



Exploring Multi-scale Deep Encoder-Decoder and PatchGAN for Perceptual Ultrasound Image Super-Resolution

Jianyong Liu¹, Heng Liu^{1(✉)}, Xiaoyu Zheng¹, and Jungong Han²

¹ Anhui University of Technology, Maanshan, Anhui, China
jianyong-liu@qq.com, hengliusky@aliyun.com, zhengxiaoyu51888@163.com

² The University of Warwick, Coventry CV4 7AL, UK
jungonghan77@gmail.com

Abstract. Ultrasound visual imaging is currently one of three mainstream image diagnosis technologies in the medical industry, but due to the limitations of sensors, transmission media and ultrasound characteristics, the quality of ultrasound imaging may be poor, especially its low spatial resolution. We incorporate, in this paper, a new multi-scale deep encoder-decoder structure into a PatchGAN (patch generative adversarial network) based framework for fast perceptual ultrasound image super-resolution (SR). Specifically, the entire algorithm is carried out in two stages: ultrasound SR image generation and image refinement. In the first stage, a multi-scale deep encoder-decoder generator is employed to accurately super-resolve the LR ultrasound images. In the second stage, we advocate the confrontational characteristics of the discriminator to impel the generator such that more realistic high-resolution (HR) ultrasound images can be produced. The assessments in terms of PSNR/IFC/SSIM, inference efficiency and visual effects demonstrate its effectiveness and superiority, when compared to the most state-of-the-art methods.

Keywords: Ultrasound image super-resolution · Deep encoder-decoder · Multi-scale · PatchGAN

1 Introduction

Ultrasonograph [6] is an effective visual diagnosis technology in medical imaging industry and has unique advantages over modalities, such as magnetic resonance imaging (MRI), X-ray, and computed tomography (CT). In the actual diagnosis through ultrasound imaging, doctors usually judge pathological changes by visually perceiving the region of interest (ROI) in the ultrasound image, such as the shape contour or edge smoothness. This means that the higher the resolution of ultrasound image, the more conducive to visual perception for better medical diagnosis. However, due to the acoustic diffraction limit of the medical industry,

it is difficult to acquire HR ultrasound data. Therefore, for enhancing the resolution of ultrasound data, image SR may become a potential solution, which is of great significance for medical clinical diagnosis based on visual perception [7, 16].

Over the past few years, the methods based on deep learning have emerged in various fields of natural image processing, ranging from image de-noising [2] to image SR [5, 10], and video segmentation [21]. Recently, they have also been applied into diverse medical image processing tasks, ranging from CT image segmentation [24] to ultrasound image SR [3, 14]. Umehara *et al.* [26] firstly apply SR convolutional neural network (SRCNN [5]) to enhance the image resolution of chest CT images. A recent work [9] demonstrates that the deeper and wider network can result in preferable image SR results and seems to have no problem of poor generalization performance. Actually, this experience may not always applicable to SR of medical data (including ultrasound images), because there may not be abundant medical image samples for training in practice. Thus, how to design an appropriate deep network structure becomes one key to improve the efficiency of medical image SR.

Other deep supervised models, especially ‘U-net’ convolutional networks [20, 25], are proposed and explored recently for bio-medical image segmentation as well as ultrasound image SR. With no fully connected layers, U-net consists of only convolution and deconvolution operations, where the former is named as encoder and the latter is called decoder. However, in such a U-net, the pooling layers and the single-scale convolutional layers may fail to take advantage of various image details and multi-range context of SR. Note that the terms of encoder and decoder in the following sections mean the convolution operation and deconvolution operations respectively.

At the same time, Ledig *et al.* [10] presented a new deep structure, namely SR generative adversarial network (SRGAN), producing photo-realistic SR images. Instead of using CNNs, Choi *et al.* [3] have applied the SRGAN model for high-speed ultrasound SR imaging. Such GANs based works claimed to obtain better image reconstruction effect with good visual quality. Unfortunately, Yochai *et al.* in their recent work [1] analyze that the visual perception quality and the distortion decreasing of an image restoration algorithm are contradictory with each other.

