imaging.

Recently, various authors have proposed image enhancement algorithms and beamformers that use deep neural networks (Yoon et al. (2019); Khan et al. (2020a); Hyun et al. (2019); Nair et al. (2018); Brickson et al. (2021); Solomon et al. (2019); Sharifzadeh et al. (2020); Vedula et al. (2017); Perdios et al. (2017, 2018); Van Sloun et al. (2019); Luchies and Byram (2018); Khan et al. (2019); Kokil and Sudharson (2020); Luijten et al. (2020); Sadeghi et al. (2021); Zuo et al. (2021); Huh et al. (2021); Khan et al. (2021a,b, 2020b); Dietrichson et al.

*Corresponding author: E-mail address : seraph377@cau.ac.kr (E.S.Lee); jong.ye@kaist.ac.kr (J.C.Ye)

2-D US imaging



3-D US imaging

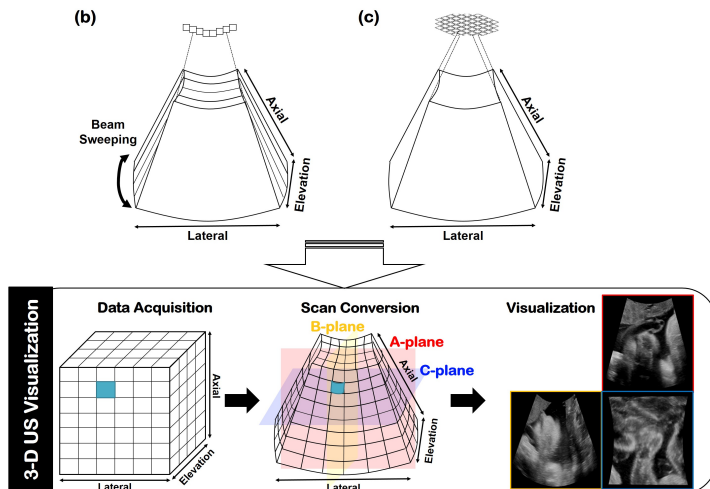


Fig. 1. US image generation steps. In the top row, various form of 2-D and 3-D probes are presented. The second row presents the three step procedure of US imaging: data acquisition, scan-conversion, and visualization. (a) Convex array probe for 2-D US imaging, (b) 3-D mechanical probe by scanning 1-D convex array, and (c) 2-D matrix array for 3-D US imaging.

(2018)). Most of the existing approaches are based on supervised learning that requires matched ground-truth data (Yoon et al. (2019); Khan et al. (2020a); Hyun et al. (2019); Nair et al. (2018); Brickson et al. (2021); Solomon et al. (2019); Sharifzadeh et al. (2020); Vedula et al. (2017); Perdios et al. (2017, 2018); Van Sloun et al. (2019); Luchies and Byram (2018); Khan et al. (2019); Kokil and Sudharson (2020); Luijten et al. (2020); Sadeghi et al. (2021); Zuo et al. (2021)). Unfortunately, this approach is not suitable for our 3-D US image enhancement problem as there are no matched high-resolution 3-D US images that can be used as ground truths for supervision. Another important limitation of existing deep learning approaches is that once the neural network is trained, it is designed to produce the same level of improvement. So if another level of improvement is required, additional training is required. This is an important disadvantage that prevents their widespread acceptance in the clinical settings, as many radiologists may have different preferences for the level of image enhancement, and the common practice is to adjust the algorithm parameters according to their preference. Therefore, it would be very useful if deep learning-based image enhancement could also provide real-time control of image quality so that radiologists can choose their preferred setup.

To address this problem, here we propose a novel 3-D ultrasound image enhancement technique that can provide real-time image enhancement and quality control based on user preferences, and that can be trained with unmatched high quality 2-D ultrasound images. In particular, our methodology is based on the recently proposed switchable CycleGAN architecture (Gu and Ye (2021); Yang et al. (2021)), which has been used successfully for various unsupervised learning problems. In contrast to the standard CycleGAN with two different generators (Zhu et al. (2017)), the switchable CycleGAN (Gu and Ye (2021); Yang et al. (2021)) has a single generator whose role can be controlled by Adaptive Instance Normalization (AdaIN) code (Huang and Belongie (2017)). In addition to the reduced complexity due to the common generator, one of the most im-

portant by-products of the switchable CycleGAN is its tuneability. In particular, once the network is trained, we can provide a real-time control of the level of image enhancement by simply interpolating the AdaIN codes (Gu and Ye (2021); Yang et al. (2021)). Therefore, this architecture successfully addresses two fundamental limitations of the existing deep learning approaches for US: lack of matched reference data and user preference-based control. Extensive experiments with clinical evaluation also confirm that the proposed method offers significantly improved image quality compared to the original 3-D scan.

This paper is structured as follows. Section 2 provides necessary backgrounds and explains the proposed method. The implementation details of the algorithm are presented in Section 3. Section 4 provides reconstruction results using various examples, followed by a discussion in Section 5 and a conclusion in Section 6.

2. Main Contribution

2.1. 3-D Ultrasound Imaging Systems

US imaging is composed of three steps: data acquisition, scan conversion and visualization. The data acquisition step is to collect the radio-frequency (RF) data with a 2-D or 3-D probe. Scan conversion refers to transforming the shape of the collected RF data into a predefined shape. The visualization step creates the final image for display.

As shown in Fig. 1(a), a 2-D system receives RF data using a 1-D array probe in a plane that has lateral and axial axes. On the other hand, a 3-D US system captures 3-D information with lateral, axial and elevation axes. In practice, obtaining 3-D RF data can be achieved by two representative probes: 2-D array probe (Yen et al. (2000); Huang and Zeng (2017)), and 3-D mechanical probe (Huang and Zeng (2017); Fenster et al. (2001); Fenster and Downey (1996); Fenster et al. (2011)). The 2-D array transducer generates a 3-D convex waveform and then collects return echoes. On the other hand, a 3-D mechanical probe

with a 1-D array transducer provides a 2-D convex waveform and sweep to acquire multiple 2-D RF data along the elevation axis.

Although the 3-D US system provides volumetric information with a single attempt, it should be noted that with a 3-D imaging system, compared to a 2-D system, less information is captured in each plane. Specifically, it should take tremendous information in a short period of time, but there is a limitation in increasing the piezo-elements due to the transmission bandwidth. For this reason, less information is used to form the image of each slice, which causes blurring throughout the image. In particular, 3-D US systems with 3-D mechanical probe, which is the target system of our study, the elevation axis consists of more sparse data, so that the axial-elevation plane and the elevation-lateral plane have a lower quality than the axial-lateral plane. This fundamental difference inevitably leads to a deterioration in the quality of the visualization process.

Specifically, as shown in Fig. 1, the red-colored plane, the axes of which are axial-lateral forms the A-plane, which is similar to the 2-D acquisition mode, but of lower quality. The plane with the axial-elevation axes, the color of which is yellow, forms the B-plane, whereas the plane of the elevation-lateral axes with blue color forms the C-plane. As mentioned before, the B-plane and C-plane are reconstructed from a smaller amount of RF data compared to A-plane, so they lose some detail and suffer from some blurring artifact.

2.2. Unsupervised 3-D US Enhancement Using 2-D US data

Due to the aforementioned fundamental limitations of 3-D scanning, image quality degradation compared to 2-D US is inevitable. To address this problem, we are interested in improving 3-D US image quality by learning image quality of 2-D US images. As shown in Fig. 2, the neural network input is a 2-D image from any plane in 3-D volume, such as A-, B-, or C-planes, and our goal is to improve their image quality by learning image quality from 2-D US images.

