CHAPTER 10

Deep multilevel contextual networks for biomedical image segmentation

Hao Chen[a], **Qi Dou**[a], **Xiaojuan Qi**[a], **Jie-Zhi Cheng**[b], **Pheng-Ann Heng**[a]
[a]The Chinese University of Hong Kong, Department of Computer Science and Engineering, Hong Kong, China
[b]Shenzhen University, School of Medicine, Shenzhen, China

Contents

10.1. Introduction

Biomedical image segmentation has been a crucial, yet challenging topic in the field of medical image computing. It serves as one of the basic components for many biomedical related applications, such as medical disease diagnosis and biological interconnection

Handbook of Medical Image Computing and Computer Assisted Intervention
https://doi.org/10.1016/B978-0-12-816176-0.00015-6

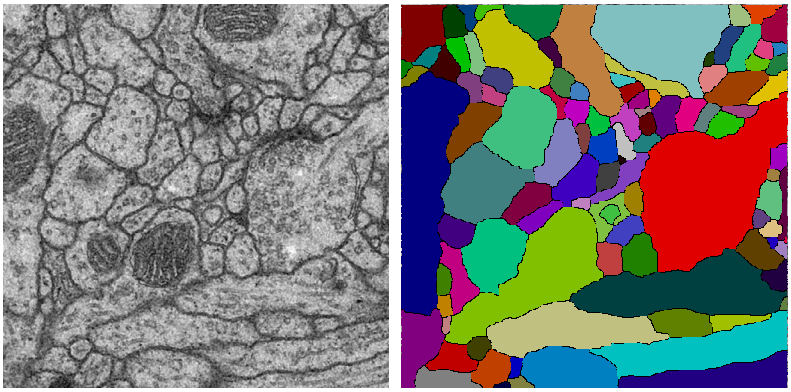

Figure 10.1 (Left) The original ssTEM image. (Right) The corresponding segmentation annotation (individual components are denoted by different colors).

interpretation. For example, the neuronal circuit reconstruction, also termed as connectome in neuroscience, from biological images can manifest the interconnections of neurons for more insightful functional analysis of the brain and other nervous systems [1, 2]. The 2D serial high resolution Electron Microscopy (EM) imaging is commonly used for the visualization of microneural circuits and hence is a very informative imaging tool for the connectome analysis. Fig. 10.1 illustrates a 2D example of serial section Transmission Electron Microscopy (ssTEM) images which are widely used for neuronal structure segmentation [3].

As can be observed in Fig. 10.1, the segmentation problem for the neuronal structures can be very challenging in three ways. First, the image deformation during the acquisition may blur the membrane boundaries between neighboring neurons as shown in Fig. 10.1 (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 [4]. Third, the presence of intracellular structures makes edge detection and region growing based methods ineffective for the identification of neuron membrane. Some confounding microstructures 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 algorithms as well.

Recently, deep learning with hierarchical feature representations has achieved promising results in various applications, including image classification [5], object detection [6–8], and segmentation [9,10]. However, the performance gap between the computerized results and human 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 biomedical

image segmentation. Second, the size of biological structures can be very diverse. Although, classification with single size subwindow 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 for biomedical image segmentation. This approach incorporates the multilevel contextual information with different receptive fields, thus it can remove the ambiguities of structural boundaries in essence that previous studies may fail to do. Inspired by previous studies [11,12], we further make the model deeper than in [11] and add auxiliary supervised classifiers to encourage the backpropagation flow. This augmented network can further unleash the power of deep neural networks for biomedical structure segmentation. Quantitative evaluation was extensively conducted on the public dataset of 2012 ISBI EM Segmentation Challenge [13] and 2015 MICCAI Nuclei Segmentation Challenge, 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 when measuring the warping error in the EM Segmentation task. In addition, the superior performance on these two benchmarks demonstrated the generalization capability of our proposed method.

10.2. Related work

10.2.1 Electron microscopy image segmentation

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 and is on 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 [14–16]. In this case, automatic segmentation methods are highly demanded to assist the parsing of the ssTEM images into concrete neurological structures for further analysis [17].

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 [4]. Although considerable progress has been made over the last decade, earlier studies achieved a limited accuracy of segmentation and often failed to suppress the intracellular structures effectively with the hand-crafted features, e.g., Radon and ray-like features [18,19,2,20].

