Multimedia Systems
https://doi.org/10.1007/s00530-017-0582-5

SPECIAL ISSUE PAPER

Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior

Xiaoming Xi[1] · Xianjing Meng[1] · Lu Yang[1] · Xiushan Nie[1] · Gongping Yang[2] · Haoyu Chen[3] · Xin Fan[4] · Yilong Yin[2] · Xinjian Chen[5]

Abstract
Automated segmentation of choroidal neovascularization (CNV) in optical coherence tomography (OCT) images plays an important role for the treatment of CNV disease. This paper proposes multi-scale convolutional neural networks with structure prior to segment CNV from OCT data. The proposed framework consists of two stages. In the first stage, the structure prior learning method based on sparse representation-based classification and the local potential function is developed to capture the global spatial structure and local similarity structure prior. The obtained prior can be used to improve the distinctiveness between CNV and background patches. In the second stage, multi-scale CNN model with incorporation of the learned structure prior is constructed for CNV segmentation. In this stage, multi-scale analysis is used to capture effective contextual information, which is robust to varying sizes of CNV. The proposed method was evaluated on 15 spectral domain OCT data with CNV. The experimental results demonstrate the effectiveness of proposed method.

Keywords Choroidal neovascularization (CNV) · Optical coherence tomography (OCT) · Segmentation · Structure prior · Convolutional neural networks (CNN)

1 Introduction

Wet AMD is most likely to cause visual loss. It is characterized by choroidal neovascularization (CNV), in which new blood vessels form and break beneath the retina. This leakage causes permanent damage to surrounding retinal tissues, distorting and destroying central vision [1].

Optical coherence tomography (OCT) has been widely employed for the evaluation of age-related macular degeneration (AMD) [2, 3]. It enables visualization of subretinal fluid, intraretinal fluid, retinal pigment epithelial detachments (RPEDs), and retinal thickening by using cross-sectional B-scans. CNV may appear on structural OCT B-scans as subretinal or sub-RPE hyperreflective material, or both of them [4]. Compared with other imaging modalities, such as fluorescein angiography (FA), indocyanine green angiography (ICGA), OCT has the following advantages [5]: (1) it is noninvasive; (2) it allows high-resolution cross-sectional images of the neurosensory retina to be obtained; (3) it has a higher speed.

Quantification of CNV would be useful to clinicians in the diagnosis of CNV disease [6]. In order to quantify the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00530-017-0582-5) contains supplementary material, which is available to authorized users.

✉ Yilong Yin
ylyin@sdu.edu.cn

✉ Xinjian Chen
xjchen@suda.edu.cn

1 School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China

2 School of Computer Science and Technology, Shandong University, Jinan, China

3 Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, Shantou, China

4 Taian Institute of Science and Technology Information, Taian, China

5 School of Electronic and Information Engineering, Soochow University, Suzhou, China

Published online: 25 January 2018

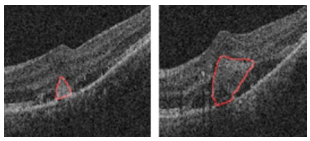

Fig. 1 Two slice image examples with CNV

CNV lesion, the lesion should be first delineated manually. However, manual delineation is subjective with observer variability and time-consuming [7]. Therefore, there is a requirement to develop a tool for automatic segmentation of CNV.

CNV segmentation in OCT images is a challenging task due to the complicated characteristics of CNV. Figure 1 shows two OCT image slices with CNV. We can observe that the CNV is a complex object with varied texture, size and irregular shape. In addition, intensity inhomogeneity and blurring boundaries also appear in CNV region. In OCT images, large numbers of noises also exist. Therefore, it is difficult to obtain accurate segmentation results by using traditional segmentation methods.

In recent years, few CNV segmentation methods were proposed. However, these methods were performed on OCT angiography images [6, 8] or FA images [9, 10].