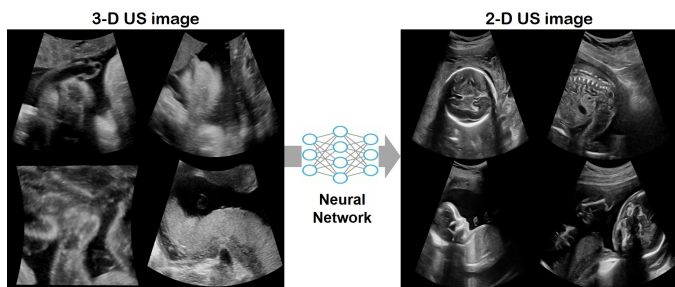


Fig. 2. Unsupervised learning to learn 2-D US image quality using A-, B-, and C- plane images from 3-D US.

Unfortunately, obtaining image from same area under the same condition using 2-D and 3-D probes is impossible, therefore supervised training with the paired dataset is not feasible. For the unsupervised training with unpaired data set, one of the most powerful tools is the CycleGAN architecture (Zhu et al. (2017)). Originally proposed for style-transfer for computer vision applications, CycleGAN has been successfully used for various bio-medical imaging fields such as CT (Kang et al. (2019); Li et al. (2019); Gu and Ye (2021); Yang et al. (2021);

Gu et al. (2021)), MRI (Oh et al. (2020); Chung et al. (2021a); Cha et al. (2020); Modanwal et al. (2020, 2021)), ultrasound (Khan et al. (2021b); Huh et al. (2021)), and optics (Chung et al. (2021b); Lim et al. (2020); Lee et al. (2019)). Moreover, it was shown that CycleGAN learns the optimal transport map that can transport one probability distribution to another (Sim et al. (2020)). Accordingly, when applied to our problem, CycleGAN can be trained to transport the distribution of A-, B-, and C-plane images from 3-D US to that of the high-resolution 2-D US images.

Furthermore, it was recently shown that CycleGAN can be synergistically combined with Adaptive Instance Normalization (AdaIN) (Huang and Belongie (2017)) so that only a single generator can be used as a forward and inverse generator (Gu and Ye (2021); Yang et al. (2021)). By combining with a light weighted AdaIN code generator, this architecture not only reduces the number of trainable parameters to make the training more stable, but also enables the tuneability which enables the inference-time control of the neural network output as demonstrated in low-dose CT denoising (Gu and Ye (2021)) and CT kernel conversion (Yang et al. (2021)). In the following, we provide more details of our switchable CycleGAN architecture.

2.3. Switchable Generator using AdaIN Layers

Fig. 3 illustrates the proposed switchable CycleGAN network for 3-D US image enhancement. The network architecture consists of a single generator G_θ and an AdaIN code generator, F_ζ , which are parameterized by θ and ζ , respectively. The main generator G_θ in the upper left part converts US images in 3-D quality into a 2-D quality domain according to the AdaIN code created from F_ζ . The right part generator returns the fake 2-D quality images to the original 3-D US quality domain, using a constant AdaIN code K that corresponds to the instance normalization. There are also two discriminators D_ϕ and D_φ , parameterized by ϕ and φ , respectively, as in the conventional CycleGAN.

The main mathematical origin of replacing the two generators with a single baseline network with two specific AdaIN codes is from the optimal transport theory (Villani (2009); Peyré et al. (2019)). Specifically, it was shown that the AdaIN transform corresponds to the optimal transport between two uncorrelated Gaussian probability distribution (Mroueh (2020)). More specifically, the AdaIN layers take a mean vector c_μ and a standard deviation vector c_σ as input:

$$\mathcal{T}(z, F_\zeta(c)) = \frac{c_\sigma \mathbf{1}}{\sigma(z)} (z - \mu(z)\mathbf{1}) + c_\mu \mathbf{1}, \quad (1)$$

with

$$(c_\mu, c_\sigma) = F_\zeta(c) \quad (2)$$

where F_ζ is the AdaIN code generator, and the c is its input vector. Then, this corresponds to the optimal transport from the distribution $z \sim \mathcal{N}(\mu(z), \sigma^2(z)\mathbf{I})$ to another distribution $\mathcal{N}(c_\mu \mathbf{1}, c_\sigma^2 \mathbf{I})$, where $\mathcal{N}(\cdot, \cdot)$ refers to the Gaussian distribution (Mroueh (2020)).

Accordingly, for the 3-D US quality images to high-quality 2-D US images, we use an auto-encoder as a baseline network

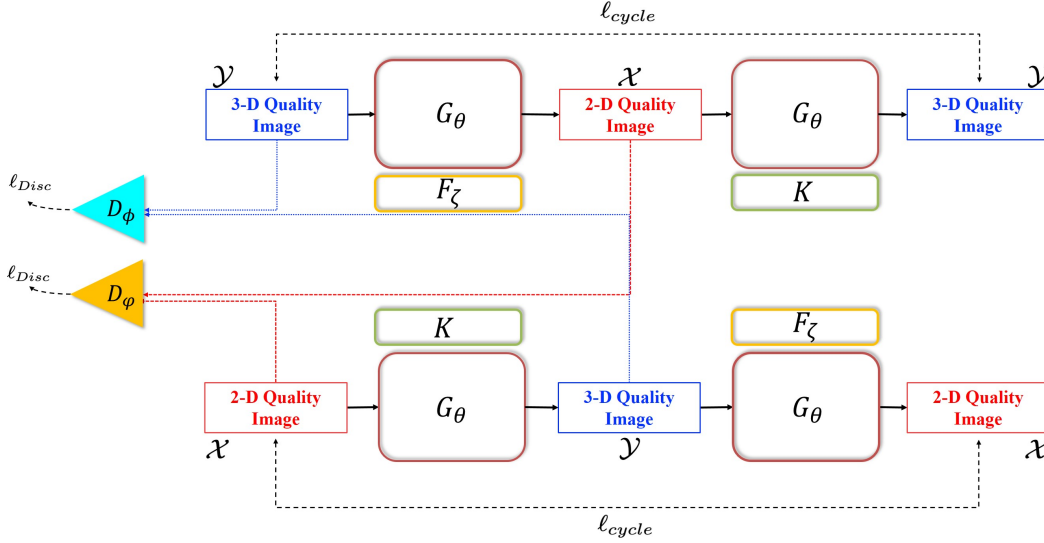


Fig. 3. Overview of the switchable CycleGAN for 3-D ultrasound image enhancement. The generator G_θ generates the target 2-D quality image from A-, B-, and C- plane images from 3-D US using the AdaIN code created from the AdaIN code generator, F_ζ . The improved images are then transformed into original image quality through G_θ with constant AdaIN code K . The discriminators D_θ, D_ϕ determine whether the image is real or fake.

and then use AdaIN transform to transport the auto-encoder features to the 2-D US image features. Similarly, for the translation from high-quality 2-D US images to 3-D US quality images, the auto-encoder features are then transported to the 3-D US quality image features using another AdaIN transform.

In particular, during the 3-D US enhancement task, the mean and variance vectors for the AdaIN layers are from AdaIN code generators, whereas for the generator to converting 2-D quality images to 3-D quality images the AdaIN code is set to a constant vector K composed of the mean and standard deviation vector set to zero and one, respectively:

$$K = (0, 1) \quad (3)$$

At the inference phase, we use the trained auto-encoder with the trained AdaIN code generator as the 3-D US image enhancement network so that it gives us an advantage of controlling image quality at the inference phase by interpolating the generated AdaIN code and the constant vector K :

$$H(c, \alpha) = (1 - \alpha)K + \alpha F_\zeta(c), \quad 0 \leq \alpha \leq 1 \quad (4)$$

where $H(\cdot, \cdot)$ is an interpolated AdaIN code between K and $F_\zeta(c)$. While the large α value implies more transition to the 2-D quality features, the small α more preserves 3-D quality features. Accordingly, the intermediate level image enhancement results along the continuous optimal transport path between the domain \mathcal{Y} and \mathcal{X} are obtained by (Gu and Ye (2021); Yang et al. (2021))

$$\mathbf{x}_\alpha = G_\theta(\mathbf{y}, H(c, \alpha)). \quad (5)$$

depending on user-preference, where \mathbf{y} denotes the 3-D US quality input images from A-, B-, or C- planes.