Recently, Ciresan et al. employed the deep convolutional neural network as a pixelwise classifier by taking a square window centered on the pixel itself as input, which

contains contextual appearance information [11]. 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 [16]. Besides, several methods worked on the probability maps produced by deep convolutional neural networks as a postprocessing step, such as learning based adaptive watershed [21], hierarchical merge tree with consistent constraints [22], and active learning approach for hierarchical agglomerative segmentation [23], to further improve the performance. These methods refined the segmentation results with respect to the measurements of Rand and warping errors [24] with significant performance boost in comparison to the results of [11].

10.2.2 Nuclei segmentation

With the advent of whole slide imaging scanners, tissue histopathology slides can be digitized and stored in the form of digital images. Meanwhile, histopathological analysis performed on these digital images has been demonstrated as an effective and reliable tool for cancer diagnosis and prognosis [25]. In the routine of histopathological examination, accurate detection and segmentation of certain histological structures, such as cancer nuclei, is one of crucial prerequisite steps to obtain reliable morphological statistics that characterize the aggressiveness of tumors. Specifically, counting of object instances such as cell nuclei has diagnostic significance for some cancerous diseases [26–28]. This requires an accurate detection and segmentation of cell nuclei. The nucleus morphism has an important diagnostic value for cancer grading [29–31].

For the nuclei detection and segmentation, various methods have been proposed to tackle this problem ranging from relatively simple approaches, such as thresholding and morphological operations [32,33], to more sophisticated methods based on handcrafted features derived from boundaries/contours [26,34], gradients [35], Laplacian-of-Gaussian [36], cytological and textural features [37], etc. Then different classifiers (e.g., Support Vector Machine (SVM), Adaboost, and Bayesian) have been employed in the literature to detect and segment nuclei from histology images [38]. However, the handcrafted features suffer from limited representation capabilities, and hence they can be vulnerable to different variations. Furthermore, the piecewise learning system separating feature extraction and classification may not be optimal or efficient for generating precise probability maps of histological structures.

Recently, stacked sparse autoencoders (SSAE) were exploited with unsupervised pretraining and following fine-tuning for nuclei detection from breast cancer histopathology images in [27]. Although along with merit of unsupervised pretraining, which can handle the situation of limited medical training data, the autoencoders usually achieved inferior performance on image recognition tasks compared to convolutional neural networks (CNNs). The success of the latter networks is mostly attributed to the more elegant structures for dealing with images. Regarding the convolutional

neural networks, the authors of [39] employed deep convolutional neural networks for mitosis detection and achieved the best performance in two grand challenges [40,41]. To further improve the efficiency and effectiveness, Hao Chen et al. [42] developed a cascaded deep learning framework, i.e., a coarse model for retrieving candidates and a fine-discrimination model for singling out mitoses from hard mimics. A spatially constrained convolutional neural network was present in [43] incorporated with neighboring ensemble prediction, demonstrating the efficacy of deep learning based features from CNNs.

10.3. Method

10.3.1 Deep multilevel contextual network

In this section, we present a deeply supervised contextual network for biomedical image segmentation. Inspired by recent studies of fully convolutional networks (FCNs) [9, 44], 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 Fig. 10.2. It basically contains two modules, i.e., downsampling 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 deconvolutional layers. The downsampling path aims at classifying the semantical meanings based on the high level abstract information, while the upsampling path is 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, this multilevel 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 maps by the backwards strided convolution [9] ($2k \times 2k$ kernel with a stride k, $k = 2, 4$ and 8 for upsampling layers, respectively). A nonlinear mapping layer (elementwise rectified linear activations) is followed for each layer that contains parameters to be trained [5].

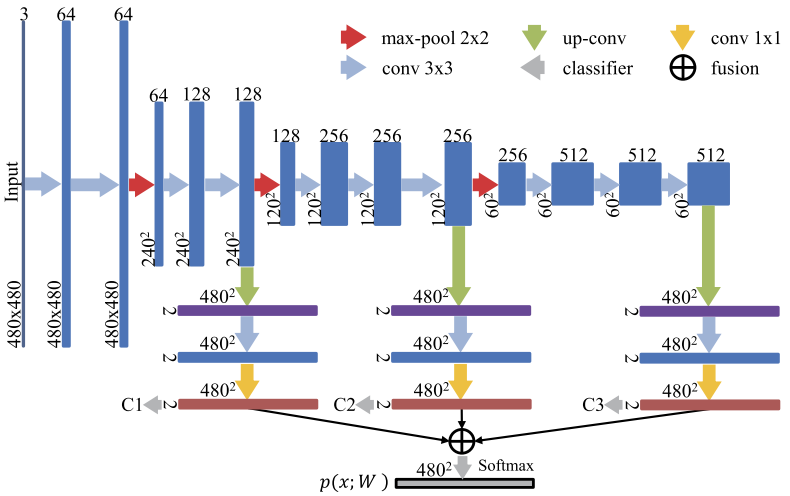

Figure 10.2 The architecture of the proposed deep contextual network.

