Ultrasound Image Super-Resolution with Two-Stage Zero-Shot CycleGAN

Jianrui Ding1*, Shili Zhao1, Fenghe Tang1 and Chunping Ning2

1School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, 150006, China
2Ultrasound Department, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, 266003, China
*Corresponding author’s e-mail: jrding@hit.edu.cn

Abstract. Medical ultrasound imaging is widely used in clinical diagnosis because of its non-invasive, convenient and quick characteristics. However, due to its low image contrast, multiple artifacts, noise and lack of paired high-resolution and low-resolution image data sets, the task of super-resolution reconstruction of medical ultrasound images is more challenging. In this paper, the Two-Stage GAN network model was adjusted by CycleGAN generation and unsupervised learning methods, and the Two-Stage ZSSR (“Zero-Shot” Super-Resolution) CycleGAN network was proposed. The objective evaluation indexes PSNR and SSIM were raised to 40.8079 and 0.9953. The visual effect was also significantly improved.

1. Introduction
Medical ultrasound imaging technology is a non-invasive, non-radiation and non-invasive detection method, which is widely used in the detection of thyroid, fetal, mammary gland and gonadal tissue [1, 2]. However, due to its low contrast, serious noise and multiple artifacts, it requires high clinical experience of doctors. Clear medical images can provide more abundant lesion information, relieve the diagnosis pressure of doctors, and improve the diagnostic efficiency and accuracy. Therefore, improving image quality is of great significance for subsequent clinical diagnosis.

At present, methods to improve the quality of ultrasonic images can be divided into two categories [3]: pre-processing technology and post-processing technology. Pre-processing technology is related to the physical properties of the signals involved, and post-processing technology is to use image processing and machine learning methods to enhance images after obtaining low-resolution images [4]. At present, deep learning is widely used in super-resolution reconstruction of ultrasound images, such as MMCNN, UNET, RESNET, SRGAN, two-stage GAN and ZSSR [5]. But due to the characteristics of ultrasound images and lack of ground truth of high-resolution images, the direct application of existing deep learning models will lead to unsatisfactory reconstruction results.

To solve the above problem, this paper takes two-stage GAN [6] network model as the baseline, proposes a two-stage ZSSR CycleGAN network model by introducing the ZSSR [7] and CycleGAN [8]. In addition, the front-end cascading U-Net network was optimized by introducing residual substructure and attention mechanism.
2. Materials and methods

2.1. Dataset and Data processing

The datasets used in this study include the private thyroid ultrasound data set collected by the Affiliated Hospital of Qingdao University, the public data set CCA-US and US-CASE. The private dataset contains a total of 135 cases and 1023 images with a resolution of approximately 710×573; The CCA-US dataset consisted of 84 carotid ultrasound images taken by medical experts with more than five years of clinical experience, with a resolution of approximately 390×330. US-CASE is a free ultrasound image database with image resolution of 300×255.

The sample images are shown in Fig. 1. The dataset was augmented by rotated at the following seven angles (45°, 90°, 135°, 180°, 225°, 270°, 315°), horizontal and vertical mirror symmetry.

![Figure 1. Sample images](image)

(a) Thyroid image  (b) CCA-US image  (c) US-CASE image

2.2. Proposed Methods

The framework proposed in this paper is shown in Fig. 2.

![Figure 2. Proposed Methods](image)

The framework utilizes multi-scale internal correlation of information contained in a single image without the need for extensive external data training.

The two-stage GAN model introduces the U-Net [9] network with strong low-frequency information reconstruction ability which is meaningful for clinical diagnosis. But it is difficult to obtain high quality HR medical ultrasound images due to the low contrast and various noises. Therefore, medical ultrasound image super-resolution reconstruction is easy to appear the blur phenomenon.

In order to achieve better performance, the original two-stage GAN network was optimized and tested on three datasets. Firstly, the optimized U-Net model is used to pre-process the input image to get the image which contains more low-frequency information. Then, the image obtained by data
augmentation is regarded as the HR set. The LR set is down sampled with different scales from HR set. Finally, HR and LR were input into the improved CycleGAN network for training.

