



Anatomy preserving GAN for realistic simulation of intraoperative liver ultrasound images

Lingyu Chen^{a,b}, Hongen Liao^c, Wentao Kong^{d,*}, Daoqiang Zhang^{a,b}, Fang Chen^{a,b,*}

^a Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing, China

^b College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

^c Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

^d Department of Ultrasound, Affiliated DrumTower Hospital, Medical School of Nanjing University, Nanjing 210008, China

ARTICLE INFO

Keywords:

High-accuracy structures
Anatomy preserving GAN
Liver ultrasound
Realistic simulation

ABSTRACT

In ultrasound-guided liver surgery, the lack of large-scale intraoperative ultrasound images with important anatomical structures remains an obstacle hindering the successful application of AI to ultrasound guidance. In this case, intraoperative ultrasound (iUS) simulation should be conducted from preoperative magnetic resonance (pMR), which not only helps doctors understand the characteristics of iUS in advance, but also expands the iUS dataset from various imaging positions, thereby promoting the automatic iUS analysis in ultrasound guidance. Herein, a novel anatomy preserving generative adversarial network (ApGAN) framework was proposed to generate simulated intraoperative ultrasound (Sim-iUS) of liver with precise structure information from pMR. Specifically, the low-rank factors based bimodal fusion was first established focusing on the effective information of hepatic parenchyma. Then, a deformation field based correction module was introduced to learn and correct the slight structural distortion from surgical operations. Meanwhile, the multiple loss functions were designed to constrain the simulation of the content, structures, and style. Empirical results of clinical data showed that the proposed ApGAN obtained higher Structural Similarity (SSIM) of 0.74 and Fréchet Inception Distance (FID) of 35.54 compared to existing methods. Furthermore, the average Hausdorff Distance (HD) error of the liver capsule structure was less than 0.25 mm, and the average relative (Euclidean Distance) ED error for polyps was 0.12 mm, indicating the high-level precision of this ApGAN in simulating the anatomical structures and focal areas.

1. Introduction

In image-guided liver surgery, intraoperative ultrasound (iUS) provides the surgeon with real-time information regarding the lesion size, location, and relation to intrahepatic vasculature that may result in an alteration in the planned surgical approach [1,2]. The information from iUS has been proven effective in identifying liver metastases, influencing surgical decision-making, and altering the surgical management in up to 30% of cases [3]. For example, iUS makes it possible to determine the respectability of primary and secondary liver tumors, and to judge the extent of resection required by the patient [5]. With the development of artificial intelligence (AI), various studies have demonstrated the efficiency of AI systems in identifying focal liver lesions in preoperative transthoracic US images captured by radiologists with high sensitivity and specificity [6,7]. However, considering the rather limited number of iUS datasets available, the use of AI for the assessment of intraoperative

liver lesions from iUS is yet challenging. Compared with the preoperative US, collecting the iUS dataset is more difficult. On the one hand, influenced by complex surgical operations, collecting iUS requires a longer learning curve [4], while on the other, the large liver volume and occasionally fat infiltration [8] make it difficult to collect US images that contain some small but important anatomical structures such as liver capsule and blood vessels in the surgery [9]. Therefore, the lack of large-scale iUS images with important anatomical structures remains a major obstacle hindering the successful application of AI to ultrasound-guided liver surgery.

In actual clinical surgery, preoperative magnetic resonance (pMR) imaging is a basic and indispensable way to provide abundant anatomical and lesion information for patients. Specially, pMR is provided with high tissue resolution, and demonstrates a full view of the anatomy including the parenchyma, internal and external vessels, etc. [10]. Conducting intraoperative ultrasound (iUS) simulation from

* Corresponding authors.

E-mail addresses: breezewen@163.com (W. Kong), chenfang@nuaa.edu.cn (F. Chen).

<https://doi.org/10.1016/j.cmpb.2023.107642>

Received 26 February 2023; Received in revised form 31 May 2023; Accepted 2 June 2023

Available online 3 June 2023

0169-2607/© 2023 Elsevier B.V. All rights reserved.

preoperative magnetic resonance (pMR) can not only facilitate doctors to understand the characteristics of iUS in advance, but also expand the iUS dataset from various imaging positions to promote the automatic iUS analysis in ultrasound guidance. Rewarding as it is, iUS simulation for liver surgery still faces great challenges [11,12]. The iUS simulation needs to guarantee that the key anatomical content provided in the pMR is also highly accurate when revealed in the Sim-iUS. For instance, intrahepatic vessels, such as the hepatic artery and portal vein, all feature a fixed range of internal diameters under normal conditions. Besides, in liver interventions, surgeons determine the severity or progression of lesions like primary liver cancer, hepatic hemangiomas, hepatic cysts, and others based on the characteristics of the anatomical structures such as the size, maximum diameter, and thickness of tumors. To this end, detailed information about crucial anatomical structures of pMR matters considerably during the iUS simulation, and should not undergo major abrupt changes.

In fact, great effort has been made to overcome the obstacles of US simulation. The published works are broadly divided into two categories: 1) **Early acoustic models with reflection or scattering.** The early approaches [13,14], obtained the slices (online) of the reconstruction volume (offline pre-processing) based on the simulated transducer location. However, in addition to discerning valid features, it also required careful consideration of specific US configuration parameters such as gain or focus, as well as the position of the transducer. Some researchers [15,16], have endeavored to reproduce the complex interactions of sound waves in vivo. They mostly produced synthetic scans by exploiting geometries extracted from other imaging modalities. For instance, Ramtin et al. [16] proposed to simulate US by combining a fixed-view scattering image that generated offline from Field II [17] and a view-dependent reflection image. Though simulated results were more realistic, employing Field II was time-consuming and computationally intensive, and it took several days for a B-mode simulation US of cyst phantoms with 50 to 128 RF scan lines, depending on the performance of the computer used. Except for the above weaknesses, earlier simulation methods, due to the lack of refraction, mirroring, range distortion, etc., also had trouble in modeling artefacts and speckle noise, as described in [18].

2) **Recent learning-based methods.** The emergence of generative adversarial networks (GAN) [19] achieved striking success on the tasks of medical image analysis [20,21]. For instance, Hu et al. [22] put forward a 3D end-to-end network for brain MR-to-PET synthesis, and designed a bidirectional mapping mechanism to synthesize perceptually realistic PET while preserving the different brain structures of any individual; Yang et al. [23] presented a structure-constrained CycleGAN for MR-to-CT by introducing a loss based on the modality independent neighborhood descriptor; Emami et al. [24] proposed a structure-aware GAN to generate synthetic CT from MR. To the best of our knowledge, most of existing works focus on the translation between modalities such as CT or MR that demonstrate clear and well-defined anatomical structures, with few attempts made to synthesize for US. For instance, Vitale et al. [25] introduced the first approach for realistic patient-specific abdominal US simulation and deep convolutional neural networks (CNN). The qualitative results showed that their work indeed improved the realism of the Sim-iUS; Immediately afterward, Tomar et al. [26] introduced a novel content-preserving image translation framework to maintain the simulation layout, which seemed like a multi-task network for translation, segmentation, and classification, extending the generator to be class-conditional and encouraging the disentanglement of the content and the style; Liang et al. [27] designed a spGAN that synthesized B-mode US images with high-resolution and customized texture editing features.

Although the above methods generated the Sim-iUS with a degree of realism and ability to display some anatomical structures, they all suffered from several non-negligible shortcomings: i) Before being fed into the GAN-based models for Sim-iUS, the raw input was preprocessed by a ray-casting simulation. For example, Vitale et al. [25] constructed a

patient-specific abdominal CT volume and its associated organ segmentations for a ray-casting-based simulator, and the resulting scan was regarded as an intermediate US simulation sent into a CycleGAN; Tomar et al. [26] employed a ray-tracing framework to render B-mode images by simulating a convex probe placed at multiple locations and orientations on the abdominal surface. However, the ray-casting-based methods are parameter-dependent and have a high budget for time and computational resources, which are more restrictive for the clinical environment; ii) Furthermore, the generated Sim-iUS is slightly distorted at the edges of the field of view (sector edges), along with large artefacts that unrelated to real anatomy, or the presence of pseudo-organs (such as pancreas or spleen) [23].

In contrast to the above methods, a novel anatomy preserving GAN (ApGAN) was hereby proposed to generate Sim-iUS directly from pMR without any preprocessing step of ray-casting. In addition, the generated Sim-iUS approximates the real US in terms of anatomical structures and style. Specially, the anatomical parenchyma contained in the hereby-proposed SimUS is highly compatible in structural properties with real US, and is provided with a superior level of precision. Extensive experiments on the real liver pMR-iUS dataset have confirmed the better performance of the proposed ApGAN than existing translation methods in superior iUS simulation.

2. Materials and methods

In this section, we describe the details of the proposed anatomy preserving GAN (ApGAN). *First*, to ensure that the Sim-iUS contains the liver parenchymal structures in pMR, we perform low-rank factors based bimodal fusion to focus on the overall structural and actual anatomical content (Section 2.3). *Second*, for reducing checkerboard artefacts, we add fixed convolutional layers in the decoder, as depicted in Section 2.4. *Third*, the structural distortion of Sim-iUS caused by modality difference, is warped by the deformation field for structural correction (Section 2.5). *Finally*, we design multiple loss functions which include content-perceptive loss, structural-correction loss, and appearance-aware loss to perceive hepatic parenchyma and structure content, notice iUS appearance and style (Section 2.6).

