2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Segmentation with Speckle Reduction and Superresolution by Deep Leaning for Human Ultrasonic Echo Image

Shuo Li, Mengfei Zhang, Yiran Li and Chikayoshi Sumi, *Senior Member, IEEE*

***Abstract*— Ultrasound (US) image diagnosis is widely used for detection and treatment of human malignant tissues. Physicians perform differentiation of tissues through interpreting ultrasound echo images morphologically. However, the ultrasound image always comes with speckles, which makes segmentation of a target tissue difficult. Recently, a deep learning (DL) approach becomes a new way for picture denoising instead of signal processing. In this report, we use the DL denoising to reduce the US speckles. Subsequently, we perform DL segmentation well known for other medical images. In order to further increase the segmentation accuracy, we also perform DL superresolution. The DL superresolution is also well known for a picture and however, not so for an echo image. The target segmentation tissue is a carotid artery, specifically a lumen. To verify the feasibilities of our approaches, simulations and *in vivo* experiments are performed.**

***Clinical Relevance*— Method effectiveness is confirmed for in vivo data.**

I. Introduction

Ultrasound (US) image diagnosis is widely used for detection and treatment of human malignant tissues. Physicians perform differentiation of tissues through interpreting ultrasound echo images morphologically. However, the ultrasound image always comes with speckles, which makes segmentation of a target tissue difficult. As far, the speckle reduction was performed by various signal processings such as incoherent compounding of steered echo images, wavelet processing, etc. Recently, a deep learning (DL) approach becomes a new way for picture denoising and recently, the reports for speckle reduction on optical, radar imaging, etc. are increasing substantially. We also previously proposed the application of DL to the speckle reduction [1]. In this report, we perform speckle reduction using an Auto-Encoder (AE) type DL. Subsequently, we perform DL segmentation using a U-net type DL. The segmentation has been performed for an X-ray CT and an MRI and however, not so for an echo image because of the existence of speckles. In ref. 2, speckle reduction and enhancing a contrast are performed via signal processing prior to the segmentation. In this report, in order to further increase the segmentation accuracy, we also perform DL superresolution. The DL superresolution is also well known for a picture and however, not so for an echo image compared to various inverse filtering on a linear model. In this report, we report the performances of well-known DL superresolutions in a trial. The target segmentation tissue is a carotid artery, specifically a lumen. To verify the feasibilities of our approaches, simulations and *in vivo* experiments are performed.

The authors are with Information Science, Graduate school of Sophia University, Tokyo 102-8554 JAPAN (corresponding author to provide phone: +81-3-3238-3415; fax: +81-3-3238-3321; e-mail: c-sumi@sophia.ac.jp).

II. Methods

We proposed two approaches I and II (Fig. 1) comprising of 3 Steps A to C to process ultrasound echo images, i.e., speckle reduction, superresolution and segmentation. In this report, in step A, the convolutional Auto-Encoder (CAE) [3] is used for reducing speckles; in step B, super-resolution using a convolutional neural network is performed for increasing a spatial resolution [4-6]; and in Step C, a traditional U-Net [7,8] is used for segmenting a target tissue or a region.

One of our proposed approaches, I, i.e., Steps A to C, is shown in Fig. 1 (left). In step A, the CAE [3], which is unsupervised, is performed for the speckle reduction. Next, in Step B, one of the Super-Resolution Convolutional Neural Network (SRCNN) [4], the Fast SRCNN (FSRCNN) [5] and the Efficient sub-pixel CNN (ESPCN) [6] is performed. For comparison, the Step B is also performed solo. Finally, in Step C, the U-Net [7,8], which processes binarized data, is used for the segmentation. The Step C is also performed solo and for the result of Step A or B.

Another approach II, of which order of Steps A and B is inversed with respect to that of approach I (Fig. 1, right), is also performed.

The Auto-Encoder (CAE) [3] used in Step A is one of the most popular approaches for unsupervised learning of complicated distributions. CAE has already shown a significant effect in generating various data such as handwritten digits, faces, object numbers, CIFAR (Canadian Institute For Advanced Research) images, segmentations; and in predicting the future from static images. However, it is still a new method for ultrasound image processing, for instance, for the speckle reduction.