Deep learning has achieved a significant success in computer vision due to its powerful learning ability. Recently, it has drawn increasing attention from medical image analysis community [11–17]. Xu et al. presented a stacked sparse autoencoder for efficient nuclei detection on high-resolution histopathological images of breast cancer [11]. Van Tulder et al. trained restricted Boltzmann machine with a generative learning objective for lung texture classification and airway detection in CT images [12]. As a classic architecture of deep neural networks, convolutional neural networks (CNNs) may be more suitable for image segmentation or classification task. Korsuk et al. proposed a spatially constrained convolutional neural network (SC-CNN) to perform nucleus detection [13]. A multi-view CNN was proposed for pulmonary nodule detection in CT images [14]. To obtain the multi-scale information about each voxel, multiple CNNs were trained based on 2D image patches with different sizes for segmentation of MR brain images [15]. In order to use multi-modality information of MR images, Zhang et al. employed CNN for segmenting isointense stage brain tissues of multi-modality MR images [16]. To obtain the training instances, 2D patches from T1, T2, and fractional anisotropy (FA) images were generated. Based on these training instances, a CNN model was trained for each modality. The final segmentation result was the combination of outputs of three CNNs. Ghesu et al. combined deep learning and marginal space learning for object detection and segmentation on a large dataset [17]. The related works have demonstrated that deep learning is an effective framework for medical images analysis. Therefore, we attempt to employ this framework for CNV segmentation in OCT images.

However, two problems arise when using CNN directly for CNV segmentation. On one hand, training patches are regarded as independent instances without considering the relationship among them, which may limit performance improvement. In OCT images, complicated characteristics of CNV may result in overlapping distributions of training patches. We use a similarity matrix to reveal the distributions of CNV and background patches approximately, as is shown in Fig. 2. In this figure, C_i denotes the CNV class from the *i*th patient while B_i denotes the background class from the *i*th patient. For each class, a center is generated by calculating the average of patches in this class. And the similarity matrix is obtained by calculating the similarity between two arbitrary centers. Euclidean distance is used as the similarity measure. In this matrix, small values indicate that two corresponding classes are similar while large values indicate that two corresponding classes are different. As shown in Fig. 2, we can infer that large inter-class similarity and intra-class differences exist in the generated patches. Based on confusing training instances, it is difficult to learn an effective CNN model without employing more useful information.

Generally speaking, images have fixed structure, and their pixels exhibit certain dependencies [18]. From the view point of the local structure, the elements in a small local region should be more similar [18]. From the point view of the global spatial structure, the elements in the same objects should be more similar than the elements in different objects. The local structure can lead to intra-class

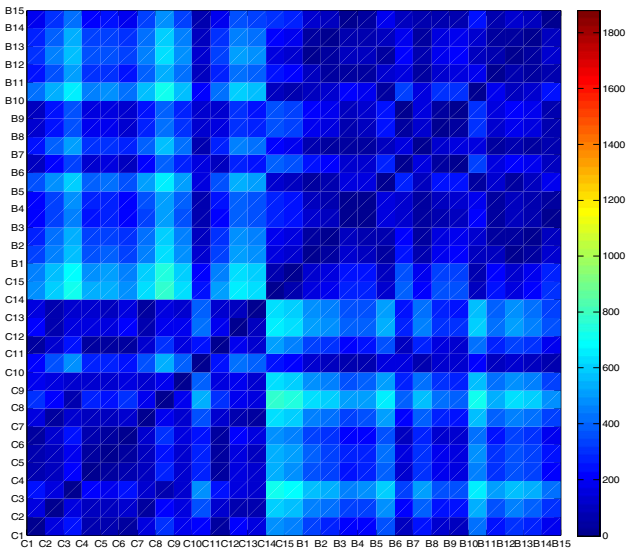

Fig. 2 Similarity matrix of patches extracted from OCT images in our database

similarity while the global structure can lead to inter-class difference. Liu et al. explicitly modeled the structure information and achieved state-of-the-art performance in their task [19]. Based on this idea, the structure information may be exploited to improve the segmentation performance of CNN.

On the other hand, traditional CNN is trained based on patches with a single size. However, size of CNV is varied (shown in Fig. 1). CNV with different size has different needs in terms of context and scale information. Unfortunately, effective contextual information may not be captured in patches with a single size, resulting in performance limitation of the CNN model.

