Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Deep Contextual Networks for Neuronal Structure Segmentation

Hao Chen[†,*], **Xiaojuan Qi**[†,*], **Jie-Zhi Cheng**[‡], **Pheng-Ann Heng**[†,§]
[†] Department of Computer Science and Engineering, The Chinese University of Hong Kong
[‡] School of Medicine, Shenzhen University, China
[§] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

Abstract

The goal of connectomics is to manifest the interconnections of neural system with the Electron Microscopy (EM) images. However, the formidable size of EM image data renders human annotation impractical, as it may take decades to fulfill the whole job. An alternative way to reconstruct the connectome can be attained with the computerized scheme that can automatically segment the neuronal structures. The segmentation of EM images is very challenging as the depicted structures can be very diverse. To address this difficult problem, a deep contextual network is proposed here by leveraging multi-level contextual information from the deep hierarchical structure to achieve better segmentation performance. To further improve the robustness against the vanishing gradients and strengthen the capability of the back-propagation of gradient flow, auxiliary classifiers are incorporated in the architecture of our deep neural network. It will be shown that our method can effectively parse the semantic meaning from the images with the underlying neural network and accurately delineate the structural boundaries with the reference of low-level contextual cues. Experimental results on the benchmark dataset of 2012 ISBI segmentation challenge of neuronal structures suggest that the proposed method can outperform the state-of-the-art methods by a large margin with respect to different evaluation measurements. Our method can potentially facilitate the automatic connectome analysis from EM images with less human intervention effort.

Introduction

In neuroscience, the neuronal circuit reconstruction, also termed as connectome, from biological images can manifest the interconnections of neurons for more insightful functional analysis of the brain and other nervous systems (Sporns, Tononi, and Kötter 2005; Laptev et al. 2012). For instance, the 2D serial high resolution Electron Microscopy (EM) imaging is commonly used for the visualization of micro neural circuits and hence is a very informative imaging tool for the connectome analysis. In this paper, we focus on the widely used serial section Transmission Electron Microscopy (ssTEM) images for neuronal structure segmentation (Cardona et al. 2010). To illustrate the image complexity, a 2D example of original ssTEM image and corresponding segmentation by expert are illustrated in Figure 1. It can be found that the neuronal structures depicted in the ssTEM images are very complex and hence require the further segmentation of each structure to elucidate the interconnection relation.

However, this is a non-trivial task. The ssTEM images can depict more than tens of thousands of neurons where each neuron may have thousands of synaptic connections. Thus, the size of ssTEM images is usually formidably large in a terabyte scale. Accordingly, the extremely complicated interconnections of neuronal structures and sheer image volume are far beyond the human capability for annotation, as the manual labeling of all neuronal structures may take decades to finish (White et al. 1986; Bock et al. 2011; Wu 2015). In this case, automatic segmentation methods are highly demanded to assist the parsing of the ssTEM images into concrete neurological structures for further analysis (Seung 2011). However, as can be observed in Figure 1, the segmentation problem for the neuronal structures can be very challenging in threefold. First, the image deformation during the acquisition may blur the membrane boundaries between neighboring neurons as shown in Figure 1 (left). Second, the variation of neuron membrane in terms of image contrast and membranal thickness can be very large. Particularly for the thickness, it can range from solid dark curves to grazed grey swaths (Jurrus et al. 2010). Third, the presence of intracellular structures makes edge detection and region growing based methods ineffective for the identification of neuron membrane. Some confounding micro-structures may also mislead the merging of regions or incorrect splitting of one region into several sections. Meanwhile, the imaging artifacts and image alignment errors can impose difficulties on the design of effective segmentation algorithm as well.

Related Work

Because of the anisotropic nature of ssTEM data, most previous methods were devised under the framework of initial 2D membrane detection and latter 3D linking process (Jurrus et al. 2010). Although considerable progress has been made over the last decade, earlier studies achieved a limited accuracy of segmentation and often failed to suppress

[*] Authors contributed equally.

