

Article

AUE-Net: Automated Generation of Ultrasound Elastography Using Generative Adversarial Network

Qingjie Zhang¹, Junjuan Zhao¹, Xiangmeng Long¹, Quanyong Luo², Ren Wang³ and Xuehai Ding^{1,*} and Chentian Shen^{2,*}

¹ School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; ivanzhang@shu.edu.cn (Q.Z.); junjuanzhao@shu.edu.cn (J.Z.); matsuko@shu.edu.cn (X.L.)

² Department of Nuclear Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China; lqyn@sh163.net

³ Department of Ultrasound Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China; ryanwang@126.com

* Correspondence: dinghai@shu.edu.cn (X.D.); qingtian@alumni.sjtu.edu.cn (C.S.)

Abstract: Problem: Ultrasonography is recommended as the first choice for evaluation of thyroid nodules, however, conventional ultrasound features may not be able to adequately predict malignancy. Ultrasound elastography, adjunct to conventional B-mode ultrasound, can effectively improve the diagnostic accuracy of thyroid nodules. However, this technology requires professional elastography equipment and experienced physicians. Aim: in the field of computational medicine, Generative Adversarial Networks (GANs) were proven to be a powerful tool for generating high-quality images. This work therefore utilizes GANs to generate ultrasound elastography images. Methods: this paper proposes a new automated generation method of ultrasound elastography (AUE-net) to generate elastography images from conventional ultrasound images. The AUE-net was based on the U-Net architecture and optimized by attention modules and feature residual blocks, which could improve the adaptability of feature extraction for nodules of different sizes. The additional color loss function was used to balance color distribution. In this network, we first attempted to extract the tissue features of the ultrasound image in the latent space, then converted the attributes by modeling the strain, and finally reconstructed them into the corresponding elastography image. Results: a total of 726 thyroid ultrasound elastography images with corresponding conventional images from 397 patients were obtained between 2019 and 2021 as the dataset (646 in training set and 80 in testing set). The mean rating accuracy of the AUE-net generated elastography images by ultrasound specialists was 84.38%. Compared with that of the existing models in the visual aspect, the presented model generated relatively higher quality elastography images. Conclusion: the AUE-net generated ultrasound elastography images showed natural appearance and retained tissue information. Accordingly, it seems that B-mode ultrasound harbors information that can link to tissue elasticity. This study may pave the way to generate ultrasound elastography images readily without the need for professional equipment.

Keywords: ultrasound elastography; neural networks; generative adversarial networks; attention module; color loss



Citation: Zhang, Q.; Zhao, J.; Long, X.; Luo, Q.; Wang, R.; Ding, X.; Shen, C. AUE-Net: Automated Generation of Ultrasound Elastography Using Generative Adversarial Network. *Diagnostics* **2022**, *12*, 253. <https://doi.org/10.3390/diagnostics12020253>

Academic Editor: Guillermo Rus

Received: 10 December 2021

Accepted: 13 January 2022

Published: 20 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The superficial location of the thyroid gland makes high-resolution ultrasound the imaging modality of choice for the evaluation of diffuse disease and nodules. Patients may be referred for ultrasonography because of palpable abnormalities. However, detection of nonpalpable nodules can be as high as 70% and varies by imaging modality [1]. In general, the prevalence of thyroid nodules is high and they may be neoplastic. Most of nodules are benign proliferative structural changes and benign follicular adenomas, with only a small percentage of malignancy [2].

Currently, further confirmation of malignancy for thyroid nodules mainly relies on Fine Needle Aspiration Cytology (FNAC). But given the high prevalence of thyroid nodules and the fact that most nodules are benign, too many unnecessary FNAC operations were possibly conducted, which results in medical resources wasting. Studies [3–6] showed that tissue elasticity is positively correlated with the likelihood of malignancy. This stimulated considerable progress in the development of a new technique, elastography, in the last decades. Strain elastography (SE), the first introduced ultrasound elastography, is a promising new tool to effectively improve diagnostic accuracy and reduce medical costs. In the last decades, SE combined with conventional ultrasound and FNAC became an important strategy for differentiating benign and malignant thyroid nodules [7].

SE is the most widely used commercial method for direct stiffness assessment during thyroid examinations, but it requires an add-on module to be combined with a conventional ultrasound transducer [8]. In addition, elastography also relies on physicians' experience. Different specialists' skillsets may cause variable tissue deformation from compression. It is difficult to standardize the pressure and strain variations. Nonuniform compression produces intra-observer and interobserver variability [4]. Extensive training and experience are crucial for obtaining reliable and reproducible elastography images and scores. For elastography scoring, the main criterion is the stiffness ratio between the hard area (e.g., blue area) and the nodule area. The second scoring element is the color distribution. These two features correspond to the spatial location relationship and the RGB three-channel relationship in computer image processing. This provides the prerequisite for using computer imaging algorithms to generate elastography images.

In the last decade of computing science, deep learning was gaining monumental traction, with neural networks leading the way. Numerous research works made important contributions in fields, such as the classification of lung and breast diseases [9,10] and the identification and segmentation of gynecological abnormality [11,12]. Since Goodfellow and others proposed generative adversarial networks (GANs) in 2014 [13], GANs were shown a promising future in building image synthesis and data generation techniques. There is an increasing amount of researches on GANs, with each iteration making further progress [14,15]. Due to the excellent performance of GANs on natural image synthesis, GANs were introduced to perform various medical imaging tasks including solving problems associated with data imbalance of virtual STIR images [16], reducing metal artifacts and the radiation dose during digital tomosynthesis [17], as well as helping to process medical image [18,19]. Although neural networks and GANs were used to extract strain images from radio frequency data [20,21], and generate shear wave elastography images [22], we attempted to directly map conventional ultrasound images towards the corresponding strain elastography (SE) images. To the best of our knowledge, this is the first work to apply the nonphysical method to generate strain elastography images from conventional ultrasound images.

The main contributions of this paper are as follows: (1) We introduced GAN networks to thyroid strain elastography image generation. (2) We proposed a whole framework for elastography image generation from raw ultrasound data. We used a preprocessing method to get image pairs from raw data and a conditional GAN-based automated generating network (AUE-net) to generate elastography images. (3) This method can generate elastography images corresponding to thyroid ultrasound images only through nonphysical computer algorithms. In this way, the need for professional equipment and manual intervention will be reduced.