To solve these two issues, this paper proposes a CNV segmentation method using multi-scale CNN with structure prior (MS-CNN-SP). The proposed segmentation framework consists of the following steps: (1) structure prior learning. First, a global structure prior learning model based on SRC is employed to obtain global spatial prior. The prior can reflect spatial location of CNV. Based on the learned global structure, a local potential function is developed to calculate the structure prior matrix which contains global spatial prior and local similarity prior. After that, the original image is transformed based on the structure prior matrix. In the transformed image, saliency of CNV is enhanced due to the introduction of effective structure prior information. Therefore, we refer to the transformed image as saliency-enhanced image in this paper. For the learned structure prior, the global structure prior is used to locate the coarse spatial position of CNV, which can reduce the similarity between CNV and background patches. While local structure prior is utilized to preserve the similarity between pixels from a small local region, result in large intra-class similarity. Therefore, the distinctiveness between CNV and background patches can be improved in the saliency-enhanced images. (2) Segmentation model training. Multi-scale analysis can be used to capture effective contextual information [20]. In order to utilize effective contextual information at different scales, multi-scale CNN model is developed. In this paper, we refer to multiple CNNs trained based on patches with different size as multi-scale CNN. The structure prior is incorporated in multi-scale CNN by using the saliency-enhanced images as the training images. The experimental results on our database demonstrate the effectiveness of the proposed method.

The contributions of this paper are summarized as follows: (1) the structure prior learning method based on SRC and local potential function is proposed to learn global spatial prior and local similarity prior, which can improve distinctiveness between CNV and background classes; (2) structure prior integrated multi-scale CNN is constructed to segment CNV in OCT images and achieves good performance on our database.

2 Method

2.1 Method overview

Figure 3 shows the framework of the proposed segmentation method. The proposed framework consists of training and testing stages. In the training stage, training images are first segmented into superpixels, and then intensity, texture and local information features are extracted for each superpixel. After that, global spatial structure is learned based on superpixels and SRC. In the learned global structure prior, spatial location of CNV can be detected. Based on the learned global structure, the local potential function is developed to calculate the structure prior matrix. After that, original images are transformed into the saliency-enhanced images based on the structure prior matrix. In order to utilize structure prior information and capture effective contextual information at different scales, training patches with different sizes are extracted from saliency-enhanced images and are used to train MS-CNN-SP models.

In the testing stage, for a test image, as same as the training stage, superpixels are generated and same features are extracted for each superpixel, and then global spatial structure is learned based on SRC. After that, structure prior matrix is calculated for saliency-enhanced image transformation. Based on the saliency-enhanced image, patches with different sizes are generated and inputted into MS-CNN-SP models. Finally, the segmentation result is obtained via fusion of segmentation results of MS-CNN-SP models.

2.2 Structure prior learning

In order to learn effective global spatial structure prior of CNV, superpixel is chosen as the elementary processing unit. After superpixel extraction, the same feature extraction method [21] is used to extract the intensity, texture and

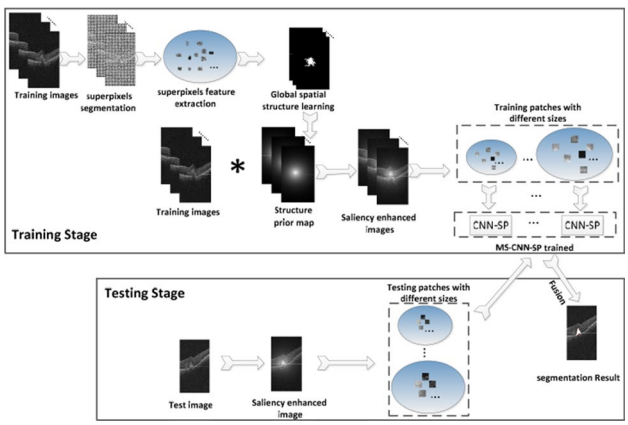

Fig. 3 Flow chart of the proposed method

local features for each superpixel. More details about feature extraction can be found in [21].

After feature extraction, global spatial structure prior of CNV is obtained in terms of superpixel classification results. In the recent years, sparse representation has been applied in medical image processing, such as image segmentation [22], feature selection [23], disease diagnose [24], and achieving promising results due to its robustness. In this paper, in order to deal with complex characteristics such as varying intensity, texture in CNV region, SRC is employed for superpixel classification. For a superpixel, sparse representation selects the atoms in the dictionary which most compactly express the input superpixel and rejects all other possible but less compact representations. As a result, similar atoms contribute more to the final superpixel classification, which is robust to the varying characteristics of CNV.

