Multi-scale Sparse Representation-Based Shadow Inpainting for Retinal OCT Images

Yaoqi Tang[a], Yufan Li[a], Hongshan Liu[b], Jiaxuan Li[a], Peiyao Jin[c], Yu Gan[b], Yuye Ling[a,*], and Yikai Su[d]

[a]John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai 200240, China
[b]Department of Electrical and Computer Engineering, The University of Alabama, AL 35487, USA
[c]Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University, Shanghai 200080, China
[d]State Key Lab of Advanced Optical Communication Systems and Networks, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ABSTRACT

Inpainting shadowed regions cast by superficial blood vessels in retinal optical coherence tomography (OCT) images is critical for accurate and robust machine analysis and clinical diagnosis. Traditional sequence-based approaches such as propagating neighboring information to gradually fill in the missing regions are cost-effective. But they generate less satisfactory outcomes when dealing with larger missing regions and texture-rich structures. Emerging deep learning-based methods such as encoder-decoder networks have shown promising results in *natural* image inpainting tasks. However, they typically need a long computational time for network training in addition to the high demand on the size of datasets, which makes it difficult to be applied on often small *medical* datasets. To address these challenges, we propose a novel multi-scale shadow inpainting framework for OCT images by synergically applying sparse representation and deep learning: sparse representation is used to extract features from a small amount of training images for further inpainting and to regularize the image after the multi-scale image fusion, while convolutional neural network (CNN) is employed to enhance the image quality. During the image inpainting, we divide preprocessed input images into different branches based on the shadow width to harvest complementary information from different scales. Finally, a sparse representation-based regularizing module is designed to refine the generated contents after multi-scale feature aggregation. Experiments are conducted to compare our proposal versus both traditional and deep learning-based techniques on synthetic and real-world shadows. Results demonstrate that our proposed method achieves favorable image inpainting in terms of visual quality and quantitative metrics, especially when wide shadows are presented.

Keywords: retinal optical coherence tomography, image inpainting, sparse representation, shadow removal, artifacts removal

1. INTRODUCTION

Retinal optical coherence tomography (OCT) images are widely used in clinical settings for diagnosing ophthalmic diseases.[1] Despite of the μm-level resolution and the high sensitivity manifested in OCT images, artifacts are often spotted including the shadows that are caused by scattering from the superficial retinal vessels. These artifacts would decrease the accuracy of automated segmentation algorithms and cause errors in clinical diagnosis.[2] Therefore, an accurate shadow inpainting algorithm that is applicable for suppressing such artifacts in retinal OCT images is of great interest to the community.

Inpainting shadow regions while simultaneously maintaining overall consistency for retinal OCT images could be categorized as an image restoration problem, which attributes to many practical applications and has been extensively advanced in the field of computer vision.[3] Traditional sequence-based methods[456]for this task that gradually fill in the missing area by diffusing neighboring information or searching for most similar patches from known regions are widely utilized. But they cannot handle complicated structures and fail to recover realistic results due to the lack of global information. Sparse representation-based methods[7] can extract features of images by decomposing data into a low-dimensional space and has shown great effectiveness in various image processing

---

tasks.[8] Besides, recent research efforts turn to deep learning-based techniques.[9] For example, Li *et al.* proposed a CNN-based method using Recurrent Feature Reasoning (RFR) module to recurrently infer missing regions.[10] Although deep learning neural networks often present superior performance over traditional counterparts, most of them are trained on *natural* images so that a domain transferring might be necessary before being directly applied to retinal OCT images.[11] More importantly, a relatively large dataset and high computational costs are required for training and testing.

Due to these challenges, incorporating sparse representation into medical image processing remains as a popular alternative. Fang *et al.* proposed a sparsity-based tomographic method for denoising Spectral-Domain OCT volumetric data.[12] Liu *et al.* proposed a dictionary-based sparse representation method recently for saturation artifacts removal in OCT images.[13] While their method works very well with saturation artifacts narrower than the patch width, it appears to be less effective when large areas of shadow are presented.