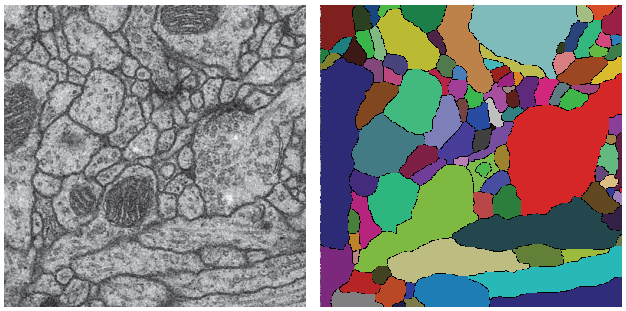


Figure 1: Left: the original ssTEM image. Right: the corresponding segmentation annotation (individual components are denoted by different colors).

the intracellular structures effectively with the hand-crafted features, e.g., radon and ray-like features (Kumar, Vázquez-Reina, and Pfister 2010; Mishchenko 2009; Laptev et al. 2012; Kaynig, Fuchs, and Buhmann 2010). Recently, deep neural networks with hierarchical feature representations have achieved promising results in various applications, including image classification (Krizhevsky, Sutskever, and Hinton 2012), object detection (Simonyan and Zisserman 2014; Chen et al. 2015) and segmentation (Long, Shelhamer, and Darrell 2014). In terms of EM segmentation, Ciresan et al. (2012) employed the deep convolutional neural network as a pixel-wise classifier by taking a square window centered on the pixel itself as input, which contains contextual appearance information. This method achieved the best performance in 2012 ISBI neuronal structure segmentation challenge. A variant version with iterative refining process has been proposed to withstand the noise and recover the boundaries (Wu 2015). Besides, several methods worked on the probability maps produced by deep convolutional neural networks as a post-processing step, such as learning based adaptive watershed (Uzunbaş, Chen, and Metaxsas 2014), hierarchical merge tree with consistent constraints (Liu et al. 2014) and active learning approach for hierarchical agglomerative segmentation (Nunez-Iglesias et al. 2013), to further improve the performance. These methods refined the segmentation results with respect to the measurements of rand error and warping error (Jain et al. 2010) with significant performance boost in comparison to the results of (Ciresan et al. 2012).

However, the performance gap between the computerized results and human neuroanatomist annotations can be still perceivable. There are two main drawbacks of previous deep learning based studies on this task. First, the operation of sliding window scanning imposes a heavy burden on the computational efficiency. This must be taken into consideration seriously regarding the large scale neuronal structure reconstruction. Second, the size of neuronal structure can be very diverse in EM images. Although, classification with single size sub-window can achieve good performance, it may produce unsatisfactory results in some regions where the size of contextual window is set inappropriately.

In order to tackle the aforementioned challenges, we propose a novel deep contextual segmentation network to demarcate the neuronal structure in EM stacks. This approach incorporates the multi-level contextual information with different receptive fields, thus it can remove the ambiguities of membranal boundaries in essence that previous studies may fail. Inspired by previous studies (Ciresan et al. 2012; Lee et al. 2014), we further make the model deeper than (Ciresan et al. 2012) and add auxiliary supervised classifiers to encourage the back-propagation flow. This augmented network can further unleash the power of deep neural networks for neuronal structure segmentation. Quantitative evaluation was extensively conducted on the public dataset of 2012 ISBI EM Segmentation Challenge (Ignacio et al. 2012), with rich baseline results for comparison in terms of pixel- and object-level evaluation. Our method achieved the state-of-the-art results, which outperformed those of other methods on all evaluation measurements. It is also worth noting that our results surpassed the annotation by neuroanatomists on the measurement of warping error.

Method

Deeply Supervised Contextual Network

In this section, we present a deeply supervised contextual network for neuronal structure segmentation. Inspired by recent studies of fully convolutional networks (FCN) (Long, Shelhamer, and Darrell 2014; Chen et al. 2014), which replace the fully connected layers with all convolutional kernels, the proposed network is a variant and takes full advantage of convolutional kernels for efficient and effective image segmentation. The architecture of the proposed method is illustrated in Figure 2. It basically contains two modules, i.e., down-sampling path with convolutional and max-pooling layers and upsampling path with convolutional and deconvolutional layers. Noting that we upsampled the feature maps with the backwards strided convolution in the upsampling path, thus we call them as deconvolutional layers. The downsampling path aims at classifying the semantical meanings based on the high level abstract information, while the upsampling path reconstructing the fine details such as boundaries. The upsampling layers are designed by taking full advantage of the different feature maps in hierarchical layers.