In this paper, the dictionary is constructed by using K-means. The superpixels from each patient are classified into two classes: CNV and background. K-means is used to generate centers to represent the complicated characteristics of each class. Therefore, $2K$ centers are obtained for K patients.

After dictionary construction, SRC is used for superpixel classification. The classification result can reveal the global spatial location of CNV. After that, a local potential function is developed to calculate structure prior matrix, as listed in Eq. (1):

$$M(i,j) = \mathrm{e}^{-\left(\frac{(\boldsymbol{cor}(i,j)-\boldsymbol{c})^2}{\sigma^2}\right)}. \tag{1}$$

In above equation, M is the structure prior matrix, $\boldsymbol{c}$ is the centroid of detected CNV while $\boldsymbol{cor}(\boldsymbol{i},\boldsymbol{j})$ denotes the coordinate of the pixel (i,j) in the image. σ can be regarded as the radius of CNV approximately. It can be obtained by calculating mean of distances between centroid and boundary points of detected CNV. In the calculated structure prior matrix, the elements in a small local region are similar because the distances between the elements and centroid are similar. This may ensure the intra-class similarity. According to the global spatial structure learned, values of elements in CNV region are larger because the distances between the CNV pixels and centroid of CNV are smaller, while the values of background elements are smaller due to the large distance between background pixels and centroid of CNV region. This can guarantee the difference between CNV and background pixels.

The obtained structure prior information is introduced into the original images and the saliency-enhanced image is transformed based on structure prior matrix according to Eq. (2):

$$I_s = MI_0. \tag{2}$$

In Eq. (2), M is the structure prior matrix, I_0 denotes the original image and I_s denotes saliency-enhanced image. Figure 4 shows several original images and corresponding saliency-enhanced images. As shown in this figure, CNV saliency is enhanced after image transformation.

2.3 CNV segmentation using multi-scale CNN with structure prior

In traditional segmentation task based on CNN, patches are extracted from training images as the training set. Information about each pixel is provided in the form of image patches where the pixel is in the center. Labels of the patches are as the same as the central pixels. The patches whose central pixels are in object regions belong to positive class, while patches whose central pixels are in the background belong to negative class. For a test image, all pixels of images are classified by using trained CNN, and object is segmented according to the pixel classification results.

In this paper, we extract the training patches from saliency-enhanced images, and CNN models are trained based on these training patches. Original OCT images are transformed into saliency-enhanced images by introducing structure prior. Therefore, patches which are extracted from saliency-enhanced images contain the structure prior. In the training stage, CNN is trained based on these training patches. We infer that the structure prior can be incorporated in the learned CNN.

CNV with different sizes has different needs in terms of context and scale information. Therefore, single scale is difficult to capture effective contexture information of CNV because size of CNV is varied. Considering complementary information on different scales may be robust to scale variations [25, 26], multi-scale analysis is employed to capture contextual information at different scales by varying the sizes of the patches. CNN models are trained on patches with different sizes, which can learn contextual information at

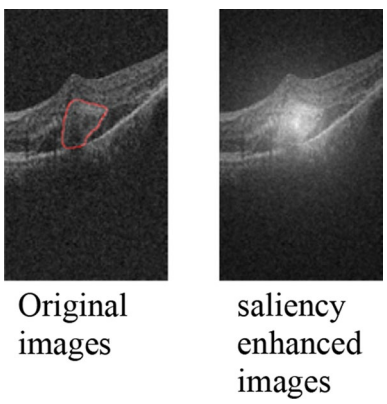

Fig. 4 Comparison of original images and saliency-enhanced images

different scales. The multi-scale CNN combines the segmentation results of CNN trained on different scales. Therefore, a multi-scale convolutional neural network is fit to deal with the problem of varied size of CNV. Based on patches with different sizes, MS-CNN-SP models are trained. Finally, the segmentation result is the fusion of the classification results of multi-scale CNN-SP models. In this paper, majority vote method is employed for segment result fusion. In our task, we train five CNN-SP models on patches with five different scales. For a pixel, if it is predicted as CNV by three or more than three CNN-SP models, it belongs to CNV in the final segmentation result.