To address these issues, we propose a multi-scale sparse representation-based shadow inpainting framework that reconstructs shadowed regions at multiple scale branches. For wide shadows, we adopt a *downsampling-inpainting-upsampling* work flow to utilize the information of other scale and reduce the shadow width for inpainting retinal OCT images. Specifically, the proposed method extracts the structural information via sparse representation, and later performs fusion with the aid of a CNN-based super-resolution module. In addition, we design a regularizing module to constraint the reconstructed output while simultaneously merging information of other scales. We validate the proposed method by conducting experiments on synthetic and real-world shadows. Both qualitative and quantitative results demonstrate the superior performance of our proposed model.

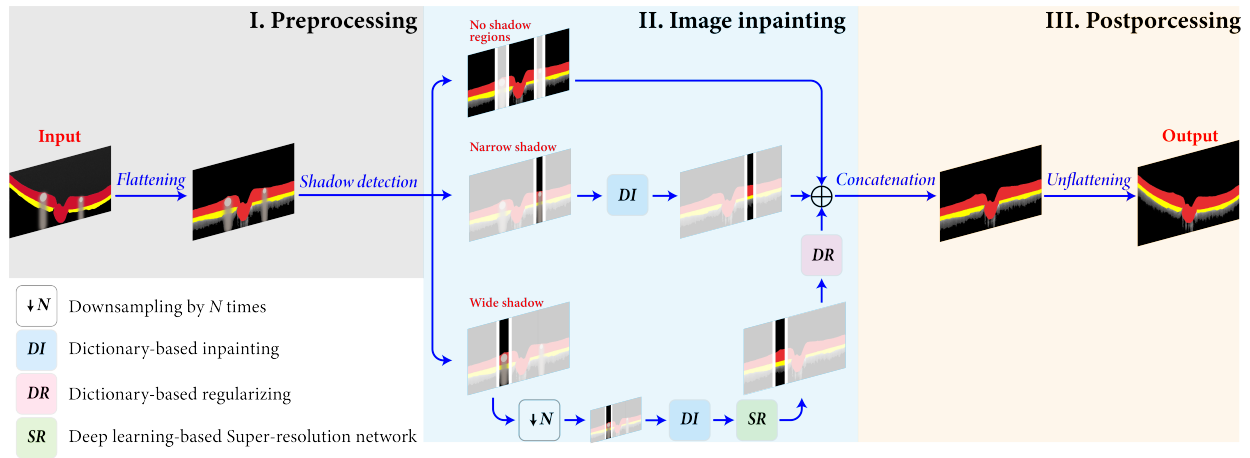## 2. METHODS

2.1 Framework

Figure 1. The framework of the proposed method.

As illustrated in Figure 1, the proposed method consists of three procedures: 1) preprocessing, 2) image inpainting, and 3) postprocessing. An input retinal OCT image that contains shadowed regions of different widths is first flattened and fed to an A-line-based shadow detection algorithm. Then, we channel shadows into three different branches based on the presence and the width of the shadows. Specifically, the flattened image is striped into the different types of strips including: (1) shadow-free regions, (2) narrowly shadowed regions and (3) widely shadowed regions. The regions with no shadow presented will be kept intact, while the narrowly shadowed regions will be processed in the dictionary-based inpainting (DI) module.

For wide shadows, we first downsample them to decrease their widths and then send them to the DI module, where they will be inpainted based on sparse representation theory. The inpainted strips will first be upsampled back to the original resolution via an super-resolution network to restore the image quality and later be regularized by a dictionary-based regularizing (DR) module to integrate complementary feature information from a higher

scales. Finally, we perform the post-processing by concatenating and unflattening the aforementioned regions to form a shadow-free output.

2.2 Preprocessing

The preprocessing process includes image flattening and shadow detection. We identify the Bruch's membrane (BM), which is assumed to be the brightest layer within all retinal tissue, by picking up the pixels with the highest intensity of each A-line. Then, the BM's location is fitted using a robust version of the locally estimated scatterplot smoothing (LOESS[14]) method. Considering that the pixel intensity could be perturbed by strong scatterers, we further detect the outliers in the fitting as good indicators of the presence of vessels and shadows. Besides, by moving the BM to a fixed depth according to the fitted curve with outliers removed, the A-lines are realigned to get the flattened image. Finally, a dilation operation is performed along with a connectivity test to group neighboring shadowed A-lines together and create a binary shadow mask where shadowed pixels are labeled as "0" and shadow-free pixels as "1". To ensure that the shadows do not coincide with the image boundary, the detected edges of shadows in the binary mask are broadened by a margin of 2 pixels.