The basic idea behind this is that global or abstract information from higher layers helps to resolve the problem of what (i.e., classification capability) and local information from lower layers helps to resolve the problem of where (i.e., localization accuracy). Finally, these multi-level contextual information are fused together with a summing operation. The probability maps are generated by inputting the fused map into a softmax classification layer. Specifically, the architecture of neural network contains 16 convolutional layers, 3 max-pooling layers for downsampling and 3 deconvolutional layers for upsampling. The convolutional layers along with convolutional kernels (3×3 or 1×1) perform linear mapping with shared parameters. The max-pooling layers downsample the size of feature maps by the max-pooling operation (kernel size 2×2 with a stride 2). The deconvolutional layers upsample the size of feature

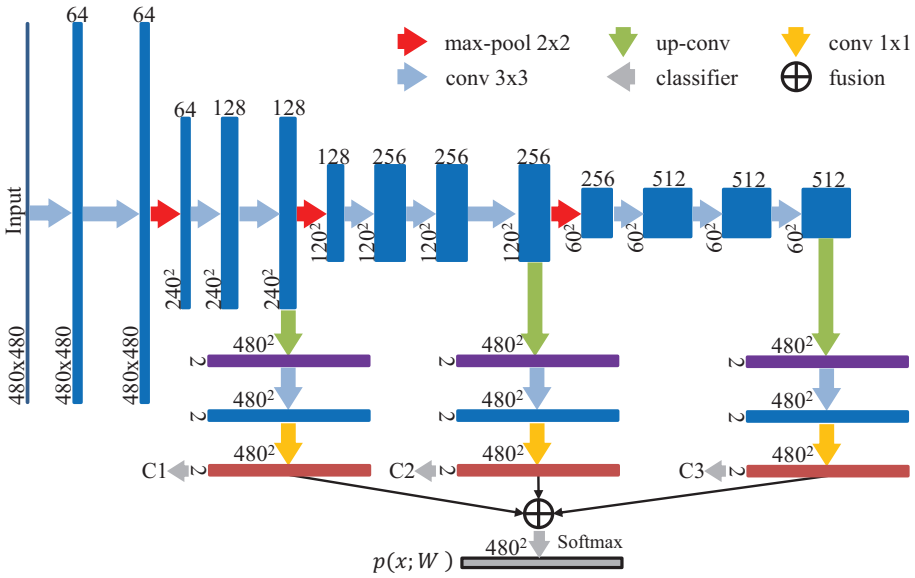


Figure 2: The architecture of the proposed deep contextual network.

maps by the backwards strided convolution (Long, Shelhamer, and Darrell 2014) ($2k \times 2k$ kernel with a stride k, $k = 2, 4$ and 8 for upsampling layers, respectively). A nonlinear mapping layer (element-wise rectified linear activations) is followed for each layer that contains parameters to be trained (Krizhevsky, Sutskever, and Hinton 2012).

In order to alleviate the problem of vanishing gradients and encourage the back-propagation of gradient flow in deep neural networks, the auxiliary classifiers C are injected for training the network. Furthermore, they can serve as regularization for reducing the overfitting and improve the discriminative capability of features in intermediate layers (Bengio et al. 2007; Lee et al. 2014; Wang et al. 2015). The classification layer after fusing multi-level contextual information produces the EM image segmentation results by leveraging the hierarchical feature representations. Finally, the training of whole network is formulated as a per-pixel classification problem with respect to the ground-truth segmentation masks, as shown following:

$$\mathcal{L}(\mathcal{X};\theta) = \frac{\lambda}{2}(\sum_{c}||W_c||_2^2 + ||W||_2^2) - \sum_{c}\sum_{x\in\mathcal{X}} w_c\psi_c(x,\ell(x)) - \sum_{x\in\mathcal{X}}\psi(x,\ell(x)) \quad (1)$$

where the first part is the regularization term and latter one including target and auxiliary classifiers is the data loss term. The tradeoff of these two terms is controlled by the hyperparameter λ. Specifically, W denotes the parameters for inferring the target output $p(x;W)$, $\psi(x,\ell(x))$ denotes the cross entropy loss regarding the true label $\ell(x)$ for pixel x in image space $\mathcal{X}$, similarly $\psi_c(x,\ell(x))$ is the loss from cth auxiliary classifier with parameters W_c for inferring the output, the parameter w_c denotes the corresponding discount weight. Finally, the parameters $\theta = \{W, W_c\}$ of deep contextual network are jointly optimized in an end-to-end way by minimizing the total loss function $\mathcal{L}$. For the testing data of EM images, the results are produced with an overlap-tile strategy to improve the robustness.

