Super-resolution technology to simultaneously improve optical & digital resolution of optical coherence tomography via deep learning

Shengting Cao[1], Xinwen Yao[2], Nischal Koirala[1], Brigitta Brott[3], Silvio Litovsky[3], Yuye Ling[4], and Yu Gan[1]

***Abstract*— Optical coherence tomography (OCT) has stimulated a wide range of medical image-based diagnosis and treatment. In cardiac imaging, OCT has been used in assessing plaques before and after stenting. While needed in many scenarios, high resolution comes at the costs of demanding optical design and data storage/transmission. In OCT, there are two types of resolutions to characterize image quality: optical and digital resolutions. Although multiple existing works have heavily emphasized on improving the digital resolution, the studies on improving optical resolution or both resolutions remain scarce. In this paper, we focus on improving both resolutions. In particular, we investigate a deep learning method to address the problem of generating a high-resolution (HR) OCT image from a low optical and low digital resolution (L^2R) image. To this end, we have modified the existing super-resolution generative adversarial network (SR-GAN) for OCT image reconstruction. Experimental results from the human coronary OCT images have demonstrated that the reconstructed images from highly compressed data could achieve high structural similarity and accuracy in comparison with the HR images. Besides, our method has obtained better denoising performance than the block-matching and 3D filtering (BM3D) and Denoising Convolutional Neural Networks (DnCNN) denoising method.**

I. INTRODUCTION

Optical coherence tomography (OCT) [1] is a non-invasive imaging modality that provides depth-resolved tissue microstructural images in real-time. Over the past decades, OCT has spawned a wide range of applications in medical image diagnosis and treatment. In cardiology, for example, OCT is considered as a suitable coronary imaging modality to assess plaques prior to stenting to ensure successful stent deployment and to assess vascular response after the intervention [2]. It is now possible to evaluate subcellular features and make a crucial clinical assessment of whether drug-eluting stents are covered with fibrin or endothelial cells using ultra-high-resolution spectral domain OCT (SD-OCT) [3]. However, such a high resolution comes at multiple costs: more demanding optical design and data storage/transmission. Conventional SD-OCT employs a line detector such as charge-coupled device (CCD) or complementary metal–oxide–semiconductor (CMOS) sensors to measure the interference signals in spectrum domain and uses Fast Fourier Transform to reconstruct the image in spatial domain. The camera must have sufficient pixels to guarantee the reconstruction quality. It thus imposes a limit on data transmission and storage capacity. It remains a challenge to increase speed without compromising image resolution.

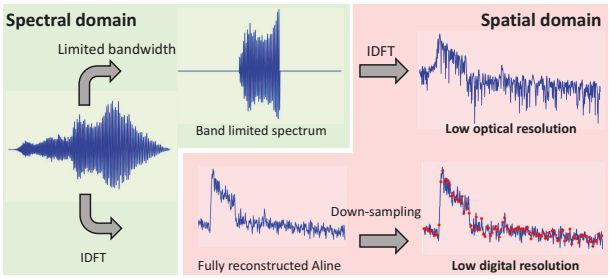

Fig. 1. Illustration of low optical resolution and low digital resolution.

Two distinct types of resolutions are used to characterize OCT systems: optical and digital resolutions. Fig. 1 represents an A-line along the axial direction, the direction where light propagates. Optical resolution [4] defines the system's capability to resolve two point sources with equal intensity. In OCT, the value of axial resolution is inversely proportional to the bandwidth of the light source [1]. A truncated measurement in spectral domain results in a lower axial resolution, as shown in the first row of Fig. 1. Lower optical resolution may allow for a higher scanning rate as only half of the spectrum is recorded. In contrast, digital resolution, defined by the distance per pixel [5], is determined by the number of pixels and imaging range. As shown in the second row of Fig. 1, low digital resolution refers to the case where the reconstructed image is numerically down-sampled. Low optical resolution is caused by bandwidth-limited measurements in spectral domain while a low digital resolution is caused by reduced data in spatial domain.

Super-resolution refers to the procedure of generating high-resolution (HR) images from low-resolution images [6]. Specifically for OCT, many techniques have been proposed to push the limit of resolution. Sparse representation methods, such as dictionary learning, have been used to reconstruct HR images from down-sampled OCT images in [7][8]. An optimization problem is posed and solved via sparse representation to alleviate the low sampling rate caused by jitter and motion artifacts [9]. However, such a sampling scheme in spectral domain does not increase the bandwidth of measured signals. Therefore, the optical

*The Graphics Processing Units (GPU), Titan V, used for this research was donated by the NVIDIA Corporation. Research reported in this paper was in part supported by National Science Foundation (PFI-1918534 and CRII-1948540), National the National Center for Advancing Translational Research of the National Institutes of Health under award number UL1TR003096.