2.1. Overview

The framework of the proposed anatomy preserving GAN is illustrated in Fig. 1. The pMR, iUS, Sim-iUS are denoted as x , y and \hat{y} , respectively. We break up our generator function \mathcal{G} into two components, an encoder $\mathcal{E}_{ApGAN}^{Enc}$ followed by a decoder $\mathcal{D}_{ApGAN}^{Dec}$, which are used sequentially to produce Sim-iUS. x is first fed into our encoder to extract the feature information $\mathcal{E}_{ApGAN}^{Enc}(x)$. A low-rank factors based bimodal fusion of the rough segmentation maps z with corresponding pMR x is also performed, aiming to extract fusion vector h for anatomical content and structure information containing the liver parenchyma. The bimodal fusion vectors h and the features $\mathcal{E}_{ApGAN}^{Enc}(x)$ are then concatenated and delivered directly into our decoder $\mathcal{D}_{ApGAN}^{Dec}$. However, during our simulation, the generated Sim-iUS presented with different, checkerboard artefacts from the real iUS, which are analyzed to be possibly due to operations such as deconvolution or interpolation during the up-sampling process. We insert fixed convolutional layers $\mathcal{F}(\cdot)$ with an order of smoothness in the decoder, to enhance the texture of \hat{y} and approximates real iUS. Owing to the surgical operations, the Sim-iUS has a slight distortion. So in the training phase, we feed output \hat{y} of decoder as an intermediate Sim-iUS \tilde{y} into the registration based network $\mathcal{R}(\cdot)$ to obtain the deformation field for structural correction, to fit the distribution of the misaligned image domain and search for a common optimal solution space for image translation and registration. Besides, during the test phase, we treat the decoder output \hat{y} as the final Sim-iUS, independent of the correction module based on registration. Ultimately,

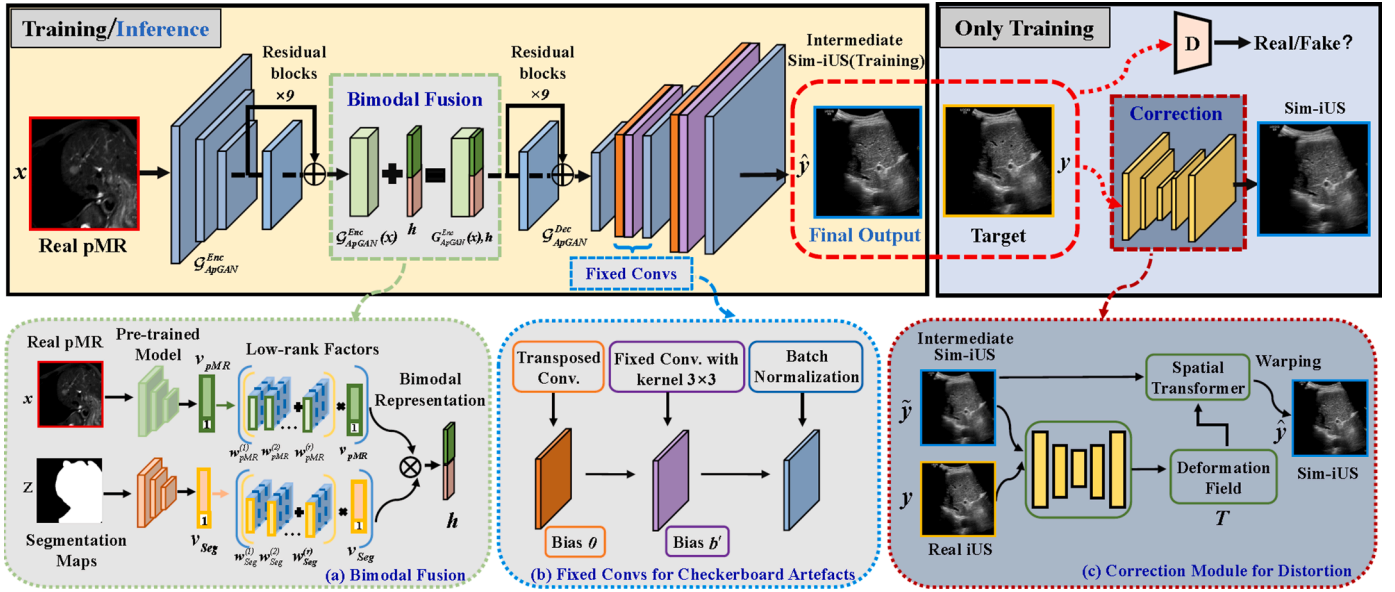


Fig. 1. The overall framework of ApGAN. First, pMR images are fed into our encoder $\mathcal{G}_{ApGAN}^{Enc}$ to extract liver information $\mathcal{G}_{ApGAN}^{Enc}(x)$. Then we establish the bimodal fusion module to obtain the fusion vectors h which focus on the whole effective parenchymal content. After that, we concatenate $\mathcal{G}_{ApGAN}^{Enc}(x)$ and h , then send merged tensors to the decoder for Sim-iUS, which consists of several fixed convolutional layers for removing checkerboard artefacts. Notably, during the training stage, Sim-iUS images are further adjusted by a correction module for structural distortion.

the final Sim-iUS is integrated into the discriminator $\mathcal{D}(\cdot)$ along with the real iUS, and adversarial training is applied to ensure that the generated Sim-iUS is highly realistic.

2.2. Baseline path

• Baseline Generator Network

The original pMR and corresponding iUS share roughly similar anatomic contours and margins (e.g., portal vein, inferior vena cava, right hepatic vein) and contain consistent anatomic content and lesion areas (e.g., liver, kidney, pancreas, polyps, etc.). Hence, we focus specifically on the anatomical content of the pMR, to constrain the substantive component of Sim-iUS. We believe that pMR and iUS have the features that share the same anatomical space. As the shared information may not be mined employing ordinary CNN networks, a baseline generator is constructed inspired by popular translation networks [28, 29, 30, 31, 32]. Moreover, the presence of multiple residual blocks in our generator improves the ability to preserve shallow structural content features.

Our generator \mathcal{G} implemented in this paper is based on the generator of CycleGAN [28], which keeps two stride-2 convolutional layers, 9 residual blocks for 256×256 images and two fractionally-strided convolutions with stride $\frac{1}{2}$. The filter kernel size of each convolutional layer and residual block is 3×3 . Each residual block only includes one convolutional layer, and we choose leaky rectified linear unit (ReLU) and instance normalization as the activation function and normalization layer, respectively.

• Patch-Level Discriminator Network

We construct a PatchGAN [32] classifier to distinguish whether the input is a real iUS, judging the degree of realism by overlapping patches with the size of 70×70 . The patch-level discriminator contains 5 typical convolutional layers with the kernel size of 4×4 , and the numbers of the filters are 64, 128, 256 and 512 for the former four layers, respectively. Among these layers, the leaky rectified linear unit with the negative slopes of 0.2 is chosen as the activation function. There also

exists sigmoid function for the last convolutional layer and instance normalization. The architecture owns fewer parameters than a discriminator that judge the whole image directly, and can receive images of arbitrary size.

2.3. Bimodal Fusion for Effective Hepatic Parenchymal Content

We consider that the structural information and lesion areas of different anatomy need to be highlighted, which is useful for the later tasks of inferencing the Sim-iUS. Thus, to enhance the exploitation of representative features about the parenchymal content or anatomy of the liver and to reduce the interference of background information, we establish a bimodal fusion module after the encoder $\mathcal{G}_{ApGAN}^{Enc}$. This module then infers the useful vector suppressing the irrelevant information as the weighted average over the real pMR and corresponding segmentation maps embedding based on the original image embedding.

Specifically, we firstly conduct a coarse segmentation of pMR to obtain segmentation maps, and then fuse the maps with the pMR. The same as most works [S1-S3] employing Low-rank Multimodal Fusion (LMF) [33], firstly, we denote the bimodal fusion module as a bilinear function $\mathcal{F} : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{Y}$, where \mathcal{Z} is the segmentation maps domain of the pMR. Given a set of vector representations, $\{v_{Seg} \in \mathbb{R}^{d_{Seg}}, v_{pMR} \in \mathbb{R}^{d_{pMR}}\}$ which are encoding unimodal information of the pMR and the segmentation maps by the pre-training VGG. Then we integrate the unimodal representations into one compact bimodal representation $h \in \mathbb{R}^{d_h}$ for the downstream simulation task. Multimodal fusion requires a transformation of the input representations into a high-dimensional tensor and then mapping it back to a lower-dimensional output vector space. Herein, the input tensor \mathcal{V} combined by the unimodal representation is computed by:

$$\mathcal{V} = v_{Seg} \otimes v_{pMR}, v_m \in \mathbb{R}^{d_m} \quad (1)$$

where \otimes denotes the tensor outer product over a set of vectors, and m represents the modality of pMR or segmentation maps. Then the input tensor $\mathcal{V} \in \mathbb{R}^{d_{Seg} \times d_{pMR}}$ is passed through a linear layer to produce an output fusion vector:

$$\begin{aligned}
h &= g(\mathcal{V}; \mathcal{W}, b) \\
&= \left(w_{pMR}^{(1)} \otimes w_{Seg}^{(1)} + w_{pMR}^{(2)} \otimes w_{Seg}^{(2)} + \dots + w_{pMR}^{(r)} \otimes w_{Seg}^{(r)} \right) \cdot (v_{Seg} \otimes v_{pMR}), h, b \\
&\in \mathbb{R}^{d_h}
\end{aligned} \tag{2}$$

where \mathcal{W} is the weight of this layer and b is the bias. The innovation of LMF is that it replaces the full tensor \mathcal{W} with a set of modality-specific low-rank factors that can be utilized to recover a low-rank weight tensor $\{w_m^{(i)}\}_{i=1}^r$, without explicitly tensorizing the unimodal representations. The minimal r that makes the decomposition valid is called the rank of the tensor, and denote $w_m^{(i)} = [w_{m,1}^{(i)}, w_{m,2}^{(i)}, \dots, w_{m,d_h}^{(i)}]$ for modality m , where $w_m^{(i)} \in \mathbb{R}^{d_m \times d_h}$ shares the same size for the second dimension. Thereafter, we take anatomical structure features extracted by the encoder $\mathcal{E}_{ApGAN}^{Enc}$, concatenate them with the feature representation obtained by performing bimodal fusion, and send them into the decoder $\mathcal{E}_{ApGAN}^{Dec}$ together.

2.4. Fixed convs to eliminate checkerboard artefacts

The deconvolution layers or the interpolation (nearest neighbor interpolation or bilinear interpolation) in the up-sampling operation distort the images, and lead to checkerboard artifacts [34]. During several attempts at simulation, we also find that Sim-iUS show abnormal artifacts, which resemble checkerboard squares. To resolve above issue, our ApGAN adopts a fixed smooth convolutional layer $\mathcal{F}(\cdot)$ with a trainable bias b' on the up-sampling course. In detail, we firstly fix the bias in the transposed convolutional layer as 0, and then insert the fixed smooth layer after the transposed one. Particularly, \mathcal{K} denotes a filter kernel of this fixed convolutional layer, and $\mathcal{K}^{(d)}$ indicates a certain degree of smoothness controlled by d . The kernel $\mathcal{K}^{(d)}$ is obtained by convolving h_0 as follows:

$$\mathcal{K}^{(d)} = \begin{cases} h_0 & (d = 0) \\ \mathcal{K}^{(d-1)} * h_0 & (d > 1) \end{cases} \tag{3}$$

where h_0 represents a kernel filter with size of $U \times U$, and d is referred to as the order of smoothness for $\mathcal{K}^{(d)}$. And the output feature maps corresponding to input \mathcal{V}' of this fixed convolutional layer can be formulated as

$$\mathcal{F} = \sum \mathcal{V}' \mathcal{K}^{(d)} + b' \tag{4}$$

2.5. Correction for structure distortion of Sim-iUS

After the decoder $\mathcal{E}_{ApGAN}^{Dec}$, the obtained Sim-iUS still differ from the real iUS in terms of the rendered anatomical structures (sharpness, missing or shifted structures, etc.) and realistic speckle noise. We infer that the major reasons are in two aspects: on the one hand, the structural features of our SimUS are derived only from pMR, and lacks the information on structural deformation and displacement caused by the surgical procedure and instruments that iUS carries. On the other hand, there exists noise, due to different modality imaging principles. Therefore, we regard Sim-iUS by the decoder as the intermediate Sim-iUS \tilde{y} . During training, we design a correction module to yield better Sim-iUS of high-accuracy structure.