2.4. Network Training

The network training was similar to the conventional CycleGAN. Specifically, the total loss function is represented by

$$\min_{\theta, \zeta} \max_{\phi, \varphi} \ell_{total}(\theta, \zeta; \phi, \varphi), \quad (6)$$

where

$$\ell_{total}(\theta, \zeta; \phi, \varphi) := \lambda_{cyc} \ell_{cycle}(\theta, \zeta) + \ell_{Disc}(\theta, \zeta; \phi, \varphi) + \lambda_{iden} \ell_{iden}(\theta, \zeta) \quad (7)$$

where the cycle-consistency term is given by

$$\ell_{cycle}(\theta, \zeta) = \mathbb{E}_{\mathbf{y} \sim P_y} [\|\mathbf{y} - G_\theta(G_\theta(\mathbf{y}, F_\zeta(c)), K)\|_1] + \mathbb{E}_{\mathbf{x} \sim P_x} [\|\mathbf{x} - G_\theta(G_\theta(\mathbf{x}, K), F_\zeta(c))\|_1], \quad (8)$$

and the discriminator term is composed of LSGAN loss (Mao et al. (2017)):

$$\begin{aligned} \ell_{Disc}(\theta, \zeta; \phi, \varphi) &= \mathbb{E}_{\mathbf{y} \sim P_y} [\|D_\phi(\mathbf{y})\|_2^2] + \mathbb{E}_{\mathbf{x} \sim P_x} [\|1 - D_\phi(G_\theta(\mathbf{x}, K))\|_2^2] \\ &+ \mathbb{E}_{\mathbf{x} \sim P_x} [\|D_\phi(\mathbf{x})\|_2^2] + \mathbb{E}_{\mathbf{y} \sim P_y} [\|1 - D_\phi(G_\theta(\mathbf{y}, F_\zeta(c)))\|_2^2]. \end{aligned} \quad (9)$$

and the identity term is given by

$$\ell_{iden}(\theta, \zeta) = \mathbb{E}_{\mathbf{y} \sim P_y} [\|\mathbf{y} - G_\theta(\mathbf{y}, K)\|_1] + \mathbb{E}_{\mathbf{x} \sim P_x} [\|\mathbf{x} - G_\theta(\mathbf{x}, F_\zeta(c))\|_1]. \quad (10)$$

During the training, the discriminator D_ϕ tries to differentiate between the real image \mathbf{x} and the generated image $G_\theta(\mathbf{y}, F_\zeta(c))$ to confirm whether it is real or fake. Similarly, D_ϕ tries to find real or fake values between \mathbf{y} and $G_\theta(\mathbf{x}, K)$. It should be noted that the proposed method has the inter-dependent G_θ and F_ζ so that they should be trained at the same time.

3. Implementation Details

3.1. Network Architecture

As shown in Fig. 4(a), we employed the simple U-Net architecture with residual learning. It is composed of three down-sampling and up-sampling steps. The convolution with stride 2 is utilized for down-sampling operation. The bilinear-upsampling with a single convolution layer is used for up-sampling operation. In the green box, we used the instance

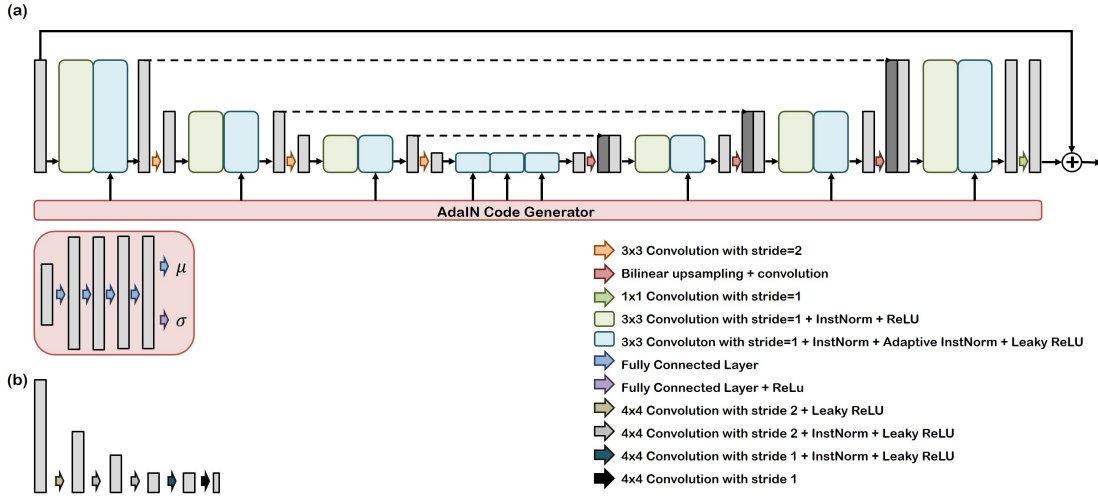


Fig. 4. The proposed network architecture. (a) The proposed generator architecture. (b) The proposed discriminator architecture.

Table 1. Image Quality Scoring Criteria

Score	Conventional US artifact	Contrast	Overall quality	Score	Blurring by denoising
5	Subtle	Excellent	Excellent	2	Substantial (negative effect on image quality)
4	Mild	Good	Good		
3	Moderate	Moderate	Moderate	1	Minimal
2	Severe	Poor	Poor		
1	Non-diagnostic	Non-diagnostic	Non-diagnostic	0	None

normalization instead of the batch normalization. The AdaIN operation is applied in each step with instance normalization where marked as blue box.

The discriminator is shown in Fig. 4(b). We used PatchGAN which classifies the local image whether it is real or fake. It is simply composed of three convolution sets of 4×4 kernel size with stride 2 and LeakyReLU. The fourth convolution is same as the previous three convolution steps except for the stride size. In the last layer, there is only 4×4 convolution with stride 1.

The AdaIN code generator, F_{ζ} , receives constant one vector of 1×128 dimension as input, which is converted into AdaIN code that has the statistical information of 2-D quality domain. The AdaIN code generator is simply composed of four fully connected layers. It has multi-head structure: one convolution layer for mean and the other layer for variance. There is an additional ReLU step for the variance to avoid the negative value. While the four fully connected layers share their parameters, the last multi-head layer has its own parameters to generate a value according to the number of channels in each step. In this study, the output of AdaIN code generator is 9 pairs of mean and variance.

3.2. Dataset

The datasets are composed of US image volumes from gynecology ($n=208$) and obstetrics ($n=115$) for routine clinical check. All subjects underwent US examinations through WS80A, HERA W10 US machines (Samsung Medison, Seoul,

Korea) using CV1-8AD(1~8 MHz), CV1-8A(1~8MHz), V4-8(4~8MHz) US probes for GYN examinations and V5-9(5~9MHz), EV3-10B(3~10MHz) US probes for OB examinations, respectively.

Because OB data showed severe heterogeneity in image characteristics according to the fetal age, position and body part, we decided to use only the GYN data as training input data and all OB volumes as test data. We divided a total of 208 GYN volumes into 100 for training, 15 for validation and 93 for test. In each volume, we chose the plane along the three axes around the center of the volume. A total of 18,424 images were used for training input and 630 images for validation. In the case of OB, we used 13 volumes with 567 images for validation and 102 volumes for test. In particular, we tried to acquire different types of images in order to cover the broad range of structural properties for the use of GYN training set and OB validation dataset. For example, we used images of bilateral ovaries and uterus for GYN training set. In terms of OB validation set, the whole body images of the first trimester were used, whereas fetal abdomen, extremities, head and spinal images were selected in case of the second trimester. However, to fair comparison, we used only ovary region for GYN and head, face region for OB to evaluate the method.