10.3.2 Regularization with auxiliary supervision

In order to alleviate the problem of vanishing gradients and encourage the backpropagation 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 [45,12,46]. The classification layer after fusing multilevel contextual information produces the 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 follows:

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

where the first part is the regularization term and latter one, including target and auxiliary classifiers, is the data loss term. The tradeoff between 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, parameter w_c denotes the corresponding discount weight. Finally, 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 biomedical images, the results are produced with an overlap–tile strategy to improve the robustness.

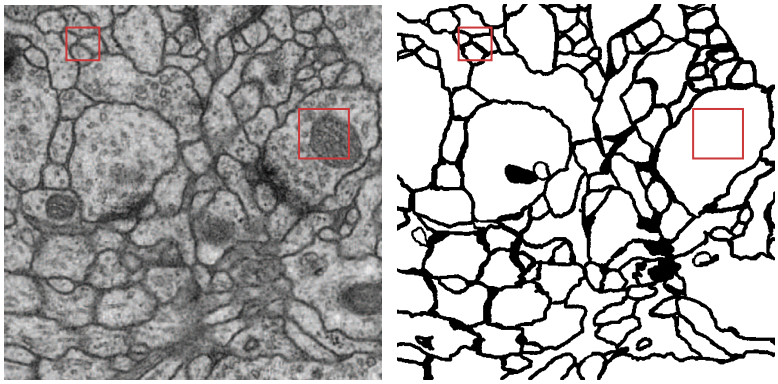

Figure 10.3 Illustration of contextual window size: (left) the original ssTEM image; (right) manual segmentation result by an expert human neuroanatomist (black and white pixels denote the membrane and nonmembrane, respectively).

10.3.3 Importance of receptive field

In the task of biomedical image segmentation, there is usually a large variation in the size of structures. Therefore, the size of a 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 Fig. 10.3, the accurate recognition of different regions from EM images may depend on different window sizes. 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.

10.4. Experiments and results

10.4.1 Dataset and preprocessing

10.4.1.1 2012 ISBI EM segmentation

We evaluated our method on the public dataset of 2012 ISBI EM Segmentation Challenge [13], which is still open for submissions. The training dataset contains a stack of 30 slices from an ssTEM dataset of the Drosophila first instar larva ventral nerve cord (VNC), which measures approximately $2 \times 2 \times 1.5$ microns with a resolution of $4 \times 4 \times 50$ nm/voxel. The images were manually annotated in the pixel-level by a hu-

man neuroanatomist using the software tool TrakEm2 [47]. 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.

10.4.1.2 2015 MICCAI nuclei segmentation

We also evaluated our method on the challenge dataset on *Segmentation of Nuclei in Digital Pathology Images* of Computational Brain Tumor Cluster of Event (CBTC) workshop in conjunction with MICCAI 2015. The training data have at least 500 manually segmented nuclei in 15 image tiles and testing data include 18 images for evaluation (the ground truth is held out by the challenge organizers). Participants are asked to detect and segment all the nuclei of testing tiles, which are extracted from whole slide tissue images. The algorithm results are compared with consensus pathologist segmented subregions. We utilize the strategy of data augmentation to enlarge the training dataset. The transformations of data augmentation include translation and rotation.

10.4.2 Details of training

The proposed method was implemented with the mixed programming technology of Matlab[1] and C++ under the open-source framework of Caffe library [48]. We randomly cropped a region (size 480×480) from the original image as the input into the network and trained it with standard backpropagation 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 2000 iterations). The parameter of corresponding discount weight w_c was set as 1 initially and decreased by a factor of 10 every 10000 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 an NVIDIA GeForce GTX Titan X GPU.