2.3 Image Inpainting

The image inpainting process is divided into different processing branches: no shadow regions, narrowly shadowed regions and widely shadowed regions. The threshold used to separate shadows is set as the patch size used in the dictionary learning stage. For shadow-free regions, the image patch will be kept intact.

2.3.1 Narrowly shadowed regions

The narrowly shadowed regions in a flattened image are cropped based on the shadow mask and then fed to a DI module, where it is divided into overlapping patches for inpainting. These patches have a size of $a \times b$, and are vectorized as $\boldsymbol{y} \in \mathbf{R}^{ab \times 1}$. For training the overcomplete dictionary $\mathbf{D} \in \mathbf{R}^{ab \times M}$ ($ab < M$, where M represents the number of dictionary atoms) used in the DI module, we select 33 high quality retinal OCT images $\mathbf{Y}$ without shadows from the dataset,[2] in which black background patches are excluded. Based on the sparsity of images, we train the dictionary as follows,

$$\min_{\mathbf{D}, \alpha} \|\boldsymbol{y}_t - \mathbf{D}\boldsymbol{\alpha}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_0 \leq L \tag{1}$$

where $\boldsymbol{\alpha} \in \mathbf{R}^{M \times 1}$ is the sparse coefficient, L is the sparsity level and $\boldsymbol{y_t}$ represents the shadow-free patch vector from the training dataset. It should be noted that the learned dictionary $\mathbf{D}$ is later used in DI model and DR model. Considering a case where we have shadowed pixels in a patch, we first remove the shadowed pixels from vector $\boldsymbol{y}$ and rows with the same index from the dictionary $\mathbf{D}$ to get $\boldsymbol{y}'$ and $\mathbf{D}'$. We assume that the remaining pixels are reliable. Sparse coefficient[15] $\boldsymbol{\alpha}$ is calculated by:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} \|\boldsymbol{y}' - \mathbf{D}'\boldsymbol{\alpha}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_0 \leq L \tag{2}$$

As an inpainting process, the inpainted patch $\hat{\boldsymbol{y}}$ for testing patch $\boldsymbol{y}$ is calculated by:

$$\hat{\boldsymbol{y}} = \mathbf{D}\hat{\boldsymbol{\alpha}} \tag{3}$$

Finally, we aggregate all the inpainted patches to form the inpainted result.

2.3.2 Widely shadowed regions

The processing of widely shadowed regions follows a similar procedure except that the flattened image is first downsampled and the inpainted output is later upsampled. For the dictionary $\mathbf{D}_d$ used in the wide shadow inpainting, the training dataset is similar except that the OCT images are downsampled first before dividing into overlapping image patches. $\downarrow_N$ is the downsampling operation with a factor of N. Specifically, the downsampling factor N is chosen depending on the statistics of the shadow width to ensure that the reduced shadow width is within the allowable range.

$$\mathbf{Y}_d = (\mathbf{Y}) \downarrow_N \tag{4}$$

For wide shadow branch, the input flattened image is first downsampled by N times (*e.g.* $N = 4$ in this study) to reduce the width of wide shadow to below the threshold and simultaneously utilize the information of this scale. The downsampled image is then fed into the DI module and inpainted,

$$\hat{\boldsymbol{\alpha}}_d = \arg\min_{\alpha_d} \|\boldsymbol{y}'_d - \mathbf{D}'_d\boldsymbol{\alpha}_d\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}_d\|_0 \leq L \tag{5}$$

The inpainted results are later upsampled by the Enhanced Deep Residual Network (EDSR)[16] $\mathcal{U}$ afterwards to restore the high resolution information.

$$\hat{\boldsymbol{y}}' = \mathcal{U}(\hat{\boldsymbol{y}}_d) = \mathcal{U}(\mathbf{D}_d\hat{\boldsymbol{\alpha}}_d) \tag{6}$$

