2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

DEEP CONTEXTUAL RESIDUAL NETWORK FOR ELECTRON MICROSCOPY IMAGE SEGMENTATION IN CONNECTOMICS

Chi Xiao[1,2], *Jing Liu*[1,2], *Xi Chen*[1], *Hua Han*[1,2,3], *Chang Shu*[1], *Qiwei Xie*[1]

[1] Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] School of Future Technology, University of Chinese Academy of Sciences, Beijing, China
[3] The Center for Excellence in Brain Science and Intelligence Technology, CAS, Shanghai, China

ABSTRACT

The goal of connectomics research is to manifest the mechanisms and functions of neural system by using electron microscopy (EM). One of the biggest challenges in connectomic reconstruction is developing reliable neuronal membranes segmentation method to reduce the burden on manual neurite labeling and validation. In this paper, we put forward an effective deep learning approach to realize neuronal membranes segmentation in EM image stacks, which utilizes spatially efficient residual network and multilevel representations of contextual cues to achieve accurate segmentation performance. Furthermore, multicut is used as post-processing to optimize the outputs of network. Experimental results on the public dataset of ISBI 2012 EM Segmentation Challenge demonstrate the effectiveness of our approach in neuronal membranes segmentation. Our method now ranks top 3 among 88 teams and yields 0.98356 Rand Score as well as 0.99063 Information Score, which outperforms most of state-of-the-art methods.

Index Terms— Connectomics, Deep Learning, Image Segmentation, Electron Microscopy

1. INTRODUCTION

The nervous system is a complex network composed of a large number of neurons, thus the study of the nervous system and its functions requires high-quality connectomics community information [1] [2]. In previous research, Takemura *et al.* [3] developed a semi-automated pipeline employing serial section transmission electron microscopy (ssTEM) to realize connectomic reconstruction in Drosophila optic medulla, the reconstruction results included 379 neurons and 8,637 synaptic contacts. Owing to the manual labeling and validation of neuronal structures, this research consumed about 15,880 human hours (containing 1,700 expert hours). In spite of this tremendous effort, the reconstructed volume was only 37 μm $\times$ 37 μm $\times$ 70 μm. Without the automation, the reconstruction of large volume (1 mm $\times$ 1 mm $\times$ 1 mm) would require $10,000\times$ more human efforts. In order to reduce the manual workload, it is critical to improve the accuracy and effectiveness of the neuronal structures segmentation and reconstruction.

To accelerate the research in automating the segmentation and reconstruction of neuronal structures, IEEE International Symposium on Biomedical Imaging (ISBI) launched a challenge for segmenting neuronal structures in EM image stacks [4]. In this challenge, a full stack of ssTEM images were provided to train machine learning algorithms for automated neuronal structures segmentation [5]. Participants could submit their segmentation results online, and then organizers measured the performance of the submissions in terms of topological metrics [6].

This challenge has attracted the majority of related research teams to participate. Ciresan *et al.* [7] utilized deep neural network as pixel classifier, which predicted the label of each pixel from the square window surrounding it. It was one of the earliest deep learning applications in EM image segmentation and won first place in this challenge. Whereas, the problems of this network were obvious, which included the selection of window size and its redundant computation. In what follows, a novel neural network, namely fully convolutional networks (FCN) [8], was proposed to solve end-to-end semantic segmentation problem for its conciseness and effectiveness. Motivated by this method, many successive variants of FCN have been proposed for EM image segmentation. Ronneberger *et al.* [9] presented a symmetric FCN network, termed as U-Net, which used skip connections to assemble the localized features and abstractive features. To overcome the segmentation difficulty of diverse neuronal structures, Chen *et al.* [10] proposed a contextual network to combine multilevel contextual information. However, above methods were limited in their shallow architectures. As a result, Quan *et al.* [11] combined U-Net with residual blocks to build a much deeper network for EM image segmentation, and then applied summation-based skip connections to obtain more accurate results. Whereas, the network might suffer from the vanishing

This research was supported by National Science Foundation of China (No. 61673381, No. 61201050, No. 61306070, No. 31472001), Special Program of Beijing Municipal Science and Technology Commission (No. Z161100000216146), Scientific Research Instrument and Equipment Development Project of Chinese Academy of Sciences (No. YZ201671) and Strategic Priority Research Program of the CAS (No. XDB02060001).