In step B, we perform the SRCNN, the FSRCNN and the ESPCN to increase the image spatial resolution, by which a decrease in spatial resolution caused by Step A is compensated and an accuracy of segmentation to be performed in the step C is increased. The SRCNN [4] is the pioneer of DL super-resolution. The network structure of SRCNN is remarkably simple, using only 3 convolutional layers. The FSRCNN [5] is an improved SRCNN mainly in 3 aspects: (1) The final deconvolution layer is used to enlarge the size, so the original low-resolution image can be directly input into the network instead of enlarging the size with the bicubic method performed in SRCNN. (2) The feature dimension is made small to use smaller convolution kernels and use more mapping layers. (3) The mapping layers can be shared, and only fine-tuning the final deconvolution layer is required if models with different upsampling multiplicities need to be trained [5]. The ESPCN [6] is efficient to extract features directly from the low-resolution image and compute the high-resolution image.

The U-Net [7] is one of the early CNN algorithms for semantic segmentation. The symmetric U-shaped structure containing compressed and expansive paths was innovative and has been influenced on the later designs of several segmentation networks.

III. Simulations

A. Echo simulation models -3 type point spread functions (PSFs)

At first, simulations were performed. Simulated echo data were generated as a convolution of models of an ultrasound pressure (PSF: point spread function) and tissues, i.e., a blood and soft tissues. 3 type PSFs were modeled with Gaussian, parabolic and rectangular envelops. The carrier frequency was set to 3.75 MHz. The pulse length and the beam width were changed. For the Gaussian-type PSF, the respective axial (x) and lateral (y) standard deviations (σ_x and σ_y) had 0.1 and 0.1, 0.2 and 0.2, 0.2 and 0.4, or 0.4 and 0.4 mm. The energies of the parabolic type PSFs were made same as those of the Gaussian-type PSFs. The pulse lengths and the beam widths of the rectangular type PSFs were made same as those of the parabolic type PSFs.

$$\text{PSFs}=\begin{cases} \exp\{-(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2})\}\cos(2\pi\frac{x}{\frac{\lambda}{2}}) & \text{Gaussian-type;} \\ A\cdot\cos(2\pi\frac{x}{\frac{\lambda}{2}}) \quad \text{if } A>0 & \\ \text{where } A=-\frac{8}{9\pi}[(\frac{x}{\sigma_x})^2+(\frac{y}{\sigma_y})^2]+1 & \text{Parabolic type;} \\ \cos(2\pi\frac{x}{\frac{\lambda}{2}}) \quad \text{if } (\frac{x}{\sigma_x})^2+(\frac{y}{\sigma_y})^2\le\frac{9\pi}{8} & \text{Rectangular type,} \end{cases} \quad (1)$$

where λ represents a wavelength (about 0.4 mm) under the assumption of an ultrasound propagation speed of 1,540 m/s. The reflectivity of blood vessel wall, RFL, was set to 2, 3, 5 and 10; and the scattering coefficients of soft tissues were modeled by a white noise with the standard deviation, 1. For the 3 superreolutions, weight data leaned for pictures [4-6] were used; and for the U-net 30 leaning data were used. Other parameters were set as follows: the epoch number = 300, the batch size = 8, and the learning rate = 0.001.

.

B. Evaluation methods

The performances of speckle reduction and super-resolution were evaluated quantitatively by using a peak-signal-to-noise ratio (PSNR); and that of segmentation was by a ratio of a segmented lumen area with respect to a labeled area for the DL (area ratio). The performances were also visually evaluated.

PSNR defined as a ratio of a peak value with respect to the root mean squared error (RMSE) estimated from input and output data [9] is often used. However, since the ground truth is unknown in real-world applications, and in order to allow performing the evaluation for various inputs or results yielded by Steps A to C, the modified PSNR was defined using a standard deviation (SD) as follows:

$$\text{PSNR} = 20\log_{10}\frac{\text{peak}}{\text{SD}} \ \text{[dB]}, \quad (2)$$

by which the evaluation can be performed for respective images independently. Since the gray-scaled images were shown in a full dynamic range, the peak value detected in vessel walls excluding artifacts equaled to 255.