[1]Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, USA ygan6@eng.ua.edu

[2]Institute for Health Technologies, Nanyang Technological University, Singapore

[3]School of Medicine, The University of Alabama at Birmingham, Birmingham, USA

[4]John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai, China

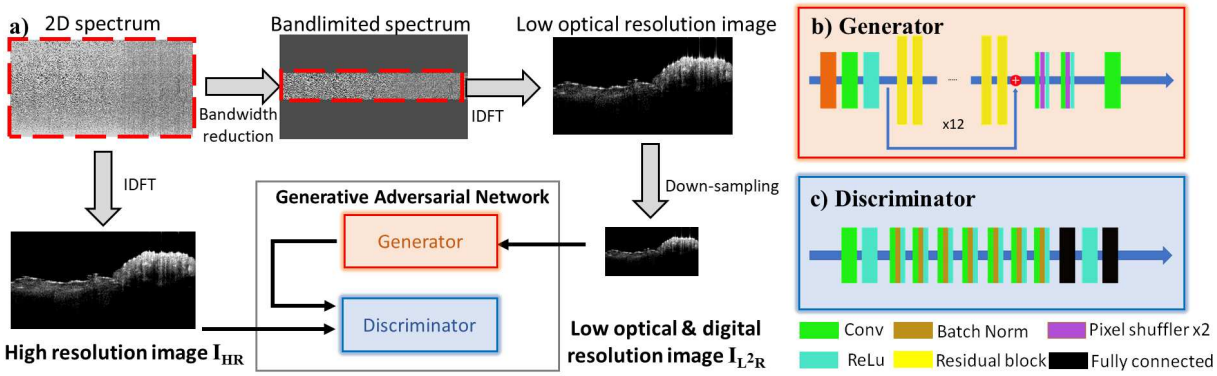

Fig. 2. a) Image pipeline for training with high resolution images (I_{HR}) and low optical, low digital image (I_{L^2R}). b) Network structure of Generator; c) Network structure of Discriminator.

resolution is not improved. Deep learning, a dominating technology in image processing and computer vision, has recently been investigated to facilitate image quality. However, the existing super-resolution-based deep learning framework in OCT [10][11][12] only improves digital resolution without considerations of the optical resolution. To the best of our knowledge, there is no existing technique to improve the optical resolution or simultaneously improves both the optical and digital resolutions via deep learning. Moreover, noise reduction is also in need as the down-sampled image is prone to noise corruption. Recently, sparse representation [13] and deep learning [11] have been proposed to address this need, but neither has been applied to optical resolution.

In this paper, we propose a deep learning framework to simultaneously improve the optical and digital resolutions from OCT images with reduced data size. We employ a generative adversarial network (GAN) to learn the relationship between the low-optical-resolution and low-digital-resolution (L^2R) OCT images and corresponding their HR images off-line. We demonstrate that the GAN network is able to generate HR images from the reduced OCT data that only contains 25% of the original spectrum information and 6.25% of the reconstructed spatial information.

II. Method

A. Data collection

Autopsy specimens were collected from the School of Medicine at the University of Alabama at Birmingham and delivered to the University of Alabama for imaging. The specimens were de-identified and not considered human subjects, according to UAB's Institutional Review Board (IRB). Specimens were imaged via a high-resolution OCT system (Thorlabs Ganymede, Newton, NJ) with an axial resolution of 3 μm and a lateral resolution of 4 μm, both in air. We imaged two segments per specimen, one at the left anterior descending (LAD) and the other at the right coronary artery (RCA). We obtained 23 volumes from three specimens, each with a volume of 1024 $\times$ 1500 $\times$ 750 voxels, corresponding to a space of 1.98 $\times$ 3 $\times$ 1.5 mm^3.

B. Problem formulation

As shown in Fig. 2 (a), we reduced the OCT data in both spectral and spatial domain. In spectral domain, we used the middle 25% of spectrum data. Zero-padding was used to maintain the same imaging depth. After reconstructing image via Inverse Discrete Fourier Transform (IDFT), we used a $\times 4$ scaling factor to obtain a low resolution optical and low digital resolution image, I_{L^2R}. In parallel, we used the full spectrum to reconstruct a high-resolution image, I_{HR}. Then, we built a generator, G, to map I_{L^2R} to I_{HR}:

$$G : I_{L^2R} \rightarrow I_{HR} \tag{1}$$

C. Network architecture

We designed our network via a modification of the super-resolution (SR)-GAN[14] which was originally for natural images. The basic idea was to use ResNet [15] as a generator G and use a discriminator D to provide feedback on the performance of G. The detailed structure is shown in Fig. 2 (b) and (c). The discriminator and generator worked together to solve the following adversarial min-max problem: [14]:

$$\begin{aligned} \min_G \max_D L(G,D) = E_{y\sim p_{I_{HR}}}[log(D(I_{HR}))] + \\ E_{x\sim p_{I_{L^2R}}}[log(1-D(G(I_{L^2R})))] \end{aligned} \tag{2}$$

where $p_{I_{HR}}$ and $p_{I_{L^2R}}$ denote the distribution of I_{HR} and I_{L^2R}.

During the optimization, network parameters were tuned by minimizing a loss function that consists of adversarial loss and content loss components. Adversarial loss was defined based on the probabilities of the discriminator over all training samples. The content loss was designed as:

$$l^{L^2R}_{content} = w_1 l^{L^2R}_{MSE} + w_2 l^{L^2R}_{MSSIM} + w_3 l^{L^2R}_{VGG} + w_4 l^{L^2R}_{TV} \tag{3}$$

where weights, from w_1 to w_4, are assigned to mean square error (MSE), multiscale structural similarity index (MSSIM), VGG feature [16], and total variation (TV), respectively.

In comparison with the SR-GAN [14], we simplified the generator structure and built a new loss function. As OCT image only has one gray-scale channel, we reduced the number of layers in the generator from 16 residual blocks to 12. MSSIM and TV were newly introduced in loss function. MSSIM was used to achieve better structural information. TV factor was used to increase the homogeneity within a tissue type. Noticeably, in addition to super-resolution, MSSIM and TV-based optimization can be also considered as a process of suppressing speckle noise, leading to an effect of desnoising.

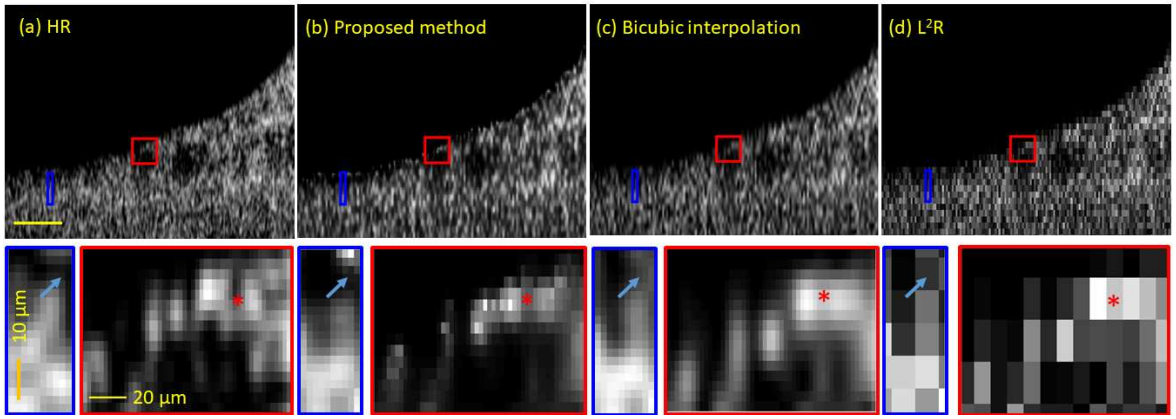

Fig. 3. Visual comparison of super-resolution performance. a) Original HR image; b) Reconstructed image from proposed method; c) Reconstructed image from bicubic interpolation; d) Original low resolution L^2R image. The second row corresponds to the insets (red and blue boxes) in the first row.

III. Experiments Results

A. Implementation

We optimized both the generator and discriminator using the Adam algorithm [17]. In our experiment, we empirically set the values of the hyperparameters as follows: $\beta_1 = 0.9$, $lr_{decay} = 0.1$. The weights in Eq. 3 were: $w_1 = 1$, $w_2 = 0.1$, $w_3 = 2 \times 10^{-6}$, $w_4 = 3 \times 10^{-6}$. The batch size is 4. We ran our scripts under Tensorflow and Tensorlayer on an Alienware Area-51 R5 computer with 48 GB of RAM and two NVIDIA GPU, 2080Ti and Titan V. Training was monitored and terminated when no further improvement was observed. We acquired a cohort of 266 images, with each one divided into patches of 384 $\times$ 384 pixels. Overall, we had 1596 patches. The training versus testing division is 3:1. Training data and testing data were shuffled based on volumes such that training and testing patches were not from the same volume.