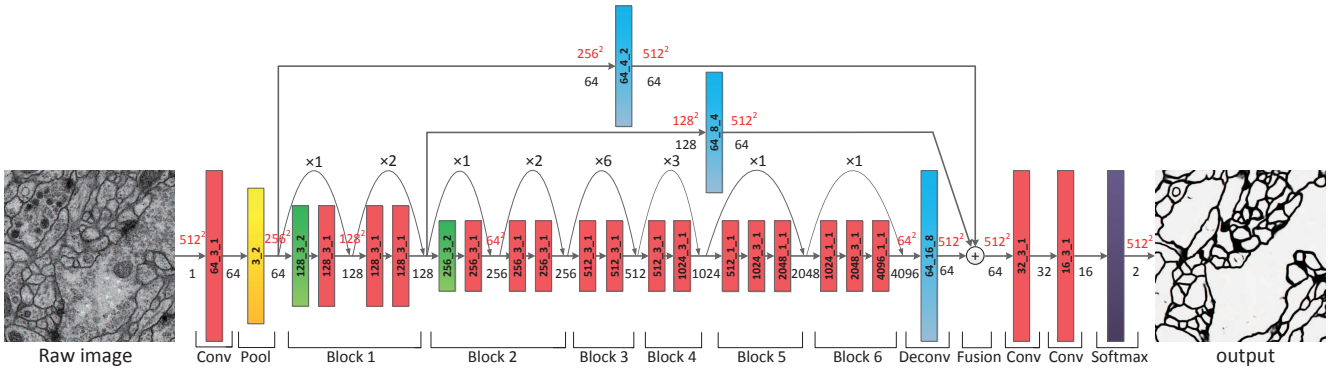


Fig. 1. The proposed network architecture. Red and green blocks annotated *M_N_S* represent convolutional layers with channels *M*, kernel size $N \times N$ and stride *S*; yellow blocks noted *N_S* imply maxpooling over $N \times N$ patches with stride *S*; blue blocks with *M_N_S* denote deconvolution layers and the parameters are similar to that of convolutional layers; purple box indicates softmax layer; the red numbers above straight arrows imply the size of feature maps, while numbers below straight arrows imply the channels of feature maps, and numbers above curved arrows represent the repetitions of residual units.

gradient problem in virtue of its complex structures. On basis of FCN segmentation, Beier *et al.* [12] used lifted multicut as post-processing method to refine the outputs of network, which greatly improved the effect of neuronal membranes segmentation. Overall, a state-of-the-art neuronal membranes segmentation algorithm requires deep and effective network to avoid the vanishing gradient problem, multilevel contextual information to differentiate neuronal membranes and other organelles, as well as exceptional post-processing step to obtain refined segmentation results.

The main contribution of our work is proposing an effective deep contextual residual network for neuronal membranes segmentation in ssTEM image stacks. Inspired by previous studies, this approach applies a more spatially efficient and better performing architecture to avert the vanishing gradient problem, and make the network more effective for neuronal membranes segmentation [13]. We further incorporate multilevel contextual cues to avoid the ambiguities in neuronal membranes and other ultrastructural objects, and then employ exceptional post-processing step to improve segmentation results. The experiment results on the public dataset of ISBI 2012 EM Segmentation Challenge suggest that our approach achieves state-of-the-art results, and it is important to note that our algorithm outperforms all published methods on several standard metrics.

2. METHOD

2.1. Network Architecture

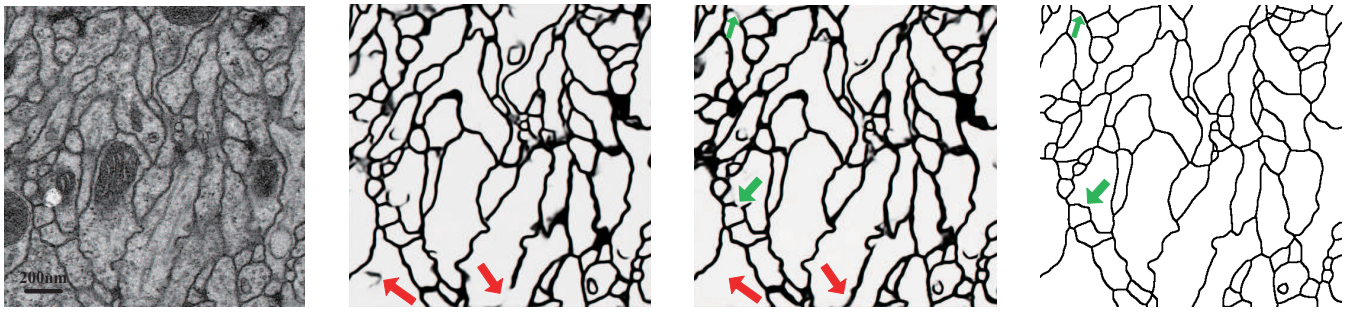
Motivated by previous studies, we propose an efficient contextual residual network to segment neuronal membranes. The overview of the proposed network architecture is illustrated in figure 1, which mainly consists of two modules: contracting path with resnet38-like structure and expansive path with deconvolutional and convolutional layers.