Actually, almost all the above-mentioned deep methods are from the perspective of single-scale spatial domain reconstruction, without a consideration of multi-scale or even frequency domain behavioral analysis of network feature learning. Motivated by the frequency analysis of neural network learning [19, 28] and the structure simulation of multi-resolution wavelet analysis [13], in this work, we present a novel deep multi-scale encoder-decoder based approach to super-resolve the LR ultrasound images. Moreover, inspired by the analysis of [1], our model integrates the PatchGAN [8] way to better tradeoff the reconstruction accuracy and the visual similarity to real ultrasound data. We perform extensive experiments on different ultrasonic data sets and the results demon-

strate our approach not only achieves high objective quality evaluation value, but also holds rather good subjective visual effect.

As far as we know, the methods that deal with the resolution enhancement of a single ultrasound image are few, let alone a thorough exploration of multi-scale and adversarial learning to achieve accurate reconstruction with perception trade-off. The contributions of the work are generalized as follows:

- By simulating the structure of multi-resolution wavelet analysis, we propose a new end-to-end deep multi-scale encoder-decoder framework that can generate a HR ultrasound image, given a LR input for $4\times$ up-scaling.
- We integrate the PatchGAN adversarial learning with the VGG feature loss and the ℓ_1 pixel-wise loss to jointly supervise the image SR process at different levels during training. The experimental results turn out that the integrated loss is good at recovering multiple levels of details of ultrasound images.
- We evaluate the proposed approach on several public ultrasound datasets. We also compare the variants of our model and analyze the performance and the differences to others, which might be useful for future ultrasound image SR research.

The rest of the paper is organized as follows. Related works are outlined in Sect. 2. Section 3 describes our proposed approach and its feasibility analysis. Lots of experimental results and analysis are shown in Sect. 4. Finally, the conclusion of this paper is summarized in Sect. 5.

2 Related Work

2.1 Natural Image SR

Given the powerful non-linear mapping, CNN based image SR methods can acquire better performance than the traditional methods. The pioneering SRCNN, proposed by Dong *et al.* [5], only utilized three convolutional layers to learn the mapping function between the LR images and the corresponding HR ones from numerous LR-HR pairs. However, the fact that SRCNN has only three convolutional layers indicates the model actually is not good at capturing image features. Considering that the prior knowledge may be helpful for the convergence, Liang *et al.* [11] used Sobel edge operator to extract gradient features to promote the training convergence. In spite of speeding up the training procedure, the improvement for reconstruction performance is very limited. Recently, based upon the structure simulation of multi-resolution wavelet analysis, Liu *et al.* [13] proposed a multi-scale deep network with phase congruency edge map guidance model (MSDEPC) to super-resolve single LR images.

Aiming to improve the reconstruction quality, Ledig *et al.* [10] applied such adversarial learning strategy to form a novel image SR model - SRGAN, of which the generator network is used to super-resolved the LR input efficiently while the discriminator network determines whether the super-resolved images approximate the real HR ones.

Recognizing that batch normalization may make the extracted features lose diversity and flexibility, Lim *et al.* [12] proposed an enhanced deep residual SR model (named as EDSR) by removing the batch normalization operation. In order to hold the flexibility of the path, they also adjusted the residual structure so that the sum of different paths will no longer pass through the ReLU layer.

Different from SRGAN [10] that only distinguishing the generated image itself, Park *et al.* [18] presented to add additional discrimination network which acts on the feature domain (called as SRFeat), so that the generator can generate high-frequency features related to the image structure. Moreover, long range jump connections were used in the generator to make information flow more easily in layers far away from each other.

2.2 Ultrasound Image SR

Compared with the flourishing situation of natural images, little attention has been paid to SR of medical images overall. Recently, Zhao *et al.* [29] explored the properties of the decimation matrix in the Fourier domain and managed to acquire an analytical solution with ℓ_2 norm regularizer for the problem of ultrasound image SR. Unlike many studies focusing on ultrasound lateral resolution enhancement, Diamantis *et al.* [4] pay their attention to axial imaging. Being conscious of the accuracy of ultrasound axial imaging mainly depends on image-based localization of single scatter, they use a sharpness based localization approach to identify the unique position of the scatter, and successfully translate SR axial imaging from optical microscopy into ultrasound imaging.