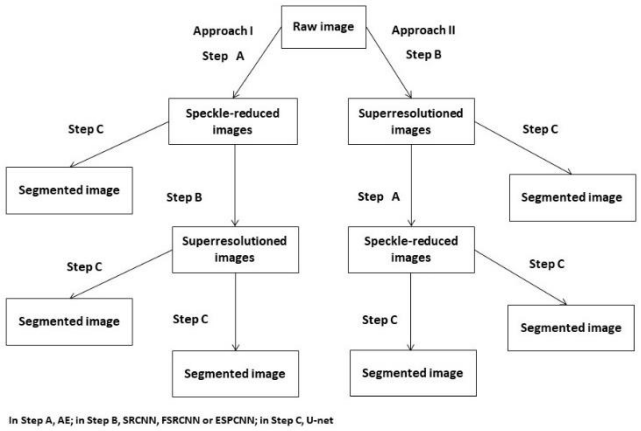


Fig. 1. Flowcharts of our proposed approaches I and II.

C. Results

The results obtained for the Gaussian-, parabolic and rectangular type PSFs with $\sigma_x = 0.2$, $\sigma_y = 0.4$ mm and RFL = 3 are shown in Figs. 2, 3 and 4. All the figures show the images obtained when using ESPCN (others obtained using SRCNN and FSRCNN omitted) in Step B; and only in Fig. 4, in addition to (a) the images, (b) and (c) summarize the PSNRs and the area ratios obtained using the ESPCN, FSNRCNN and SRCNN, respectively (other similar results omitted). Regarding the spatial resolution [(a)], the PSNR [(b)] and the area ratio [(c)], the results obtained for the rectangular type PSF (Fig. 4) was the best of all, and the order of a high performance was rectangular; parabolic (Fig. 3); Gaussian types (Fig. 2). For all the PSF types, the order of a high performance in Step B was ESPCN; FSRCNN; SRCNN. Correspondingly, ESPCN yielded the highest performances in both Approaches I and II.

Specifically, at first, speckle reductions were successfully performed with Step A in the surrounding of the blood vessel for all the PSF types (Figs. 2, 3 and 4a). The PSNR increased, for instance, for the rectangular PSF, from 29.5 up to 33.3 by Step A solo (Fig. 4b). However, simultaneously the spatial resolution substantially decreased (Figs. 2, 3 and 4a). Moreover, for the Gaussian and parabolic type PSFs, artifacts were generated in the lumen, which degraded the subsequent results of Step B or C. This will be discussed in the next subsection also with other axial and lateral deviations and RFLs.

The spatial resolution increased well in Step B (Figs. 2, 3 and 4a) except for the speckle-reduced images. Speckles were slightly reduced simultaneously. The PSNR increased, for instance, for the rectangular PSF, from 29.5 up to 36.7 > 35.6 > 34.7 with ESPCN, FSRCNN and SRCNN solo (Fig. 4b). For the speckle-reduced images, Step B did not recover nor increase the spatial resolution visually regardless the 3 methods.

Segmented lumen was colored by red (Figs. 2, 3 and 4a). Step C worked well, which was confirmed quantitatively and visually (Fig. 4c). The segmentations were successfully performed particularly with Step B. The area ratio increased,

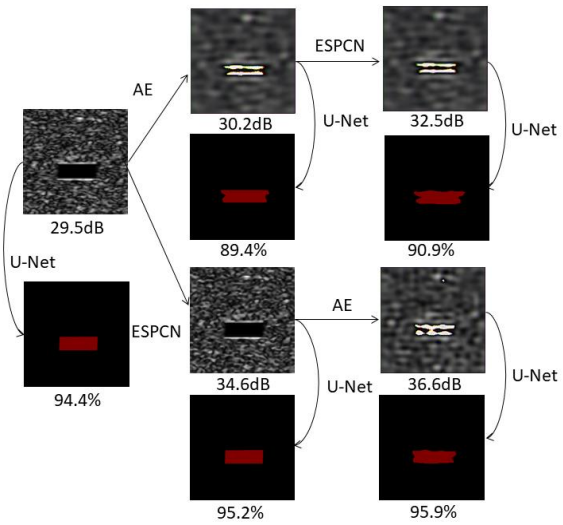


Fig. 2. For Gaussian-type PSF, images obtained with ESPCN network in Step B.