2.2.1 Unsupervised ZSSR model and CycleGAN embedding

The ZSSR model does not need to provide the HR image in advance, through the information repeatability within a single image [10], and trains a dedicated small convolutional neural network for each image during the test. It can solve the problem of medical ultrasound image without high resolution image.

GAN network uses adversarial loss to make the generated image infinitely close to the real scene image. However, for ultrasound images lacking ground-truth, only learning the process from LR to HR cannot reconstruct images well. CycleGAN [11] proposed cyclic consistency loss, which can reliably generate images and solve the problem of unpaired data sets. Previous studies have shown that GAN-based super-resolution reconstruction of medical images may generate unrealistic artificial details, thus affecting clinical diagnosis. In order to suppress the addition of artificial details, the basic model of CycleGAN network is improved in this paper.

The improved CycleGAN network model is shown in Fig. 3. Fig. 3(a) is the HR cycle consistency loss GAN network model, and Fig. 3(b) is the LR cycle consistency loss GAN network model. The CycleGAN model proposed in this paper consists of two sets of GAN using two generators and one discriminator each. The loss is not only considered for HR-LR and LR-HR, but also makes full use of the cycle loss of LR-HR-LR and HR-LR-HR as well as the confrontational characteristics of the discriminator to promote the generator to produce better perception-consistent super-resolution reconstruction results.

2.2.2 Multi-scale generator

In order to further improve the performance of SR, a multi-scale generator is introduced. The structure of multi-scale HR generator is shown in Fig. 4. At the same time, Gaussian noise is added to the LR generator to further simulate the image degradation process. The structure is shown in Fig. 5.

2.2.3 Loss function

Although the original antagonistic loss can force the distribution of the generated SR image to approximate that of the target HR image, there is no guarantee that the learned mapping can map a single input X to the natural scene image Y for a dataset lacking HR image. Cycle loss were introduced in this paper to ensure perceptual consistency and achieve accurate reconstruction of ultrasonic images. The cyclic consistency loss can be defined as Formula 1.
where \( G(x) \) is a mapping function from LR to HR and \( F(x) \) is a mapping function from HR to LR. LR and HR image pairs are defined as \( \{x_i, y_i\} \).

### 3. Experiment and results

#### 3.1. Performance measurement

In the experiment, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were used as quantitative evaluation indexes to analyse the contribution of our different networks, as shown in Equations (2) - (7).

\[
L_{cyc} = \frac{1}{N} \left( \sum_{i=1}^{N} \left( \| F \left( G(x_i) \right) - x_i \|_1 + \| G \left( F(y_i) \right) - y_i \|_1 \right) \right)
\]

(1)

where \( G(x) \) is a mapping function from LR to HR and \( F(x) \) is a mapping function from HR to LR. LR and HR image pairs are defined as \( \{x_i, y_i\} \).

#### 3.2. Experimental results

The performance of two-stage ZSSR CycleGAN is shown in Table 1.

| Method               | Private Dataset (PSNR/SSIM) | CCA-US (PSNR/SSIM) | US-CASE (PSNR/SSIM) |
|----------------------|-----------------------------|-------------------|---------------------|
| SRCNN                | 35.8125/0.9484              | 36.8753/0.8508    | 22.1508/0.8270      |
| Two-Stage GAN       | 35.9014/0.9500              | 36.9502/0.8510    | 22.0324/0.8279      |
| ZSSR                 | 38.5622/0.9921              | 36.8775/0.9661    | 31.1691/0.9661      |
| Two-Stage ZSSR CycleGAN | 40.6834/0.9947              | 40.7600/0.9918    | 37.1074/0.9842      |

The performance evaluation indexes PSNR and SSIM have been improved by about 4.87dB and 0.15, respectively. When the two-stage GAN is trained without ground-truth, PSNR and SSIM improved by about 4.78dB and 0.14 dB, respectively. It has been proved that two-stage ZSSR CycleGAN network can obtain medical ultrasound images with good reconstruction quality.