We expect the correction module to obtain a deformation field to address the slight structural distortions and displacements of the intermediate Sim-iUS and then get final Sim-iUS with high accuracy structural information through spatial transformation. The procedure is expressed as: $\hat{y} = \tilde{y} \circ T$, where T represents a deformation field that produces random displacements for each pixel. To obtain the optimal Sim-iUS, the following objective function needs to be minimized:

$$\hat{\mathcal{T}} = \underset{\mathcal{T}}{\operatorname{argmin}} - \sum \mathcal{L}(T \circ \mathcal{G}(x), y) \tag{5}$$

Thus, we establish a registration module $\mathcal{R}(\cdot)$ to acquire the deformation field T . In brief, the module takes \tilde{y} and y to a U-Net based registration network, then generates the random deformation field and performs a resampling operation, and outputs the corrected Sim-iUS \hat{y} at last. Meanwhile, we propose suitable structural-correction loss functions for correction module to constrain correction process, which is covered in the Section 2.6. The correction module acquires the deformation field by the U-Net based encoder/decoder backbone, in combination with, residual connections, atrous convolutions and pyramid scene parsing pooling.

2.6. Objective Functions

We devise multiple loss functions to further constrain the structures and appearance of our Sim-iUS, which include content-perceptive loss, structural-correction loss, appearance-aware loss, and the least-square adversarial loss to perceive hepatic parenchyma and structure content, perceive iUS appearance and style.

• Content-perceptive Loss

We observe that during intraoperative ultrasound simulation, not only the entire image content of the paired pMR and Sim-iUS should be consistent, but also the corresponding patches should share content between the two. For example, liver capsules in the Sim-iUS are supposed to be more strongly correlated with the corresponding structural patches in the pMR than with the rest of the pMR, as shown in Fig. 2. Since the paired pMR and iUS share the same hypothetical anatomical space, this paper employs patch-wise contrastive learning to strengthen the correlation between structural patches in pMR and patches at corresponding locations in iUS. Specifically, as shown in Fig. 2, the yellow, blue and green boxes represent query patches, positive patches and negative patches, respectively. We sample a query patch u from the Sim-iUS generated by ApGAN, and then expect u to be more semantically similar to the input patches u^+ which are sampled from the same input pMR at different locations, in comparison to N negative patches u^- which are sampled from the same input pMR at different locations.

To further maximize the mutual information between corresponding input and output patches in Sim-iUS and pMR, we construct a multi-layer, patch-wise contrastive loss to compose the content perceptive loss with the encoder part of our generator and a two-layer MLP network. Concretely, in order to enable ApGAN to maximize the correlation between patch pairs of Sim-iUS and pMR on the multi-scale feature space, ApGAN selects the features of the L -layer in the encoder $\mathcal{E}_{ApGAN}^{Enc, (l)}$ and then feeds them into a two-layer shallow perception \mathcal{M}_L , yielding feature stacks $\{u_l\}_L = \{\mathcal{M}_L(\mathcal{E}_{ApGAN}^{Enc, (l)}(x))\}_L$ for pMR, where $\mathcal{E}_{ApGAN}^{Enc, (l)}$ represents the output of the l -th chosen layer. We denote the index of the layers as $l \in \{1, 2, \dots, L\}$ and $s \in \{1, \dots, S_l\}$, where S_l is the number of spatial localizations in each layer. This paper regards the corresponding positive features as $u_l^s \in \mathbb{R}^{C_l}$ and the rest features as $u_l^{s'} \in \mathbb{R}^{(S_l-1) \times C_l}$, where C_l is the number of channels at each layer. Meanwhile, we encode the output Sim-iUS into $\{\hat{u}_l\}_L = \{\mathcal{M}_L(\mathcal{E}_{ApGAN}^{Enc, (l)}(\mathcal{G}(x)))\}_L$. The definition of patch-wise contrastive loss is:

$$\mathcal{L}_{contrastive}(\hat{u}_l^s, u_l^s, u_l^{s'}) = -\log \left[\frac{\exp(\hat{u}_l^s \cdot u_l^s / \tau)}{\exp(\hat{u}_l^s \cdot u_l^s / \tau) + \sum_{k=1}^{S_l-1} \exp(\hat{u}_l^s \cdot u_l^{s_k} / \tau)} \right] \tag{6}$$

We then enhance the similarity of the patches at multiple scales to build the content-perceptive loss:

$$\mathcal{L}_{content}(\mathcal{G}) = \mathbb{E}_{x \sim pMR} \sum_{l=1}^L \sum_{s=1}^{S_l} \mathcal{L}_{contrastive}(\hat{u}_l^s, u_l^s, u_l^{s'}) \tag{7}$$

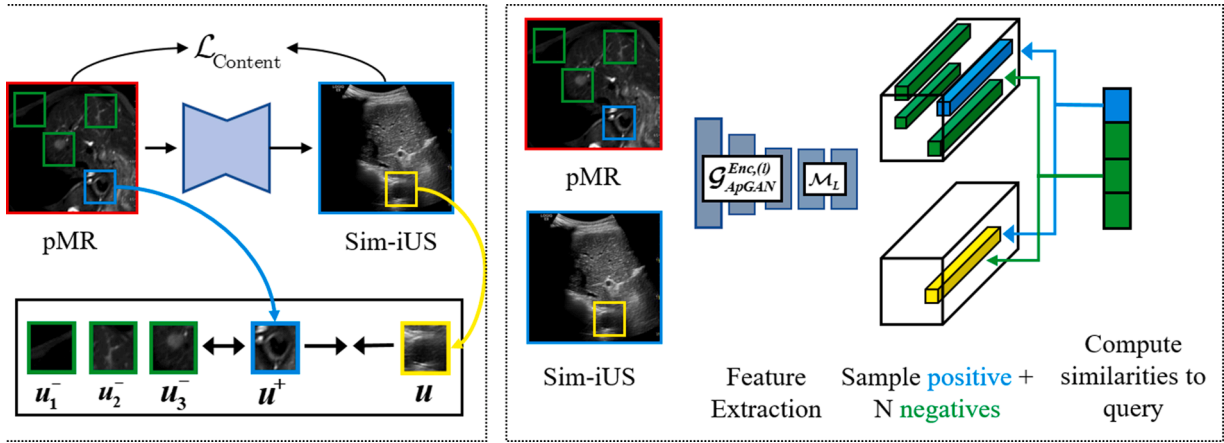


Fig. 2. The overview of content-perceptive loss.

To efficiently compute the content-perception loss of multi scales, based on patch-wise contrastive learning, ApGAN acquires semantic features from five layers, which are RGB pixels and two downsampling convolutional layers and two residual blocks. We select features from five layers with corresponding perceptual region sizes of 1×1 , 9×9 , 15×15 , 35×35 , and 99×99 . For each feature map in the different layers, 256 locations are randomly sampled, and then 256-D features are obtained by 2-layer shallow MLP. More specifically, for a query patch sampled in the output Sim-iUS of ApGAN, 1 positive sample and 255 negatives are captured in the features encoded by pMR. We follow the setting of CUT to use the temperature value 0.07, and enqueue 256 patches per image per iteration.

- Structural-correction Loss

The correction module \mathcal{R} is proposed to alter structural distortion, based on an additional registration network. To constrain the correction process for superior Sim-iUS, we devise a structural-correction loss that is comprised of $\mathcal{L}_{correct}(\mathcal{R})$ and $\mathcal{L}_{Smooth}(\mathcal{R})$, as shown in Fig. 3.

$\mathcal{L}_{Smooth}(\mathcal{R})$ is inherently a regularization term to evaluate the smoothness of the deformation field. And $\mathcal{L}_{correct}(\mathcal{R})$ narrows the distribution distance between the Sim-iUS and the real iUS, which is formulated as follows:

$$\mathcal{L}_{correct}(\mathcal{R}) = \mathbb{E}_{\tilde{y}, \tilde{y}} [\|y - \tilde{y} \circ \mathcal{R}(\tilde{y}, y)\|_1] \quad (8)$$

where $\mathcal{R}(\tilde{y}, y)$ signifies the deformation field, \circ represents resampling operations. Therefore, our total structural-correction loss is comprised of the above two components in equation (11). After several five-fold cross-validation, we choose the best experimental Sim-iUS by setting $\lambda_{correct} = 20$ and $\lambda_{Smooth} = 10$.

$$\mathcal{L}_{structure}(\mathcal{G}) = \lambda_{correct} \cdot \mathcal{L}_{correct}(\mathcal{R}) + \lambda_{Smooth} \cdot \mathcal{L}_{Smooth}(\mathcal{R}) \quad (9)$$

- Appearance-Aware Loss

The most conspicuous appearance feature of US is the presence of large artifacts and speckle noise, which we believe to be its unique