Importance of Receptive Field

In the task of EM image segmentation, there is a large variation on the size of neuronal structures. Therefore, the size of receptive field plays a key role in the pixel-wise classification given the corresponding contextual information. It's approximated as the size of object region with surrounding context, which is reflected as the intensity values within the window. As shown in Figure 3, different regions may depend on a different window size. For example, the cluttered neurons need a small window size for clearly separating the membranes between neighboring neurons, while a large size is required for neurons containing intracellular structures so as to suppress the false predictions. In the hierarchical structure of deep contextual networks, these upsampling layers have different receptive fields. With the depth increasing, the size of receptive field is becoming larger. Therefore, it can handle the variations of reception field size properly that different regions demand for correct segmentation while taking advantage of the hierarchical feature representations.

Morphological Boundary Refinement

Although the probability maps output from the deep contextual network are visually very good, we observe that the membrane of ambiguous regions can sometimes be discontinued. This is partially caused by the averaging effect of probability maps, which are generated by several trained models. Therefore, we utilized an off-the-shelf watershed algorithm (Beucher and Lantuejoul 1979) to refine the contour. The final fusion result $p_f(x)$ was produced by fusing

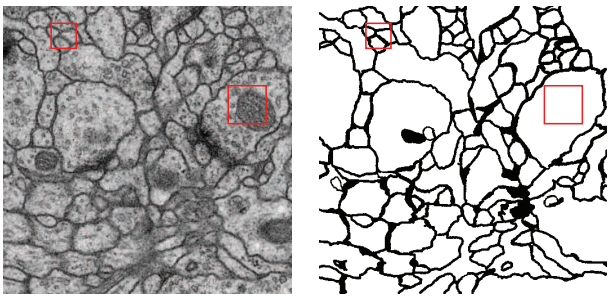


Figure 3: Illustration of contextual window size. Left: the original ssTEM image. Right: manual segmentation result by an expert human neuroanatomist (black and white pixels denote the membrane and non-membrane, respectively).

the binary contour $p_w(x)$ and original probability map $p(x)$ with linear combination:

$$p_f(x) = w_f p(x) + (1 - w_f) p_w(x) \quad (2)$$

The parameter w_f is determined by obtaining the optimal result of rand error on the training data in our experiments.

Experiments and Results

Data and Preprocessing

We evaluated our method on the public dataset of 2012 ISBI EM Segmentation Challenge (Ignacio et al. 2012), which is still open for submissions. The training dataset contains a stack of 30 slices from a ssTEM dataset of the Drosophila first instar larva ventral nerve cord (VNC), which measures approximately 2x2x1.5 microns with a resolution of 4x4x50 nm/voxel. The images were manually annotated in the pixel-level by a human neuroanatomist using the software tool TrakEm2 (Cardona et al. 2012). The ground truth masks of training data were provided while those of testing data with 30 slices were held out by the organizers for evaluation. We evaluated the performance of our method by submitting results to the online testing system. In order to improve the robustness of neural network, we utilized the strategy of data augmentation to enlarge the training dataset (about 10 times larger). The transformations of data augmentation include scaling, rotation, flipping, mirroring and elastic distortion.

Details of Training

The proposed method was implemented with the mixed programming technology of Matlab and C++ under the open-source framework of Caffe library (Jia et al. 2014). We randomly cropped a region (size 480×480) from the original image as the input into the network and trained it with standard back-propagation using stochastic gradient descent (momentum = 0.9, weight decay = 0.0005, the learning rate was set as 0.01 initially and decreased by a factor of 10 every two thousand iterations). The parameter of corresponding discount weight w_c was set as 1 initially and decreased by a factor of 10 every ten thousand iterations till a negligible value 0.01. The training time on the augmentation dataset took about three hours using a standard PC with a 2.50 GHz Intel(R) Xeon(R) E5-1620 CPU and a NVIDIA GeForce GTX Titan X GPU.