The structure of this paper is organized as following. Section 1 introduces relevant research background and our study. In Section 2, we first introduce the base generative adversarial networks. Second, the datasets, network architecture, experiment details, and evaluation metrics are presented. Section 3 presents the results of the USE-GAN method in fields of visual comparisons and quantitative metrics. Lastly, Section 4 details the conclusion.

2. Materials and Methods

2.1. Generative Adversarial Networks and Pix2pix

Since Goodfellow and others proposed GAN in 2014 [13], many studies explored very important paths using GANs in a variety of medical application scenarios. Bermudez and other researchers [18] trained a GAN to synthesize new T1-weighted brain MRIs with high quality compared to that of real images. In the paper by Gomi and others [17], the authors showed that GAN-based networks could achieve promising results in LDCT denoising. GANs were also used to generate more training data. Shin and his team [23] trained a GAN to generate synthetic brain tumor MRIs and evaluated the performance by a segmentation network subsequently trained with the generated MRI data. Kossen and others [16] successfully generated anonymized and labeled TOF-MRA patches that retained generalizable information and showed good performance for vessel segmentation.

In the fields of the generative adversarial network, the goal of image translation is to translate an input image from one domain to another by using given input-output image pairs as training data. Current state-of-the-art approaches [14,15,24–26] typically employ conditional GANs to optimize the networks either with explicitly labeled paired data or by forcing cyclic consistency with unpaired data. In contrast to L1 loss or MSE loss, the adversarial loss has become a popular choice for many image-to-image tasks and can often produce higher image quality. The discriminator is like a trainable loss function which can be automatically adapted to the differences between the generated and real images in the target domain while training. Image translation is one of the most promising applications in computer vision.

The pix2pix [26] proposed by Isola and his team in 2017 is a common conditional GAN framework for image translation. In this framework, the goal of the generator G is to translate the semantic label graph into a real image, while the goal of the discriminator D is to discriminate between the real image and the translated image. The framework is a supervised learning neural network based on the U-Net architecture. Following this framework, Wang and others [14] proposed a coarse-to-fine generation method in 2018 called pix2pixHD for synthesizing higher resolution images from semantic label graphs. The method first learns low-resolution translations, and then gradually scales up to high-resolution. The quality of synthesis images is effectively improved by applying new adversarial loss, multiscale generator, and discriminator architecture. To reduce the heavy computational burden of convolution of high-resolution feature maps, Liang and others [27] proposed a Laplacian Pyramid Translation Network (LPTN) in 2021 to perform the two tasks of attribute and content detail transformation simultaneously.

2.2. Datasets

The data used in this paper were obtained from Shanghai Jiao Tong University Affiliated Sixth People's Hospital specializing in thyroid disorders in Shanghai, China. The original size of all ultrasound images was 800×555 (in Figure 1a). We collected a total of 726 thyroid ultrasound elastography images with corresponding conventional images (pressed B model ultrasound images obtained when elastography was performed) from 397 patients between 2019 and 2021. The patients included 98 males and 299 females, with a minimum age of 20 and a maximum age of 82, and a mean age of 43.6 (standard deviation of 13.4). Under the specialists' recommendations, thyroid nodules with sizes between 5 and 30 mm were included, while nodules with coarse calcification and predominantly cystic nodules were excluded. To enhance the model generalization capability among different types of equipment, only simple RGB images in the range of 0–255 are used. The training set has 646 images and the testing set has 80 images. This study was approved by the ethics committee of Shanghai Jiao Tong University Affiliated Sixth People's Hospital.

2.3. Network Architecture

To generate strain elastography images from conventional ultrasound images, three key requirements have to be considered: (1) The original input data does not just con-

tain input ultrasound images. They are the focused regions of interest (ROI) manually circled by the specialist (the green dash rectangle in Figure 1) to generate the corresponding elastography images. (2) The ultrasound images are grayscale images, which contain less information. The information of focused features such as nodule locations and tissues need to be extracted to generate elastography images. (3) The elastography image mainly relies on color to distinguish tissue elasticity (shown in Figure 1). Therefore, it is important to consider color information in the generation process, the network structure, and loss function.

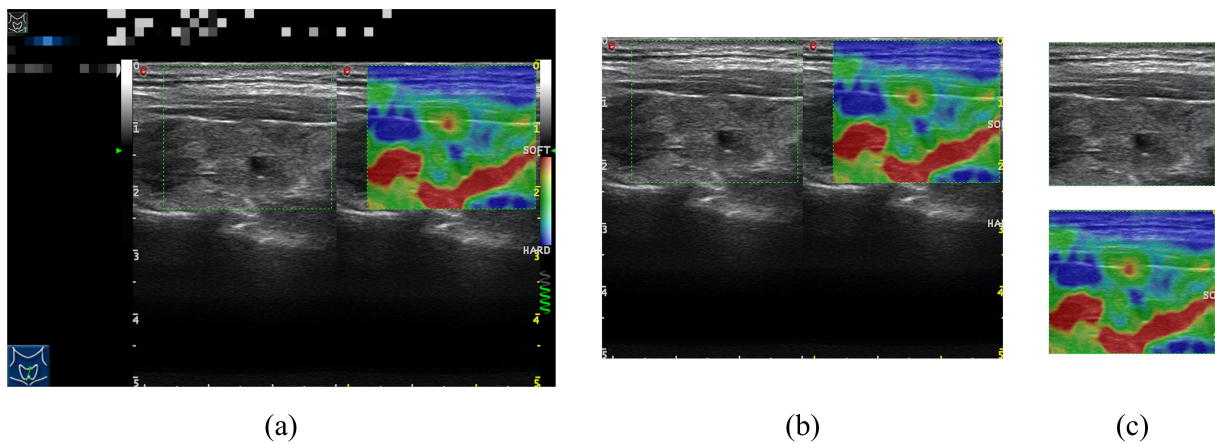


Figure 1. Example image of three steps in data preprocessing. (a) Original images obtained from same ultrasound equipment, which have same dimensions of 800×555 . (b) First cropped image using fixed positions according to equipment information. (c) Final cropped ultrasound image-elastography image pair obtained by color gradient cropping algorithm.