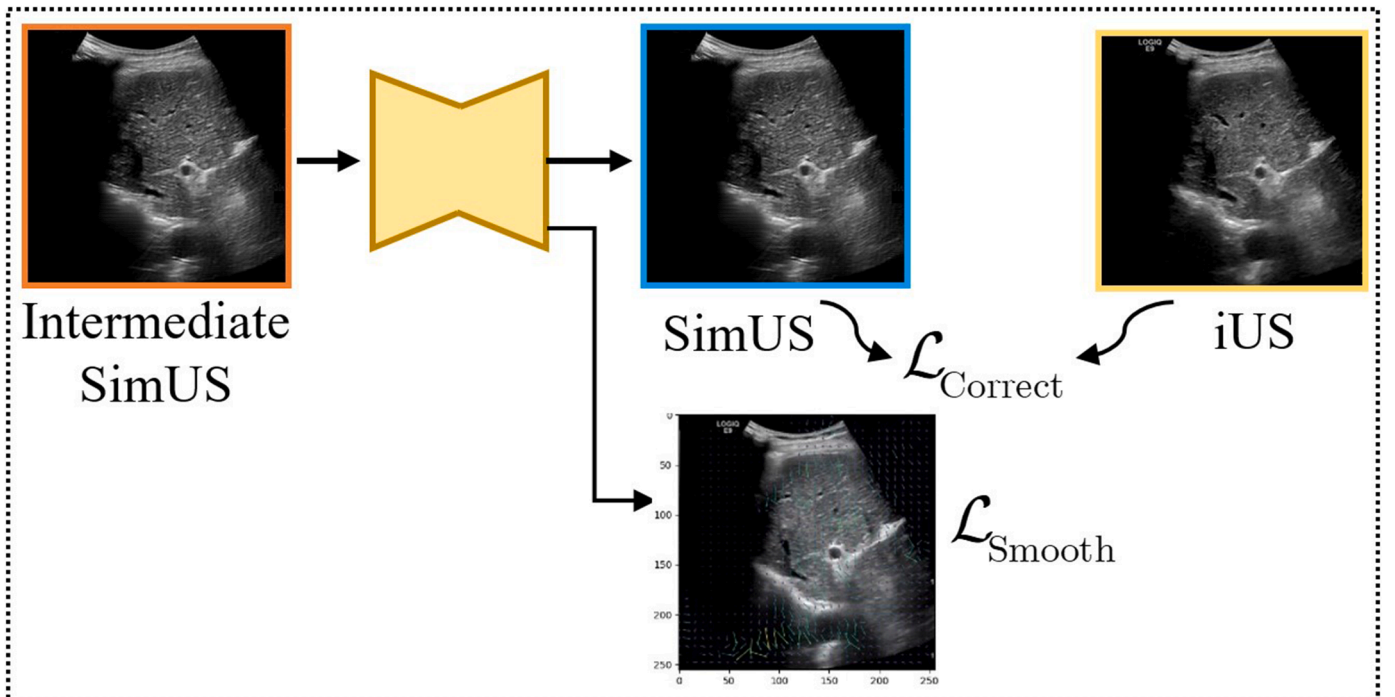


Fig. 3. The overview of structural-correction loss.

texture properties. Whereas both content-perceptive loss and structural-correction loss spotlight the preservation of anatomical structures, we design an appearance-aware loss that includes two parts to encourage texture and style similarity between Sim-iUS and iUS. The first part is a texture-enhancement loss, as shown in Fig. 4, which minimizes Manhattan distances between Gram matrices of intermediate feature maps:

$$\mathcal{L}_{\text{texture}}(\mathcal{G}) = \mathbb{E} \left[\sum_i \| G_i^{\mathcal{O}}(\text{iUS}) - G_i^{\mathcal{O}}(\text{SimUS}) \|_1 \right] \quad (10)$$

where $G_i^{\mathcal{O}}$ denotes Gram matrix of feature map ϕ_i and distributes spatial information of texture and style. We extract feature maps from relu4, relu12, relu14, and relu16 layers of our encoder $\mathcal{E}_{\text{ApGAN}}^{\text{Enc}}$. Besides, Sajjadi et al. [35] has proved that checkerboard artifacts are mitigated effectively by leveraging Gram matrices. The second part is a style-capture loss. We reinforce our encoder $\mathcal{E}_{\text{ApGAN}}^{\text{Enc}}$ to learn the overall stylistic features of real iUS by performing random mask, and implement L2 loss to compute the differences between the simulated results of the masked image and real iUS, as shown in Fig. 5:

$$\mathcal{L}_{\text{style}}(\mathcal{G}) = \| (1 - \widehat{\mathcal{B}}) \odot (x - \mathcal{E}_{\text{ApGAN}}^{\text{Enc}}(\widehat{\mathcal{B}} \odot x)) \|_2^2 \quad (11)$$

where \odot implies the element-wise product operation, and the size of the mask block is 24×24 . We set the number of mask blocks to 32 according to the size of the pMR images and the mask block, and suppose $\widehat{\mathcal{B}}$ is the binary mask corresponding to mask operation, where the value of the discarded pixel is 1 and the value of the original input pixel is 0. Thus, we summarize the appearance-aware loss is the sum of $\mathcal{L}_{\text{texture}}$ and $\mathcal{L}_{\text{style}}$:

$$\mathcal{L}_{\text{appear}}(\mathcal{G}) = \gamma_{\text{texture}} \mathcal{L}_{\text{texture}}(\mathcal{G}) + \gamma_{\text{style}} \mathcal{L}_{\text{style}}(\mathcal{G}) \quad (12)$$

In all experiments, we follow [35] to set $\gamma_{\text{texture}} = 250$ and $\gamma_{\text{style}} = 20$.

• Adversarial Loss

We adopt the least-square adversarial loss to train the generator and discriminator with the following standard GAN equations. And the loss attempts to make the prediction Sim-iUS appear realistic and can select a particular pattern from the multiple domains, which is as close as possible to the target iUS domain.

$$\mathcal{L}(\mathcal{D}) = \mathbb{E}_{x \sim \text{pMR}} [(\mathcal{D}(\mathcal{G}(x)))^2] + \mathbb{E}_{y \sim \text{iUS}} [(\mathcal{D}(y) - 1)^2] \quad (13)$$

$$\mathcal{L}(\mathcal{G}) = \mathbb{E}_{x \sim \text{pMR}} [(\mathcal{D}(\mathcal{G}(x)) - 1)^2] \quad (14)$$

• Final Objective Function

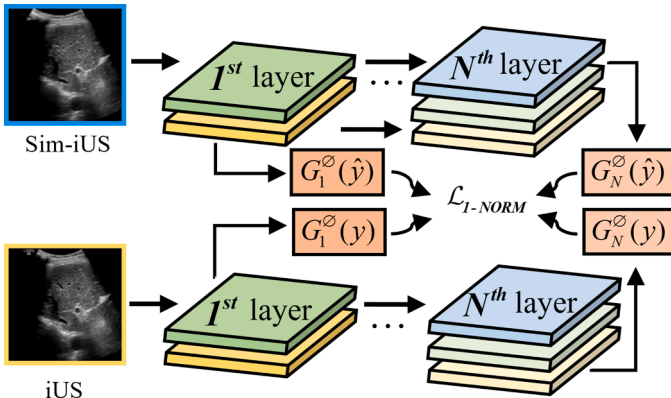


Fig. 4. The overview of texture loss.

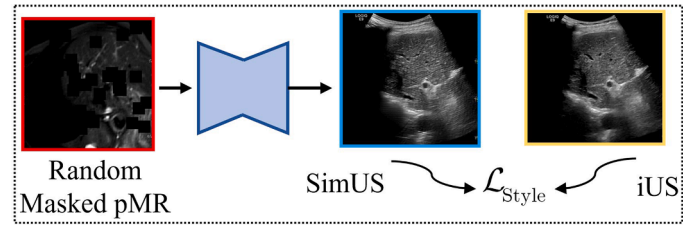


Fig. 5. The overview of style loss.

We settle the task of iUS simulation from pMR through a decoupled joint loss function. Our anatomy preserving GAN combines structural and style information, and employs a perceptive encoder, a decoder with bimodal fusion, fixed convolutional layers and correction modules to generate realistic results. Eventually, the overall objective function for our generator is a weighted combination of all the above losses:

$$\mathcal{L}_{\text{Ours}}(\mathcal{G}) = \alpha_{\text{content}} \mathcal{L}_{\text{content}}(\mathcal{G}) + \lambda_{\text{structure}} \mathcal{L}_{\text{structure}}(\mathcal{G}) + \gamma_{\text{appear}} \mathcal{L}_{\text{appear}}(\mathcal{G}) + \eta \mathcal{L}(\mathcal{G}) \quad (15)$$

3. Experiments and Results

In this section, the dataset collected in a real clinical scene was firstly introduced. Next, qualitative and quantitative comparison was conducted between the proposed ApGAN with the state-of-the-art simulation algorithms. Afterwards, the ablation study was conducted to ensure the validity of each proposed module and loss function. Specially, the function of the correction module was also separately evaluated. Furthermore, in order to prove that the Sim-iUS simulated from pMR have no loss of structural details and generate reliable anatomical information like the iUS, a set of experiments and analyses were carried out on specific anatomical structures and lesion areas of the liver.

The proposed ApGAN was trained using the Adam optimizer based on the final objective function, and hyper-parameters were set as follows: constant learning rate of 2×10^{-4} for the first 100 epochs then linearly decaying to 0, β_1 of 0.5 and β_2 of 0.999; 200 maximum epochs; batch size of $\alpha_{\text{content}}, \lambda_{\text{structure}}, \gamma_{\text{appear}}$ and η were set equal to 1, 5, 5, and 1, respectively. Specifically, while training, when the corresponding weight was increased from 1 for structural-correction loss, Sim-iUS containing key anatomical structures was demonstrated more completely, and the SSIM metrics reached the maximum value at $\lambda_{\text{structure}}$ taken 5. In terms of appearance-aware loss, in order to make Sim-iUS more similar to the real US in terms of texture and artifacts, the γ_{appear} setting was adjusted to 5 according to the variation of FID and KID, so as to achieve the best overall simulation effect. It is worth mentioning that the correction module of the proposed ApGAN trained and updated the parameters by gradient backpropagation together with the generator. The model and all experiments were implemented in PyTorch 1.6.0 and developed on two NVIDIA RTX GPUs with 2*12 GB memory.

3.1. Training Dataset

An institutional research ethics board-approved retrospective review of patients with focal liver lesions was performed, and the requirement for informed consent was waived. The dataset was collected from the Nanjing Drum Tower Hospital between April 2020 and December 2021.

Preoperative MR. The preoperative T1 MR scanning was performed with Gd-EOB-DTPA (Primovist, Bayer-Schering Pharma, Osaka Japan) as the contrast agent. Gd-EOB-DTPA was injected into the vein and distributed with blood to normal or abnormal tissues.