Qualitative Evaluation

Two examples of qualitative segmentation results without morphological boundary refinement are demonstrated in Figure 4. We can see that our method can generate visually smooth and accurate segmentation results. As the red arrows shown in the figure, it can successfully suppress the intracellular structures and produce good probability maps that classify the membrane and non-membrane correctly. Furthermore, by utilizing multi-level representations of contextual information, our method can also close gaps (contour completion as the blue arrows shown in Figure 4) in places where the contrast of membrane is low. Although there still exist ambiguous regions which are even hard for human experts, the results of our method are more accurate in comparison to those generated from previous deep learning studies (Stollenga et al. 2015; Ciresan et al. 2012). This evidenced the efficacy of our proposed method qualitatively.

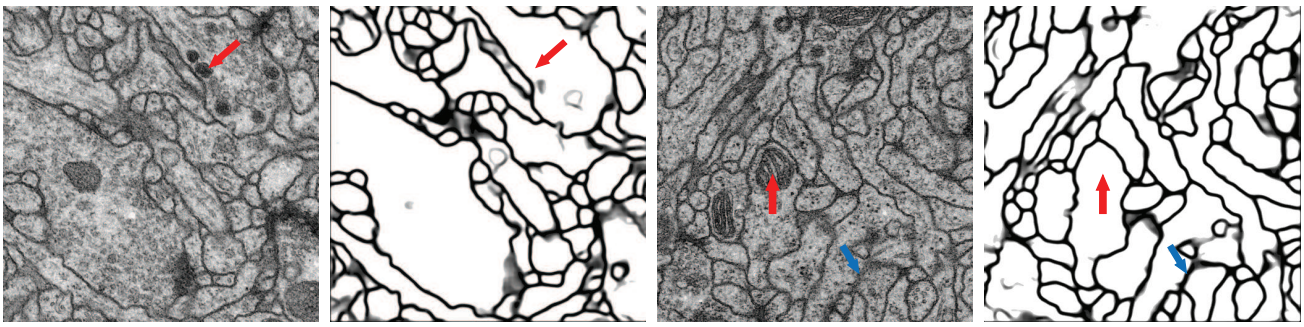
Quantitative Evaluation and Comparison

In the 2012 ISBI EM Segmentation Challenge, the performance of different competing methods is ranked based on their pixel and object classification accuracy. Specifically, the 2D topology-based segmentation evaluation metrics include rand error, warping error and pixel error (Ignacio et al. 2012; Jain et al. 2010), which are defined as following:
Rand error: 1 - the maximal F-score of the foreground-restricted rand index (Rand 1971), a measure of similarity between two clusters or segmentations. For the EM segmentation evaluation, the zero component of the original labels (background pixels of the ground truth) is excluded.
Warping error: a segmentation metric that penalizes the topological disagreements (object splits and mergers).
Pixel error: 1 - the maximal F-score of pixel similarity, or squared Euclidean distance between the original and the result labels.

The evaluation system thresholds the probability maps with 9 different values (0.1-0.9 with an interval 0.1) separately and return the minimum error for each segmentation metric. The quantitative comparison of different methods can be seen in Table 1. Noting that the results show the best performance for each measurement across all submissions by each team individually. More details and results are available at the leader board[1]. We compared our method with the state-of-the-art methods with or without post-processing separately. Furthermore, we conducted extensive experiments with ablation studies to probe the performance gain in our method and detail as following.

Results Comparison without Post-Processing Preliminary encouraging results were achieved by *IDSIA* team (Ciresan et al. 2012), which utilized a deep convolutional neural network as a pixel-wise classifier in a sliding window way. The best results were obtained by averaging

[1]Please refer to the leader board for more details: http://brainiac2.mit.edu/isbi_challenge/leaders-board



Slice #06 Slice #13

Figure 4: Examples of original EM images and segmentation results by our method (the darker color of pixels denotes the higher probability of being membrane in neuronal structure).