B. Experiment setup

In our experiment, we generate HR image I_{gen} from I_{L^2R}, and compare with its I_{HR} over similarity and accuracy. We evaluate the similarity using structural similarity index (SSIM)[18] and Pearson correlation coefficient (Corr)[19]:

$$SSIM = \frac{(2\mu_{I_{gen}}\mu_{I_{HR}} + C_1) + (2\sigma_{I_{gen}I_{HR}} + C_2)}{(\mu_{I_{gen}}^2 + \mu_{I_{HR}}^2 + C_1)(\sigma_{I_{gen}}^2 + \sigma_{I_{HR}}^2 + C_2)} \quad (4)$$

$$\rho = \frac{Cov(I_{gen}, I_{HR})}{\sigma_{I_{gen}}\sigma_{I_{HR}}} = \frac{E[(I_{gen} - \mu_{I_{gen}})(I_{HR} - \mu_{I_{HR}})]}{\sigma_{I_{gen}}\sigma_{I_{HR}}} \quad (5)$$

where μ, σ are the local means, standard deviations. C_1 and C_2 are regularizers determined by the dynamic range. We evaluate the accuracy based on MSE and PSNR [20]:

$$MSE = \frac{\sum_{M,N}[I_{gen}(m,n) - I_{HR}(m,n)]^2}{M \times N} \quad (6)$$

$$PSNR = 10log_{10}(\frac{R^2}{MSE}) \quad (7)$$

where m = 1, 2, .., M and n = 1, 2, ..., N are indexes of row and column in image; R is the max fluctuation of images.

Based on the four metrics, we compare the super-resolution performance of our method with bicubic and orginal SR-GAN [14]. Moreover, we compare the denoising performance with Block-matching and 3D filtering (BM3D) method [21] and Denoising Convolutional Neural Networks (DnCNN) [22]. We add Gaussian noise to L^2R images with different σ^2 value, then feed the noisy L^2R images to our network to get the generated images. As BM3D and DnCNN work on fully-sampled images, we first add noise to interpolated L^2R images and then feed those images to each denoising frameworks to obtain reconstructed images.

C. Results

We evaluate the performance of super-resolution. Table I lists the mean values of PSNR, MSE, SSIM, and Corr obtained from proposed method, SR-GAN, and bicubic method. Our method outperforms other two methods with a lower MSE and higher values of SSIM, Corr, and PSNR. It indicates superiority in reconstructing fine morphological details for low optical and digital resolution images. Fig. 3 shows the visual comparison of a representative image processed by our proposed method and bicubic interpolation, the method with higher PSNR/Corr and lower MSE than SR-GAN, using the central 25% spectrum in spectrum domain and 6.25% reconstructed data in spatial domain. First, regarding optical resolution, our method preserves and enhances the blurry boundary region over the axial direction (highlighted as arrow region). It resolves a clear boundary in the generated image. Second, regarding digital resolution, L^2R is not able to resolve the two peaks column around the star marks in the red box of Fig. 3. Bicubic interpolation resulted in an over-smooth transition between the two peaks. In contrast, our method is able to highlight the value change on the left and right columns close to the star mark.

To evaluate the denoising performance, we plot the PSNR over three levels of Gaussian noise in Fig. 4, using proposed method, BM3D, and DnCNN. We observe higher PSNR values from our methods than values from the other two. The BM3D method has a higher PSNR than DnCNN method in two corrupted scenarios (σ^2 = 0.04 and 0.06) while DnCNN only has better performance when the noise is mild

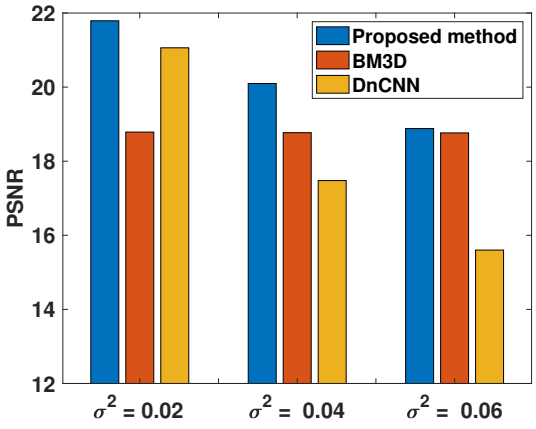

Fig. 4. Denoising performance.

(σ^2 = 0.02). Our method is computationally efficient. The runtime of our method (∼2.48 sec/image) is less than that of BM3D(∼5.97 sec/image) and DnCNN(∼20.16 sec/image).

TABLE I
QUANTITATIVE COMPARISON AMONG THREE METHODS IN SUPER-RESOLUTION.