Paired pMR and iUS datasets. The paired pMR and iUS images were collected using the LOGIQ E9 ultrasound-navigation system (GE Healthcare, Milwaukee, WI, United States) during liver puncture surgery. LOGIQ E9 ultrasound-navigation system furnished with the electromagnetic tracking unit trakSTARTM was adopted to ensure the same

imaging plane for pMR and iUS scanning. A total of 4,130 pMR and iUS image pairs were obtained, and the pMR and iUS images were resized to 256×256 . For model training, the paired set was split into training and test sets by 80-20% ratio for a five-fold cross-validation by using a patient-specific separation in order to ensure images from the same patient are not be used both in training and testing. This five-fold cross-validation approach was employed in the training data to acquire the means and standard deviations of the evaluation metrics.

3.2. Comparison with State-of-the-art Methods

The proposed ApGAN has been compared with several state-of-the-art simulation methods. Primarily, the obtained Sim-iUS is presented by CycleGAN [28], Pix2Pix [29], SASAN [30], StarGAN [31], Pix2Pix GAN [37], Conditional GAN [38] and our ApGAN. Then, five widespread-used metrics, including Structural Similarity (SSIM), Peak Signal to Noise Ratio (PSNR), Fréchet Inception Distance (FID), Kernel Inception Distance (KID) and Mean Absolute Error (MAE), are picked to measure the quality of Sim-iUS. Through comprehensive experiments,

the superiority of the ApGAN is proved. The pMR and iUS pairs are shown in the first and second row of Fig. 6, and StarGAN, CycleGAN, SASAN, Pix2Pix, Pix2Pix GAN, Conditional GAN and the trained ApGAN to produce Sim-iUS are presented in the third to ninth rows, respectively. Besides, broadly speaking, Sim-iUS from the ApGAN has finer internal detail and reveals more anatomically layered features (indicated by green arrows).

As is seen from Fig. 6, StarGAN, a scalable model that can be simulated across domains, generates Sim-iUS close to the style of real US. However, while analyzing the details of Sim-iUS, apart from the anatomical content and the vascular boundaries of the liver being too blurred (indicated by the red arrow in the third row), areas of distribution of highlighted structures, such as the liver capsule, extending over a wide area, were found, which might overlap with other focal areas and cover the liver lesions such as polyps.

CycleGAN failed to simulate the beam divergence, but produced angular contents with low sharpness (red arrows in the fourth row), which is not consistent with US imaging principles. It was thus inferred that it lacked enough constraints on structures and lost detailed

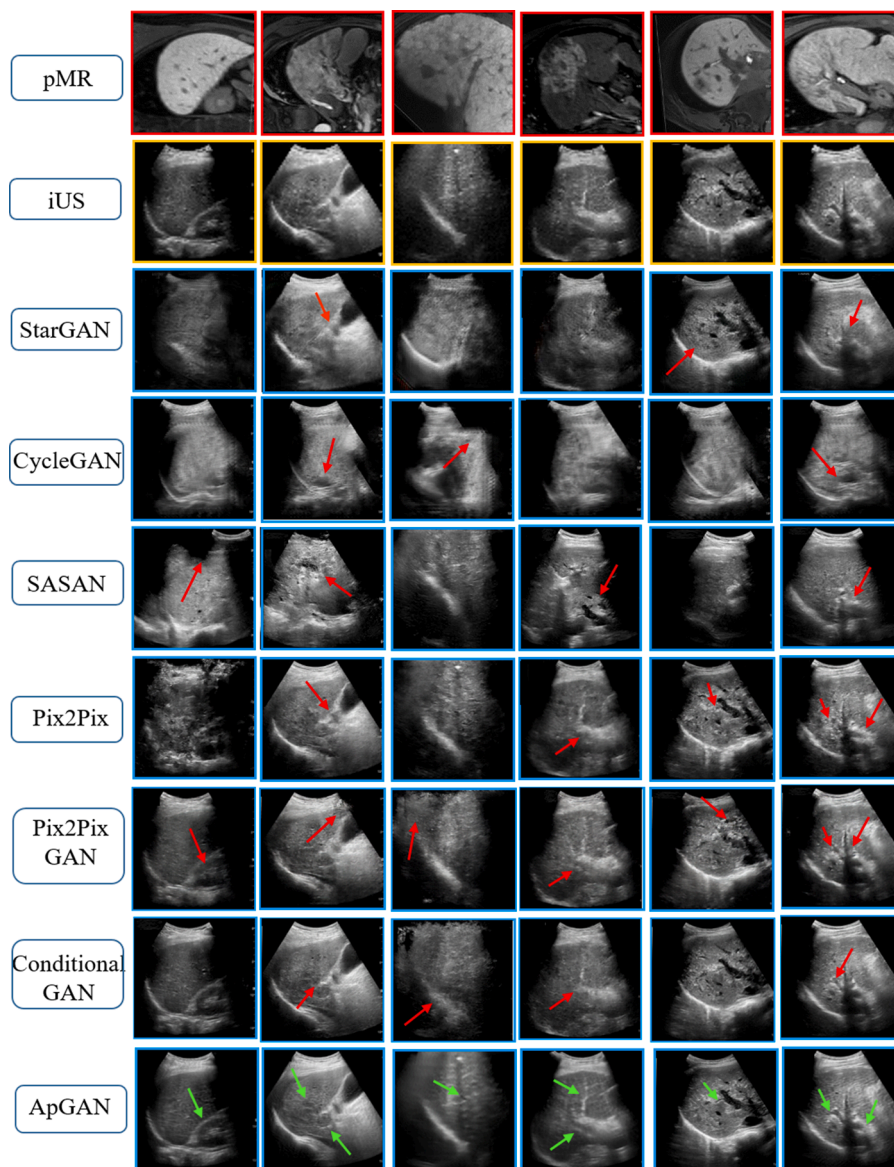


Fig. 6. The first two rows are the real pMR and iUS, and then the corresponding Sim-iUS obtained by StarGAN, CycleGAN, SASAN, Pix2Pix, Pix2Pix GAN, Conditional GAN and our ApGAN are listed. Red arrows indicate the missing, distorted, or blurry areas, and green arrows illustrate our high-precision structures by ApGAN.

information. However, the bimodal fusion and structural correction modules of the proposed ApGAN were both dedicated to mining the valid parenchyma and structural information of liver in pMR. Besides, CycleGAN yielded anatomical structures that did not present in pMR (e.g., kidneys, red arrows in the second and third columns).

SASAN proposed to learn the spatial normalization of self-attention for the intermediate activation states of the generator and to enhance the use of semantic information. It was observed during training that the attention maps did not focus on the presence of important anatomical structures, but were more evenly targeted to the whole image. As a result, the Sim-iUS provided little detailed anatomical content and adjacent spatial relationships, or even no shape to speak of (red arrows in the fifth row).

Pix2Pix is generally a classical generation model that combines L1 loss to render highly-structured outputs as the grounding truth. Surprisingly, herein, it also yielded Sim-iUS that were blurry in certain regions, leaving adjacent structures without clear demarcation, or even turning into connected regions (red arrows in the sixth row). Though the features were close to the speckle noise, the blurring of boundaries still had an impact on the surgeons doing image analysis.

Additionally, as the coarse segmentation maps from pMR are utilized in ApGAN, in order to ensure fairness in the experiments, we conduct comparisons with Pix2Pix GAN and Conditional GAN which both take segmentation maps as input additionally.

Pix2Pix GAN chooses Pix2Pix as the backbone, and the difference is that it has two paths to update the weights, which adds an external path to provide the comparison results between ground truth and fake images from the discriminator. We combine paired pMR and corresponding segmentation maps, and feed them into the model for US simulation.

Conditional GAN aims to learn a non-linear mapping between the input image and a random noise vector to yield the output. Thus, we extract feature embedding of segmentation maps as conditions to facilitate US simulation tasks.

It can be seen that the ApGAN obtains better results than Pix2Pix GAN and Conditional GAN in various quantitative metrics. Meanwhile, it can be observed in Fig. 6 that the generated Sim-iUS in Conditional GAN are completer and more accurate in terms of content and structures, compared to Pix2Pix GAN. However, compared to ApGAN, the Sim-iUS of Conditional GAN is blurry or lacks anatomical structures in some cases, which further proves the superiority of the ApGAN in terms of preserving high precision anatomical structures.

The last row showed that the ApGAN outperformed any of the above models. ApGAN generated Sim-iUS that could be observed and utilized by physicians, with clear anatomical details of the liver and well-defined boundaries of structures such as the portal veins (indicated by green arrows), capable of simulating different grey levels due to diverse acoustic parameters of human tissues, thereby resulting in highlighted areas representing different anatomical contents. Besides, the generated artefacts were more realistic and more closely resemble the qualities of real iUS.

The corresponding quantitative results are listed in TABLE 1, where it could be observed that the ApGAN received the best evaluation results. Concretely, ApGAN acquired a 19.4% improvement in SSIM relative to Pix2Pix; and had a higher PSNR from 23.92 to 26.76. Besides, FID was also reduced by about 40%. Meanwhile, ApGAN achieved a reduction in KID from 42.48 to 0.86 relative to CycleGAN, and SSIM boasted a 270% increase. In conclusion, the proposed ApGAN obtained the most realistic Sim-iUS.

3.3. Ablation Study

In ablation experiments, comparison was drawn between the baseline (CUT) and several progressively improved models, which own the sequential addition of a correction module to the baseline for structural correction of the intermediate Sim-iUS (CUT+C), a bimodal fusion module (CUT+CB), fixed convolution layers in the decoder (CUT+CBF)

Table 1

The quantitative metrics of state-of-the-art and our ApGAN. Best results are marked bold.

	SSIM(\uparrow)	PSNR(\uparrow)	FID(\downarrow)	100 \times KID(\downarrow)	MSE(\downarrow)
StarGAN [27]	0.42 ± 0.11	14.90 ± 3.87	126.65 ± 0.69	8.62 ± 0.03	0.17 ± 0.13
CycleGAN [24]	0.20 ± 0.11	11.47 ± 2.09	351.89 ± 0.28	42.58 ± 0.17	0.32 ± 0.20
SASAN [26]	0.41 ± 0.07	12.95 ± 1.98	88.26 ± 0.28	6.16 ± 0.22	0.22 ± 0.08
Pix2Pix [25]	0.62 ± 0.06	23.92 ± 2.17	55.52 ± 0.28	3.86 ± 0.11	0.01 ± 0.01
Pix2Pix GAN [37]	0.63 ± 0.06	23.94 ± 1.91	75.64 ± 0.17	5.91 ± 0.01	0.02 ± 0.02
Conditional GAN [38]	0.69 ± 0.08	24.91 ± 2.42	48.89 ± 0.21	2.28 ± 0.01	0.01 ± 0.02
Our ApGAN	0.74 ± 0.08	26.76 ± 2.90	35.54 ± 0.12	0.86 ± 0.06	0.01 ± 0.01

to resolve checkerboard artifacts, and style loss (CUT+CBF+Style) and texture loss (CUT+CBF+Texture) that constrain the Sim-iUS to be closer to the real US style.