Different from resnet38 [13], we utilize a max-pooling operation (pool size 3×3 with stride 2) after the first convolutional layer to reduce the number of model parameters. The residual unit consists of two convolutional layers with kernel size 3×3 (same padded), each followed by batch normalization and exponential linear units (ELU) nonlinearity (to mitigate the internal covariate shift) [12]. To generate feature maps at 1/8 resolution, the first convolutional layers with kernel size 3×3 and stride 2 in residual block 1 and 2 are used for downsampling. Whereas, residual block 3, 4, 5 and 6 adopt dilated network strategy, which does not reduce the size of feature maps.

For expansive path, the deconvolutional layers upsample the feature maps by fractional strided convolution with channels 64, kernel size $2N \times 2N$ and stride *N* ($N = 2$, 4 and 8 for upsampling layers, respectively) [8]. In what follows, several summation-based skip connections are utilized to incorporate global information from higher layers and local cues from lower layers. Being different from concatenation-based skip connection, summation-based skip connection fuses multilevel contextual information more thoroughly and helps deal with the vanishing gradient problem. In closing, two convolutional layers and dropout ($p = 0.5$) are employed to refine the per-pixel prediction and avoid overfitting.

2.2. Training

We use all 30 slices of the training dataset with 512×512 resolution to train the network, whereas, the samples are not sufficient for deep learning training. To avoid overfitting, we exploit augmentation strategy as rotation, flipping and elastic distortion to enlarge the training dataset. After data enrichment, the number of the training samples is up to 9,000, which is satisfied for our network training.



(a) (b) (c) (d)

Fig. 2. The qualitative comparison of different segmentation methods. (a) Raw image of slice 16/30 on test dataset; (b) Resnet38 result; (c) Our result without post-processing; (d) Our result with multicut post-processing.

The proposed deep network is implemented using Keras deep learning library and TensorFlow backend. In training process, our network is optimized by Adaptive Moment Estimation (Adam) with the following optimization hyperparameters: *learning rate* = 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and *epsilon* = 10^{-8} for numerical stability. Pixel-wise mean squared error is chosen as the loss function. It takes nearly 36 hours to train our network for 10 epochs with the batch size of 2 on a K40 GPU.

2.3. Post-processing of Network Outputs

Our deep contextual residual network produces decent segmentation results, however, some neuronal membranes are discontinuous in ambiguous regions, which might be attributed to overfitting or inappropriate loss function. For further improving the accuracy of segmentation results, we utilize multicut algorithm to refine the boundary probability maps [12].

The post-processing method contains three parts: (1) aggregating boundary probability maps into superpixels by applying distance transform watershed superpixels algorithm; (2) connecting superpixels into a 3D region adjacency graph, and then training a random forest classifier to predict the score of edges in 3D superpixel maps; (3) merging 3D superpixels into neurites by solving the multicut graph partitioning problem [14]. It is plausible to suggest that multicut improves segmentation results since it learns potentials from the inter-slice edges.

3. EXPERIMENTS AND RESULTS

We validate our approach on the public dataset of ISBI 2012 EM Segmentation Challenge, which were taken from Drosophila larva ventral nerve cord (VNC). The training dataset consists of a stack of 30 slices from ssTEM, which measures around 2 $\mu m \times 2\ \mu m \times 1.5\ \mu m$ with a voxel resolution of 4 $nm \times 4\ nm \times 50\ nm$. Equally, the testing dataset with 30 slices is another volume obtained from VNC. The ground truth masks were annotated by human neuroanatomists by using the software tool TrakEm2 [15]. The performance is measured by metrics Rand Score Thin (V^{rand}) and Information Score Thin (V^{info}) as defined in Ref. [6]. V^{rand} is similar to "Rand Index" [16], which measures the accuracy with which pixels are associated to their corresponding segmentation. V^{info} calculates the similarity between predicted segmentation and ground truth segmentation, which is related to "Variation of Information" [17].

Figure 2 exhibits the qualitative comparison of different segmentation methods. It leads to the obvious conclusion that our approach is more accurate than resnet38 (see red arrows). In addition, multicut as post-processing method refines discontinuous boundaries in probability map (see green arrows). The quantitative comparison is summarized in table 1, and more details of the leader board are available at the following web site[1]. Note that our method (without post-processing) achieves better results than CUMedVision [10] and Masters [18]. After refined by multicut, our approach now ranks 3rd on the leader board and yields 0.98356 V^{rand} as well as 0.99063 V^{info}, which surpasses the performance of lifted multicut and other state-of-the-art methods.