10.4.3 2012 ISBI neuronal structure segmentation challenge

10.4.3.1 Qualitative evaluation

Two examples of qualitative segmentation results without morphological boundary refinement are demonstrated in Fig. 10.4. We can see that our method can generate visually smooth and accurate segmentation results. As the red arrows show in the figure, it can successfully suppress the intracellular structures and produce good probability

[1] MATLAB® is a trademark of The MathWorks.

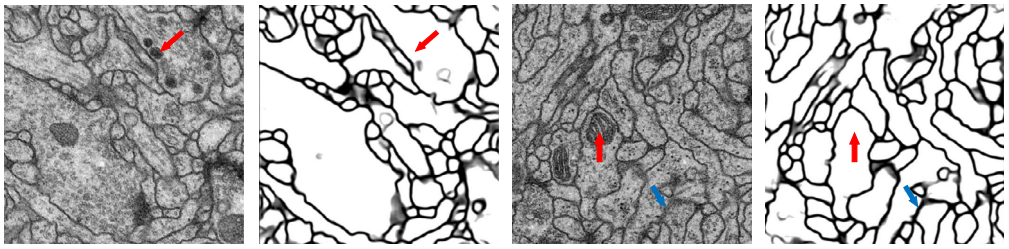

Figure 10.4 Examples of original EM images and segmentation results by our method (the darker color of pixels denotes the higher probability of being membrane in neuronal structure).

maps that classify the membrane and nonmembrane correctly. Furthermore, by utilizing multilevel representations of contextual information, our method can also close gaps (contour completion as the blue arrows shown in Fig. 10.4) 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 [49,11]. This evidenced the efficacy of our proposed method qualitatively.

10.4.3.2 Quantitative evaluation metrics

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, warping, and pixel errors [13,24], which are defined as follows:

Rand error: 1 – the maximal F-score of the foreground-restricted Rand index [50], 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 of 0.1) separately and return the minimum error for each segmentation metric. The quantitative comparison of different methods can be seen in Table 10.1. Note that the results show the best performance for each measurement across all submissions by each team individually. More details and results are available at

Table 10.1 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 [21]	0.022777620	0.000807953	0.110460288	4
CUMedVision-motif	0.025540655	0.000321579	0.057912350	5
motif [16]	0.026326384	0.000426483	0.062739851	6
SCI [22]	0.028054308	0.000515747	0.063349324	7
Image Analysis Lab Freiburg [51]	0.038225781	0.000352859	0.061141279	8
Connectome	0.045905709	0.000478999	0.062029263	9
PyraMiD-LSTM [49]	0.046704591	0.000462341	0.061624006	10
DIVE	0.047680695	0.000374222	0.058205303	11
IDSIA [11]	0.048314096	0.000434367	0.060298549	12
INI	0.060110507	0.000495529	0.068537199	13
MLL-ETH [2]	0.063919883	0.000581741	0.079403258	14
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 a total of 38 teams participating in this challenge.

the leader board.[2] We compared our method with the state-of-the-art methods with or without postprocessing separately. Furthermore, we conducted extensive experiments with ablation studies to probe the performance gain in our method and detail as follows.

10.4.3.3 Results comparison without postprocessing

Preliminary encouraging results were achieved by the *IDSIA* team [11], which utilized a deep convolutional neural network as a pixelwise classifier in a sliding window way. The best results were obtained by averaging 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 multisize windows (i.e., different receptive fields in upsampling layers) into one unified framework. This can help to generate more accurate probability maps by leveraging multilevel contextual informa-

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

tion. The *Image Analysis Lab Freiburg* team [51] designed a deep U-shaped network by concatenating features from lower layers and improved the results of [11]. This further demonstrated the effectiveness of contextual information for accurate segmentation. However, with such a deep network (i.e., 23 convolutional layers), the backpropagation 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 multidimensional long short-term memory model for fast volumetric segmentation [49]. Unfortunately, a relatively inferior performance was achieved by this method. From Table 10.1, 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)* averaging 4 models (the same number of models as in [11]) achieved much smaller Rand and warping errors than other teams also employing deep learning methods without sophisticated postprocessing steps, such as *DIVE*, *IDSIA*, and *Image Analysis Lab Freiburg*. This corroborates the superiority of our approach by exploring multilevel contextual information with auxiliary supervision.