The target datasets consist of 2-D US images acquired from WS80A, HERA W10 using C2-6(2~6MHz), CA1-7A(1~7MHz). Only OB datasets are used as these are more suitable for the purpose of this research such as artifact suppression and sharp boundary. The images are captured by various

structures and status. We selected 14,995 images for the target dataset.

3.3. Training details

To train the network, we used the Adam optimizer with $\beta = 0.5$. The learning rate was started from $1e-4$ and decreased linearly after 10 epochs. We set the total epoch as 50 and selected the 15 epoch which show the best result. We implemented with patch-processing whose size is 256 and the batch size was 1. To avoid the over-fitting problem, we used data augmentation technique like flipping, and rotating. The hyper-parameters were set as 10, 5 for λ_{cyc} and λ_{iden} , respectively. We implemented all work on the Python 3.6.12 with Tensorflow 1.14.0 and MATLAB 2017a. We trained the network with the NVIDIA GeForce GTX 1080 Ti GPU.

3.4. Evaluation metric

A 15-year experienced board-certified abdominal radiologist (ESL) evaluated the result of the proposed method using a test set in a blinded manner. Note that the test set was completely invisible during training. We acquired 93 volumes for quality assessment, which confined to the ovaries for the consistency of evaluation regardless of subjects' factor as possible. Evaluation statistics was implemented for input and result from $\alpha = 0.6, 0.7, 0.8, 0.9$.

There are four categories for assessing image quality: 1) conventional US artifacts, including reverberation, side-lobes, beam thickness and etc; 2) image contrast, for example, inter-organ contrast (ovaries and adjacent structures) and intra-organ contrast (focal lesion contrast within ovaries); 3) blurring by denoising process, which means blurred textures and boundaries frequently seen after usual noise reduction; 4) overall image quality, determined subjectively. The simple evaluation criteria and scoring system are listed in the Table. 1. All statistical analysis were carried out with MedCalc version 20.015 (MedCalc software).

4. Experimental Results

4.1. Gynecology Results

To validate our proposed method, we applied our algorithms to real gynecology (GYN) clinical data sets. The results are shown in Fig. 5. The first row shows the input image. From the second to the last row, each row refers to the outputs of the proposed network with different α values. The left column is the A-plane, which is cut along the elevation axis, the second and third are the B-plane and the C-plane, respectively, which are cut along the lateral axis and axial axis, respectively. We have presented the images from $\alpha = 0.6$ to $\alpha = 0.9$, the change being clearly visible.

As shown in the Fig. 5, the input image is gradually changed to the target style depending on α , which can be again verified from the magnified area. The follicle boundaries are not clear from the input image. However, the proposed method connects the structures smoothly. The contrast is increased so that the follicle borders can be clearly seen. In addition, tearing artifacts in the B-plane is well suppressed and their resolution appears to

be increased. The side lobe artifacts, which is one of the basic artifact in US image, are clearly visible in the input image. On the other hand, the proposed method has suppressed it and its boundaries become sharp and easily discernible as the value of α is increased.

4.2. Computational Time

Our method is based on a deep neural network, so that after training the network, the improved images can be generated in real time. The average computation time of our method is 0.139 seconds when it reconstructs the 968×968 image. The time is only calculated with GPU NVIDIA GeForce GTX 1080 Ti. We obtained the mean value of the computing time with 270 test images, half of which consist of GYN and half of OB.

4.3. Clinical Evaluation

A professional abdominal radiologist evaluated the proposed method according to the criteria of image analysis. We took about 15 images of the A-, B- and C- planes around the ovarian region. It was processed with the proposed method using $\alpha = 0.6, 0.7, 0.8, 0.9$. Then we randomly shuffled the set. We used 93 volumes for clinical test, so there were 1395 mixed sets for clinical evaluation.

We implemented the Friedman test with the Bonferroni correction (Friedman (1937); Bonferroni (1936)). The evaluation results are shown in Fig. 6. The first and second graphs show the artifact and contrast rating. The third and fourth graphs represent degree of blurring and the overall quality score. According to artifact and contrast graphs, it can be easily seen that the higher the value of α , the higher the score. In particular, the entire proposed method has a high score compared to the input image. What should be noted here is that even though blurring effect occurs, the overall image quality from the proposed method overwhelms the input images.

To check how meaningful the statistical information is, we implemented a p -test with Wilcoxon paired test, which is shown in the Table. 2 (Conover (1999)). Each a, b, c, d, e denotes the input and from $\alpha = 0.6$ to $\alpha = 0.9$. Each row is the scoring category. We calculated a p -value that shows how significant the statistical differences are. The part marked in red is the statistically significant one. It should be noted that all the results of the proposed method in overall quality categories show a significant difference comparing with the input image.

4.4. Obstetric Results

We tested our algorithm with the OB images that were not used in network training. The goal was to demonstrate that our method can be robustly applied to a wide range of datasets.

In Fig. 7, we have visualized the head bone of the fetus. Each row denotes the input and its output of the proposed method with different α values. Each column shows the A-, B- and C-planes. As shown in Fig. 7, the input image has ambiguous boundaries, in particular its structure in the head bone is not easy to see. However, the result of the proposed method is that the internal structure is more visible than the input image. In addition, its boundary becomes sharp as the value of α is increased. In particular, the side-lobe artifact is well suppressed according to the α variation.