Table 1: Results of 2012 ISBI Segmentation Challenge on Neuronal Structures

Group name	Rand Error	Warping Error	Pixel Error	Rank
** human values **	0.002109173	0.000005341	0.001041591	
CUMedVision (Our)	**0.017334163**	**0.000000000**	**0.057953485**	**1**
DIVE-SCI	0.017841947	0.000307083	0.058436986	2
IDSIA-SCI	0.018919792	0.000616837	0.102692786	3
optree-idsia (Uzunbaş, Chen, and Metaxsas 2014)	0.022777620	0.000807953	0.110460288	4
motif (Wu 2015)	0.026326384	0.000426483	0.062739851	5
SCI (Liu et al. 2014)	0.028054308	0.000515747	0.063349324	6
Image Analysis Lab Freiburg (Ronneberger, Fischer, and Brox 2015)	0.038225781	0.000352859	0.061141279	7
Connectome	0.045905709	0.000478999	0.062029263	8
PyraMiD-LSTM (Stollenga et al. 2015)	0.046704591	0.000462341	0.061624006	9
DIVE	0.047680695	0.000374222	0.058205303	10
IDSIA (Ciresan et al. 2012)	0.048314096	0.000434367	0.060298549	11
INI	0.060110507	0.000495529	0.068537199	12
MLL-ETH (Laptev et al. 2012)	0.063919883	0.000581741	0.079403258	13
CUMedVision-4(C3)	0.043419035	0.000342178	0.060940140	
CUMedVision-4(C2)	0.046058434	0.000421524	0.061248112	
CUMedVision-4(C1)	0.258966855	0.001080322	0.102325669	
CUMedVision-4(with C)	0.035134666	0.000334167	0.058372960	
CUMedVision-4(w/o C)	0.040492503	0.000330353	0.062864362	
CUMedVision-6(with C)	0.040406591	0.000000000	0.059902422	
CUMedVision-4(with fusion)	0.017334163	0.000188446	0.057953485	

There are total 38 teams participating this challenge till Sep 2015.

the outputs from 4 deep neural network models. Different from this method by training the neural network with different window sizes (65 and 95) separately, our approach integrates multi-size windows (i.e., different receptive fields in upsampling layers) into one unified framework. This can help to generate more accurate probability maps by leveraging multi-level contextual information. The *Image Analysis Lab Freiburg* team (Ronneberger, Fischer, and Brox 2015) designed a deep U-shaped network by concatenating features from lower layers and improved the results than those of (Ciresan et al. 2012). This further demonstrated the effectiveness of contextual information for accurate segmentation. However, with such a deep network (i.e., 23 convolutional layers), the back-propagation of gradient flow may be a potential issue and training took a long time (about 10 hours). Instead of using the convolutional neural network, the *PyraMiD-LSTM* team employed a novel parallel multi-dimensional long short-term memory model for fast volumetric segmentation (Stollenga et al. 2015). Unfortunately, a relatively inferior performance was achieved by this method. From Table 1, we can see that our deep segmentation network (with 6 model averaging results, i.e., *CUMedVision-6(with C)*) without watershed fusion achieved the best performance in terms of warping error, which outperformed other methods by a large margin. Notably it's the only result that surpasses the performance of expert neuroanatomist annotation. Our submitted entry *CUMedVision-4(with C)* on averaging 4 models (the same number of models as (Ciresan et al. 2012)) achieved much smaller rand and warping errors than the results of other teams also employing deep learning methods without sophisticated post-processing steps, such as *DIVE*, *IDSIA*, and *Image Analysis Lab Freiburg*. This cor-

roborates the superiority of our approach by exploring multi-level contextual information with auxiliary supervision.

Results Comparison with Post-Processing In order to further reduce the errors, we fused the results from watershed method as illustrated in the method section, which can reduce the rand error dramatically while increasing the warping error unfortunately. This is reasonable since these two errors consider the segmentation evaluation metric from different aspects. The former one could penalize even slightly misplaced boundaries while the latter one disregards non-topological errors. Different from our simple post-processing step, the *SCI* team post-processed the probability maps generated by the team *DIVE* and *IDSIA* with a sophisticated post-processing strategy (Liu et al. 2014). The post-processed results were evaluated under the team name of *DIVE-SCI* and *IDSIA-SCI*, respectively. Although it utilized a supervised way with hierarchical merge tree to achieve structure consistency, the performance is relatively inferior compared to ours, in which only an unsupervised watershed method was used for post-processing. In addition, our method also outperformed other methods with sophisticated post-processing techniques including *optree-idsia* and *motif* by a large margin. This further highlights the advantages of our method by exploring multi-level contextual information to generate probability maps with better likelihood. We released the probability maps including training and testing data of our method for enlightening further sophisticated post-processing strategies[2].