Based on the above considerations, we designed a framework (see Figure 2) to generate elastography images from conventional ultrasound images by first preprocessing and then using the AUE-net (see Figure 3). As shown in Figure 1a, the original data are not a simple ultrasound image-elastography image pair, but only a green box manually circled by a specialist to mark the focus area, and the green box is not uniform in color, size, and position. Therefore, we coiffed the images by an image gradient algorithm.

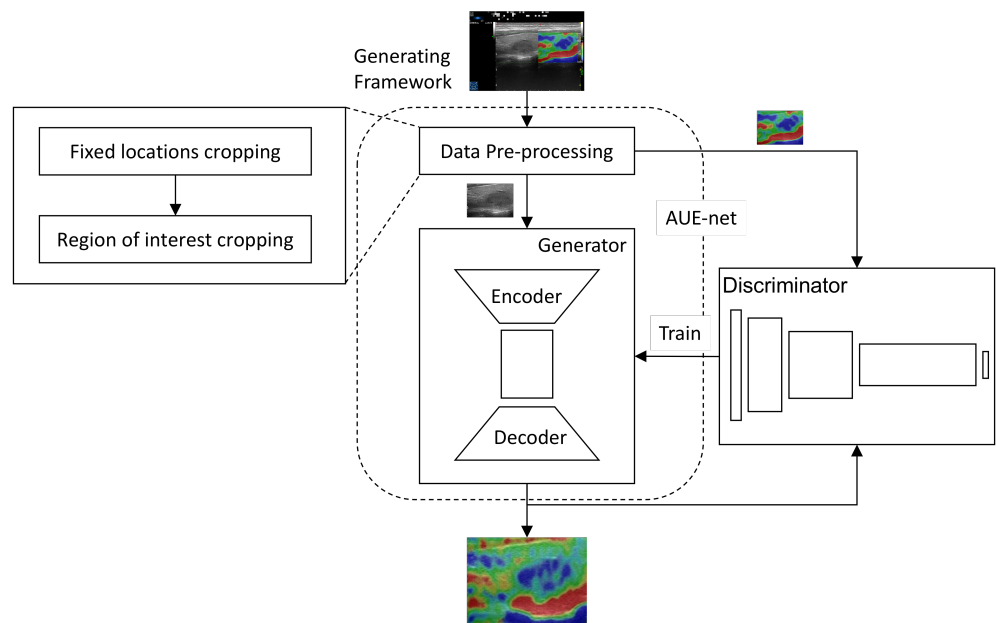


Figure 2. Whole flow of our generation framework.

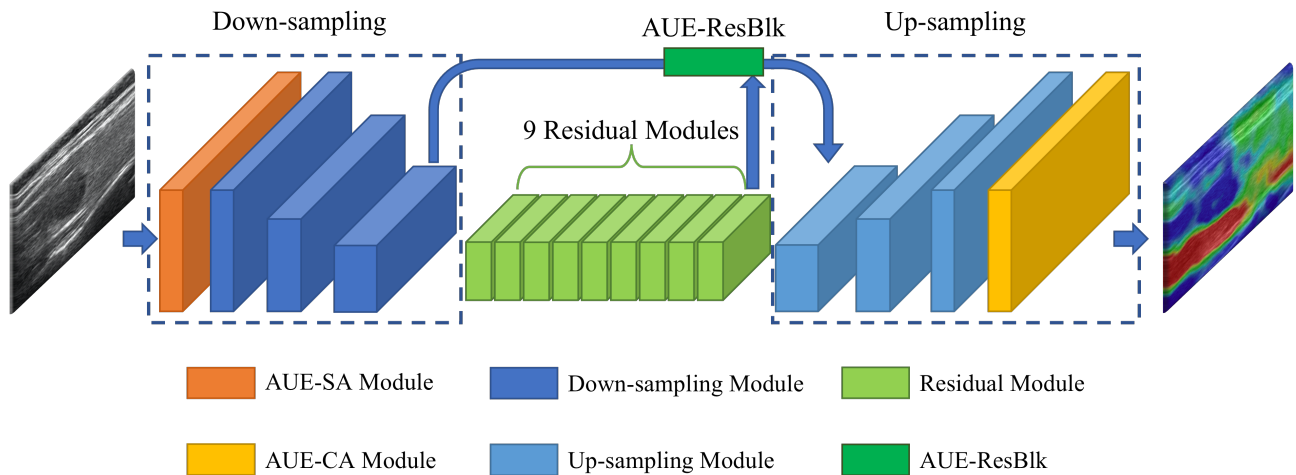


Figure 3. Architecture of AUE-net, which consists of 1 AUE-SA module, three down-sampling modules, nine residual blocks, three up-sampling modules, one AUE-CA module, and one AUE-ResBlk.

First, we cropped the raw image at fixed locations (left, right, upper, and lower boundaries) in the image according to different ultrasound devices:

$$I_{flc} = FLC(I_{ori}), \tag{1}$$

where I_{ori} and I_{flc} are original images and fixed locations cropped images respectively. $FLC(\cdot)$ is the fixed locations cropping operation. Then, the image obtained from the first crop (Figure 1b) is traversed in the vertical direction in row units, where the number of points with nongray pixels is counted within each row. After that, the gradients of the data in all rows are calculated and the two largest gradients are selected as the upper and lower boundaries of the ROI region. Finally, the above operation is repeated in the horizontal direction to obtain the left and right boundaries of the ROI area.

$$I_{ROI} = CBC\left(I_{flc}, \underset{x}{\operatorname{argmax}} \nabla^1 M_{color}\right), \tag{2}$$

where I_{ROI} and I_{flc} are output ROI images and fixed locations cropped images respectively. $CBC(\cdot)$ is the color boundary cropping operation. The matrix M_{color} contains the number of color pixels. x is a boundary position of ROI. We obtained the final cropped data by the above algorithm (Figure 1c).

As shown in Figure 3, the AUE-net consists of two main components: a generator and a discriminator (omitted in Figure 3). The generator and discriminator have several submodules, including the nodule position attention module (AUE-SA), the AUE residual block (AUE-ResBlk), and the color attention module (AUE-CA), as shown in Figure 4. We will discuss them separately in this section.