10.4.3.4 Results comparison with postprocessing

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 [52] to refine the contour. The final fusion result $p_f(x)$ was produced by fusing the binary contour $p_w(x)$ and original probability map $p(x)$ with a linear combination

$$p_f(x) = w_f p(x) + (1 - w_f) p_w(x). \tag{10.1}$$

The parameter w_f is determined by obtaining the optimal result of Rand error on the training data in our experiments. After fusing the results from watershed method (i.e., CUMedVision-4 (with fusion)), the Rand error can be reduced dramatically while unfortunately increasing the warping error. This is reasonable since these two errors consider the segmentation evaluation metric from different aspects. The former could penalize even slightly misplaced boundaries while the latter disregards nontopological errors. Different from our simple postprocessing step, the *SCI* team postprocessed the probability maps generated by the team *DIVE* and *IDSIA* with a sophisticated postprocessing strategy [22]. The postprocessed 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 postprocessing. In addition, our method also outperformed other methods with sophisticated postprocessing techniques including *optree-idsia* and *motif* by a large margin. This further highlights the advantages of our method by exploring multilevel 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 postprocessing strategies.[3]

10.4.3.5 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 10.1, compared with methods using single contextual information including *CUMedVision-4 (C3/C2/C1)*, the deep contextual model harnessing the multilevel 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 and pixel errors 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 backpropagation of gradient flow. By fusing the results from the watershed method, we achieved the result with Rand error of 0.017334, warping error of 0.000188, and pixel error of 0.057953, outperforming 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 real-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.

10.4.4 2015 MICCAI nuclei segmentation challenge

10.4.4.1 Qualitative evaluation

Some segmentation examples of testing data from 2015 MICCAI nuclei segmentation challenge can be seen in Fig. 10.5. We can see that our method can accurately segment the nuclei from pathology images. Some touching nuclei can be further split with postprocessing steps such as a watershed algorithm.

10.4.4.2 Quantitative evaluation metrics

The nuclei segmentation challenge employed two metrics for evaluation: traditional Dice coefficient and object-level Dice coefficient. The Dice metric was applied to measure the amount of overlap between the results of algorithms and human annotations

[3] See http://appsrv.cse.cuhk.edu.hk%7Ehchen/research/2012isbi_seg.html.

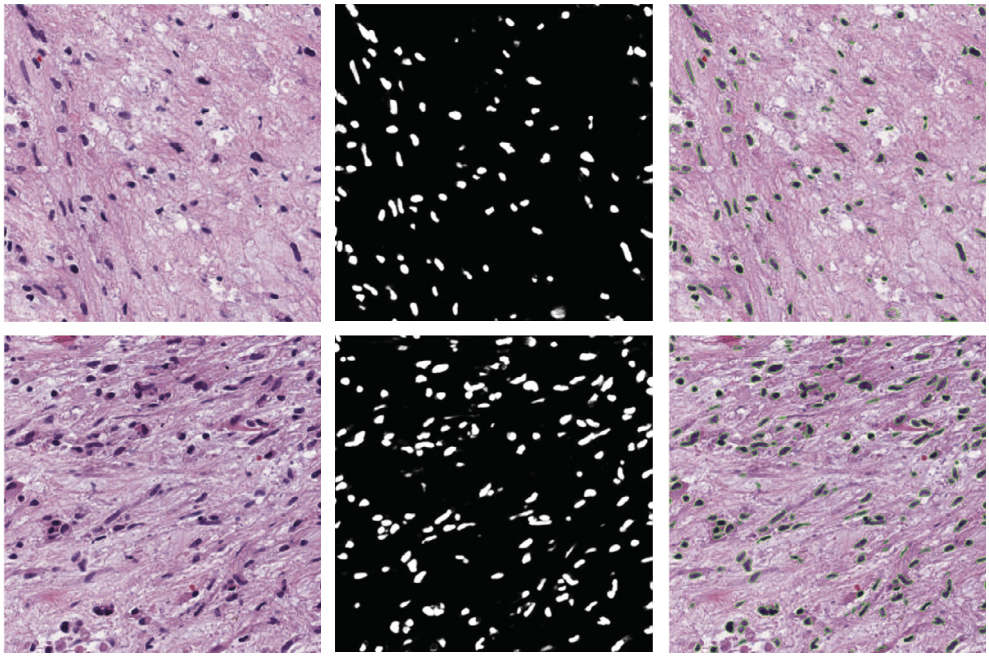

Figure 10.5 Examples of nuclei segmentation results: original images (left), probability maps (middle), and segmentation results by our method (right).

in terms of the nuclei regions that were detected and segmented. Dice metric does not take into account the cases of split and merge. A split is the case in which the human segments a region in a single nucleus, but the algorithm segments the same region in multiple nuclei. A merge is the case in which the algorithm segments a region into a single nucleus, but the human segments the same region into multiple nuclei. Object-level Dice coefficient is calculated based on the object-level segmentation, which provides a measure of splits and merges. Readers can refer to the challenge website[4] to learn more details of Dice and object-level Dice coefficients. The final ranking score was made by considering the average of Dice and object-level Dice coefficients.