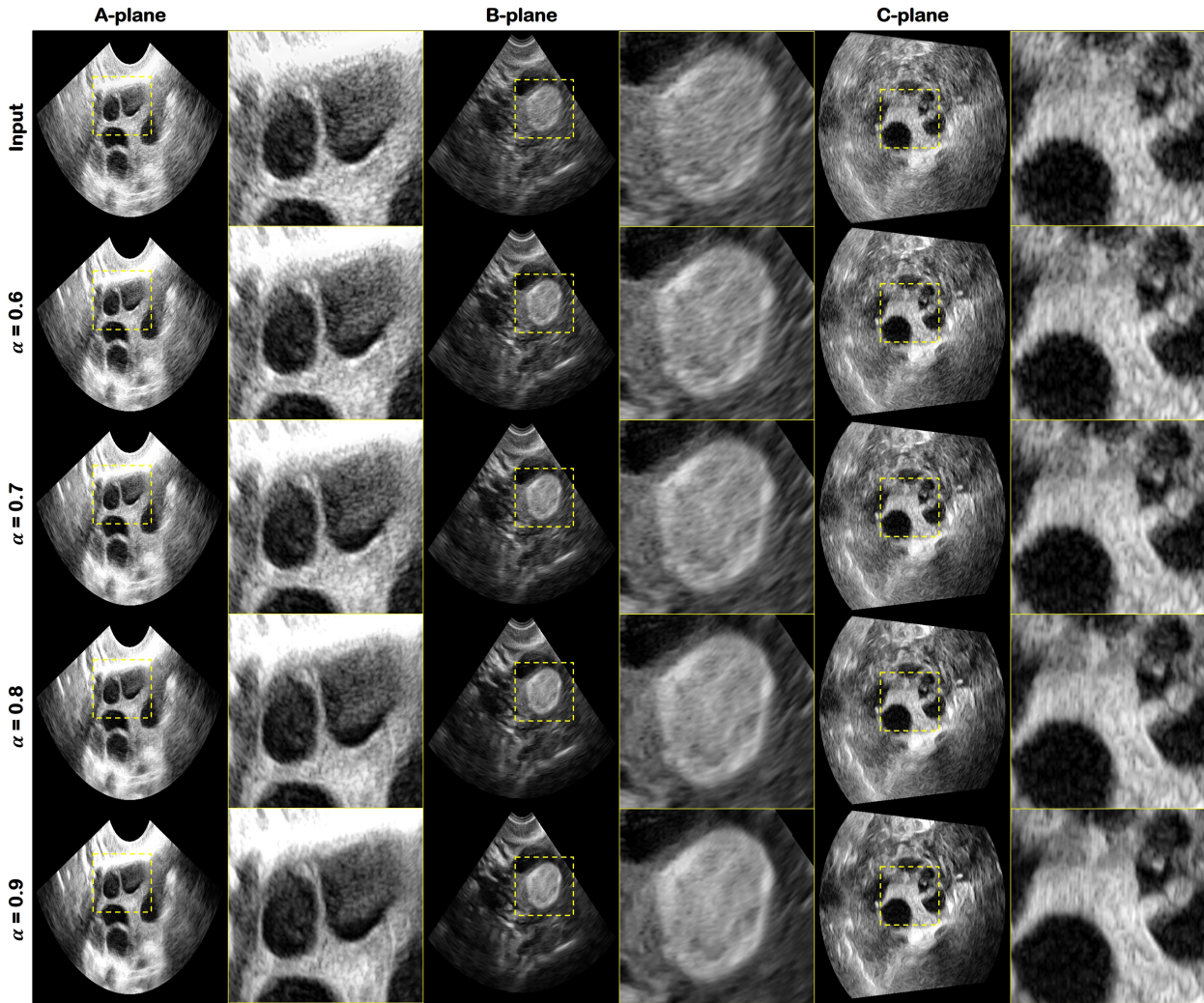


Fig. 5. Results of the proposed method for gynecology test set. The first row is the input image. The result image using the proposed method is displayed from the second to the last row according to the α value variation. The yellow box is the magnified region.

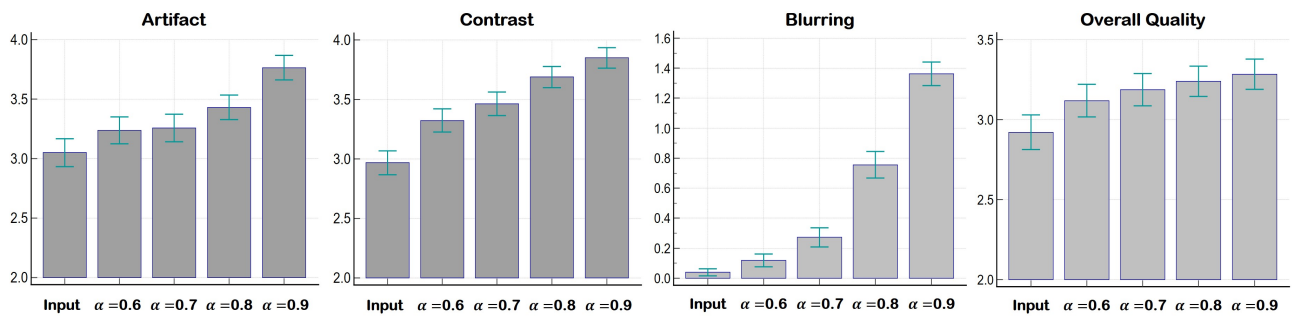


Fig. 6. Clinical evaluation results. Each statistic is calculated using the Friedman test with Bonferroni correction. The first and second graphs are artifact and contrast values. The third and fourth graphs are the blurring effect and the overall quality score. Each bar in the chart shows the input and results of the proposed method from $\alpha = 0.6$ to $\alpha = 0.9$.

Table 2. Wilcoxon paired test result

Score Category	a & b	a & c	a & d	a & e	b & c	b & d	b & e	c & d	c & e	d & e
Artifact	0.0471	0.0290	< 0.0001	< 0.0001	0.8233	0.0091	< 0.0001	0.0294	< 0.0001	< 0.0001
Contrast	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0468	< 0.0001	< 0.0001	0.0017	< 0.0001	0.0139
Blurring	0.0046	< 0.0001	< 0.0001	< 0.0001	0.0003	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Overall Quality	0.0214	0.0011	< 0.0001	< 0.0001	0.3854	0.1223	0.0191	0.4081	0.2319	0.6175

a : Input, b : $\alpha = 0.6$, c : $\alpha = 0.7$, d : $\alpha = 0.8$, e : $\alpha = 0.9$.

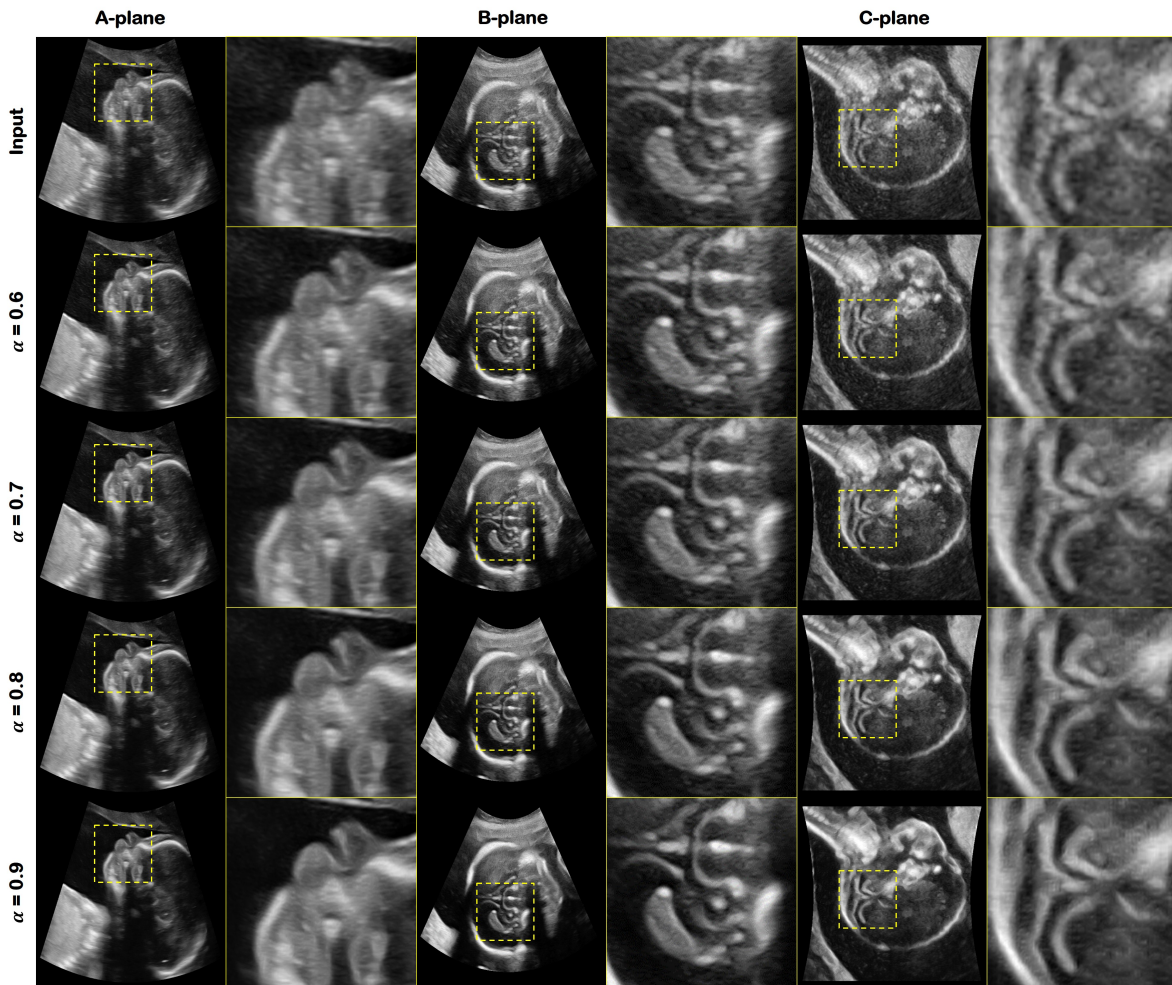


Fig. 7. Results of the proposed method using obstetric test set. The first row is the input image. The result images using the proposed method are displayed from the second to the last row as a α value change. The yellow box is the magnified region.