(1) AUE-ResBlk: our AUE residual module was inspired by the work of Spatially Adaptive Denormalization (SPADE). However, unlike SPADE using semantic masks as inputs, we extended it to the underlying module where global features are extracted. We took the previously extracted feature maps as input and fused them with the newly extracted features (see Figure 4c):

$$OUT = \gamma_i(PF) \cdot \frac{pf_i - \mu_i}{\sigma_i} + \beta_i(PF), \tag{3}$$

where PF are previously extracted feature maps, pf_i denote the activations of the i th layer of a deep convolutional network. μ_i and σ_i are the mean and standard deviation of the pf_i activations. γ_i and β_i denote the conversion of PF to the scaling and bias

values. Through this approach, it is possible to make full use of the organizational feature information extracted by the upper layer network.

(2) Generator: our coarse-to-fine image generator was based on the architecture proposed by Johnson and others, which has proven to be successful for neural style migration for images up to 512×512 . We first analyzed the characteristics of the elastography image generation task. And combining with these, we added spatial attention (AUE-SA) module (see Figure 4b) to find the nodal positions at the front end, adding an AUE residual block before the up-sampling module for better feature fusing, and adding a channel attention (AUE-CA) module (see Figure 4a) for analyzing the color distribution at the end. However, our model demonstrated that the AUE-CA module will better generate elastography images by learning to reasonably assign the weights of the red, blue, and yellow channels at the final output, while the AUE-SA module can learn the tissue location features in the input ultrasound images to better extract features and generate elastography images of the target region (i.e., thyroid nodules). The generator can also be described by common GAN:

$$\min_G \max_D \mathcal{L}_{GAN}(G, D), \quad (4)$$

where the loss function $\mathcal{L}_{GAN}(G, D)$ is:

$$\mathbb{E}_{(s,x)}[\log D(s, x)] + \mathbb{E}_x[\log(1 - D(G(x), x))], \quad (5)$$

where G for generator, D for discriminator, x for inputs, and s for labels.

(3) Discriminator: a multiscale, patch-based discriminator with the InstanceNorm (IN) [28] was utilized from the pix2pixHD method (omitted in Figure 3). Furthermore, we applied the Spectral Normalization [29] to all the convolutional layers of the discriminator. And to enhance the nonlinear mapping capability, we used the Leaky-ReLU activation function instead of the ReLU activation function. At different scales, the discriminator can act as a feature extractor for the generator to optimize feature matching losses. These discriminators were trained to use authentic and synthetic images at different scales.

(4) Color-Loss: since elastography images use color to represent different hardness, generating a more natural elastography image means developing an image with a more realistic color distribution. To obtain a better color distribution of the elastography images, a loss function that measures the color difference between the generated image and the real image is needed. We need to eliminate the effect of texture and content in the image and measure only the differences in brightness, contrast, and primary colors in the image. Hence, we processed the image with Gaussian blur to ignore the small range of pixel differences by an additional convolution layer, and then we computed the distance between the obtained feature maps to express the color differences. This color loss between X and Y could be written as:

$$\mathcal{L}_{color}(X, Y) = \text{dis}(X_b, Y_b), \quad (6)$$

where $\text{dis}(\cdot)$ denotes the function for computing the distance. After comparing the actual experimental results by using different distance functions, the sum of Euclidean distance and L1 distance are used. X_b and Y_b denote the two images after Gaussian blur respectively, for example:

$$X_b(i, j) = \sum_{k,l} X(i+k, j+l) \cdot G(k, l), \quad (7)$$

where $G(k, l)$ denotes the Gaussian kernel with the size of $k \times l$. Our full GAN loss combined with the generator loss \mathcal{L}_{GAN} , feature matching loss \mathcal{L}_{FM} and color loss \mathcal{L}_{color} is as follows:

$$\min_G \left(\left(\max_{D_i \in D} \sum \mathcal{L}_{GAN}(G, D_i) + \beta \cdot \mathcal{L}_{color}(G, D_i) \right) + \alpha \cdot \sum \mathcal{L}_{FM}(G, D_i) \right), \quad (8)$$

where α controls the weight of the feature loss and β presents the importance of color loss. D_i is the result of i th discriminator.

In summary, AUE-net fully considers and leverages the characteristics of the input ultrasound image and the elastography image generation task. It is implemented and optimized by adding a nodal position attention module, a color channel attention module, an ultrasound feature fusion module, and a loss function optimization to optimize color distribution.

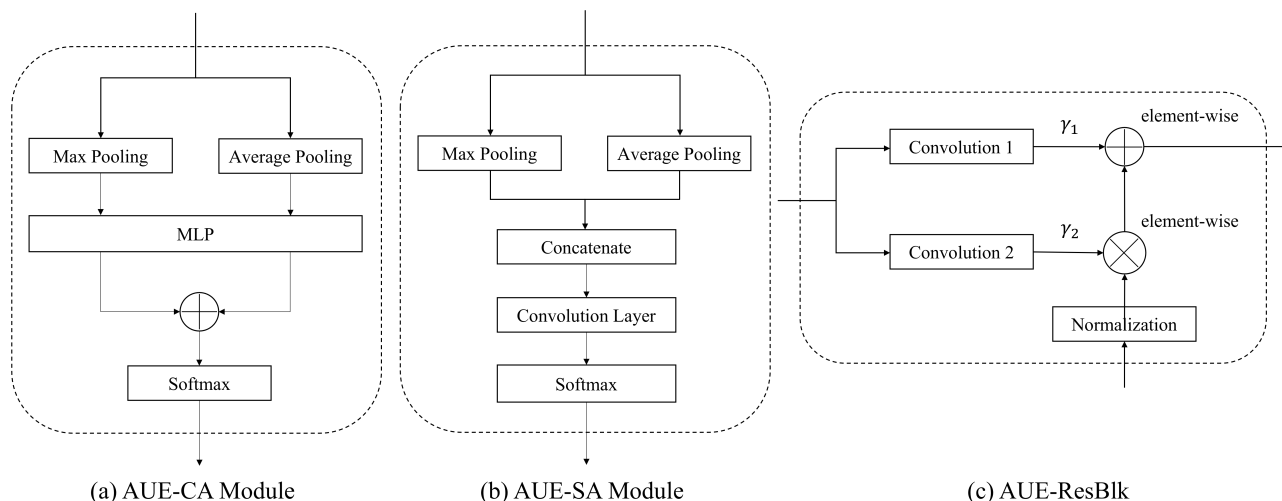


Figure 4. Submodules of our AUE-net. (a) is used to create a color attention map for elastography image color. (b) helps to locate nodule in elastography image. (c) helps making full use of feature information extracted before.