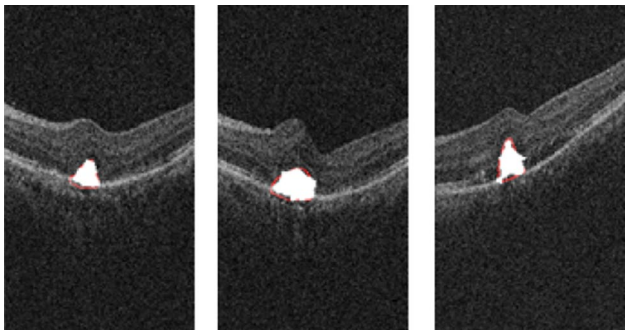

Fig. 5 CNV segmentation results of several slice images

3 Experimental results

3.1 Evaluation metrics

SD-OCT scans of 15 eyes diagnosed with CNV were acquired using Topcon 3D-OCT-1000 (Topcon Corporation, Tokyo, Japan). Each SD-OCT volume contains $512 \times 1024 \times 128$ voxels. This study was approved by the Intuitional review board of Joint Shantou International Eye Center and adhered to the tenets of the Declaration of Helsinki. Because of its retrospective nature, informed consent was not required from subjects. The ground truth of CNV region in all B-scans is manually delineated by retinal specialists.

To evaluate the performance of the proposed method, performance metrics such as Dice similarity coefficient (DSC), true positive volume fraction (TPVF) and false positive volume fraction (FPVF) were used as performance indices. The Dice similarity coefficient was used to measure the accuracy of the automatic segmentation result as compared against reference standard delineation; TPVF indicates the fraction of the total amount of CNV in the true segmentation by the proposed method; FPVF denotes the amount of CNV falsely identified by the proposed method. They are calculated as follows:

$$\mathrm{DSC} = 2 \times \frac{|V_\mathrm{A} \cap V_\mathrm{M}|}{|V_\mathrm{A} \cup V_\mathrm{M}|},$$

$$\mathrm{TPVF} = \frac{|V_\mathrm{A} \cap V_\mathrm{M}|}{|V_\mathrm{M}|},$$

$$\mathrm{FPVF} = \frac{|V_\mathrm{A}| - |V_\mathrm{A} \cap V_\mathrm{M}|}{|V - V_\mathrm{M}|},$$

where $|\cdot|$ denotes volume, V_A denotes the CNV region segmented by the proposed method, V_M denotes the CNV region delineated by retinal specialist, V denotes the total volume of the OCT data.

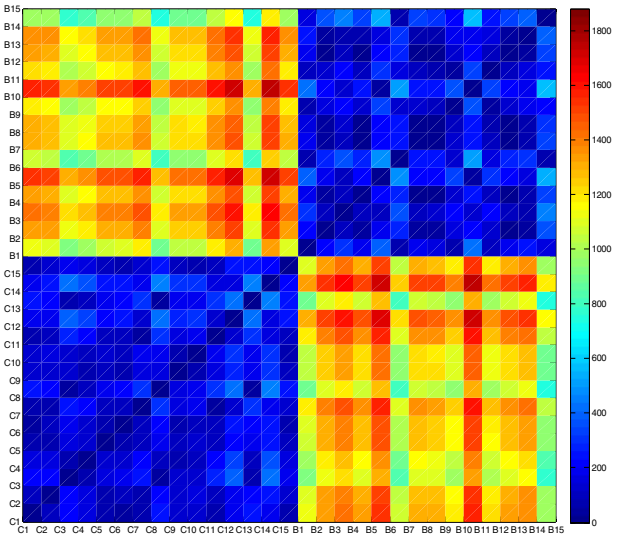

Fig. 6 Similarity matrix of saliency-enhanced images

3.2 Experiment settings

Caffe [27] is implemented in our experiment and Alexnet is used as the training networks. Patches with sizes of 13×13, 15×15, 17×17, 25×25, 35×35 are extracted, respectively. In the experiment, we denote CNN with structure prior trained based on patches with different sizes as 13-CNN-SP, 15-CNN-SP, 17-CNN-SP, 25-CNN-SP, and 35-CNN-SP, respectively.

3.3 Effectiveness of structure prior evaluation

In this experiment, we compare multi-scale CNN (MS-CNN) and MS-CNN-SP to demonstrate the effectiveness of structure prior. Figure 5 gives segmentation result of several slice image examples.