5. Discussion

5.1. Comparison with conventional CycleGAN

The proposed method is based on the CycleGAN. However, one of the big problems with applying traditional CycleGAN to this task is that the level of translation in 2-D quality cannot be controlled as desired. This means that once the network is trained, the input image is translated into 2-D quality directly. Unfortunately, there is a lot of concern that too much changes in image quality makes sense of disparity for most radiologists. It also leads to unacceptable changes in the image that hinder an accurate diagnosis. For this reason, we believe that our Switchable CycleGAN is very useful.

Specifically, the comparison results are shown in Fig. 8. We visualized OB A-, C- planes and GYN A-, C- planes. The first column is the input image and the second, third column are the results of the proposed method and CycleGAN. Both the proposed method and the CycleGAN results show an improvement in image quality such as noise reduction and contrast enhancement. However, as can be seen in the first row where the red arrow is pointing, the bone is emphasized too much by CycleGAN. It leads to misinterpretations in the case of biparietal diameter assessment. In the second row, the red arrow points to

the float in the amniotic fluid. It contains important information to interpret the status of the fetus. For example, the movement of the particles can be crucial criteria to diagnose the digestive system disorders. However, the CycleGAN recognized this as a noise and suppressed it. On the other hand, the proposed method retains it with high contrast, in particular it can be more visible by controlling the α value. In the third and fourth rows, the red arrow points to the ovarian structure, which can determine the status of the follicle. Since the 2-D quality emphasizes the contrast too much, it sometimes suppresses the weak signal. As you can see in the third and fourth rows, the structure becomes invisible in the CycleGAN result. On the contrary, the proposed method shows sharp boundaries compared to the input image.

5.2. Spatial Control

Another great advantage of the proposed method is that the user can select the area to be improved. Since AdaIN changes its content feature map using the style statistics generated by the AdaIN code generator, the user can control it spatially. Assume there is a mask like in the first row of Fig. 9. There are four areas marked as green, purple, blue, and orange. Instead of applying the masking procedure four times, our method can

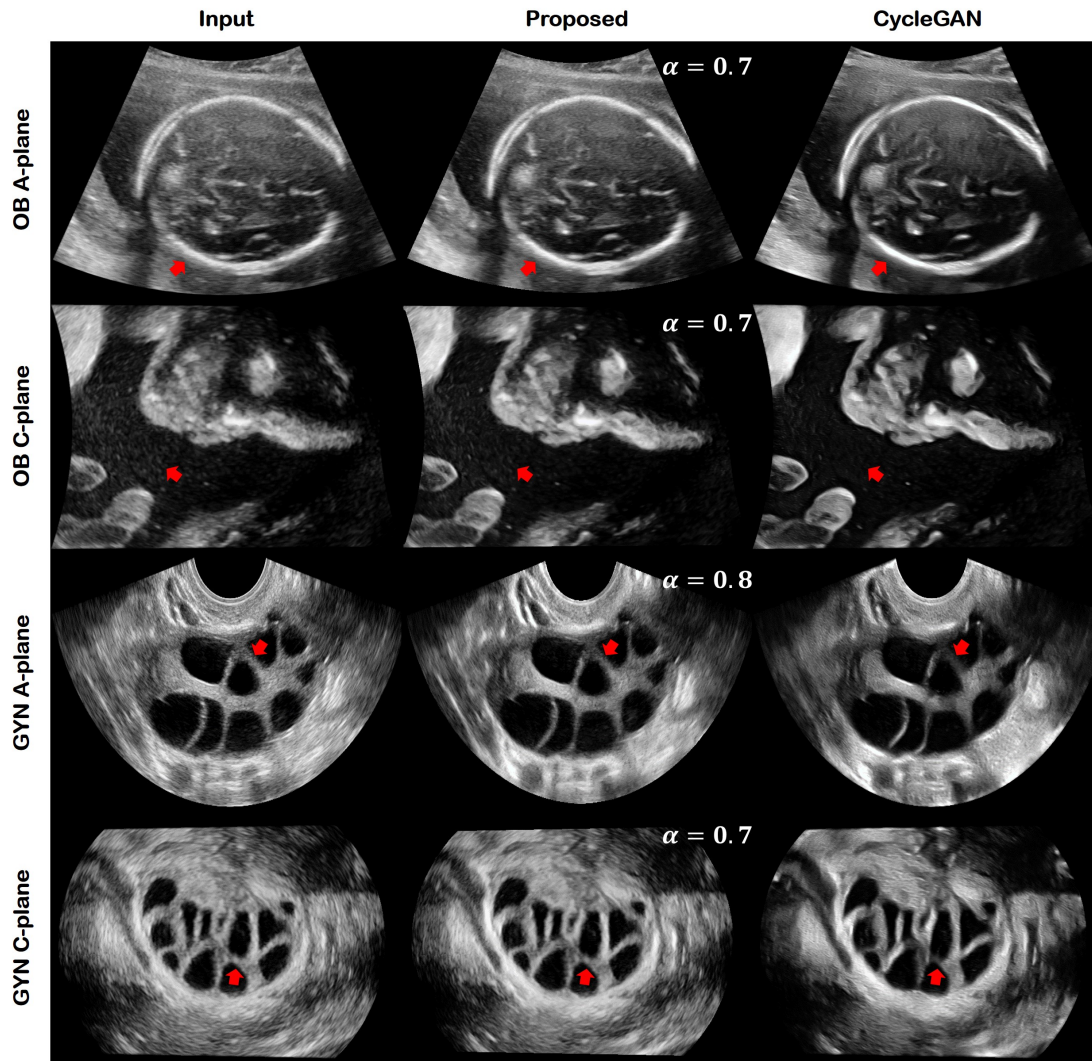


Fig. 8. The comparison result between the proposed method and standard CycleGAN. The first and second rows are obstetric A- and C- planes. The third and fourth rows are gynecology A- and C- planes. The first column denotes the input image. The second and third columns are the result from the proposed method and the CycleGAN, respectively. The differences seen at the red arrows are discussed in the main text.

easily generate a spatially enhanced image with a single feed-forward process. Specifically, the mask enters the network with the image. Then it is multiplied by the content feature map to normalize only the selected area of with a certain α value through AdaIN code interpolation. Accordingly, we can apply four different α values to four different regions with a single feed-forward network.

As shown in Fig. 9, when the mask area in the top row is applied to each image, the proposed method naturally produces spatially varying quality improvement. This provides more nature views with only enhancing area with diagnostic interest.

5.3. 3-D Volume Rendering

In order to verify the effect of the proposed method in the 3-D volume image, we generated a 3-D rendering view with the system Paraview 5.10.0. In Fig. 10, we have shown two GYN and OB volume data sets. The first row is an input image and the second row is output by the proposed method. We have all generated A-plane images that are cut along the elevation axis and fed into the trained network. Then the output planes are

concatenated along the elevation axis to acquire original volume data. All inputs and proposed paired volume images were visualized using the same intensity scale.