4. CONCLUSION

In this paper, we raise a robust, efficient and applicable approach based on deep neural network for accurate end-to-end neuronal membranes segmentation. The proposed method exploits effective residual network and incorporates multilevel contextual information by using summation-based skip connections, which is capable of improving the adaptability of the network in diverse neuronal membranes segmentation. Furthermore, multicut as post-processing step is employed to refine the segmentation results. Experimental results on ISBI 2012 EM Segmentation Challenge demonstrate the effectiveness of the proposed method and confirm that our method achieves state-of-the-art results on standard quality metrics.

[1] http://brainiac2.mit.edu/isbi_challenge/leaders-board-new

Table 1. Leading Groups of ISBI 2012 EM Segmentation Challenge on Neuronal Structures.

Group name	V^{rand}	V^{info}	Rank
** human values **	0.997847778	0.998997659	
IAL - Steerable Filter CNN	0.986800916	0.991438892	1
HVCL@UNIST	0.983651122	0.991303595	2
CASIA_MIRA (Our)	**0.983563573**	**0.990630782**	**3**
IAL MC/LMC [12]	0.982616131	0.989461939	4
IAL LMC [12]	0.982240005	0.988448278	5
PolyMtl [11]	0.980582825	0.988163049	6
KUnet	0.980222514	0.988967601	7
M2FCN	0.979527600	0.989627989	8
IAL IC	0.977345721	0.989240736	9
Masters [18]	0.977141154	0.987534429	10
CUMedVision [10]	0.976824580	0.988645822	11
CASIA_MIRA (without post-processing)	0.977312412	0.987823374	
Resnet38 [13]	0.973820341	0.987658762	

A total of 88 teams participated in ISBI 2012 EM Segmentation challenge till October 10th, 2017.

References

[1] J. W. Lichtman and W. Denk, "The big and the small: challenges of imaging the brains circuits," *Science*, vol. 334, no. 6056, pp. 618-623, 2011.

[2] Q. W. Xie, X. Chen, L. J. Shen, *et al.*, "Micro reconstruction system for brain," *Systems Engineering - Theory and Practice*, vol. 37, no. 11, pp. 3006-3017, 2017.

[3] S. Y. Takemura, A. Bharioke, Z. Lu, *et al.*, " A visual motion detection circuit suggested by Drosophila connectomics," *Nature*, vol. 500, no. 7461, pp. 175-181, 2013.

[4] I. Arganda-Carreras, S. Seung, A. Cardona, *et al.*, "2012 ISBI Challenge: Segmentation of neuronal structures in EM stacks," 2012, http://brainiac2.mit.edu/isbi_challenge/.

[5] A. Cardona, S. Saalfeld, S. Preibisch, *et al.*, "An integrated micro- and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy," *PLoS biology*, vol. 8, no. 10, pp. e1000502, 2010.

[6] I. Arganda-Carreras, S. C. Turaga, D. R. Berger, *et al.*, "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Frontiers in neuroanatomy*, vol. 5, no. 9, pp. 142, 2015.

[7] D. Ciresan, A. Giusti, L. M. Gambardella, *et al.*, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. of NIPS*, pp. 2843-2851, 2012.

[8] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of CVPR*, pp. 3431-3440, 2015.

[9] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. of MICCAI*, pp. 234-241, 2015.

[10] H. Chen, X. Qi, J. Z. Cheng, *et al.*, "Deep contextual networks for neuronal structure segmentation," in *Proc. of AAAI*, pp. 1167-1173, 2016.

[11] T. M. Quan, D. G. C. Hildebrand and W. K. Jeong, "FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics," *arXiv*:1612.05360, 2016.

[12] T. Beier, C. Pape, N. Rahaman, *et al.*, "Multicut brings automated neurite segmentation closer to human performance," *Nature Methods*, vol. 14, no. 2, pp. 101-102, 2017.

[13] Z. Wu, C. Shen and A. Hengel, "Wider or deeper: Revisiting the resnet model for visual recognition," *arXiv*:1611.10080, 2016.

[14] M. Keuper, E. Levinkov, N. Bonneel, G. Lavoué, T. Brox and B. Andres, "Efficient decomposition of image and mesh graphs by lifted multicuts," in *Proc. of ICCV*, pp. 1751-1759, 2015.

[15] A. Cardona, S. Saalfeld, J. Schindelin, *et al.*, "TrakEM2 Software for Neural Circuit Reconstruction," *PLoS ONE*, vol. 7, no. 6, pp. e38011, 2012.

[16] R. Unnikrishnan, C. Pantofaru and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 929-944, 2007.

[17] M. Meilă, "Comparing clusterings: an axiomatic view," in *Proc. of ICML*, pp. 577-584, 2005.

[18] S. Wiehman and H. D. Villiers, "Semantic segmentation of bioimages using convolutional neural networks," in *Proc. of IJCNN*, pp. 624-631, 2016.