Figure 6 shows the similarity matrix of saliency-enhanced images. Compared with distributions of original images (shown in Fig. 2), distinctiveness between CNV

and background is improved after saliency-enhanced image transformation. For example, for original images, distance between C1 and B9 is smaller than distance between C1 and C3. The difference between inter-classes is smaller than difference between intra-class, which may result in performance degradation of trained CNN. On the contrary, for saliency-enhanced images, the difference between different classes is enlarged while the variance of the same class is reduced due to introduction of the structure prior information, resulting in performance improvement.

Table 1 lists the performance of MS-CNN and MS-CNN-SP. As observed in this table, MS-CNN-SP outperforms MS-CNN significantly, especially for DSC, about 28 percentage points are increased. The reason is that the learned structure prior can capture global spatial information about CNV and preserve the local similarity of pixels. The global structure prior can reduce the similarity between CNV and background patches while local structure prior can lead to large similarity of intra-class patches. Therefore, the learned structure prior is useful to improve the distinctiveness between the patches extracted from CNV region and background. MS-CNN-SP can learn the discriminative structure information from saliency-enhanced images. Therefore, MS-CNN-SP is more effective than MS-CNN.

Table 1 CNV segmentation result, compared with MS-CNN

	MS-CNN-SP	MS-CNN	*p* values
DSC	0.7806 ± 0.067	0.5092 ± 0.074	0.005
TPVF	0.8024 ± 0.081	0.7862 ± 0.045	0.066
FPVF	0.0036 ± 0.001	0.0183 ± 0.005	0.003

3.4 Effectiveness of multi-scale analysis evaluation

Figure 7 shows the average DSC, TPVF, and FPVF of CNN-SP at different scales. The performance of trained CNN-SP is different according to variance of scales. DSC of CNN-SP with size 25×25 is about 0.76, which is better than other CNN-SP models trained based on different scales. However, the DSC of MS-CNN-SP is about 0.78, which is improved by about 2% via fusion of multi-scale information.

From this experiment, we can observe that MS-CNN-SP outperforms CNN-SP at different scales. Since the size of CNV is varied, it is difficult to capture effective contextual information by using single scale. However, multi-scale