	PSNR	MSE	SSIM	Corr
Proposed method	**24.7905**	**241.9376**	**0.8056**	**0.9315**
Bicubic	24.6012	255.8057	0.7789	0.9293
Original SR-GAN	24.4781	259.28897	0.7966	0.9258

IV. DISCUSSION AND CONCLUSION

We presented a deep learning framework to reconstruct high resolution images from low optical and low digital resolution images. The proposed method has shown superiority in comparison with the conventional interpolation methods and the denoising methods.

The proposed framework has three major contributions. First, our method considers an OCT system with both the low optical and low digital resolutions. It is one step further than current deep learning-based OCT super-resolution methods [10][11][12] that solely consider digital resolution. Second, we demonstrate the feasibility of using highly compressed OCT data in spectral measurement to achieve the same optical resolution, making it possible to maintain a high resolution while reducing the bandwidth (less spectrum information to be saved). Third, our method can also be used as a denoising method to enhance the image quality.

We did not compare our method with sparse representation in this study because sparse representation has a strict requirement on the sparsity of compressed signals. Due to the fact that the image data has been reduced twice, in spectral and spatial domain sequentially, the sparsity property of the signal is not well investigated. In the future, we will perform a theoretical study on its property and conduct the comparison accordingly. With a goal of verifying the enhancement in clinical decision-making, we will also correspond our OCT images with histology images to identify tissue types such as cholesterol crystals, activated macrophages, etc.

REFERENCES

[1] W. Drexler and J. G. Fujimoto, *Optical coherence tomography: technology and applications*. Springer Science & Business Media, 2008.

[2] W. Wijns, J. Shite, M. Jones, S. Lee, M. Price, F. Fabbiocchi, E. Barbato, T. Akasaka, H. Bezerra, and D. Holmes, "Optical coherence tomography imaging during percutaneous coronary intervention impacts physician decision-making: ILUMIEN I study," *European heart journal*, vol. 36, 08 2015.

[3] L. Liu, J. Gardecki, S. Nadkarni, J. Toussaint, Y. Yagi, B. Bouma, and G. Tearney, "Imaging the subcellular structure of human coronary atherosclerosis using 1-μm resolution optical coherence tomography," *Nature medicine*, vol. 17, pp. 1010–4, 07 2011.

[4] A. J. Den Dekker and A. Van den Bos, "Resolution: a survey," *JOSA A*, vol. 14, no. 3, pp. 547–557, 1997.

[5] R. Gonzalez and R. Woods, *Digital Image Processing*. Pearson, 2017.

[6] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *arXiv preprint arXiv:1902.06068*, 2019.

[7] D. Zermeno, P. Mayo, L. Nicholson, and A. Achim, "Super-resolution oct using sparse representations and heavy-tailed models," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2019, 07 2019, pp. 5585–5588.

[8] L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Transactions on Medical Imaging*, vol. 32, no. 11, pp. 2034–49, 2013.

[9] X. Liu and J. U. Kang, "Compressive SD-OCT: the application of compressed sensing in spectral domain optical coherence tomography," *Opt. Express*, vol. 18, no. 21, pp. 22 010–22 019, Oct 2010.

[10] T. Zhou, K. Zhou, J. Yang, L. Fang, Y. Hu, Y. Zhao, J. Cheng, X. Chen, S. Gao, and J. Liu, "Digital resolution enhancement in low transverse sampling optical coherence tomography angiography using deep learning," *arXiv preprint arXiv:1910.01344*, 2019.

[11] Y. Huang, Z. Lu, Z. Shao, M. Ran, J. Zhou, L. Fang, and Y. Zhang, "Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network," *Opt. Express*, vol. 27, no. 9, pp. 12 289–12 307, Apr 2019.

[12] Q. Hao, K. Zhou, J. Yang, L. Fang, Z. Chai, Y. Ma, Y. Hu, S. Gao, and J. Liu, "High signal-to-noise ratio reconstruction of low bit-depth optical coherence tomography using deep learning," 2019.

[13] R. Zhao, Y. Zhao, Z. Chen, Y. Zhao, J. Yang, Y. Hu, J. Cheng, and J. Liu, "Speckle reduction in optical coherence tomography via super-resolution reconstruction," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2019, pp. 5589–5592.

[14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *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*, 2017, pp. 4681–4690.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR, San Diego, CA, USA, May 7-9*, 2015.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, April 2004.

[19] J. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, ser. Statistics, textbooks and monographs. Marcel Dekker Incorporated, 2003.

[20] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 2366–2369.

[21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, 2007.

[22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, 2017.