Having seen the power of deep learning in natural image processing, Umehara *et al.* [26] firstly applied the SRCNN to enhance the resolution of some chest CT images and the results demonstrated the CNN based SR model is also suitable for medical images. Moreover, in order to ease the problem of lacking numerous training samples in common medical datasets, Lu *et al.* [14] utilized dilated CNNs and presented a new unsupervised SR framework for medical ultrasound images. However, this is not a real unsupervised method in the sense that their model still needs LR patches as well as their corresponding HR labels.

Very recently, Van Sloun *et al.* [25] applied U-Net [20] deep model to improve upon standard ultrasound localization microscopy (Deep-ULM), and obtained SR vascular images from high-density contrast-enhanced ultrasound data. Their deep-ULM model is observed to be suitable for real time applications, resolving about 1250 HR patches per second. Aiming to improve the texture reconstruction of ultrasound image SR, Choi *et al.* [3] slightly modified the architecture of SRGAN [10] to improve the lateral resolution of ultrasound images. Despite its surprisingly good performance, some evidence [17] (including our corresponding observations in Fig. 3 and Fig. 4) showed that the produced super-resolution image is easy to contain some linear aliasing artifacts.

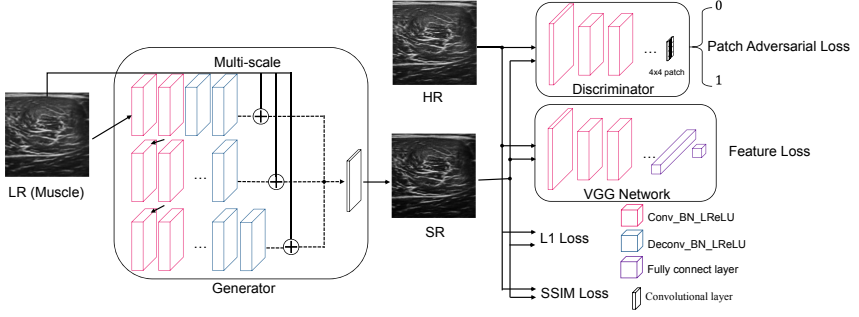


Fig. 1. The proposed ultrasound image SR model: multi-scale ultrasound super-resolved image generator (left) and patches discriminator based on four levels losses (right).

3 Methodology

According to the wavelet and multi-resolution analysis (MRA) [15], an image $f(x)$ is expressed as

$$f(x) = \sum_{k \in Z} a_k^{j_0} \phi_k^{j_0}(x) + \sum_{j=j_0}^J \sum_k b_k^j \psi_k^j(x), \quad (1)$$

in which j varies from j_0 to J , k indexes the basis function, and $\{a_k^{j_0}\}$, $\{b_k^j\}$ act as weighting parameters to associate to the scale function $\phi(x)$ and the wavelet function $\psi(x)$, respectively. Concretely, the image $f(x)$ consists of two components (see Eq. (1)), which are the approximation (the first item, low frequency component) and the details (the second item, high frequency components). From the point of view of deep learning, Eq. (1) may be looked on as a combination reconstruction of different scales branches and each scale reconstruction can be realized by network deconvolution (decoder). Moreover, in Eq. (1), the low frequency (approximation) coefficients a_k^j and the high frequency (detail) coefficients b_k^j can be calculated as:

$$\begin{aligned} a_k^j &= \langle f(x), \phi_k^j(x) \rangle = \sum_i p_{ik}^j f_i \\ b_k^j &= \langle f(x), \psi_k^j(x) \rangle = \sum_i q_{ik}^j f_i \end{aligned} \quad (2)$$

Here, the image $f(x)$ can be represented as $f = \{f_1, f_2, \dots, f_i, \dots\}$, the scale function ϕ_k^j is relaxed to $\{p_{1k}^j, p_{2k}^j, \dots, p_{ik}^j, \dots\}$, and the detail function ψ_k^j can be loosened to $\{q_{1k}^j, q_{2k}^j, \dots, q_{ik}^j, \dots\}$. Obviously, if regarding the weights p_{ik}^j and q_{ik}^j as the convolution kernels at scale j and using the inner projection as the feature encoding, Eq. (2) can be easily implemented by one scale convolution (encoder) on the image $f(x)$. Thus, based on such structure simulation analysis,