Furthermore, to constraint the generated inpainted results, a sparse representation-based regularizing module is designed to refine the contents after feature aggregation. The regularization is conducted by Equation 7 for all patches having inpainted pixels.

$$\hat{\boldsymbol{y}}'' = \mathbf{D}\hat{\boldsymbol{\alpha}}_c \tag{7}$$

where $\hat{\boldsymbol{\alpha}}_c$ is calculated by:

$$\hat{\boldsymbol{\alpha}}_c = \arg\min_{\alpha_c} \|\hat{\boldsymbol{y}}' - \mathbf{D}\boldsymbol{\alpha}_c\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}_c\|_0 \leq L \tag{8}$$

The final results are concatenated by the outputs of three aforementioned branches. After unflattening, the shadows in the output image are replaced by inpainted results and the other reliable A-lines remain unchanged.

2.4 Experimental Settings

We evaluate our proposed method on retinal OCT images with both synthetic and real-world shadows. For synthetic shadows, we randomly generate 468 shadows of different widths ranging from 7 to 24 pixels in 9 OCT B-scans. For real-world shadows, 5 additional OCT B-scans which are not included in the training are tested.

2.4.1 Dictionary learning

For dictionary learning, two dictionaries each containing 128 atoms are trained with the same parameters for different scales based on Equation (1) with $L = 2$. The training dataset consists of 33 OCT B-scans selected from the dataset,[2] which are collected at the Ophthalmology Department of Shanghai General Hospital with a size of 1024×992 pixels, corresponding to a field of view of 20.48 mm×7.94 mm. For the dictionary of the other scale, images are downsampled by a factor of 4 correspondingly. After cropping, 10,000 overlapping patches are extracted per B-scan with a size of 8×8 pixels, which makes the total 330,000 patches.

2.4.2 Deep learning-based super-resolution network

To train the deep learning-based super-resolution network, 800 OCT B-scans from OCTA-500 dataset[17] are randomly picked for pretraining the EDSR, centered on the macula or the optic nerve. We then finetune the pretrained model by using the collected dataset[2] with 544 B-scans for training and 181 B-scans for validation. It should be noted that the training dataset for super-resolution network and DI module are mutually exclusive. Each B-scan is cropped into overlapping patches with a size of 48×48 pixels.

2.4.3 Deep learning-based inpainting network

To validate the proposed method, we compared our model with existing traditional methods together with state-of-the-art deep learning-based inpainting methods. Specifically, SI, TV,[18] the CNN-based network RFR[10] and RN[19] are chosen. RFR and RN are pretrained in *natural* image datasets. We use their original implementation and the models trained on Paris StreetView[20] training set provided by the authors. We then finetuned the pretrained model on OCT images dataset which contains 180,800 randomly selected images from OCTA-500 and 11,208 images from our training dataset. Specifically, the training images include missing regions, which are generated by applying synthetic binary shadow masks with a normally distributed mask width. The images and the masks are then cropped into patches with a size of 256×256 to conform the input format of the networks with no further resizing.