2.4. Implementation Details

All experiments are conducted in the following computer environment as shown in Table 1.

Table 1. Experimental environment information.

Items	Information
Operating System	Linux Red Hat 4.8.5-36
CPU	IBM(R) POWER9 (3.8 GHz)
GPU	NVIDIA Tesla V100 32 G

We trained a total of 1500 epochs using the batch size of 8. The number of down-sampling modules and up-sampling modules were both three. Each module had a convolutional layer (deconvolutional layer in the up-sampling module), a normalization layer and an activation layer. The residual module consist of nine blocks. These values were chosen after performing a preliminary grid-search-based optimization procedure. Due to the specificity of dynamic games in the GAN training, the quality of generated images could not be simply estimated by the loss function [30]. According to specialists, we selected 1500 epochs as the best result from the generation quality of multiple training epochs such as 500, 1000 and 1500 (in Figure 5). The perceptual loss [24] consists of feature maps of the *Relu1_1*, *Relu2_1*, *Relu3_1*, *Relu4_1*, *Relu5_1* layers of the pretrained VGG-19 model with weights [1/32, 1/16, 1/8, 1/4, 1]. Spectral normalization [29] was applied to the whole network. We utilized IN [28] as both generator and multiscale discriminator. The Adam [31] optimizer was used with two time-scale update rules with setting $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rates for the generator and discriminator were 0.0002 and 0.0001, respectively.

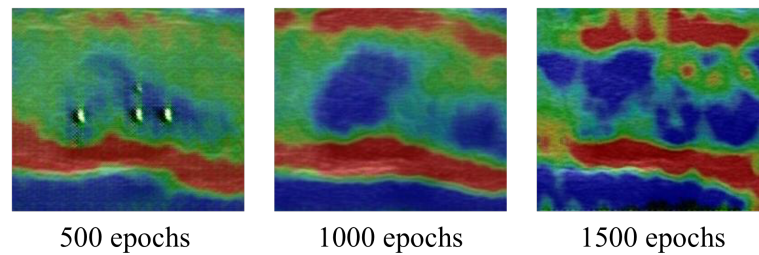


Figure 5. Generation results after 500, 1000, and 1500 epochs.

To enhance the robustness and generalization of the model, there is a certain probability (e.g., 0.5) that different data augmentation operations will be performed at input data, including ± 20 pixels of horizontal, vertical, or diagonal translation, horizontal and vertical mirror inversion, and $\pm 15^\circ$ of rotation.

Since the elastography image generated from each ultrasound image needs to be reproducible, we compared our method with paired I2IT methods using conditional GANs without a random noise map input, i.e., pix2pix [25], pix2pixHD [14], and LPTN [27] (in Table 2).

Table 2. Comparison among our method and state-of-the-art methods.

Methods	Year	Highlights
pix2pix [25]	2017	General-purpose solution to image-to-image translation based on cGAN
pix2pixHD [14]	2018	Synthesizing high-resolution photo-realistic images A novel adversarial loss, new multiscale generator, and discriminator architectures
LPTN [27]	2021	Laplacian pyramid decomposition and reconstruction Speeding-up the high-resolution photo-realistic I2IT tasks
AUE-net	2021	AUE attention modules, AUE residual blocks and color loss Generating strain elastography images from conventional ultrasound images

2.5. Evaluation Metrics

In terms of evaluation metrics, we used traditional image similarity metrics for evaluation, such as peak signal-to-noise ratio (PSNR) [32] and structural similarity index measure (SSIM) [33]:

$$\text{PSNR}(I, G) = \frac{10}{C} \sum \log_{10} \left(\frac{\text{MAX}_I^2}{\frac{1}{mn} \|I - G\|_2^2} \right), \quad (9)$$

$$\text{SSIM}(I, G) = \frac{(2\mu_I\mu_G + c_1)(2\sigma_{IG} + c_2)}{(\mu_I^2 + \mu_G^2 + c_1)(\sigma_I^2 + \sigma_G^2 + c_2)}, \quad (10)$$

where I and G are two images with $m \times n$, C is the number of channels. MAX_I^2 is the maximum pixel value of the image which is 255 here. $\|\cdot\|_2$ stands for the Euclidean norm. μ is the mean. σ^2 is the variance. σ_{IG} the covariance of I and G . c_1 and c_2 are two variables to stabilize the division with weak denominator. The paper also uses the Frechet Inception Distance (FID), first proposed by Heusel and others [34]:

$$\text{FID}(I, G) = \|\mu_I - \mu_G\|_2^2 + \text{tr}(C_I + C_G - 2\sqrt{C_I C_G}), \quad (11)$$

where μ is the mean, tr is the trace, and C is the covariance matrix. The FID utilizes a pretrained Inception network on the ImageNet dataset to evaluate the quality of the GAN-generated images. The generated and real images are first fed into the pretrained Inception network. Then, the activation means of the final layer and the covariance of the assumed Gaussian distribution are extracted. Finally, the Frechet distance between them is calculated. The FID is computed over the learned feature space and was shown to correlate well with human visual perception [35].

Finally, to measure the clinical value of our method, we asked thyroid specialists to provide scores. First, two specialists scored randomly disrupted real and generated images according to the Ueno–Rago criterion [36,37]. Then two scores of the same image were matched, with the same being 1 and different being 0. The scoring accuracy of the entire test set was tallied. This index is consistent with the clinical prediction process for thyroid nodules and can be applied for clinical use.

3. Results

3.1. Visual Comparisons

We compared the visual performance of the elastography image generated by our model with that of three networks (pix2pix, pix2pixHD, and LPTN). As shown in Figure 6, the AUE-net proposed in this paper outperformed all other three methods in terms of imaging realism and quality.