10.4.4.3 Quantitative results and comparison

The quantitative results of our method and comparison with other methods can be seen in Table 10.2. Our method achieved the highest Dice score and outperformed other methods by a large margin, which demonstrates the efficacy and generalization capability of our proposed method quantitatively.

[4] 2015 MICCAI nuclei segmentation challenge: http://miccai.cloudapp.net:8000/competitions/37.

Table 10.2 Results of testing data in 2015 MICCAI nuclei segmentation challenge.

Team	Dice	Object-level dice	Score	Ranking
Our team	**0.877**	**0.722**	**0.799**	**1**
Team2	0.826	0.694	0.760	2
Team3	0.792	0.642	0.717	3
Team4	0.813	0.617	0.715	4

10.4.5 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 in previous studies [11,16] 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), 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.

10.5. Discussion and conclusion

In this paper we have presented a deeply supervised contextual neural network for biomedical image segmentation. By harnessing the multilevel contextual information from the deep hierarchical feature representations, it can have better discrimination and localization abilities, which are key to biomedical image segmentation related tasks. The injected auxiliary classifiers can encourage the backpropagation of gradient flow in training the deep neural network, thus further improving the segmentation performance. Extensive experiments on the public dataset of 2012 ISBI EM Segmentation Challenge and 2015 MICCAI Nuclei Segmentation Challenge corroborated the effectiveness and generalization capability of our method. In addition, our approach is general and can be easily extended to other biomedical applications. Future work will include further refining of the segmentation results with other sophisticated postprocessing techniques [21–23] and more biomedical applications.

Acknowledgment

This work is supported by Hong Kong RGC General Research Fund (No. CUHK412513) and Shenzhen-Hong Kong Innovation Circle Funding Program (No. SGLH20131010151755080 and GHP/002/13SZ). The authors gratefully thank the challenge organizers for providing datasets and helping the evaluation.

References

[1] Olaf Sporns, Giulio Tononi, Rolf Kötter, The human connectome: a structural description of the human brain, PLoS Computational Biology 1 (4) (2005) e42.

[2] Dmitry Laptev, Alexander Vezhnevets, Sarvesh Dwivedi, Joachim M. Buhmann, Anisotropic ssTEM image segmentation using dense correspondence across sections, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, Springer, 2012, pp. 323–330.

[3] Albert Cardona, Stephan Saalfeld, Stephan Preibisch, Benjamin Schmid, Anchi Cheng, Jim Pulokas, Pavel Tomancak, Volker Hartenstein, An integrated micro- and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy, PLoS Biology 8 (10) (2010) 2564.

[4] Elizabeth Jurrus, Antonio R.C. Paiva, Shigeki Watanabe, James R. Anderson, Bryan W. Jones, Ross T. Whitaker, Erik M. Jorgensen, Robert E. Marc, Tolga Tasdizen, Detection of neuron membranes in electron microscopy images using a serial neural network architecture, Medical Image Analysis 14 (6) (2010) 770–783.

[5] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[6] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv:1409.1556, 2014.

[7] Hao Chen, Chiyao Shen, Jing Qin, Dong Ni, Lin Shi, Jack CY Cheng, Pheng-Ann Heng, 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, 2015, pp. 515–522.

[8] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A.W.M. Van Der Laak, Meyke Hermsen, Quirine F. Manson, Maschenka Balkenhol, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, JAMA 318 (22) (2017) 2199–2210.

[9] Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully convolutional networks for semantic segmentation, preprint, arXiv:1411.4038, 2014.

[10] Korsuk Sirinukunwattana, Josien P.W. Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Gu, Li Yang Wang, Bogdan J. Matuszewski, Elia Bruni, Urko Sanchez, et al., Gland segmentation in colon histology images: the GlaS challenge contest, Medical Image Analysis 35 (2017) 489–502.