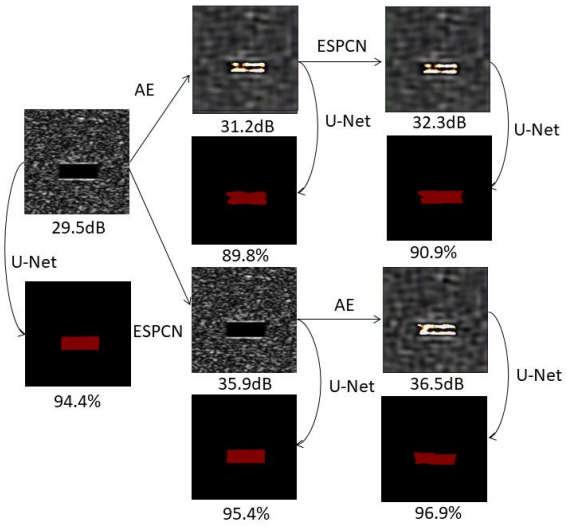
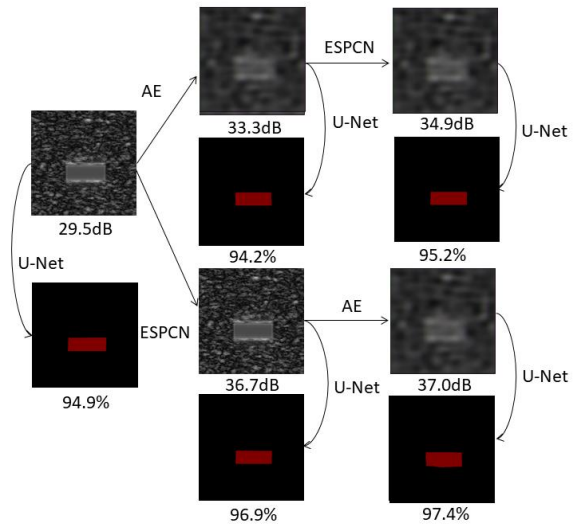


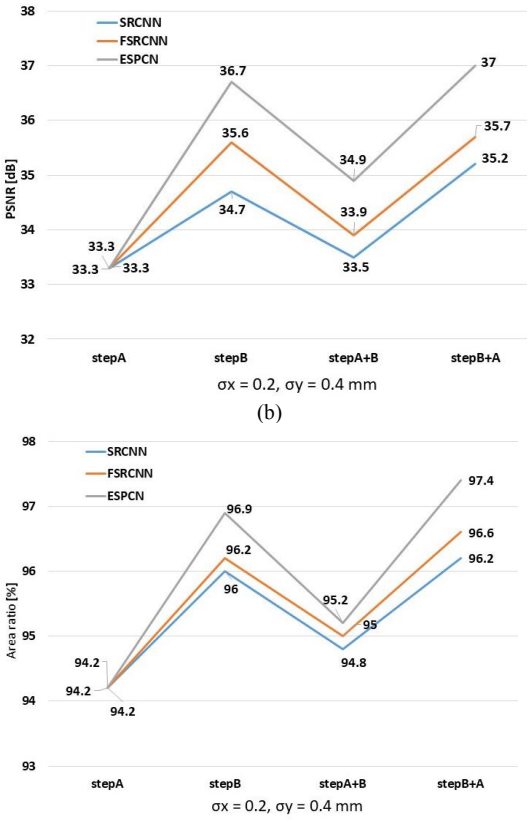
Fig. 3. Results obtained for parabolic type PSF. See caption of Fig.2.

for instance, for the rectangular PSF, from 94.9 to 96.9% with ESPCN (Step B) solo.

For all the PSF types, Approach II with performing Step B prior to Step A was the best of all visually and quantitatively (Figs. 2 to 4). Approach II achieved the higher PSNR and area ratio than the respective Steps solo and Approach I with performing Step A prior to Step B, e.g., for the rectangular PSF, 37.0 vs 29.5 and 34.9; 97.4 vs 94.9 and 95.2%. As mentioned above, Step B solo was effective in increasing the PSNR and the area ratio as well as the spatial resolution and however, ineffective when performed after performing Step A in Approach I, i.e., the lower PSNR and area ratio than with Step B solo. The reason will also be discussed in the next subsection. Actually, the speckle shape and size had an influence on the speckle reduction, the superresolution and the segmentation (corresponding specific results omitted). The better PSNR and area ratio were obtained from the higher spatial resolution images, i.e., with smaller σ_x and σ_y such as 0.1 mm. In the same sense, the order of a high performance was rectangular; parabolic; Gaussian type PSFs.



(a)



(b)

(c)

Fig. 4. Results obtained for rectangular type PSF. (a) Images obtained with ESPCN network in Step B; and (b) PSNRs and (c) area ratios obtained when performing 3 methods.