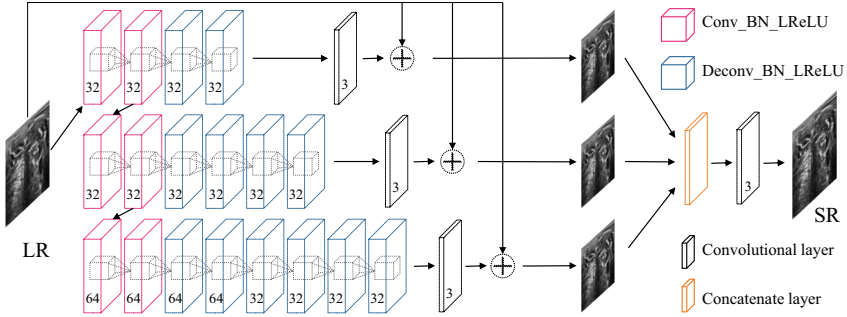


Fig. 2. The pipeline of multi-scale deep encoder-decoder SR network.

obviously we can construct a multi-scale deep encoder and decoder network to renew some lost high-frequency details for image SR.

The pipeline of our proposed ultrasound image SR approach is shown in Fig. 1. Our overall model can be seen as a GAN framework, which consists of two parts: one is a multi-scale encoder-decoder based SR generator to enhance the resolution of the ultrasound images with four levels losses; the other is patches discriminator, which is used to further recover the ultrasound images.

3.1 Multi-scale Encoder-Decoder SR

Inspired by above multi-resolution analysis and the structure simulation, in this work, we specially construct a multi-scale encoder-decoder deep network for the task of the ultrasound image SR and treat it as a generator. The detailed architecture of the proposed multi-scale encoder-decoder SR network is shown in Fig. 2. The specific configurations of the network can be found in Table 1. Here, according to Eq. (1), if regarding the LR image as the approximation component of the HR one, The optimization goal of multi-scale encoder-decoder learning can be treated as:

$$\tilde{f} = \arg \min_f (\|y + \sum_j F_j(y, \Theta_j) - f\|_1), \quad (3)$$

where y and f represent the LR image and the corresponding HR image, and $F(\cdot)$ indicates the reconstruction function. Θ is the learned parameter of the network and the symbol j denotes a specific scale.

In the multi-scale structure, the LR image I_{LR} is firstly sent to three scales encoder-decoder branches to obtain the image details of different scales. Then, these detail maps are directly added to LR image input to get three reconstruction images at different scale thanks to that the LR image can be regarded as the approximation (low frequency) component of the HR one. Finally, we concatenate three reconstruction images and warp them to acquire the super-resolved ultrasound image I_{SR} .

Table 1. The configuration of three scales encoder-decoder streams

Scale1	(conv3-32) ^a × 2 (deconv3-32) × 2
Scale2	(conv3-32) × 2 (conv3-32) × 2 (deconv3-32) × 4
Scale3	(conv3-32) × 2 (conv3-32) × 2 (conv3-64) × 2 (deconv3-64) × 2 (deconv3-32) × 4

^aThe convolution and deconvolution layers are denoted as conv/deconv (kernel size)-(number of filters)

3.2 Patches Discrimination and Loss Function

Recent works [8, 10] show that only using MSE pixel wise loss to supervise SR generation tends to produce over-smooth results. Therefore, in our proposed model, we incorporate four levels loss functions (one pixels loss and three details loss) to supervise the generated SR images to approach the ground-truth HR ones at all levels of details. Moreover, to encourage high-frequency structure, our model takes the similar structure of PatchGAN [8] to discriminate the patches between the generated SR images and the true HR ones.

The MSE loss (ℓ_2 loss) between the generated version $G(y)$ and the real HR one f can be calculated and treated as an objective function for minimization during training. However, due to the energy average characteristics, MSE loss will lead to over-smooth phenomenon. Thus, we replace MSE (ℓ_2) loss with ℓ_1 loss. Here the ℓ_1 loss becomes a measure of the proximity of all the corresponding pixels between such two images. Actually, the ℓ_1 loss is used in the proposed multi-scale encoder and decoder structure (see Eq. (3)).