2.4.4 Segmentation network

To evaluate the inpainting performance of different methods on machine analysis, we fed the inpainted image to a widely used segmentation network UNet.[21] The Dice score of the segmented results are calculated to analyze the performance of different inpainting methods.
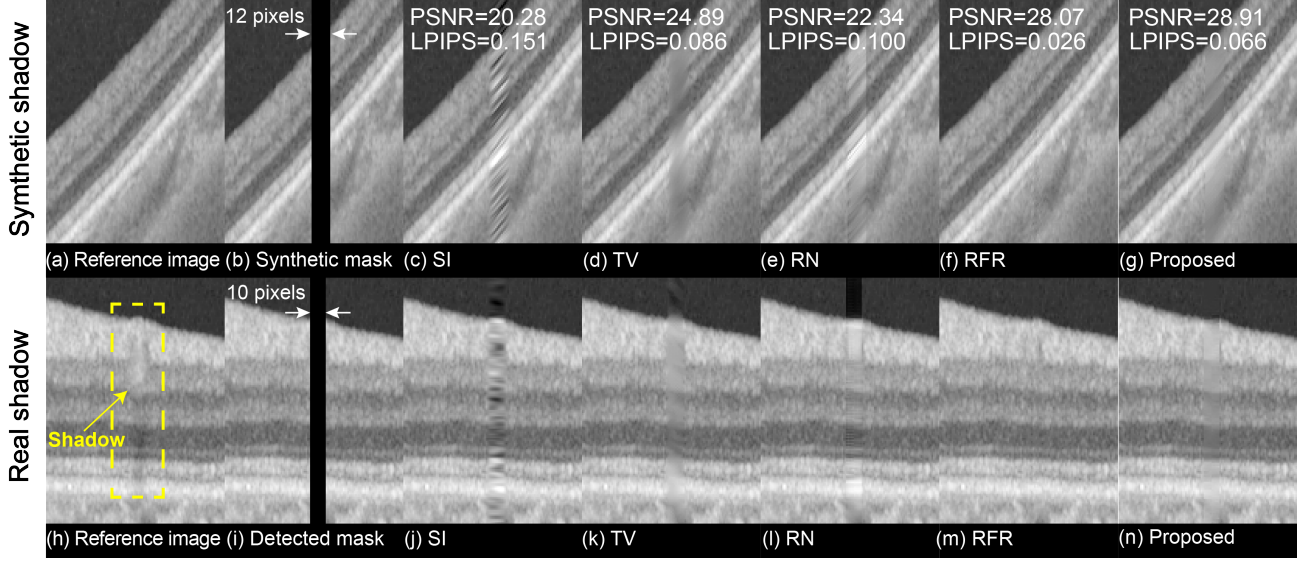
3. RESULTS

Figure 2. Visual comparison between different methods for synthetic shadow and real shadow.

A qualitative comparison between the results obtained by different techniques for synthetic and real-word shadows is shown in Figure 2. The black vertical bars shown in Figure 2(b) and (i) are the synthetic and the detected shadow mask with a shadow width of 12 pixels and 10 pixels, respectively. In Figure 2(c) and (j), the shadows inpainted by SI show severe discontinuities within each layer. On the other hand, TV generates a roughly smooth structure. However, the results still contain blurs and over-smoothing textures inside the shadow region as shown in Figure 2(d) and (k). For CNN-based RN network, the resultant patches appear more realistic and less blurry as given in Figure 2(e) and (l). However, it generates some unexpected contrast artifacts which lead to a very low PSNR and SSIM. Another deep learning-based network RFR generates the most visually pleasing results as presented in Figure 2(f) and (m) while the inpainted images manifest the lowest LPIPS as well. Nonetheless, it possesses a slightly lower PSNR than our proposed method shown in Figure 2(g) and (n). Nonetheless, the visual performance of our proposed method preserves layer continuity but is not as good as RFR. Nonetheless, the proposed method has the highest PSNR and SSIM, but a higher LPIPS than RFR.

To quantitatively evaluate the inpainting quality, the mean and standard deviation of PSNR, LPIPS and SSIM are calculated for different techniques on the synthetic testing set. Table 1 shows the results with best performance value written in bold. It can be observed that our model outperforms other methods in both PSNR and SSIM. The efficacy of the proposed method could partially attribute to the sparsity-promoting loss function in DI module and the proposed multi-scale scheme. As to the LPIPS metric, the proposed method underperforms the RFR network, which is largely consistent with the visual evaluation mentioned above.

To further evaluate the capability of the proposed technique on handling wide shadows, we repeat the inpainting experiment on various shadow widths (ranging from 7 pixels to 24 pixels) by using different techniques and calculating the corresponding PSNR, LPIPS, and SSIM. The results are plotted in Figure 3. One of the major advantages of the proposed method is that the inpainting performance possesses a relatively lower dependency on the shadow width. As shown in Figure 3, the proposed method and RFR remains almost constant for PSNR and SSIM, while the performance of other traditional methods drops dramatically as the shadow width increases

Table 1. Objective quantitative metrics and running time compared with other methods on synthetic shadows.