D. Discussions

At first, the artifact generated by CAE is discussed with respect to the pulse length and beam width; and the reflectivity.

The Pulse length and beam width was one of factors leading to the artifacts as shown in Fig. 5, for instance, RFL = 3 when (a) $\sigma_x = \sigma_y = 0.1$, (b) $\sigma_x = 0.1$, $\sigma_y = 0.2$, (c) $\sigma_x = 0.2$, $\sigma_y = 0.4$ and (d) $\sigma_x = 0.4$, $\sigma_y = 0.6$ mm. The less artifacts were obtained with the smaller pulse length and beam width. Moreover, the order of less artifacts was the rectangular;

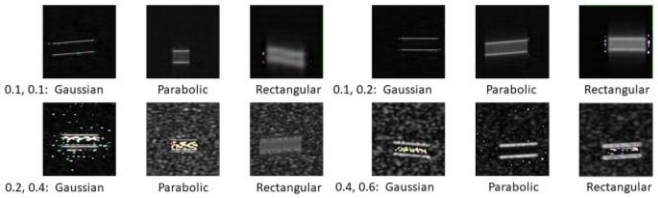


Fig. 5. Artifacts caused by speckle reduction for RFL = 3 with various σ_x and σ_y.

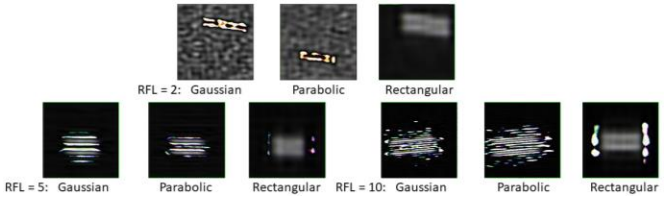


Fig. 6. Artifacts caused by speckle reduction when RFL = 2, 5 and 10.

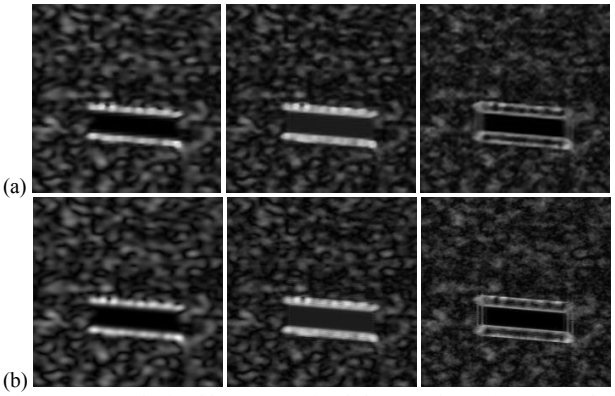


Fig. 7. Images obtained by ESPCN for (left) Gaussian-, (center) parabolic and (right) rectangular type PSFs with $\sigma_x = 0.4$ and $\sigma_y = 0.6$ mm. (a) Raw images and (b) results.

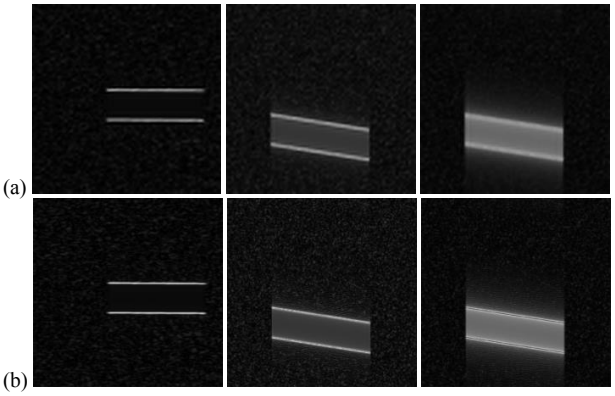


Fig. 8. Images obtained with $\sigma_x = \sigma_y = 0.1$ mm. See caption of Fig. 7.

parabolic; Gaussian-type PSFs. Thus, the less artifacts can be obtained from the higher spatial resolution image.

The artifact also depended on the reflectivity. Fig. 6 shows for RFL = 2, 5 and 10 with $\sigma_x = 0.2$, $\sigma_y = 0.4$ mm the speckle-reduced images. Also see the results with RFL = 3 (Figs. 2 to 4). As shown, the smaller reflectivity the tissue mode had, the less artifacts.