Given a set of LR and HR image pairs $\{f_i, y_i\}_{i=1}^N$ and assuming the components of network reconstruction at multiple scales can be obtained, then the ℓ_1 pixel-wise loss function for the proposed network can be denoted as:

$$\ell_{pixel} = \sum_{i=1}^N \|(y_i + \sum_j \lambda_j F_j(y_i, \Theta_j)) - f_i\|_1, \quad (4)$$

where λ is regulation coefficient for different details reconstruction term (usually can be set as the reciprocal of the number of scales). Here and in the following, $G(y_i) = y_i + \sum_j \lambda_j F_j(y_i, \Theta_j)$.

We also use the feature loss to guarantee edge features when acquiring super-resolved ultrasound images. Different from the pixel-wise loss, we firstly transform y_i and f_i into certain common feature space or manifold with a mapping function $\phi(\cdot)$. Then we can calculate the distance between them in such feature space. Usually, the feature loss can be described as:

$$\ell_{feature} = \sum_{i=1}^N \|\phi(G(y_i)) - \phi(f_i)\|_2, \quad (5)$$

For the mapping function $\phi(\cdot)$ in our proposed model, in practice we use the combination of the output of the 12th and the 13th convolution layers of the VGG [23] network to realize it. By this way, we can recover the clear edge feature details in super-resolved ultrasound images.

Since SSIM [27] measures the structure similarity in a neighborhood of certain pixel between the generated image and the ground truth one, we may directly apply this measure as one kind of loss:

$$\ell_{ssim} = \sum_{i=1}^N \|SSIM(G(y_i)) - SSIM(f_i)\|_2, \quad (6)$$

Based on the adversarial mechanism of PatchGAN [8], the adversarial loss of our multi-scale SR generator can be defined as:

$$\ell_{adv} = \sum_{i=1}^N -\log(D(G(y_i))) \quad (7)$$

where $D(\cdot)$ is the discriminator of PatchGAN.

The loss of the patches discriminator can be described as the following:

$$\mathcal{L}_{dis} = \sum_{i=1}^N \log(D(f_i)) + \sum_{i=1}^N \log(1 - D(G(y_i))), \quad (8)$$

Finally, the total generator loss utilized to supervise the training of the network can be expressed as:

$$\mathcal{L}_{gen} = \alpha \ell_{pixel} + \beta \ell_{feature} + \gamma \ell_{ssim} + \eta \ell_{adv}, \quad (9)$$

where the α , β , γ , η are the weighting coefficients, which in our experiments are set with 0.16, $2e-6$, 0.84 and $1e-4$, respectively.

4 Experimental Results and Analysis

4.1 Datasets and Training Details

The reconstruction experiments and the performance comparisons are performed on two ultrasound image datasets: CCA-US¹ and US-CASE². The CCA-US

¹ <http://splab.cz/en/download/databaze/ultrasound>.

² <http://www.ultrasoundcases.info/Cases-Home.aspx>.

dataset contains 84 B-mode ultrasound images of common carotid artery (CCA) of ten volunteers (mean age 27.5 ± 3.5 years) with different body weight (mean weight 76.5 ± 9.7 kg). The US-CASE dataset contains 115 ultrasound images of liver, heart and mediastinum, etc. For the objective quality measurement on the super-resolved images, the well-known PSNR [dB], IFC [22], and SSIM [27] metrics are used. The running code of our work is publicly available at https://github.com/hengliusky/UltraSound_Image_Super_Resolution.

For all ultrasound images in CCA-US and US-CASE, by flipping and rotating, each one is enhanced to 32 images, resulting in a training set of 5760 images and a test set of 640 images. Training images are cropped into small overlapped patches with a size of 64×64 pixels and a stride of 14. The cropped ground truth patches are treated as the HR patches. The corresponding LR ones are obtained by two bi-cubic interpolation of the ground truth.

Table 2. Performance comparisons for $4\times$ SR on two ultrasound datasets. The best results are indicated in Bold.