Metric	SI	TV	RN	RFR	Proposed
PSNR (dB)↑	21.12±2.38	24.04±4.13	17.91±2.09	29.99±0.97	**31.21±1.08**
SSIM↑	0.30±0.11	0.59±0.08	0.47±0.08	0.69±0.07	**0.79±0.06**
LPIPS↓	0.306±0.034	0.245±0.047	0.249±0.026	**0.064±0.018**	0.170±0.063
Time(Train+Test) (s)	0+139	0+6,226	71,200+8,236	128,400+977	4,206+1,203

(compared with the patch size). We suggest it is because the proposed method utilizes a multi-scale framework to handle the wide shadows. Instead of being directly inpainted, the wide shadows are first narrowed by downsampling to ensure a satisfactory inpainting performance at a reduced resolution. The following procedures including a deep learning-based upsampling and a dictionary-based regularization minimize the impairment brought by the downsampling to the resolution. Since the patch size and threshold are set to be 8 pixels, there are performance improved when the width of shadow reaches 8 pixels and 16 pixels. Furthermore, the LPIPS for RFR is the best while our method has the similar trend with traditional methods.
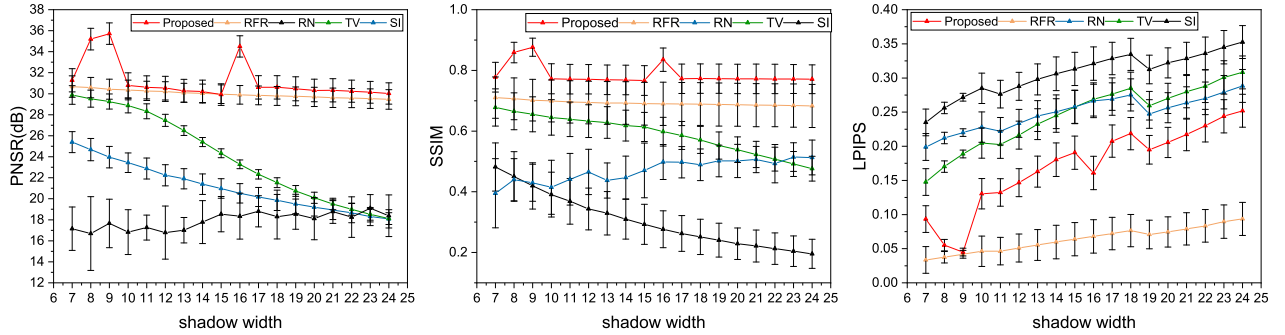

Figure 3. PSNR, SSIM and LPIPS curve of different methods over shadow width.

We further compare the effectiveness of shadow inpainting for machine diagnosis. Specifically, raw OCT images containing real shadows and the corresponding inpainted results by different methods are sent into a retinal OCT image segmentation network UNet for retinal layers segmentation. The results are provided in Figure 4 with nine layers annotated by different colors. The Dice score between ground truth and prediction is computed for the quantitative evaluation of segmentation performance. Due to the presence of the large blood vessels, the segmentation results obtained from raw images suffer from obvious distortion as shown in Figure 4. From Figure 4 (c), the poor inpainted results of SI causes segmentation results to create scattered labels and layer discontinuities emerge. Different from SI, TV and RN decrease the scattered labels but still exists bending of retinal layers influenced by less satisfactory inpainted results. As a comparison, the proposed method achieves a better Dice score and less visual error except for the choroid layer. Finally, it is worth noting that RFR achieves the best performance and has the highest Dice score, which is consistent with the above analysis.

To sum up, our proposed multi-scale framework can achieve expected inpainting results and has slightly higher PSNR and SSIM metrics, while the deep learning method RFR achieves the best visual performance and LPIPS score.