Specifically, AUE-net translated the input strained ultrasound image into the corresponding elastography image with little texture distortion (Figure 6). The results showed a more realistic and natural color distribution, and the masked part of the original image without ultrasound information was partially removed. Among these methods, pix2pixHD was the second-best network. However, it introduced more hardened features to the nodule locations, i.e., the blue area of the nodule, and over-representation of hardened features, due to insufficient reconstruction capability of the decoder for the nodule. In contrast, AUE-net enhanced the hardened information generation in the nodule region during the encoding-decoding process by the spatial attention module and the channel attention module, producing results closer to the actual image.

In general, most of the existing I2IT methods are based on the autoencoder generation framework (LPTN is based on the Laplace pyramid) with three main modules: (1) decomposing content information and attributes on the low-dimensional latent space by the encoding process; (2) transforming the attributes within the latent space by the residual module; (3) reconstructing the image from the transformed attributes by the decoding process. The ability to reconstruct elastography images was simulated by the network parameters of the encoder-decoder, but the characteristics of elastography images were not taken into account in these methods, which led to mediocre results. Our proposed AUE-net was an encoder-decoder network that took into account the characteristics of nodule position correlation and color distribution correlation of elastography images. Better results could be generated through the spatial and channel attention modules. The elastography images generated from the perspective of image algorithms were more nodule location correlated and could assist specialists with nodule localization. The image algorithm helped to reduce human efforts and decreased the need for equipment.

3.2. Quantitative Results

This subsection compared the AUE-net with the I2IT methods described above using three image similarity metrics (PSNR, SSIM and FID) and scores from the specialists. The results of our experiments in terms of image quality were shown in Table 3.

Table 3. PSNR, SSIM, FID, and scoring accuracy of our method compared to that of state-of-the-art methods.

Methods	PSNR ↑	SSIM ↑	FID ↓	Score Acc ↑
pix2pix [25]	12.688	0.281	205.38	30.8%
pix2pixHD [14]	28.651	0.456	56.71	75.4%
LPTN [27]	10.879	0.364	155.24	32.3%
Ours	28.736	0.499	51.09	84.4%

↑ and ↓ denote higher is better and lower is better, respectively.

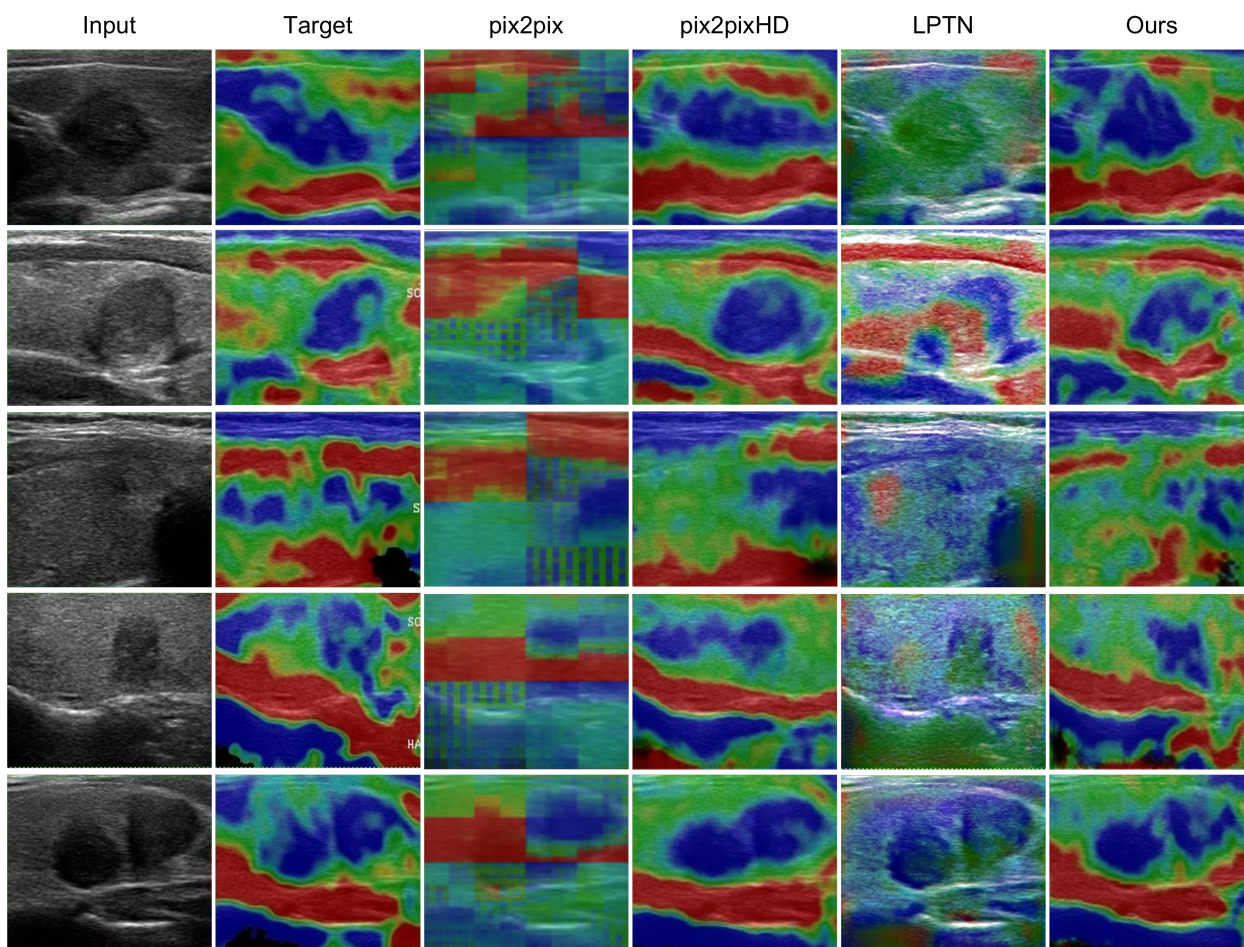


Figure 6. Generation results compared to pix2pix, pix2pixHD and LPTN. First two columns show input ultrasound data and target elastography image.