TABLE 2 shows the results of the ablation experiments at the pixel level (SSIM and MAE) and the feature level (FID and 100 \times KID), respectively. It could be observed that CUT+C increased SSIM by 0.09, while MAE and FID decreased substantially by 0.12 and 58.91, respectively. The adopted approach of learning jointly with the correction module seemed to be effective in resolving the structural distortion and approaching the real US style, which was evidenced by the fact that the performance of SSIM and FID was greatly improved. In addition, CUT+CB increased by 0.1 over CUT+C on SSIM by mining the effective parenchymal content of liver with the bimodal module. Then, compared with CUT+CB, CUT+CBF for checkerboard artefacts reduced FID and KID by 15.3 and 2.05, respectively. Meanwhile, the addition of two loss functions (CUT+CBF+Style+Texture) in the last row that perceive the stylistic and textural features improved the proposed ApGAN in all metrics, especially on FID, KID and PSNR, which decreased by 41.98, 3.15 and increased by 5.92 from CUT+CBF, respectively. It was further demonstrated that the image quality of the proposed Sim-iUS was gradually improved, with a high similarity to the real US at both the pixel and the feature level. The results confirmed the validity of all components of the model, and ApGAN generated better Sim-iUS to help with ultrasound-guided liver surgery.

3.4. Analysis for the simulation of specific anatomical structures

To more closely match the liver surgery in the real clinical trial scenario, and convince the physicians of the authenticity of the Sim-iUS

Table 2

The quantitative metrics (means \pm standar) of ablation experiments on our ApGAN.

	SSIM(\uparrow)	PSNR(\uparrow)	FID(\downarrow)	100 \times KID(\downarrow)	MSE(\downarrow)
CUT (Baseline)	0.38 ± 0.06	13.55 ± 1.59	166.30 ± 0.52	16.45 ± 0.11	0.18 ± 0.07
CUT+C	0.47 ± 0.06	18.09 ± 1.46	107.39 ± 0.43	7.56 ± 0.06	0.06 ± 0.02
CUT+CB	0.57 ± 0.06	19.97 ± 1.55	92.89 ± 0.37	6.06 ± 0.08	0.04 ± 0.01
CUT+CBF	0.60 ± 0.09	20.84 ± 3.13	77.52 ± 0.28	4.01 ± 0.06	0.04 ± 0.02
CUT+CBF+Style	0.61 ± 0.08	21.12 ± 2.88	69.74 ± 0.13	3.50 ± 0.10	0.03 ± 0.02
CUT+CBF+Texture	0.64 ± 0.07	23.00 ± 2.22	67.77 ± 0.29	3.39 ± 0.05	0.02 ± 0.01
Our ApGAN	0.74 ± 0.08	26.76 ± 2.90	35.54 ± 0.12	0.86 ± 0.06	0.01 ± 0.01

simulated from pMR, several unique anatomical structures and focal areas involved in the adopted dataset were individually analyzed. The anatomical structures and lesion areas in the Sim-iUS were assessed for consistency with the relevant image information in the real iUS. On the one hand, the anatomical structures (e.g. liver capsule, kidney, blood vessels, etc.) present in Sim-iUS were compared with those in real iUS, with the offset errors counted. On the other, the consistency of the real iUS with whether the simulated lesion areas (distances of calculi, polyps, cysts, etc. from blood vessels, liver capsule, etc.) was further analyzed. During the evaluation, the Hausdorff Distance (HD) and Euclidean Distance (ED) of these structures between the Sim-iUS and real iUS were computed.

• Simulation of linear and annular anatomical structures

The presence of distinct linear and annular anatomical structures in the Sim-iUS provided surgeons more detailed information for identification and localization of typical liver parenchyma or lesions, making it necessarily important to ensure the accuracy of these two types of structures.

For the linear structures, the simulation effect of the liver capsules was investigated, as shown in Fig. 7(a). In this case, experts marked points in both Sim-iUS and real iUS that were able to depict the contours of the liver capsules (e.g. the blue and red points in Fig. 7(a)), and the HD error of linear offset for each sample was then calculated. Meanwhile, to observe the overall simulation effect of ApGAN on the linear structures, the average HD errors of the liver capsule for each patient were counted, as shown in Fig. 7(b). The average HD errors of all patients were less than 0.25 mm, suggesting the high-level efficiency of the proposed ApGAN in simulating the linear US anatomy.

The annular structures were divided into two categories according to the type of anatomical structure: one type of analysis was for the annular structures of the gallbladder or kidney itself (as shown in Fig. 8(a)). The HD error of annular offset for gallbladder and kidney structures was calculated, and the average annular error for each of the two anatomical structures was again assessed (as shown in Fig. 8(b)). The best results were achieved with an error of 0.14 mm in the gallbladder and a minimum error of 0.21 mm in the kidney, both of which indicated the good simulation performance of the proposed ApGAN in various annular structures.

Another type of annular structure was the ring diagram formed by calculating the relative distances between various anatomical structures (as shown in Fig. 9(a)), and the focus was hereby placed on analyzing the relative distance errors between vessel sections in this part. The ED errors between adjacent points were calculated based on the marks of experts, and the mean ED error across all marker points for each sample was then counted. Besides, the average relative ED errors across all Sim-iUS containing vessel cross-sections for each patient were also plotted, as shown in Fig. 9(b). The average relative ED errors were below 0.17mm for all patients except patient 1, which was only around 0.27mm. This

implied that the relative positions of the anatomical structures in the Sim-iUS did not change significantly and remained relatively stable. In conclusion, the proposed ApGAN works efficiently in simulating anatomical structures with high accuracy, whether in the form of linear or annular shapes

• Simulation of focal areas

During ultrasound-guided liver surgery, surgeons will focus on the areas of the lesion to accordingly determine the condition of patients. Therefore, the accuracy of the structural information of the lesion should also be ensured in the simulation. A total of three types of focal structures were covered in this paper, i.e., cysts, polyps, and calculi. The distances of the cross-sections of the focal structures relative to other anatomical structures (vascular cross-sections, liver capsule and gallbladder) were calculated to determine whether the focal structures in Sim-iUS matched the true iUS.

For cysts, a total of 10 samples were selected to calculate the average relative ED error from the cross-sections of cysts to the obvious vessel cross-section that was 0.14 mm (two examples are shown in Fig. 10(a) and Fig. 10(b)). Thus, it confirmed that the proposed ApGAN had a high precision in guaranteeing the invariance of the position of the cysts relative to other anatomical structures during the simulation. For polyps, the ED errors from the polyp to the anatomical structure (liver capsule or gallbladder) were calculated based on the marker points of expert (as shown in Fig. 10(c) and Fig. 10(d)). During the statistical process, the average ED error for 10 labelled samples from the polyp to the liver capsule and 10 labelled samples from the polyp to the gallbladder was computed, which were 0.07 mm and 0.12 mm, respectively.

The expert-labelled samples were divided into two categories for calculi, i.e., from calculus to vascular cross-section and from calculus to gallbladder, and the relative ED errors were then calculated separately. Similarly, the average relative errors for 10 samples selected for each category were also calculated, which were 0.18 mm and 0.15 mm, respectively. Hence, the proposed ApGAN was also highly accurate in simulating lesion structures like polyps.

4. Discussion

An anatomy preserving GAN was proposed for performing liver intraoperative US simulation from pMR. The empirical results demonstrated that ApGAN retained highly accurate information about the focal structures, without altering their relative positions in relation to the other anatomical structures. The proposed method could assist the US guidance in liver surgery from two aspects, including facilitating the surgeons to understand the characteristics of US in advance, and expanding the iUS dataset to promote automatic iUS analysis.

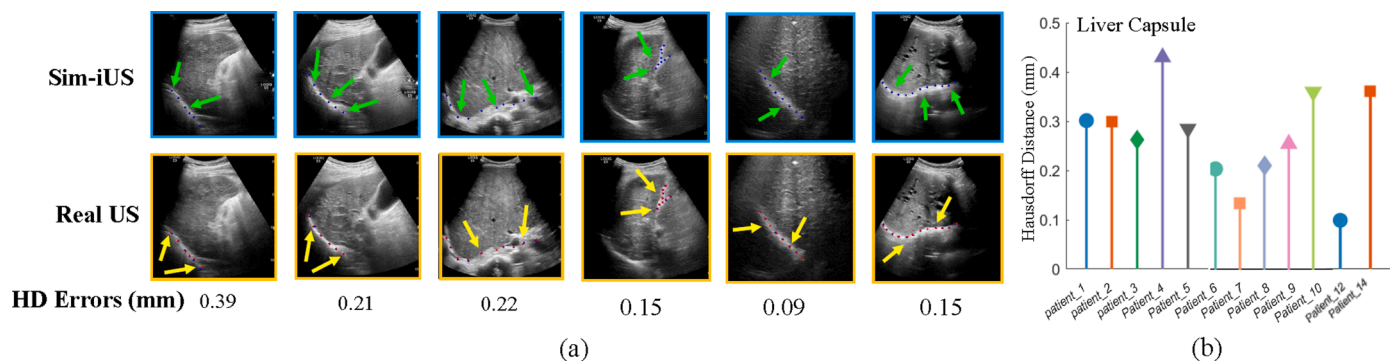


Fig. 7. Calculation of HD errors for linear structures (liver capsule) and statistical plots of average HD errors.

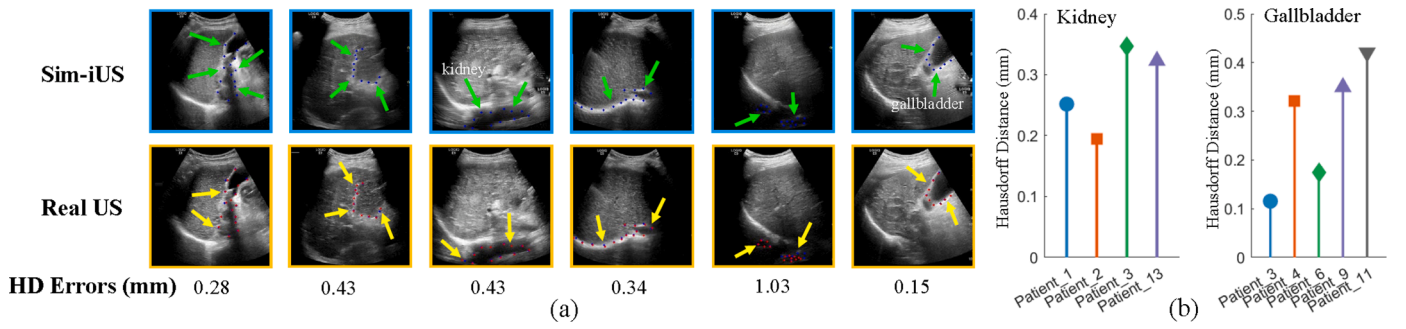


Fig. 8. Calculation of HD errors for various annular structures and statistical plots of average HD errors.