[11] Dan Ciresan, Alessandro Giusti, Luca M. Gambardella, Jürgen Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in Neural Information Processing Systems, 2012, pp. 2843–2851.

[12] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, Zhuowen Tu, Deeply-supervised nets, preprint, arXiv:1409.5185, 2014.

[13] Arganda-Carreras Ignacio, Seung Sebastian, Cardona Albert, Schindelin Johannes, ISBI challenge: segmentation of neuronal structures in EM stacks, http://brainiac2.mit.edu/isbi_challenge/, 2012.

[14] J.G. White, E. Southgate, J.N. Thomson, S. Brenner, The structure of the nervous system of the nematode Caenorhabditis elegans: the mind of a worm, Philosophical Transactions of the Royal Society of London 314 (1986) 1–340.

[15] Davi D.Bock, Wei-Chung Allen Lee, Aaron M. Kerlin, Mark L. Andermann, Greg Hood, Arthur W. Wetzel, Sergey Yurgenson, Edward R. Soucy, Hyon Suk Kim, R. Clay Reid, Network anatomy and in vivo physiology of visual cortical neurons, Nature 471 (7337) (2011) 177–182.

[16] Xundong Wu, An iterative convolutional neural network algorithm improves electron microscopy image segmentation, preprint, arXiv:1506.05849, 2015.

[17] H. Sebastian Seung, Neuroscience: towards functional connectomics, Nature 471 (7337) (2011) 170–172.

[18] Ritwik Kumar, Amelio Vázquez-Reina, Hanspeter Pfister, Radon-like features and their application to connectomics, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, IEEE, 2010, pp. 186–193.
[19] Yuriy Mishchenko, Automation of 3D reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs, Journal of Neuroscience Methods 176 (2) (2009) 276–289.
[20] Verena Kaynig, Thomas J. Fuchs, Joachim M. Buhmann, Geometrical consistent 3D tracing of neuronal processes in ssTEM data, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010, Springer, 2010, pp. 209–216.
[21] Mustafa Gökhan Uzunbaş, Chao Chen, Dimitris Metaxsas, Optree: a learning-based adaptive watershed algorithm for neuron segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014, Springer, 2014, pp. 97–105.
[22] TingLiu, Cory Jones, Mojtaba Seyedhosseini, Tolga Tasdizen, A modular hierarchical approach to 3D electron microscopy image segmentation, Journal of Neuroscience Methods 226 (2014) 88–102.
[23] Juan Nunez-Iglesias, Ryan Kennedy, Toufiq Parag, Jianbo Shi, Dmitri B. Chklovskii, Xi-Nian Zuo, Machine learning of hierarchical clustering to segment 2D and 3D images, PLoS ONE 8 (8) (2013) 08.
[24] VirenJain, Benjamin Bollmann, Mark Richardson, Daniel R. Berger, Moritz N. Helmstaedter, Kevin L. Briggman, Winfried Denk, Jared B. Bowden, John M. Mendenhall, Wickliffe C. Abraham, et al., Boundary learning by optimization with topological constraints, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2010, pp. 2488–2495.
[25] Metin N. Gurcan, Laura E. Boucheron, Ali Can, Anant Madabhushi, Nasir M. Rajpoot, Bulent Yener, Histopathological image analysis: a review, IEEE Reviews in Biomedical Engineering 2 (2009) 147–171.
[26] Shivang Naik, Scott Doyle, Shannon Agner, Anant Madabhushi, Michael Feldman, John Tomaszewski, Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology, in: 5th IEEE International Symposium on Biomedical Imaging, IEEE, 2008, pp. 284–287.
[27] Jun Xu, Lei Xiang, Qinshan Liu, Hannah Gilmore, Jianzhong Wu, Jinghai Tang, Anant Madabhushi, Stacked Sparse Autoencoder (SSAE) for nuclei detection on breast cancer histopathology images, IEEE Transactions on Medical Imaging 35 (2016) 119–130.
[28] Hao Chen, Xiaojuan Qi, Lequan Yu, Pheng-Ann Heng, DCAN: deep contour-aware networks for accurate gland segmentation, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2016, pp. 2487–2496.
[29] Michael Stierer, Harald Rosen, Renate Weber, Nuclear pleomorphism, a strong prognostic factor in axillary node-negative small invasive breast cancer, Breast Cancer Research and Treatment 20 (2) (1991) 109–116.
[30] B. Dunne, J.J. Going, Scoring nuclear pleomorphism in breast cancer, Histopathology 39 (3) (2001) 259–265.
[31] Christopher W. Elston, Ian O. Ellis, Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up, Histopathology 19 (5) (1991) 403–410.
[32] Humayun Irshad, et al., Automated mitosis detection in histopathology using morphological and multi-channel statistics features, Journal of Pathology Informatics 4 (1) (2013) 10.
[33] Chanho Jung, Changick Kim, Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization, IEEE Transactions on Biomedical Engineering 57 (10) (2010) 2600–2604.
[34] Stephan Wienert, Daniel Heim, Kai Saeger, Albrecht Stenzinger, Michael Beil, Peter Hufnagl, Manfred Dietel, Carsten Denkert, Frederick Klauschen, Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach, Scientific Reports 2 (2012).