4. DISCUSSION

4.1 Time efficiency

The last row of Table 1 lists the training and testing time of different methods. Traditional methods like TV need to iteratively propagate information from the neighboring regions to the missing area, which leads to a relatively longer testing time compared to that of SI. As for deep learning-based methods, they spend most of
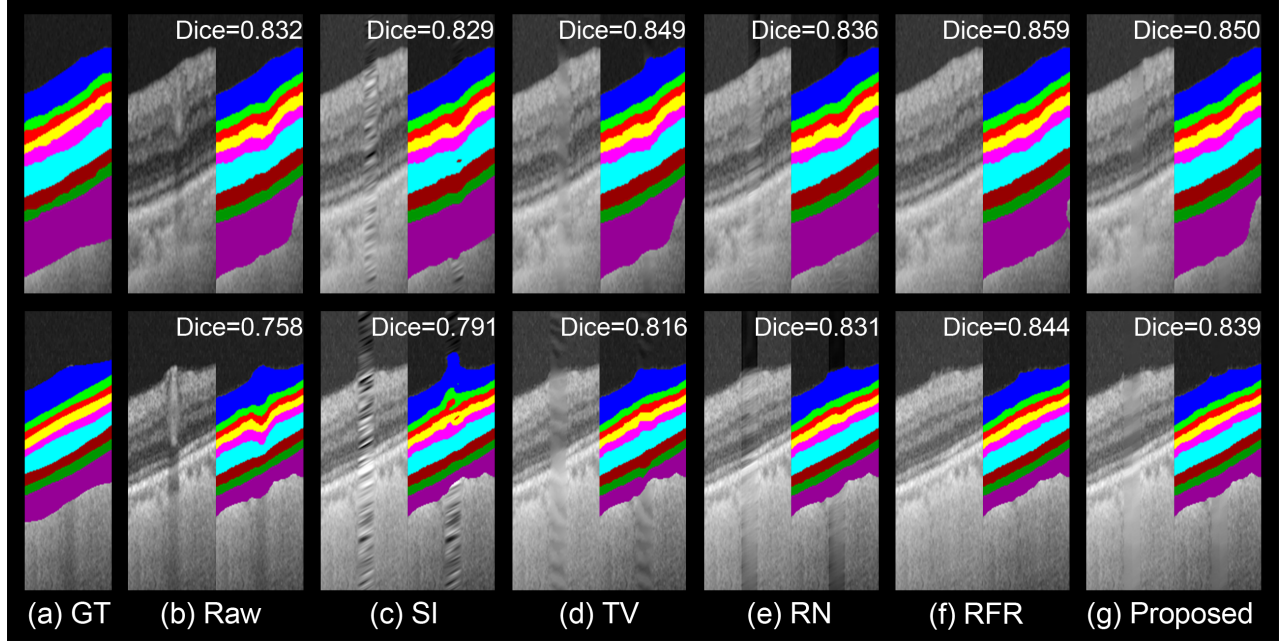
Figure 4. Retinal OCT image segmentation results before and after inpainting.

the time on training the network parameters. In contrast, the training time for the proposed method is much smaller than deep neural network, since the dictionary learning procedure does not include the computational-heavy convolution operations in the data-driven deep learning network. Besides, it should be noted that the downsample and upsample operation add additional time to the test procedure.

4.2 Performance of deep learning network

In this paper, we select two deep learning-based networks for comparison: RN and RFR. RN adopts an encoder-decoder framework for solving the mean and variance shift problem and improving the training. RFR recurrently infers the missing boundaries using the convolutional feature maps. They are both widely used *natural* image inpainting networks. When deep learning networks work properly, they present a better performance than most traditional methods and sparse representation-based methods. However, there is a chance that some deep learning networks might overfit the data and breakdown during the testing.

For RN network, the inpainting results for OCT images are not good as we expected. RN is designed and pretrained on *natural* images, which have the different characteristics of *medical* images. It might be necessary to adjust the network architecture or use a transferring network in future work.

For RFR network, the results has the most realistic visual appearance and the lowest LPIPS value, which indicates its powerful performance. RFR can capture more high-level structure features and shows promising results for images with distinct patterns. Though the RFR has a smaller PSNR, we conjecture that it captures the noise feature of coherent noise which has a Poisson distribution. Thus, in the inpainting process, RFR may add some unexpected noise into the inpainted region while our method manifests an over smoothing effect.

5. CONCLUSIONS

In conclusion, we propose a novel multi-scale sparse representation-based shadow inpainting framework for retinal OCT images. A *downsampling-inpainting-upsampling* work flow is employed to handle wide shadows cast by large vessels. Moreover, a CNN-based super-resolution network is implemented for upsampling and simultaneously enhancing spatial resolution, which minimizes the penalties imposed by the downsampling procedure. Experimental results obtained by our methods outperform other existing methods on both the synthetic and

real-world shadow inpainting tasks. In the future, this proposed framework will be extended to other medical image inpainting or reconstruction processing tasks by exploring structure characteristics in medical images through sparse representation.