Next, for the superresolution, the 3 type PSFs with other pulse lengths and beams widths (standard deviations) from 0.2 and 0.4 mm are discussed. Fig. 7 shows for the 3 PSF-echo images (left, Gaussian; center, parabolic; and right, rectangular types) with $\sigma_x = 0.4$ and $\sigma_y = 0.6$ mm (a) the raw

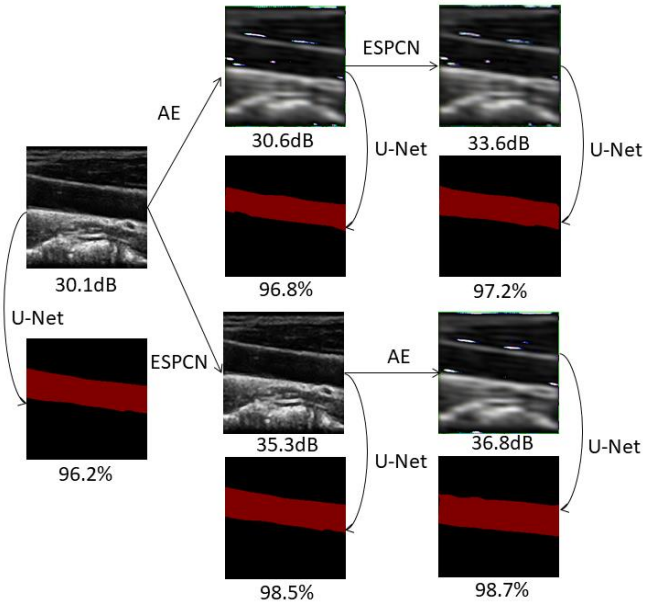


Fig. 9. For human *in vivo* carotid artery, images obtained with ESPCN network in Step B.

images and (b) the ESPCN results; Fig. 8 shows those with $\sigma_x = \sigma_y = 0.1$ mm.

As mentioned above, when the pulse length and beam width were smaller such as with $\sigma_x = \sigma_y = 0.1$ mm than such as with 0.2 and 0.4 mm and 0.4 and 0.6 mm, the PSNR became higher and the segmentation area became more correct. The pulse length and beam width also had influences on the superresolution results as shown. When the pulse length and beam width were larger such as with 0.4 and 0.6 mm, the spatial resolution did not increase so much except for the rectangular PSF originally with a high spatial resolution (Fig. 7).

However, for the originally high spatial resolution images such as with 0.1 and 0.1 mm, the superresolution was remarkably effective in increasing the spatial resolution, particularly for the rectangular PSF (Fig. 8). Thus, the super-resolution works better when the spatial resolution was higher. Since the speckle reduction (step A) decreases the spatial resolution substantially, the superresolution should be performed on the raw image prior to the speck reduction. This is the reason why the approach II yielded the better results than the approach I.

IV. Human in Vivo Carotid Experiments

A. In vivo data

The distributed echo data [10] of human *in vivo* carotid arteries were used, of which pixel sizes were 390 × 330 pixels. Different linear-array-type transducers with two frequencies (10 and 14 MHz) were used. In Step B, the same 3 methods were used. All the same parameters were set as those for the simulation data except for 32 U-net learning data.

B. Results

The order of a high performance was the same as that confirmed in the simulations, i.e., with ESPSN network; FSRCNN; SRCNN in Step B. For the 3 methods, almost the

same PSNR results as those obtained in the simulations were obtained (omitted). Fig. 9 shows the images, with no artifacts which can be generated by Step A, obtained with ESPCN. The speckles were reduced successfully and the artery wall was sharpened substantially. The lumen was also segmented correctly. The best PSNR (36.8) and area-ratio (98.7%) were obtained by Approach II with ESPCN method.

V. Conclusions

We proposed 2 approaches for achieving high accuracy segmentation (step C) through (step A) speckle reduction and (step B) superresolution. Approach I uses the auto-encoder (CAE) to reduce a speckle noise first, then performs super-resolution to recover the image spatial resolution using 3 methods, and finally implements the U-Net to segment a target lumen, while Approach II performs the super-resolution to increase the image spatial resolution in advance, then uses the CAE to reduce a speckle noise, and finally using the U-Net to acquire segmented target lumen. Simulation and human *in vivo* carotid artery data showed that Approach II is superior to Approach I visually and regarding the PSNR and area ratio.