Ablation Studies of Our Method In order to probe the performance gain of our proposed method, extensive ablation studies were conducted to investigate the role of each component. As illustrated in Table 1, compared with methods using single contextual information including *CUMedVision-4(C3/C2/C1)*, the deep contextual model harnessing the multi-level contextual cues achieved significantly better performance on all the measurements. Furthermore, we compared the performance with (*CUMedVision-4(with C)*) and without (*CUMedVision-4(w/o C)*) the injection of auxiliary classifiers C, the rand error and pixel error from method with C were much smaller while the warping error with C is competitive compared to the method without C. This validated the efficacy of auxiliary classifiers with deep supervision for encouraging back-propagation of gradient flow. By fusing the results from the watershed method, we achieved the result with rand error 0.017334, warping error 0.000188, and pixel error 0.057953, which outperforms those from other teams by a large margin. To sum up, our method achieved the best performance on different evaluation measurements, which demonstrates the promising possibility for read-world applications. Although there is a tradeoff with respect to different evaluation metrics, the neuroanatomists can choose the desirable results based on the specific neurological requirements.

Computation Time Generally, it took about 0.4 seconds to process one test image with size 512×512 using the same configuration of training. Taking advantage of fully convolutional networks, the computation time is much less than previous studies (Ciresan et al. 2012; Wu 2015) utilizing a sliding window way, which caused a large number of redundant computations on neighboring pixels. With new imaging techniques producing much larger volumes (terabyte scale) that contain thousands of neurons and millions of synapses, the automatic methods with accurate and fast segmentation capabilities are of paramount importance. The fast speed and better accuracy of our method make it possible for large scale image analysis.

[2]Results: http://appsrv.cse.cuhk.edu.hk\%7Ehchen/research/2012isbi_seg.html

Conclusion

In this paper we have presented a deeply supervised contextual neural network for neuronal structure segmentation. By harnessing the multi-level contextual information from the deep hierarchical feature representations, it can have better discrimination and localization abilities, which are key to image segmentation related tasks. The injected auxiliary classifiers can help to encourage the back-propagation of gradient flow in training the deep neural network, thus further improve the segmentation performance. Extensive experiments on the public dataset of 2012 ISBI EM Segmentation Challenge corroborated the effectiveness of our method. We believe the promising results are a significant step towards automated reconstruction of the connectome. In addition, our approach is general and can be easily extended to other biomedical applications. Future work will include further refining the segmentation results with other sophisticated post-processing techniques (Uzunbaş, Chen, and Metaxsas 2014; Liu et al. 2014; Nunez-Iglesias et al. 2013) and investigating on more biomedical applications.

Acknowledgements This work is supported by National Basic Research Program of China, 973 Program (No. 2015CB351706) and a grant from Ministry of Science and Technology of the People's Republic of China under the Singapore-China 9th Joint Research Program (No. 2013DFG12900). The authors also gratefully thank the challenge organizers for helping the evaluation.

References

Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H.; et al. 2007. Greedy layer-wise training of deep networks. *Advances in neural information processing systems* 19:153.

Beucher, S., and Lantuejoul, C. 1979. Use of watersheds in contour detection. In *International Conference on Image Processing*.

Bock, D. D.; Lee, W.-C. A.; Kerlin, A. M.; Andermann, M. L.; Hood, G.; Wetzel, A. W.; Yurgenson, S.; Soucy, E. R.; Kim, H. S.; and Reid, R. C. 2011. Network anatomy and in vivo physiology of visual cortical neurons. *Nature* 471(7337):177–182.

Cardona, A.; Saalfeld, S.; Preibisch, S.; Schmid, B.; Cheng, A.; Pulokas, J.; Tomancak, P.; and Hartenstein, V. 2010. An integrated micro-and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. *PLoS biology* 8(10):2564.

Cardona, A.; Saalfeld, S.; Schindelin, J.; Arganda-Carreras, I.; Preibisch, S.; Longair, M.; Tomancak, P.; Hartenstein, V.; and Douglas, R. J. 2012. Trakem2 software for neural circuit reconstruction. *PloS one* 7(6):e38011.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.