DataSets	Bicubic	SRCNN [5]	SRGAN [10]	Our variant (w/o PatchGAN)	Our proposed
	PSNR/IFC	PSNR/IFC	PSNR/IFC	PSNR/IFC	PSNR/IFC
CCA-US	20.913/1.213	20.673/0.972	25.331/1.127	22.035/1.226	25.711/2.290
US-CASE	26.299/1.055	25.636/1.009	29.069 /1.102	26.373/1.080	29.038/ 2.127

Table 3. Performance comparisons on average PSNR and SSIM for $4\times$ SR. The best results are indicated in Bold.

DataSets	EDSR [12]	SRFeat [18]	Our variant (w/o PatchGAN)	Our proposed
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
CCA-US	25.290/0.740	25.602/0.721	24.755/ 0.834	25.714 /0.814
US-CASE	27.432/0.804	28.864/0.808	28.922/0.840	29.038/0.881

The comparative experiments are to perform different ultrasound image SR methods for fulfilling $4\times$ SR task. In Table 2 and Table 3, we provide the quantitative evaluation comparisons, and in Fig. 3 and Fig. 4, some visual comparison examples with PSNR/IFC or PSNR/SSIM measures are also provided. In addition, to compare the efficiency of our approach with other methods, we also show the inference speed, the model capacity and the throughput of data processing of all methods in Table 4.

4.2 Experimental Comparisons and Analysis

From Table 2 and Table 3, we can see that our proposed model can achieve the best or the second best PSNR measure on different ultrasound image datasets. As for IFC and SSIM measures, our approach will always get the best evaluation value even both on such two ultrasound datasets. Moreover, even without the help of PatchGAN, the multi-scale deep encoder-decoder also demonstrates quite good SR performance (see Table 2), which indicates that it does recover multiple scales image details. While according to Table 3, we see that the feature loss makes sense in enhancing the SR performance of ultrasound data.

Table 4. The comparisons of model inference efficiency. Blue text indicates the best performance and green text indicate the second best performance

	SRCNN [5]	SRGAN [10]	SRFeat [18]	EDSR [12]	Our proposed
Platform	MATLAB	TensorFlow	TensorFlow	TensorFlow	TensorFlow
Test image size	600 * 448	150 * 112	150 * 112	150 * 112	600 * 448
Inference time	188 ms	53 ms	136 ms	49 ms	176 ms
Throughout (Kb/ms)	4.189	0.929	0.362	1.0	4.474
Model capacity	270 KB	9.1 MB	37.2 MB	9.1 MB	1.6 MB

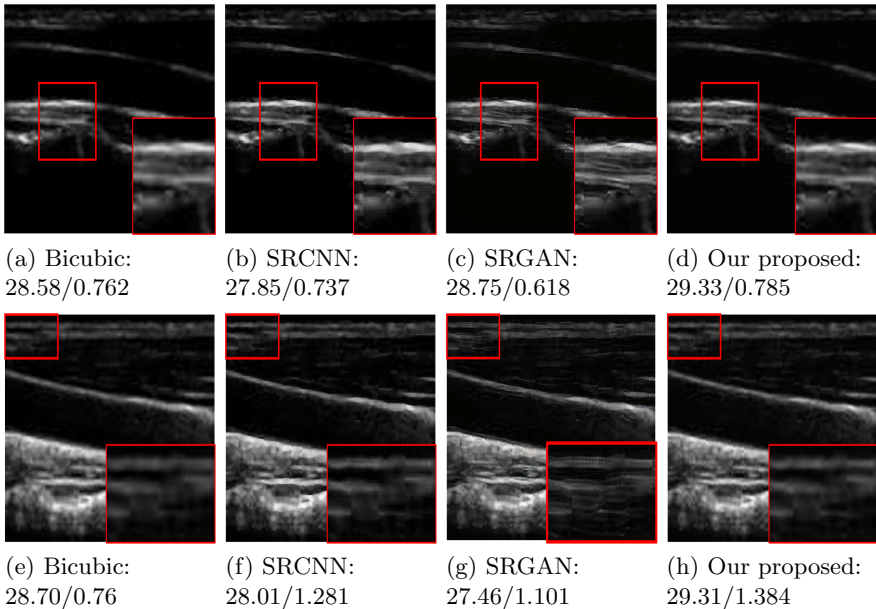


Fig. 3. The Visual and PSNR/IFC comparisons of super-resolved (4 \times) ultrasound images from CCA-US dataset by (a, e) Bicubic, (b, f) SRCNN, (c, g) SRGAN, and (d, h) Our proposed.