As shown in Tables 3 and 4, the results generated by our model were the best among the four models. The mean accuracy of scores given by two specialists was 84.38%. The pix2pixHD achieved an accuracy with 75.4%. These results demonstrated that the AUE-net proposed in this study showed better performance for generating elastography image than the general I2IT methods.

Table 4. Accuracy of scores from two specialists.

	Correct Amount of 80 ↑	Score Accuracy ↑
Specialist 1	67	83.75%
Specialist 2	68	85.00%
Mean	67.5	84.38%

↑ and ↓ denote higher is better and lower is better, respectively.

Based on specialists’ opinions, the elastography images generated by our model could meet the needs of clinical diagnostic applications and provide practical value. We reviewed the cases that showed errors from the specialists. Specifically, the scoring accuracy was 66.67% for grade I, 78.57% for grade II, 82.61% for grade III, 90.48% for grade IV, and 100.00% for grade V. The lower accuracy of grade I was probably because of its small number. Most of the errors were between grade II and grade III. These errors were due to visual observation. In the Ueno–Rago criterion [36,37], grade II is with green predominating, while grade III is with blue dominating. It is therefore quite a challenge to determine accurately when the lesion area is mixed with blue and green.

There were also a few other error cases that seem to result from unknown circumstances. For example, in Figure 7, the true elastography image showed that most tissue was hardened. There could be two possibilities: either the device was biased which resulted in incorrect calculation or most of the targeted tissue became hardened. Special cases like this should be excluded when assessing the validity and performance of our model.

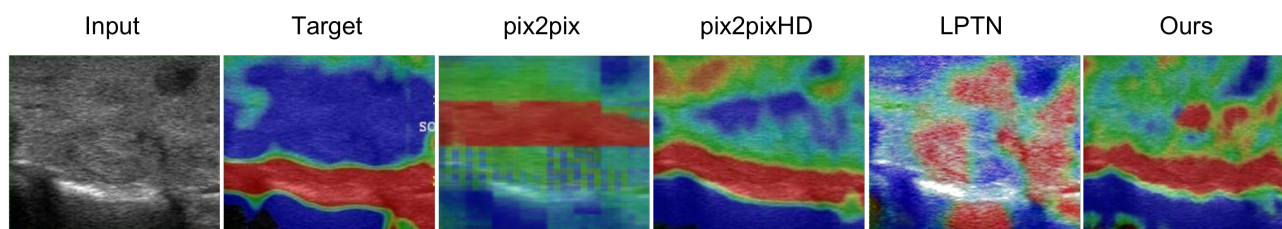


Figure 7. Special example in test images shows that most tissue is hardened, while generated result is not good enough.

Although this paper showed promising results in terms of image quality and scoring accuracy, our current study was limited in several aspects. First, due to the inherent difference between image algorithms and physical methods, the proposed AUE-net could not generate the same strain elastography images as the actual images. Our method would not be used as a substitute for the strain elastography technique, but was considered as a prefractionated reference for tissue. Second, our method had not yet investigated generalization among different devices due to the significant variation of elastography images produced.

4. Conclusions

With strain elastography becoming a promising new tool that can effectively improve diagnostic accuracy for thyroid nodules, we built and proposed in this paper a new method for generating elastography images based on the characteristics of thyroid elastography images. The AUE-net method used a base pix2pix architecture with a fused residual module and combined with a spatial attention module, a channel attention module, and a color loss feature. Results from quantitative and qualitative evaluation showed that the elastography images generated by presented model produced better quality compared to that of other reported models. Accordingly, conventional ultrasound image harbors information that can be linked to tissue elasticity. However, our work has some limitations. Since the current study used postpressure ultrasound images to generate the corresponding elastography images, the experience of specialists was still required. For our next step, we will continue to study the generation of elastography images by using direct prediction of plain ultrasound images to eliminate the influence of human factors.

Author Contributions: Conceptualization, X.D., C.S., Q.Z.; methodology, X.D., C.S., Q.Z., R.W.; software, Q.Z., X.L.; formal analysis, X.D., C.S., Q.Z., J.Z., X.L.; resources, R.W., Q.L.; data curation, R.W., Q.L.; writing—original draft preparation, Q.Z.; writing—review and editing, X.D., C.S., J.Z., Q.Z., R.W., Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by Shanghai Sixth People’s Hospital scientific research project (NO.YNLC201903), the Shanghai Municipal Health Commission scientific research project (NO.20204Y0256) and the Interdisciplinary Program of Shanghai Jiao Tong University (No. YG2019ZDA09).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Shanghai Jiao Tong University Affiliated Sixth People’s Hospital (Approval code: 2021-KY-115(K); Approval date: 7 December 2021).

Informed Consent Statement: Informed consent was waived due to the retrospective nature of this study.

Acknowledgments: We would like to thank Helen Li and Wang Yang for English revisions. We appreciate the High Performance Computing Center of Shanghai University and Shanghai Engineering Research Center of Intelligent Computing System (No. 19DZ2252600) for providing the computing resources and technical support.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

STIR	Short Tau Inversion Recovery
GAN	Generative adversarial network
ROI	regions of interest
PSNR	Peak signal-to-noise ratio
FNAC	Fine-needle aspiration cytology
SE	Strain elastography
I2IT	Image-to-image translation
SSIM	Structural similarity index measure
MSE	Mean-square error
FID	Frechet Inception distance