The CAE succeeded in speckle reduction. However, the spatial resolution decreased. Moreover, the *in vivo* and simulation data come with artifacts, being dependent on, a reflectivity and US pulse bandwidths, i.e., a pulse shape, and a pulse length and a beam width. Raw echo data should have high spatial resolutions and the more effective methods may exist. Moreover, the 3 super-resolutions were effective visibly and quantitatively for a high spatial resolution image but not visibly for a low one although the PSNR and area ratio increased. The superresolution should be performed on a high spatial resolution. Since the speckle reduction performed in Step A decreases the spatial resolution, the super resolution should be performed on the raw image prior to the speck reduction processing. This is the reason why the approach II yielded the better results than the approach I in this report. The comparison between Approaches I and II are being continued. We have already been searching for more efficient DL speckle reduction and superresolution methods. The newest results obtained with other methods such as TecoGAN (Temporal Coherence Generative adversarial networks) [11], DDSRCN (Deep Denoising SRCN), ResNet (Residual Network) types [12], etc. will also be presented in accompanying papers [13,14] in addition to those with the 3 superresolution methods learned for the same echo data. Our previously performed inverse signal filtering with the maximum a posteriori [15] and the regularization [16] will also be used together. The contrast enhancement will also be performed together. The U-net also succeeded in segmentation, particularly through Steps A and B. However, to make the segmentation the more efficient, the parameters in the DL such as a leaning data number and an epoch number should be larger and a batch size should be optimized. Particularly, the possible occurrence of overfitting should be removed.

Thus, the proposed approaches show a high potential to be clinically useful for automatic detection of blood flow and a blood vessel. Next targets will also be a heart, a tumor, tumor nodules, etc.

To yield more accurate results, the more echo image data should be used than in this report. In addition, the ultrasound parameters such as a frequency, a pulse length, a beam width, axial and lateral bandwidths and focus position etc., and the tissue parameters such as RFL, scattering density and attenuation, etc. should also be dealt with in detail. When the ground truth is required for DL in real-world applications, using the simulation data will be performed. Not detected but raw rf-echo data will also be targets. These will be reported elsewhere with the omitted results.

References

[1] Sumi, C. (2016). JP patent application no. 2016-222598, Nov 15, 2016 (in Japanese).

[2] Almajalid, R., Shan, K., Du1, Y., Zhang, M. (2018). Development of a Deep-Learning-Based Method for Breast Ultrasound Image Segmentation. Proc of 17th IEEE International Conference on Machine Learning and Applications. pp. 1103-1108.

[3] Mao, X., Shen, C., Yang, Y. B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In Advances in neural information processing systems. pp. 2802-2810.

[4] Dong, C., Loy, C. C., He, K., Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, pp. 295-307.

[5] Dong, C., Loy, C. C., Tang, X. (2016). Accelerating the Super-Resolution Convolutional Neural Network. arXiv:1608.00367.

[6] Shi, W., Caballero1, J., Huszar´, F., Totz, J., Aitken1, A. P., Bishop, R., Rueckert, D., Wang, Z. (2016). Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874-1883.

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1.

[8] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3-11, Springer, Cham.

[9] https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

[10] http://splab.cz/en/download/databaze/ultrasound

[11] Chu, M., Xie, Y., Mayer, J., Leal-Taixe, L., Thuerey, N. (2020). Learning temporal coherence via self-supervision for GAN-based video generation. ACM Trans. Graph., 39(4), Article 75.

[12] GitHub - titu1994/Image-Super-Resolution: Implementation of Super Resolution CNN in Keras.

[13] Li, Y., Zhang, M., Sumi, C. (2021). Superresolution of echo image using DDSRCNN and TecoGAN. Proc of 43rd Ann Int Conf IEEE EMBS to be published (1 page).

[14] Zhang, M., Li, Y., Sumi, C. (2021). Ultrasound echo speckle reduction with superresolution using DDSRCNN and TecoGAN. Proc of 43rd Ann Int Conf IEEE EMBS to be published (1 page).

[15] Sumi, C. (2018). IEICE Technical Report. vol. US2018-4, pp. 23-28 (in Japanese), of which version up to be published in English.

[16] Sumi, C., Kawami, H., Hamada, R., Shionoya, S., Muraoka, K. (2017). IEICE Technical Report. vol. US2017-10, pp. 11-16 (in Japanese), of which version up to be published in English.