As shown in Fig. 10, the structure boundaries are clearer in the proposed volume than in the input volume. In particular, the side-lobe artifact has been suppressed and its structure is then better visible. While artifacts are shown in input volume images which is highlighted as a yellow box at the top of the image, the proposed method suppresses well. The arm and head of the fetus are more clearly visible compared to the input volume image. In the fourth column in particular, the spine of the fetus can hardly be found on the input volume image, but it is easy to identify with the proposed method.

6. Conclusion

Compared to the 2-D systems, the 3-D US can acquire significant rich information in a single trial. However, due to their image generation process, the 3-D US imaging suffers from deteriorated image quality. To overcome this problem, we proposed

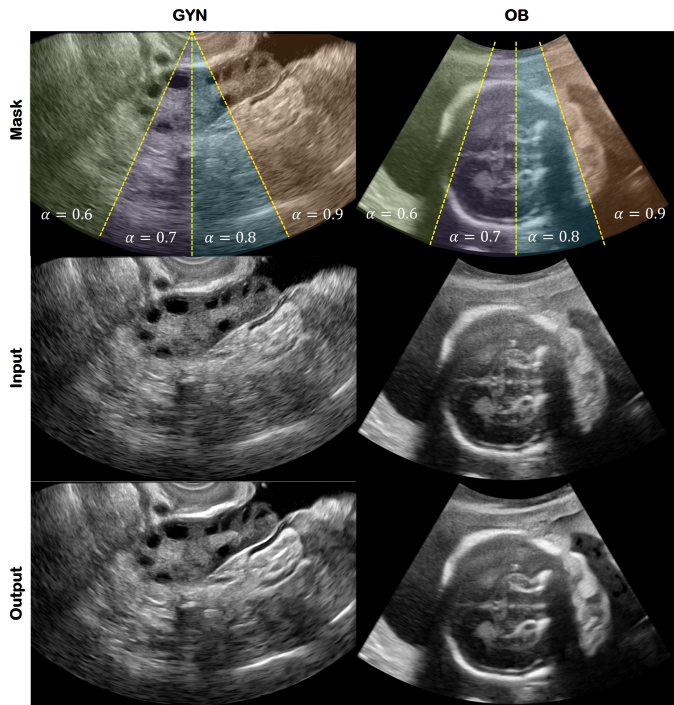


Fig. 9. Spatial control results. The first row is the mask that shows where and which α value was applied. The second and third rows are the input and output, respectively. The first column is gynecology dataset and the second column is obstetric dataset.

a 3-D US image enhancement method based on the Switchable CycleGAN that translates the A-, B-, and C- planes of 3-D US into high-resolution 2-D image quality. Thanks to the use of AdaIN, a single network could translate images at any planes in 3-D US into the high-quality images, and also provided continuous translation between the input and target domain. Using extensive experiments with clinical evaluation, we have confirmed that the proposed method can improve 3-D US from low quality to high quality. In addition, we have certified the robustness of the proposed method through tests with completely invisible data sets such as obstetric datasets. Furthermore, we confirmed that the tunable nature of our method is very useful for clinical purpose by avoiding over-smoothing clinically useful features and provide a user-centric control of image quality depending on user's preference. Thanks to the outstanding performance and flexibility, we believe that the proposed method can be a useful platform for clinical 3-D US systems.

Acknowledgement

This work was supported by the National Research Foundation of Korea under Grant NRF-2020R1A2B5B03001980.

References

Bonferroni, C., 1936. Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8, 3–62.

Brickson, L.L., Hyun, D., Jakovljevic, M., Dahl, J.J., 2021. Reverberation noise suppression in ultrasound channel signals using a 3d fully convolutional neural network. *IEEE Transactions on Medical Imaging* 40, 1184–1195.

Cha, E., Chung, H., Kim, E.Y., Ye, J.C., 2020. Unpaired training of deep learning tMRA for flexible spatio-temporal resolution. *IEEE Transactions on Medical Imaging* 40, 166–179.

Chung, H., Cha, E., Sunwoo, L., Ye, J.C., 2021a. Two-stage deep learning for accelerated 3D time-of-flight MRA without matched training data. *Medical Image Analysis* 71, 102047.

Chung, H., Huh, J., Kim, G., Park, Y.K., Ye, J.C., 2021b. Missing Cone Artifact Removal in ODT Using Unsupervised Deep Learning in the Projection Domain. *IEEE Transactions on Computational Imaging* 7, 747–758.

Conover, W.J., 1999. Practical nonparametric statistics. volume 350. John Wiley & Sons.

Dietrichson, F., Smistad, E., Ostvik, A., Lovstakken, L., 2018. Ultrasound speckle reduction using generative adversarial networks, in: 2018 IEEE International Ultrasonics Symposium (IUS), IEEE. pp. 1–4.

Fenster, A., Downey, D.B., 1996. 3-d ultrasound imaging: A review. *IEEE Engineering in Medicine and Biology Magazine* 15, 41–51.

Fenster, A., Downey, D.B., Cardinal, H.N., 2001. Three-dimensional ultrasound imaging. *Physics in medicine & biology* 46, R67.

Fenster, A., Parraga, G., Bax, J., 2011. Three-dimensional ultrasound scanning. *Interface focus* 1, 503–519.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* 32, 675–701.

Gu, J., Yang, T.S., Ye, J.C., Yang, D.H., 2021. CycleGAN denoising of extreme low-dose cardiac CT using wavelet-assisted noise disentanglement. *Medical Image Analysis* 74, 102209.

Gu, J., Ye, J.C., 2021. Adain-based tunable cyclegan for efficient unsupervised low-dose ct denoising. *IEEE Transactions on Computational Imaging* 7, 73–85.

Huang, Q., Zeng, Z., 2017. A review on real-time 3d ultrasound imaging technology. *BioMed research international* 2017.

Huang, X., Belongie, S., 2017. Arbitrary style transfer in real-time with adaptive instance normalization, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510.

Huh, J., Khan, S., Ye, J.C., 2021. Unsupervised learning for acoustic shadowing artifact removal in ultrasound imaging, in: 2021 IEEE International Ultrasonics Symposium (IUS), IEEE. pp. 1–4.

Hyun, D., Brickson, L.L., Looby, K.T., Dahl, J.J., 2019. Beamforming and speckle reduction using neural networks. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 66, 898–910.

Kang, E., Koo, H.J., Yang, D.H., Seo, J.B., Ye, J.C., 2019. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. *Medical physics* 46, 550–562.

Khan, S., Huh, J., Ye, J.C., 2019. Deep learning-based universal beamformer for ultrasound imaging, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 619–627.

Khan, S., Huh, J., Ye, J.C., 2020a. Adaptive and compressive beamforming using deep learning for medical ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*.

Khan, S., Huh, J., Ye, J.C., 2020b. Unsupervised deconvolution neural network for high quality ultrasound imaging, in: 2020 IEEE International Ultrasonics Symposium (IUS), IEEE. pp. 1–4.

Khan, S., Huh, J., Ye, J.C., 2021a. Switchable deep beamformer for ultrasound imaging using adain, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 677–680.

Khan, S., Huh, J., Ye, J.C., 2021b. Variational formulation of unsupervised deep learning for ultrasound image artifact removal. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 68, 2086–2100.

Kokil, P., Sudharson, S., 2020. Despeckling of clinical ultrasound images using deep residual learning. *Computer Methods and Programs in Biomedicine*, 105477.

Lee, S., Han, S., Salama, P., Dunn, K.W., Delp, E.J., 2019. Three dimensional blind image deconvolution for fluorescence microscopy using generative adversarial networks, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE. pp. 538–542.

Li, Z., Huang, J., Yu, L., Chi, Y., Jin, M., 2019. Low-dose ct image denoising using cycle-consistent adversarial networks, in: 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE. pp. 1–3.