References

1. Wilhelm, S. Evaluation of thyroid incidentaloma. *Surg. Clin. N. Am.* **2014**, *94*, 485–497. [[CrossRef](#)] [[PubMed](#)]
2. Alexander, L.F.; Patel, N.J.; Caserta, M.P.; Robbin, M.L. Thyroid ultrasound: Diffuse and nodular disease. *Radiol. Clin.* **2020**, *58*, 1041–1057. [[CrossRef](#)] [[PubMed](#)]
3. Barr, R.G. Shear wave liver elastography. *Abdom. Radiol.* **2018**, *43*, 800–807. [[CrossRef](#)] [[PubMed](#)]
4. Barr, R.G.; Ferraioli, G.; Palmeri, M.L.; Goodman, Z.D.; Garcia-Tsao, G.; Rubin, J.; Garra, B.; Myers, R.P.; Wilson, S.R.; Rubens, D. Elastography assessment of liver fibrosis: Society of radiologists in ultrasound consensus conference statement. *Radiology* **2015**, *276*, 845–861. [[CrossRef](#)]
5. Correias, J.M.; Tissier, A.M.; Khairoune, A.; Khoury, G.; Eiss, D.; Hélénon, O. Ultrasound elastography of the prostate: State of the art. *Diagn. Interv. Imaging* **2013**, *94*, 551–560. [[CrossRef](#)]
6. Onur, M.R.; Poyraz, A.K.; Bozgeyik, Z.; Onur, A.R.; Orhan, I. Utility of semiquantitative strain elastography for differentiation between benign and malignant solid renal masses. *J. Ultrasound Med.* **2015**, *34*, 639–647. [[CrossRef](#)]
7. Zhao, C.K.; Xu, H.X. Ultrasound elastography of the thyroid: Principles and current status. *Ultrasonography* **2019**, *38*, 106. [[CrossRef](#)]
8. Sigrist, R.M.; Liau, J.; El Kaffas, A.; Chammas, M.C.; Willmann, J.K. Ultrasound elastography: Review of techniques and clinical applications. *Theranostics* **2017**, *7*, 1303. [[CrossRef](#)]
9. Soni, M.; Gomathi, S.; Kumar, P.; Churi, P.P.; Mohammed, M.A.; Salman, A.O. Hybridizing Convolutional Neural Network for Classification of Lung Diseases. *Int. J. Swarm Intell. Res. (IJSIR)* **2022**, *13*, 1–15. [[CrossRef](#)]
10. Mohammed, M.A.; Al-Khateeb, B.; Rashid, A.N.; Ibrahim, D.A.; Abd Ghani, M.K.; Mostafa, S.A. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Comput. Electr. Eng.* **2018**, *70*, 871–882. [[CrossRef](#)]
11. Hussein, I.J.; Burhanuddin, M.A.; Mohammed, M.A.; Benameur, N.; Maashi, M.S.; Maashi, M.S. Fully-automatic identification of gynaecological abnormality using a new adaptive frequency filter and histogram of oriented gradients (HOG). *Expert Syst.* **2021**, e12789. doi: 10.1111/exsy.12789 [[CrossRef](#)]
12. Hussein, I.J.; Burhanuddin, M.; Mohammed, M.A.; Elhoseny, M.; Garcia-Zapirain, B.; Maashi, M.S.; Maashi, M.S. Fully Automatic Segmentation of Gynaecological Abnormality Using a New Viola–Jones Model. *Comput. Mater. Contin.* **2021**, *66*, 3161–3182. [[CrossRef](#)]
13. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*. [[CrossRef](#)]
14. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8798–8807.
15. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4401–4410.
16. Haubold, J.; Demircioglu, A.; Theysohn, J.M.; Wetter, A.; Radbruch, A.; Dörner, N.; Schlosser, T.W.; Deuschl, C.; Li, Y.; Nassenstein, K.; et al. Generating Virtual Short Tau Inversion Recovery (STIR) Images from T1-and T2-Weighted Images Using a Conditional Generative Adversarial Network in Spine Imaging. *Diagnostics* **2021**, *11*, 1542. [[CrossRef](#)]

17. Gomi, T.; Sakai, R.; Hara, H.; Watanabe, Y.; Mizukami, S. Usefulness of a Metal Artifact Reduction Algorithm in Digital Tomosynthesis Using a Combination of Hybrid Generative Adversarial Networks. *Diagnostics* **2021**, *11*, 1629. [[CrossRef](#)]
18. Bermudez, C.; Plassard, A.J.; Davis, L.T.; Newton, A.T.; Resnick, S.M.; Landman, B.A. Learning implicit brain MRI manifolds with deep learning. In Proceedings of the SPIE—Medical Imaging 2018: Image Processing, Houston, TX, USA, 2 March 2018; Volume 10574, p. 105741L.
19. Nishiyama, D.; Iwasaki, H.; Taniguchi, T.; Fukui, D.; Yamanaka, M.; Harada, T.; Yamada, H. Deep generative models for automated muscle segmentation in computed tomography scanning. *PLoS ONE* **2021**, *16*, e0257371. [[CrossRef](#)]
20. Wu, S.; Gao, Z.; Liu, Z.; Luo, J.; Zhang, H.; Li, S. Direct reconstruction of ultrasound elastography using an end-to-end deep neural network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 374–382.
21. Kibria, M.G.; Rivaz, H. Gluenet: Ultrasound elastography using convolutional neural network. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 21–28.
22. Wildeboer, R.R.; van Sloun, R.J.; Mannaerts, C.K.; Moraes, P.; Salomon, G.; Chammas, M.; Wijkstra, H.; Mischi, M. Synthetic Elastography Using B-Mode Ultrasound through a Deep Fully Convolutional Neural Network. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2020**, *67*, 2640–2648. [[CrossRef](#)]
23. Shin, H.C.; Tenenholtz, N.A.; Rogers, J.K.; Schwarz, C.G.; Senjem, M.L.; Gunter, J.L.; Andriole, K.P.; Michalski, M. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *International Workshop on Simulation and Synthesis in Medical Imaging*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–11.
24. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.
25. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
26. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 2337–2346.
27. Liang, J.; Zeng, H.; Zhang, L. High-Resolution Photorealistic Image Translation in Real-Time: A Laplacian Pyramid Translation Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 9392–9400.
28. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* **2016**, arXiv:1607.08022.
29. Wang, S.Y.; Bau, D.; Zhu, J.Y. Sketch Your Own GAN. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 19–25 June 2021; pp. 14050–14060.
30. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 214–223.
31. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
32. Regmi, K.; Borji, A. Cross-view image synthesis using conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3501–3510.
33. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier gans. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 2642–2651.
34. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [[CrossRef](#)]
35. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 586–595.
36. Rago, T.; Vitti, P. Potential value of elastosonography in the diagnosis of malignancy in thyroid nodules. *Q. J. Nucl. Med. Mol. Imaging* **2009**, *53*, 455. [[PubMed](#)]
37. Ueno, E.; Ito, A. Diagnosis of breast cancer by elasticity imaging. *Eizo Joho Med.* **2004**, *36*, 2–6.