ACKNOWLEDGMENTS

REFERENCES

[1] Petzold, A., Balcer, L. J., Calabresi, P. A., Costello, F., Frohman, T. C., Frohman, E. M., et al., "Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis," *Lancet Neurology* **16**(10), 797–812 (2017).

[2] Li, J., Jin, P., Zhu, J., Zou, H., Xu, X., Tang, M., Zhou, M., Gan, Y., He, J., Ling, Y., and Su, Y., "Multi-scale gcn-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary oct images," *Biomed. Opt. Express* **12**(4), 2204–2220 (2021).

[3] Elharrouss, O., Almaadeed, N., Al-Maadeed, S., and Akbari, Y., "Image inpainting: A review," *Neural Processing Letters* **51**(2), 2007–2028 (2020).

[4] Xu, Z. and Sun, J., "Image inpainting by patch propagation using patch sparsity," *IEEE Transactions on Image Processing* **19**(5), 1153–1165 (2010).

[5] Xiaoqun Zhang, T. F. C., "Wavelet inpainting by nonlocal total variation," *Inverse Problems & Imaging* **4**(1), 191–210 (2010).

[6] Girard, M., Strouthidis, N. G., Ethier, C. R., and Mari, J. M., "Shadow removal and contrast enhancement in optical coherence tomography images of the human optic nerve head," *Investigative Ophthalmology & Visual Science* **52**(10), 7738–7748 (2011).

[7] Mo, J. and Zhou, Y., "The research of image inpainting algorithm using self-adaptive group structure and sparse representation," *Cluster Computing* **22** (05 2019).

[8] Zhang, Z., Xu, Y., Yang, J., Li, X., and Zhang, D., "A survey of sparse representation: Algorithms and applications," *IEEE Access* **3**, 490–530 (2015).

[9] Hong, X., Xiong, P., Ji, R., and Fan, H., "Deep fusion network for image completion," in [*Proceedings of the 27th ACM International Conference on Multimedia*], *MM '19*, 2033–2042, Association for Computing Machinery, New York, NY, USA (2019).

[10] Li, J., Wang, N., Zhang, L., Du, B., and Tao, D., "Recurrent feature reasoning for image inpainting," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], (June 2020).

[11] Razzak, M. I., Naz, S., and Zaib, A., "Deep learning for medical image processing: Overview, challenges and the future," in [*Classification in BioApps: Automation of Decision Making*], 323–350, Springer International Publishing, Cham (2018).

[12] Fang, L., Li, S., Nie, Q., Izatt, J. A., Toth, C. A., and Farsiu, S., "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Express* **3**, 927–942 (May 2012).

[13] Liu, H., Cao, S., Ling, Y., and Gan, Y., "Inpainting for saturation artifacts in optical coherence tomography using dictionary-based sparse representation," *IEEE Photonics Journal* **13**(2), 1–10 (2021).

[14] Cleveland, W. S., "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association* **74**(368), 829–836 (1979).

[15] Zhang, Z., Xu, Y., Yang, J., Li, X., and Zhang, D., "A survey of sparse representation: Algorithms and applications," *IEEE Access* **3**, 490–530 (2015).

[16] Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M., "Enhanced deep residual networks for single image super-resolution," in [*2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*], 1132–1140 (2017).

[17] Li, M., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., and Chen, Q., "IPN-V2 and OCTA-500: Methodology and dataset for retinal image segmentation," (2020).

[18] Getreuer, P., "Total Variation Inpainting using Split Bregman," *Image Processing On Line* **2**, 147–157 (2012).

[19] Yu, T., Guo, Z., Jin, X., Wu, S., Chen, Z., Li, W., Zhang, Z., and Liu, S., "Region normalization for image inpainting," in [*Proceedings of the AAAI Conference on Artificial Intelligence*], **34**(07), 12733–12740 (2020).

[20] Doersch, C., Singh, S., Gupta, A., Sivic, J., and Efros, A. A., "What makes paris look like paris?," *ACM Trans. Graph.* **31** (jul 2012).

[21] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*], Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., eds., 234–241, Springer International Publishing, Cham (2015).