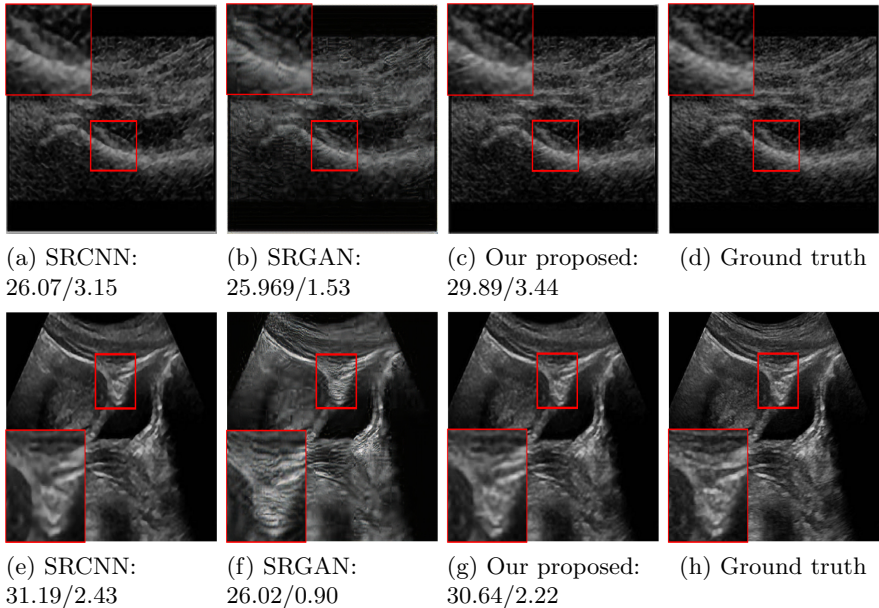


Fig. 4. The Visual and PSNR/IFC comparisons of super-resolved ($4\times$) ultrasound images from US-CASE dataset by (a, e) SRCNN, (b, f) SRGAN, (c, g) Our proposed, and (d, h) Ground truth.

From Fig. 3 and Fig. 4, it is obvious that comparing with other approaches, our presented method can not only acquire the clearer SR visual results but also not to introduce the unwanted aliasing artifacts. In addition, if comparing the structure of our model with that of SRGAN [10], we can see that integrating four levels losses and adopting the local patches discrimination really play a crucial role on accurately super-resolve the LR ultrasound images. Therefore, all the quantitative comparisons and the visual results on different ultrasound image datasets illustrate the excellent performance and the wide effectiveness.

Furthermore, according to Table 4, this demonstrates that our proposed approach is most efficient and hereby to some extent is practically valuable for ultrasonic imaging visual diagnosis in medical industry.

5 Conclusion

In this work, different from the previous approaches, we propose a novel multi-scale deep encoder-decoder and PatchGAN based approach for ultrasound image SR of medical industry. Firstly, an end-to-end multi-scale deep encoder-decoder structure is employed to accurately super-resolve the LR ultrasound images with a comprehensive imaging loss, including the pixel-wise loss, the feature loss, the SSIM loss and the adversarial loss. And then we utilize the discriminator to

urge the generator to get more realistic ultrasound images. Based on the evaluations on two different ultrasound image datasets, our approach demonstrates the best performance not only in objective qualitative measures and the inference efficiency but also in visual effects.

It is noted that the SR of ultrasound image requires higher reconstruction accuracy than that of natural image. Thus, our future work will focus on exploring the relationship between the reconstruction accuracy and the perception clearness for medical image SR.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant No. 61971004 and by the Key Project of Natural Science of Anhui Provincial Department of Education under Grant No. KJ2019A0083.