Chen, H.; Shen, C.; Qin, J.; Ni, D.; Shi, L.; Cheng, J. C.; and Heng, P.-A. 2015. Automatic localization and identification of vertebrae in spine ct via a joint learning model with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*. Springer. 515–522.

Ciresan, D.; Giusti, A.; Gambardella, L. M.; and Schmidhuber, J. 2012. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, 2843–2851.

Ignacio, A.-C.; Sebastian, S.; Albert, C.; and Johannes, S. 2012. 2012 ISBI Challenge: Segmentation of neuronal structures in EM stacks. http://brainiac2.mit.edu/isbi_challenge/.

Jain, V.; Bollmann, B.; Richardson, M.; Berger, D. R.; Helmstaedter, M. N.; Briggman, K. L.; Denk, W.; Bowden, J. B.; Mendenhall, J. M.; Abraham, W. C.; et al. 2010. Boundary learning by optimization with topological constraints. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2488–2495. IEEE.

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; and Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

Jurrus, E.; Paiva, A. R.; Watanabe, S.; Anderson, J. R.; Jones, B. W.; Whitaker, R. T.; Jorgensen, E. M.; Marc, R. E.; and Tasdizen, T. 2010. Detection of neuron membranes in electron microscopy images using a serial neural network architecture. *Medical image analysis* 14(6):770–783.

Kaynig, V.; Fuchs, T. J.; and Buhmann, J. M. 2010. Geometrical consistent 3d tracing of neuronal processes in sstem data. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*. Springer. 209–216.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105.

Kumar, R.; Vázquez-Reina, A.; and Pfister, H. 2010. Radon-like features and their application to connectomics. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 186–193. IEEE.

Laptev, D.; Vezhnevets, A.; Dwivedi, S.; and Buhmann, J. M. 2012. Anisotropic sstem image segmentation using dense correspondence across sections. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer. 323–330.

Lee, C.-Y.; Xie, S.; Gallagher, P.; Zhang, Z.; and Tu, Z. 2014. Deeply-supervised nets. *arXiv preprint arXiv:1409.5185*.

Liu, T.; Jones, C.; Seyedhosseini, M.; and Tasdizen, T. 2014. A modular hierarchical approach to 3d electron microscopy image segmentation. *Journal of neuroscience methods* 226:88–102.

Long, J.; Shelhamer, E.; and Darrell, T. 2014. Fully convolutional networks for semantic segmentation. *arXiv preprint arXiv:1411.4038*.

Mishchenko, Y. 2009. Automation of 3d reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs. *Journal of neuroscience methods* 176(2):276–289.

Nunez-Iglesias, J.; Kennedy, R.; Parag, T.; Shi, J.; Chklovskii, D. B.; and Zuo, X.-N. 2013. Machine learning of hierarchical clustering to segment 2d and 3d images. *PloS one* 8(8):08.

Rand, W. M. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association* 66(336):846–850.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*.

Seung, H. S. 2011. Neuroscience: towards functional connectomics. *Nature* 471(7337):170–172.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sporns, O.; Tononi, G.; and Kötter, R. 2005. The human connectome: a structural description of the human brain. *PLoS Comput Biol* 1(4):e42.

Stollenga, M. F.; Byeon, W.; Liwicki, M.; and Schmidhuber, J. 2015. Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation. *arXiv preprint arXiv:1506.07452*.

Uzunbaş, M. G.; Chen, C.; and Metaxsas, D. 2014. Optree: a learning-based adaptive watershed algorithm for neuron segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. Springer. 97–105.

Wang, L.; Lee, C.-Y.; Tu, Z.; and Lazebnik, S. 2015. Training deeper convolutional networks with deep supervision. *arXiv preprint arXiv:1505.02496*.

White, J.; Southgate, E.; Thomson, J.; and Brenner, S. 1986. The structure of the nervous system of the nematode caenorhabditis elegans: the mind of a worm. *Phil. Trans. R. Soc. Lond* 314:1–340.

Wu, X. 2015. An iterative convolutional neural network algorithm improves electron microscopy image segmentation. *arXiv preprint arXiv:1506.05849*.