Some of the results are shown in Fig. 6.
It is found that in the training network without U-Net cascade, the network model may degenerate into a more general two-stage GAN on some images, and the effect is almost the same as that of ZSSR, and the reconstructed image is relatively fuzzy. Therefore, we further improve the cascaded U-Net and expect to further improve the performance.

Table 2 shows the experimental results of the improved U-Net strategy based on residual and attention on private thyroid dataset.

| Model                        | PSNR  | SSIM  |
|------------------------------|-------|-------|
| U-Net                        | 40.6834 | 0.9947 |
| With residual block         | 40.7243 | 0.9952 |
| With residual and attention | **40.8079** | **0.9953** |

It is proved that U-Net had a strong performance ability in the reconstruction of the tissue structure information of ultrasonic images.

Some results are shown in Fig. 7.
It can be seen that Fig. 7(c) has obvious organizational structure and clearer edges, which intuitively shows that the improved U-Net in this paper can better reconstruct low-frequency information from the visual effect.

4. Discussions and conclusions
The characteristics of low contrast and high noise of ultrasound images pose challenges for radiologists' clinical diagnosis. Improving the resolution and contrast of ultrasound images is not only can help radiologist identify lesions clearly, but also provides guarantee for computer automatically detection and identification lesions. In this paper, a two-stage ZSSR CycleGAN model with improved U-Net was proposed to ultrasound image super-resolution. It is designed to solve the problem of no ground-truth HR dataset and good results have been achieved through experiment verification with private thyroid dataset and public dataset.

Further research can be carried out in the following aspects in the future: 1) Modify the model structure and parameters by introducing radiologists' prior knowledge. 2) To study the influence of different quality images collected by different ultrasonic equipment on the super resolution reconstruction technology, and further improve the generalization performance of the super resolution reconstruction model; 3) Focus on the reconstruction images with poor visual effect, find their correlation, and further optimize the model.

Acknowledgments
This work is supported, in part, by Shandong Natural Science Foundation of China; the Grant numbers is ZR2020MH290.

References
[1] Elangovan A, Jeyaseelan T. (2016). Medical imaging modalities: A survey, 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS). 1-4.
[2] Huber M, Long J, Flint K, et al. (2020). Implementation of adaptive transmit parameter adjustment in ultrasound imaging, 2020 IEEE International Ultrasonics Symposium (IUS). IEEE, 1-4.
[3] Ortiz S H C, Chiu T, Fox M D. (2012). Ultrasound image enhancement: A review. Biomedical Signal Processing and Control, 7(5): 419-428.
[4] Rundo L, Tanguerloni A, Nobile M S, et al. (2019). MedGA: a novel evolutionary method for image enhancement in medical imaging systems. Expert Systems with Applications, 119: 387-399.
[5] Wang, Z., Chen, J., & Hoi, S. (2020). Deep learning for image super-resolution: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP (99), 1-1.
[6] Zhou, Z., Wang, Y., Guo, Y., Qi, Y., & J Yu. (2019). Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network[J]. IEEE Transactions on Biomedical Engineering, 67(1): 298-311.
[7] Shocher, A., Cohen, N., & Irani, M. (2018). Zero-Shot Super-Resolution Using Deep Internal Learning. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 3118-3126.
[8] Bulat, A., Jing, Y., & Tzimiropoulos, G. (2018). To learn image super-resolution, use a gan to learn how to do image degradation first. European Conference on Computer Vision. 185-200.
[9] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 234-241.
[10] Lu, J., & Liu, W. (2018). Unsupervised Super-Resolution Framework for Medical Ultrasound Images Using Dilated Convolutional Neural Networks. 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE.
[11] Yuan, Y., Liu, S., J Zhang, Zhang, Y., Dong, C., & Lin, L.. (2018). Unsupervised Image Super-Resolution using Cycle-in-Cycle Generative Adversarial Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE. 701-710.

[12] Jie, H., Li, S., Gang, S., & Albanie, S.. (2017). Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP (99).