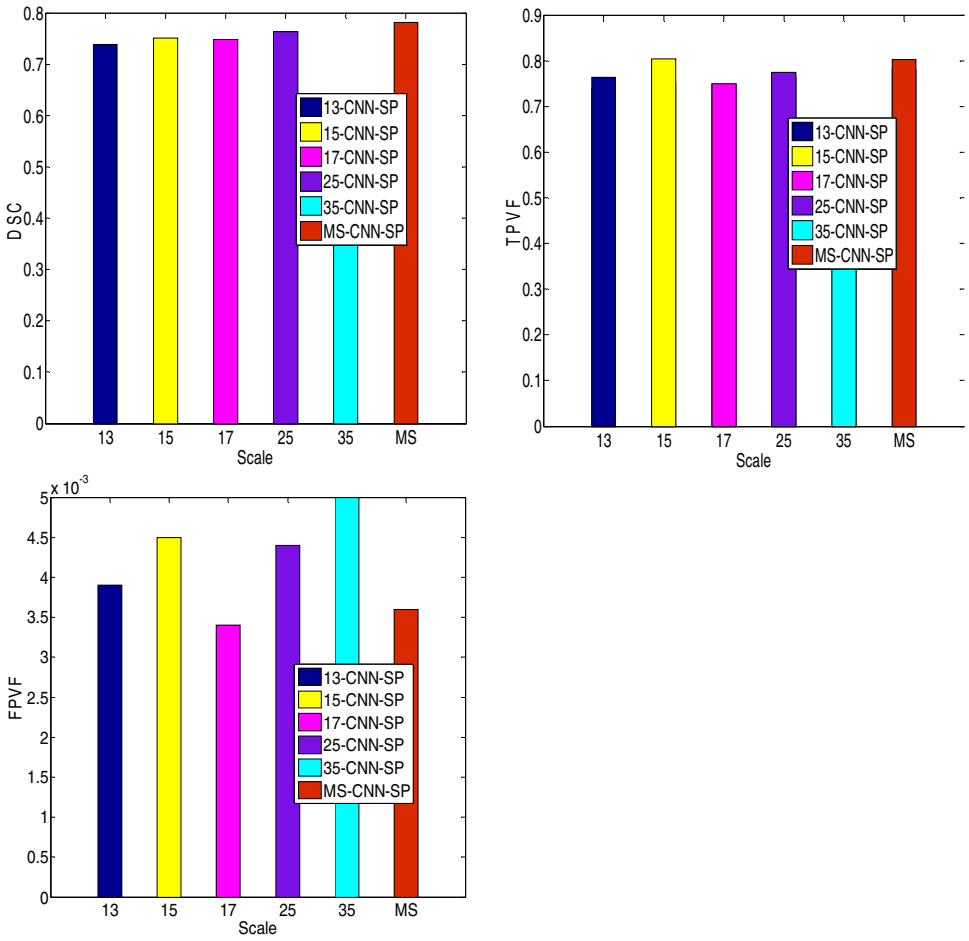

Fig. 7 Performance of CNN-SP at different scales

analysis can combine complementary information between different scales, which is robust to scale variations. Therefore, MS-CNN-SP can boost the performance.

4 Conclusion

In this paper, we propose a CNV segmentation method using multi-scale CNN with structure prior. In the proposed framework, structure prior learning based on SRC and local potential function is first developed to capture the global spatial structure and local similarity structure prior, which can lead to large inter-class difference and intra-class similarity. Considering that complementary information between different scales may be robust to CNV size variation, multi-scale analysis is employed to capture effective contextual information at different scales. Finally, multi-scale CNN with the incorporation of learned structure prior is constructed to segment CNV in OCT images. The experimental results demonstrate the effectiveness of proposed method.

The proposed method improves segmentation performance due to the introduction of structure prior and multi-scale information. However, the proposed algorithm is 2-D, ignoring useful information between slices which is important for CNV segmentation on 3D-OCT data. Therefore, our future work will focus on further improving segmentation performance by extending our method to 3D CNN framework.

Acknowledgements This work is supported by Natural Science Foundation of China (61701280), Natural Science Foundation of Shandong Province (ZR2016FQ18, ZR2017QF009), National Basic Research Program of China (973 Program) under Grant 2014CB748600, National Science Fund for Outstanding Young Scholars (61622114), Natural Science Foundation of China (61573219, 81371629, 61671274, 61703235, 61701281), the Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions, Shandong Provincial Key Research and Development Plan (Grant no. 2017CXGC1504). The Fostering Project of Dominant Displine and Talent Team of SDUFE. The authors would like to greatly thank the editors and the reviewers for their valuable comments and suggestions.