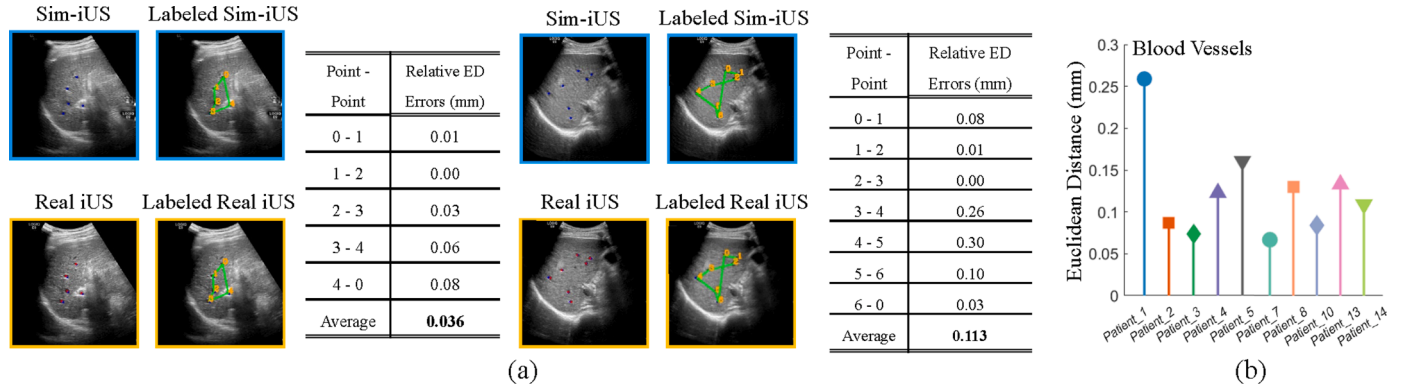


Fig. 9. Average relative ED errors between multiple vessel cross-sections and average ED error per patient.

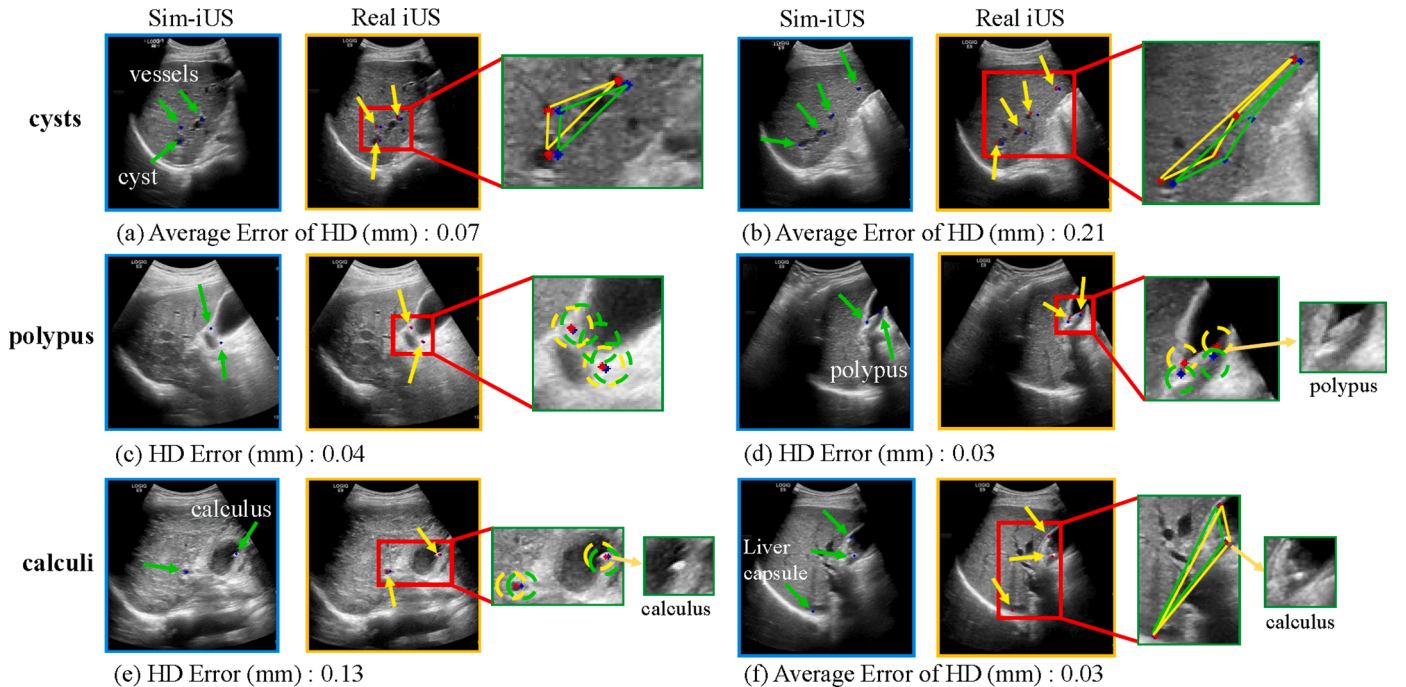


Fig. 10. Two cases that containing cysts, polyps, and calculi lesions to calculate average relative HD errors.

4.1. User Study

To validate whether the acquired Sim-iUS could be employed as training materials to facilitate the surgeons to understand the characteristics of US in advance, five experts from the ultrasound department with extensive diagnostic and surgical experience were invited to

evaluate the quality of Sim-iUS. A total of 160 images of four different types were randomly selected: Sim-iUS generated by three models (our ApGAN, Pix2Pix and StarGAN, the first three best models described in Section III-B) and the corresponding real iUS, with 40 images of each type. Each expert was required to view each image and determine whether it was a real or fake liver iUS image.

Ideally, the accuracy for real iUS should be close to 1, while a higher authenticity of Sim-iUS indicated more difficulties in distinguishing between the real and the fake. Among three models, ApGAN achieved lower accuracy than the other two methods, which made it rather difficult for the experts to distinguish the Sim-iUS from real iUS images, and could thus be employed as training materials to facilitate the surgeons to understand the characteristics of iUS in advance. The corresponding radar plots are shown in Fig. 11. For clear demonstration, we retain the value of real iUS, and plot the GAN-based images using (1-Accuracy) as the quantity. It can be seen that the value range of ApGAN is between 0.6 and 1.0, while there are results below 0.6 in Pix2Pix and StarGAN, indicating that some of the images generated by these two methods are easily distinguishable.

4.2. Promoted effect to iUS classification task

To explore whether the proposed ApGAN method could expand the iUS dataset with high-precision anatomical structures to promote the AI-based structure classification task of iUS, the results of three dataset settings were compared. The baseline dataset contained all real iUS, including 378 positive samples containing gallbladder structure, and 557 negative samples without. The second Sim-iUS expanding dataset added 40% positive and negative samples generated from the proposed ApGAN to the baseline data. The third Real-iUS expanding dataset added 40% samples from additional real iUS. The well-established AlexNet was adopted as the downstream structure classification model, since AlexNet was well recognized with low false-positive and good accuracy in medical US classification [36].

The structure classification results are presented in TABLE 3, where it can be observed that the proposed method not only acquired significant performance improvement (ACC: 0.9139 vs. 0.7950), but also achieved similar evaluation results compared with the dataset expanding method of using the additional real iUS. In this case, ApGAN method could effectively expand the iUS dataset with precise anatomical structures to

Table 3

Comparison of the structure classification performance with different datasets.

Training Data	Accuracy	Precision	Recall	F1
Baseline dataset	0.8278	0.9047	0.3220	0.4750
Sim-iUS expanding dataset	0.9139	0.9750	0.6610	0.7878
Real-iUS expanding dataset	0.9221	0.9166	0.7457	0.8224

promote the structure identification task.

According to empirical results and the user study, the task of iUS simulation accomplished by the proposed ApGAN is rather promising for clinical applications: (i) Sim-iUS can be utilized as training materials for doctors to learn and interpret iUS; (ii) it facilitates the surgeons to understand the characteristics of iUS in advance and to better synchronize the analysis of iUS and pMR intraoperatively; (iii) the Sim-iUS is beneficial to expand iUS datasets and promote the application of AI in ultrasound guidance. Last but most importantly, ApGAN, which generates Sim-iUS with highly accurate structures, further increases the confidence of surgeons and the potentials of AI technology in the ultrasound navigation system.