References

1. Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6228–6237 (2018)
2. Chen, J., Chen, J., Chao, H., Yang, M.: Image blind denoising with generative adversarial network based noise modeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3155–3164 (2018)
3. Choi, W., Kim, M., HakLee, J., Kim, J., BeomRa, J.: Deep CNN-based ultrasound super-resolution for high-speed high-resolution B-mode imaging. In: Proceedings of the IEEE International Ultrasonics Symposium, pp. 1–4, October 2018. <https://doi.org/10.1109/ULTSYM.2018.8580032>
4. Diamantis, K., Greenaway, A.H., Anderson, T., Jensen, J.A., Dalgarno, P.A., Sboros, V.: Super-resolution axial localization of ultrasound scatter using multifocal imaging. *IEEE Trans. Biomed. Eng.* **65**(8), 1840–1851 (2018)
5. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016)
6. Hoskins, P.R., Martin, K., Thrush, A.: *Diagnostic Ultrasound: Physics and Equipment*. Cambridge University Press, Cambridge (2010)
7. Hudson, J.M.: Dynamic contrast enhanced ultrasound for therapy monitoring. *Eur. J. Radiol.* **84**(9), 1650–1657 (2015)
8. 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, pp. 1125–1134 (2017)
9. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
10. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690 (2017)
11. Liang, Y., Wang, J., Zhou, S., Gong, Y., Zheng, N.: Incorporating image priors with deep convolutional neural networks for image super-resolution. *Neurocomputing* **194**, 340–347 (2016)
12. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 136–144 (2017)

13. Liu, H., Fu, Z., Han, J., Shao, L., Hou, S., Chu, Y.: Single image super-resolution using multi-scale deep encoder-decoder with phase congruency edge map guidance. *Inf. Sci.* **473**, 44–58 (2019)
14. Lu, J., Liu, W.: Unsupervised super-resolution framework for medical ultrasound images using dilated convolutional neural networks. In: Proceedings of the IEEE 3rd International Conference on Image, Vision and Computing, pp. 739–744. IEEE (2018)
15. Mallat, S.: *A Wavelet Tour of Signal Processing*. Academic Press, San Diego (1999)
16. Morin, R., Bidon, S., Basarab, A., Kouamé, D.: Semi-blind deconvolution for resolution enhancement in ultrasound imaging. In: Proceedings of the IEEE International Conference on Image Processing, pp. 1413–1417. IEEE (2013)
17. Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts. *Distill* **1**(10), e3 (2016)
18. Park, S.J., Son, H., Cho, S., Hong, K.S., Lee, S.: SRFeat: single image super-resolution with feature discrimination. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 439–455 (2018)
19. Rahaman, N., et al.: On the spectral bias of neural networks. In: Proceedings of the 36th International Conference on Machine Learning, pp. 5301–5310. PMLR (2019)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
21. Sakkos, D., Liu, H., Han, J., Shao, L.: End-to-end video background subtraction with 3D convolutional neural networks. *Multimedia Tools Appl.* **77**(17), 23023–23041 (2017). <https://doi.org/10.1007/s11042-017-5460-9>
22. Sheikh, H.R., Bovik, A.C., De Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **14**(12), 2117–2128 (2005)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
24. Skourt, B.A., El Hassani, A., Majda, A.: Lung CT image segmentation using deep neural networks. *Procedia Comput. Sci.* **127**, 109–113 (2018)
25. van Sloun, R.J., Solomon, O., Bruce, M., Khaing, Z.Z., Eldar, Y.C., Misch, M.: Deep learning for super-resolution vascular ultrasound imaging. In: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1055–1059. IEEE (2019)
26. Umehara, K., Ota, J., Ishida, T.: Application of super-resolution convolutional neural network for enhancing image resolution in chest CT. *J. Digit. Imaging* **31**(4), 441–450 (2018)
27. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
28. Xu, Z.-Q.J., Zhang, Y., Xiao, Y.: Training behavior of deep neural network in frequency domain. In: Gedeon, T., Wong, K.W., Lee, M. (eds.) ICONIP 2019. LNCS, vol. 11953, pp. 264–274. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-36708-4_22
29. Zhao, N., Wei, Q., Basarab, A., Kouamé, D., Tourneret, J.Y.: Single image super-resolution of medical ultrasound images using a fast algorithm. In: Proceedings of the IEEE 13th International Symposium on Biomedical Imaging, pp. 473–476. IEEE (2016)