Lim, S., Park, H., Lee, S.E., Chang, S., Sim, B., Ye, J.C., 2020. CycleGAN with a blur kernel for deconvolution microscopy: Optimal transport geometry. *IEEE Transactions on Computational Imaging* 6, 1127–1138.

Luchies, A.C., Byram, B.C., 2018. Deep neural networks for ultrasound beam-

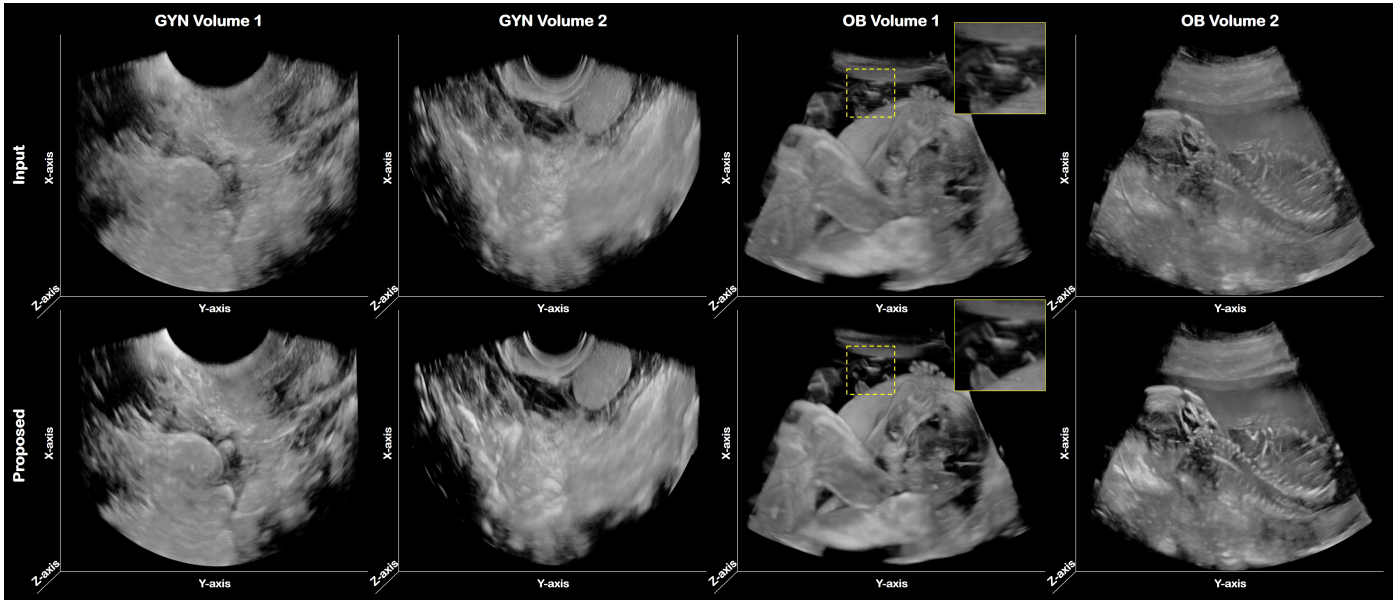


Fig. 10. 3-D volume rendering results. The first and second row denote the input and the results from the proposed method, respectively. The first and second column are two different test gynecology volume sets, and the third and fourth column are obstetric volume test sets.

- forming. *IEEE transactions on medical imaging* 37, 2010–2021.
- Luijten, B., Cohen, R., de Bruijn, F.J., Schmeitz, H.A., Mischi, M., Eldar, Y.C., van Sloun, R.J., 2020. Adaptive ultrasound beamforming using deep learning. *IEEE Transactions on Medical Imaging* 39, 3967–3978.
- Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S., 2017. Least squares generative adversarial networks, in: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2794–2802.
- Modanwal, G., Vellal, A., Buda, M., Mazurowski, M.A., 2020. Mri image harmonization using cycle-consistent generative adversarial network, in: *Medical Imaging 2020: Computer-Aided Diagnosis*, International Society for Optics and Photonics. p. 1131413.
- Modanwal, G., Vellal, A., Mazurowski, M.A., 2021. Normalization of breast mris using cycle-consistent generative adversarial networks. *Computer Methods and Programs in Biomedicine* , 106225.
- Mroueh, Y., 2020. Wasserstein style transfer, in: *International Conference on Artificial Intelligence and Statistics*, PMLR. pp. 842–852.
- Nair, A.A., Tran, T.D., Reiter, A., Bell, M.A.L., 2018. A deep learning based alternative to beamforming ultrasound images, in: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE. pp. 3359–3363.
- Oh, G., Sim, B., Chung, H., Sunwoo, L., Ye, J.C., 2020. Unpaired deep learning for accelerated mri using optimal transport driven cyclegan. *IEEE Transactions on Computational Imaging* 6, 1285–1296.
- Perdios, D., Besson, A., Arditi, M., Thiran, J.P., 2017. A deep learning approach to ultrasound image recovery, in: *2017 IEEE International Ultrasonics Symposium (IUS)*, Ieee. pp. 1–4.
- Perdios, D., Vonlanthen, M., Besson, A., Martinez, F., Arditi, M., Thiran, J.P., 2018. Deep convolutional neural network for ultrasound image enhancement, in: *2018 IEEE International Ultrasonics Symposium (IUS)*, IEEE. pp. 1–4.
- Peyré, G., Cuturi, M., et al., 2019. Computational optimal transport. *Foundations and Trends® in Machine Learning* 11, 355–607.
- Sadeghi, A., Apostolakis, I., Meral, C., Vignon, F., Shin, J.S., Robert, J.L., 2021. Improving contrast of fundamental ultrasound imaging using a deep neural network, in: *2021 IEEE International Ultrasonics Symposium (IUS)*, IEEE. pp. 1–3.
- Sharifzadeh, M., Benali, H., Rivaz, H., 2020. Phase aberration correction: A convolutional neural network approach. *IEEE Access* 8, 162252–162260.
- Sim, B., Oh, G., Kim, J., Jung, C., Ye, J.C., 2020. Optimal transport driven cyclegan for unsupervised learning in inverse problems. *SIAM Journal on Imaging Sciences* 13, 2281–2306.
- Solomon, O., Cohen, R., Zhang, Y., Yang, Y., He, Q., Luo, J., van Sloun, R.J., Eldar, Y.C., 2019. Deep unfolded robust pca with application to clutter suppression in ultrasound. *IEEE transactions on medical imaging* 39, 1051–1063.
- Van Sloun, R.J., Solomon, O., Bruce, M., Khaing, Z.Z., Eldar, Y.C., Mischi, M., 2019. Deep learning for super-resolution vascular ultrasound imaging, in: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE. pp. 1055–1059.
- Vedula, S., Senouf, O., Bronstein, A.M., Michailovich, O.V., Zibulevsky, M., 2017. Towards CT-quality ultrasound imaging using deep learning. *arXiv preprint arXiv:1710.06304* .
- Villani, C., 2009. *Optimal transport: old and new*. volume 338. Springer.
- Yang, S., Kim, E.Y., Ye, J.C., 2021. Continuous conversion of ct kernel using switchable cyclegan with adain. *IEEE Transactions on Medical Imaging* .
- Yen, J.T., Steinberg, J.P., Smith, S.W., 2000. Sparse 2-d array design for real time rectilinear volumetric imaging. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 47, 93–110.
- Yoon, Y.H., Khan, S., Huh, J., Ye, J.C., 2019. Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning. *IEEE transactions on medical imaging* 38, 325–336.
- Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232.
- Zuo, H., Zhang, J., Luo, J., Peng, B., 2021. Phase constraint improves ultrasound image quality reconstructed using deep neural network, in: *2021 IEEE International Ultrasonics Symposium (IUS)*, IEEE. pp. 1–4.