[35] Mitko Veta, A. Huisman, Max A. Viergever, Paul J. van Diest, Josien P.W. Pluim, Marker-controlled watershed segmentation of nuclei in H&E stained breast cancer biopsy images, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2011, pp. 618–621.
[36] Yousef Al-Kofahi, Wiem Lassoued, William Lee, Badrinath Roysam, Improved automatic detection and segmentation of cell nuclei in histopathology images, IEEE Transactions on Bio-Medical Engineering 57 (4) (2010) 841–852.
[37] Kien Nguyen, Anil K. Jain, Bikash Sabata, et al., Prostate cancer detection: fusion of cytological and textural features, Journal of Pathology Informatics 2 (2) (2011) 3.
[38] Humayun Irshad, Antoine Veillard, Ludovic Roux, Daniel Racoceanu, Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential, IEEE Reviews in Biomedical Engineering 7 (2014) 97–114.
[39] Dan C. Cireşan, Alessandro Giusti, Luca M. Gambardella, Jürgen Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Springer, 2013, pp. 411–418.
[40] Ludovic Roux, Daniel Racoceanu, Nicolas Loménie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frédérique Capron, Catherine Genestie, Gilles Le Naour, Metin N. Gurcan, Mitosis detection in breast cancer histological images, an ICPR 2012 contest, Journal of Pathology Informatics 4 (2013).
[41] Mitko Veta, Paul J. Van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders B.L. Larsen, Jacob S. Vestergaard, Anders B. Dahl, et al., Assessment of algorithms for mitosis detection in breast cancer histopathology images, Medical Image Analysis 20 (1) (2015) 237–248.
[42] Hao Chen, Qi Dou, Xi Wang, Jing Qin, Pheng Ann Heng, Mitosis detection in breast cancer histology images via deep cascaded networks, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[43] Korsuk Sirinukunwattana, Shan Raza, Yee-Wah Tsang, David Snead, Ian Cree, Nasir Rajpoot, Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images, IEEE Transactions on Medical Imaging (2016) 1196–1206.
[44] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected crfs, preprint, arXiv:1412.7062, 2014.
[45] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al., Greedy layer-wise training of deep networks, Advances in Neural Information Processing Systems 19 (2007) 153.
[46] Liwei Wang, Chen-Yu Lee, Zhuowen Tu, Svetlana Lazebnik, Training deeper convolutional networks with deep supervision, preprint, arXiv:1505.02496, 2015.
[47] Albert Cardona, Stephan Saalfeld, Johannes Schindelin, Ignacio Arganda-Carreras, Stephan Preibisch, Mark Longair, Pavel Tomancak, Volker Hartenstein, Rodney J. Douglas, TrakEM2 software for neural circuit reconstruction, PLoS ONE 7 (6) (2012) e38011.
[48] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell, Caffe: convolutional architecture for fast feature embedding, preprint, arXiv:1408.5093, 2014.
[49] Marijn F. Stollenga, Wonmin Byeon, Marcus Liwicki, Juergen Schmidhuber, Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation, preprint, arXiv:1506.07452, 2015.
[50] William M. Rand, Objective criteria for the evaluation of clustering methods, Journal of the American Statistical Association 66 (336) (1971) 846–850.
[51] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: convolutional networks for biomedical image segmentation, preprint, arXiv:1505.04597, 2015.
[52] S. Beucher, C. Lantuejoul, Use of watersheds in contour detection, in: International Conference on Image Processing, 1979.