5. Conclusion

In this paper, we propose a novel anatomy preserving GAN, namely ApGAN for liver US simulation from pMR. Extensive experiments on clinical data validate that our model generates more realistic US images with precise anatomical structures and lesion areas. The generated US images can be employed as training materials to facilitate the surgeons to understand the characteristics of iUS in advance and promote the iUS intelligent analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial

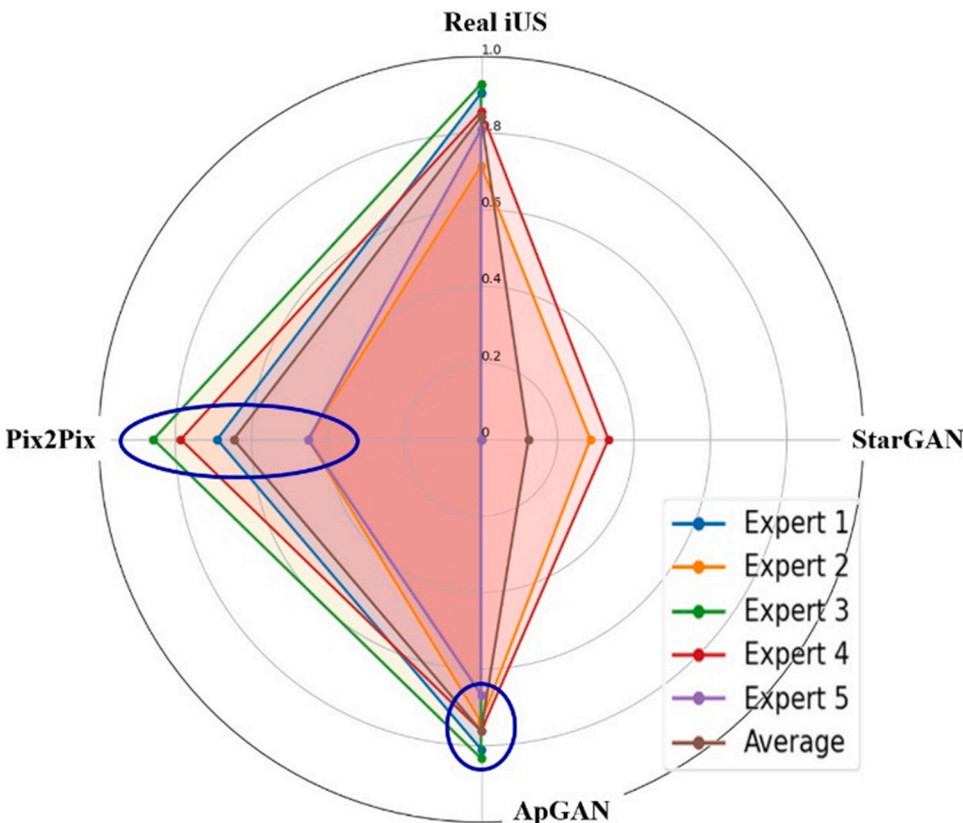


Fig. 11. Radar charts showing the results of the user study. Four different types of images are used: real iUS and Sim-iUS obtained by StarGAN, Pix2Pix and our ApGAN. The accuracy for real iUS should be close to 1, while for the three GAN-Based models, the higher the vertex value of the radar chart is, the more difficult expert is to distinguish between true and false, reflecting the authenticity of simulated ultrasound. The blue circle shows the changing areas of expert interpretation accuracy. Even though Pix2Pix achieves relatively realistic results, there is a significant difference in interpretation accuracy among experts (with a large elliptical range). Comparatively, the ApGAN corresponding results are relatively uniform, with a small range of interpretation accuracy, indicating that the Sim-iUS obtained in this paper has higher fidelity.

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Fang Chen is the corresponding author. The authors acknowledge supports from National Nature Science Foundation of China grants (U20A20389, 62271246, 62136004), Beijing Hospitals Authority Clinical Medicine Development of special funding support (XMLX202141).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.cmpb.2023.107642](https://doi.org/10.1016/j.cmpb.2023.107642).

References

- [1] M.G. Lubner, L. Mankowski Gettle, D.H. Kim, T.J. Ziemlewicz, N. Dahiya, P. Pickhardt, Diagnostic and procedural intraoperative ultrasound: technique, tips and tricks for optimizing results, *The British journal of radiology* 94 (1121) (2021), 20201406 no.
- [2] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S.X. Li, D. Ni, T. Wang, Deep learning in medical ultrasound analysis: a review, *Engineering* 5 (2) (2019) 261–275.
- [3] C. Sietses, S. Meijer, M. Van den Tol, The impact of intraoperative ultrasonography on the surgical treatment of patients with colorectal liver metastases, *Surgical endoscopy* 24 (8) (2010) 1917–1922.
- [4] Y. Barash, E. Klang, A. Lux, E. Konen, N. Horesh, R. Pery, N. Zilkha, R. Eshkenazy, I. Nachmany, N. Pencovich, Artificial intelligence for identification of focal lesions in intraoperative liver ultrasonography, *Langenbeck's Archives of Surgery* (2022) 1–8.
- [5] G.A. Parker, W. Lawrence Jr, J. Horsley 3rd, J.P. Neifeld, D. Cook, M.J. Koretz, Intraoperative ultrasound of the liver affects operative decision making, *Annals of surgery* 209 (5) (1989) 569.
- [6] H. Zhou, T. Jiang, Q. Li, C. Zhang, C. Zhang, Y. Liu, J. Cao, Y. Sun, P. Jin, J. Luo, et al., Us-based deep learning model for differentiating hepatocellular carcinoma (hcc) from other malignancy in cirrhotic patients, *Frontiers in oncology* (2021) 1640.
- [7] W. Xue, C. Cao, J. Liu, Y. Duan, H. Cao, J. Wang, X. Tao, Z. Chen, M. Wu, J. Zhang, et al., Modality alignment contrastive learning for severity assessment of covid-19 from lung ultrasound and clinical information, *Medical image analysis* 69 (2021), 101975.
- [8] C. Dietrich, A. Ignee, C. Greis, X. Cui, D. Schreiber-Dietrich, M. Hocke, Artifacts and pitfalls in contrast-enhanced ultrasound of the liver, *Ultraschall in der Medizin-Europäischer Journal of Ultrasound* 35 (02) (2014) 108–128.
- [9] E.J. Hagopian, Liver ultrasound: A key procedure in the surgeon's toolbox, *Journal of Surgical Oncology* 122 (1) (2020) 61–69.
- [10] G.P. Penney, J.M. Blackall, D. Hayashi, D.J. Hawkes, et al., Overview of an ultrasound to ct or mr registration system for use in thermal ablation of liver metastases, *Proc. Medical Image Understanding and Analysis* 1 (2001) 6568.
- [11] L. Zhang, T. Portenier, O. Goksel, Deep image translation for enhancing simulated ultrasound images, in: *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis* (2020) 85–94.
- [12] J. Liang, X. Yang, H. Li, Y. Wang, M.T. Van, H. Dou, C. Chen, Z. Mai, et al., Synthesis and edition of ultrasound images via sketch guided progressive growing gans, in: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2020, pp. 1793–1797.
- [13] D. Aiger, D. Cohen-Or, Real-time ultrasound imaging simulation, *Real-Time Imaging* 4 (4) (1998) 263–274.
- [14] D. Henry, J. Troccaz, J.-L. Bosson, O. Pichot, Ultrasound imaging simulation: Application to the diagnosis of deep venous thromboses of lower limbs, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 1998, pp. 1032–1040.
- [15] O. Kutter, R. Shams, N. Navab, Visualization and gpu-accelerated simulation of medical ultrasound from ct images, *Computer methods and programs in biomedicine* 94 (3) (2009) 250–266.
- [16] R. Shams, N. Navab, Real-time simulation of medical ultrasound from ct images, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2008, pp. 734–774.
- [17] J.A. Jensen, Field: A program for simulating ultrasound systems, in: *10TH NORDIC/BALTIC CONFERENCE ON BIOMEDICAL IMAGING* 4, Citeaser, 1996, pp. 351–353. VOLSUPPLEMENTPART 1.
- [18] P. Rubi, E.F. Vera, J. Larrabide, M. Calvo, J. D'Amato, I. Larrabide, Comparison of real-time ultrasound simulation models using abdominal ct images, in: *12th international symposium on medical information processing and analysis* 10160, SPIE, 2017, pp. 55–63.
- [19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, *Communications of the ACM* 63 (11) (2020) 139–144.
- [20] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, S. Chen, T. Wang, S. Wang, Skin lesion segmentation via generative adversarial networks with dual discriminators, *Medical Image Analysis* 64 (2020), 101716.
- [21] B. Lei, Y. Zhao, Z. Huang, X. Hao, F. Zhou, J. Qin, H. Lei, Adaptive sparse learning using multi-template for neurodegenerative disease diagnosis, *Medical Image Analysis* 61 (2020), 101632.
- [22] S. Hu, B. Lei, S. Wang, Y. Wang, Y. Shen, Bidirectional mapping generative adversarial networks for brain mr to pet synthesis, *IEEE Transactions on Medical Imaging* 41 (1) (2021) 145–157.
- [23] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, and Z. Xu, “Unsupervised mr-to-ct synthesis using structure-constrained cyclegan,” *IEEE transactions on medical imaging*, vol. 39, no. 12, pp. 4249–4266.
- [24] H. Emami, M. Dong, S.P. Nejad-Davaran, C.K. Glide-Hurst, Sa-gan: Structure-aware gan for organ-preserving synthetic ct generation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 471–481.
- [25] S. Vitale, J.I. Orlando, E. Iarussi, I. Larrabide, Improving realism in patient-specific abdominal ultrasound simulation using cyclegans, *International journal of computer assisted radiology and surgery* 15 (2) (2020) 183–192.
- [26] D. Tomar, O. Goksel, Content-preserving unpaired translation from simulated to realistic ultrasound images, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 659–669.
- [27] J. Liang, X. Yang, Y. Huang, H. Li, S. He, X. Hu, Z. Chen, W. Xue, J. Cheng, D. Ni, Sketch guided and progressive growing gan for realistic and editable ultrasound image synthesis, *Medical Image Analysis* 79 (2022), 102461.
- [28] J.-Y. Zhu, T. Park, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [29] P. Isola, J.-Y. Zhu, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [30] D. Tomar, G. Vray, J.-P. Thiran, Self-attentive spatial adaptive normalization for cross-modality domain adaptation, *IEEE Transactions on Medical Imaging* 40 (10) (2021) 2926–2938.
- [31] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, J. Choo, Stargan: Unified generative adversarial networks for multi-domain image-to-image translation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.
- [32] T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image-to-image translation, *European conference on computer vision* (2020) 319–345.
- [33] Z. Liu, Y. Shen, V.B. Lakshminarasimhan, P.P. Liang, A. Zadeh, L.-P. Morency, arXiv preprint, 2018.
- [34] A. Odena, V. Dumoulin, C. Olah, Deconvolution and checkerboard artifacts, *Distill* 1 (10) (2016) e3.
- [35] M.S. Sajjadi, M. Hirsch, Enhancenet: Single image super-resolution through automated texture synthesis, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4491–4500.
- [36] M.H. Yap, G. Pons, J. Marti, S. Ganau, M. Sents, R. Zwigglelaar, A.K. Davison, R. Marti, Automated breast ultrasound lesions detection using convolutional neural networks, *IEEE journal of biomedical and health informatics* 22 (4) (2017) 1218–1226.
- [37] D. Popescu, M. Deaconu, L. Ichim, et al., Retinal blood vessel segmentation using pix2pix gan, in: *2021 29th Mediterranean Conference on Control and Automation (MED)*, IEEE, 2021, pp. 1173–1178.
- [38] N. Molinier, G. Painchaud-April, A. Le Duff, et al., Ultrasonic imaging using conditional generative adversarial networks, *Ultrasonics* (2023), 107015.