References

1. Dewan, A., Liu, M., Hartman, S., Zhang, S.S., Liu, D.T., Zhao, C., Tam, P.O., Chan, W.M., Lam, D.S., Snyder, M., et al.: HTRA1 promoter polymorphism in wet age-related macular degeneration. Science. **314**, 5801, 989–992 (2006)
2. Chen, X., Zhang, L., Sohn, E.H., Lee, K., Niemeijer, M., Chen, J., Sonka, M., Abramoff, M.D.: Quantification of external limiting membrane disruption caused by diabetic macular edema from SD-OCT. Invest. Ophthalmol. Vis. Sci. **53**(13), 8042–8048 (2012)
3. Chen, X., Niemeijer, M., Zhang, L., Lee, K., Abràmoff, M.D., Sonka, M.: 3D segmentation of fluid-associated abnormalities in retinal OCT: probability constrained graph–search–graph-cut. IEEE Trans. Med. Imaging. **31**(8), 1521–1531 (2012)
4. de Carlo, T.E., Bonini Filho, M.A., Chin, A.T., Adhi, M., Ferrara, D., Baumal, C.R., Witkin, A.J., Reichel, E., Duker, J.S., Waheed, N.K.: Spectral-domain optical coherence tomography angiography of choroidal neovascularization. Ophthalmology. **122**, 6, 1228–1238 (2015)
5. Shi, F., Chen, X., Zhao, H., Zhu, W., Xiang, D., Gao, E., Sonka, M., Chen, H.: Automated 3-D retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments. IEEE Trans. Med. Imaging **34**, 2, 441–452 (2015)
6. Liu, L., Gao, S.S., Bailey, S.T., Huang, D., Li, D., Jia, Y.: Automated choroidal neovascularization detection algorithm for optical coherence tomography angiography. Biomed. Opt. Express. **6**, 9, 3564–3576 (2015)
7. Zhang, J., Gao, Y., Gao, Y., Brent, M., Shen, D.: Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. IEEE Trans. Med. Imaging. **35**(12), 2524–2533 (2016)
8. Gao, S.S., Liu, L., Bailey, S.T., Flaxel, C.J., Huang, D., Li, D., Jia, Y.: Quantification of choroidal neovascularization vessel length using optical coherence tomography angiography. J. Biomed. Opt. **21**(7), 076010–076010 (2016)
9. Abdelmoula, W., Shah, S., Fahmy, A.S.: Segmentation of choroidal neovascularization in fundus fluorescein angiograms. IEEE Trans. Biomed. Eng. **60**(5), 1439–1445 (2013)
10. Tsai, C.L., Yang, Y.L., Chen, S.J., Lin, K.S., Chan, C.H., Lin, W.Y.: Automatic characterization of classic choroidal neovascularization by using adaboost for supervised learning. Invest. Ophthalmol. Vis. Sci. **52**(5), 2767–2774 (2011)
11. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., Madabhushi, A.: Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Trans. Med. Imaging. **35**(1), 119–130 (2016)
12. van Tulder, G., de Bruijne, M.: Combining generative and discriminative representation learning in convolutional restricted Boltzmann machines. IEEE Trans. Med. Imaging, **35**, 5, 1262–1272 (2016)
13. Sirinukunwattana, K., Raza, S., Tsang, Y.W., Snead, D., Cree, I., Rajpoot, N.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans. Med. Imaging. **35**(5), 1196–1206 (2016)
14. Setio, A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S., Winkler Wille, M., Naqibullah, M., Sanchez, C., van Ginneken, B.: Pulmonary nodule detection in CT images using multiview convolutional networks. IEEE Trans. Med. Imaging. **35**(5), 1160–1169 (2016)
15. Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J., Isgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans. Med. Imaging. **35**(5), 1252–1261 (2016)
16. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D.: Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage **108**, 214–224 (2015)
17. Ghesu, F., Krubasik, E., Georgescu, B., Singh, V., Zheng, Y., Hornegger, J., Comaniciu, D.: Marginal space deep learning: efficient architecture for volumetric image parsing. IEEE Trans. Med. Imaging. **35**(5), 1217–1228 (2016)
18. 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)
19. Liu, M., Zhang, D., Shen, D.: Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. IEEE Trans. Med. Imaging. **35**(6), 1463–1474 (2016)
20. Zhang, J., Gao, Y., Wang, L., Tang, Z., Xia, J.J.: D Shen. Automatic craniomaxillofacial landmark digitization via segmentation-guided partially-joint regression forest model and multi-scale

statistical features. IEEE Trans. Biomed. Eng. **63**, 9, 1820–1829 (2016)
21. Xi, X., Shi, H., Han, L., Wang, T., Ding, H.Y., Zhang, G., Tang, Y., Ying, Y.: Breast tumor segmentation with prior knowledge learning. Neurocomputing **237**, 145–149, (2017)
22. Wang, L., Chen, K.C., Gao, Y., Shi, F., Liao, S., Li, G., Shen, S.G., Yan, J., Lee, P.K., Chow, B., Liu, N.X., Xia, J.J., Shen, D.: Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization. Med. Phys. **41**, 4, 6372–6387 (2014)
23. Liu, M., Zhang, D.: Pairwise constraint-guided sparse learning for feature selection. IEEE Trans. Cybern. **46**, 1, 298–310, (2016)
24. Liu, M., Zhang, J., Yap, P.-T., Shen, D.: View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data. Med. Image Anal. **36**, 2, 123–134 (2017)
25. Alvarez, J.M., LeCun, Y., Gevers, T., Lopez, A.M.: Semantic road segmentation via multi-scale ensembles of learned features. European conference on computer vision workshop, pp. 586–595 (2012)
26. Yi, D., Lei, Z., Li, S.Z.: Age estimation by multi-scale convolutional network. Asian conference on computer vision, pp. 144–158 (2014)
27. Jia, Y.C.: An open source convolutional architecture for fast feature embedding. http://caffe.berkeleyvision.org/ (2016)