# DENSELY CONNECTED LARGE KERNEL CONVOLUTIONAL NETWORK FOR SEMANTIC MEMBRANE SEGMENTATION IN MICROSCOPY IMAGES

*Dongnan Liu*[⋆], *Donghao Zhang*[⋆], *Siqi Liu*[⋆], *Yang Song*[⋆], *Haozhe Jia*[†], *Dagan Feng*[⋆], *Yong Xia*[†], *Weidong Cai*[⋆]

[⋆] School of Information Technologies, University of Sydney, Australia
[†] School of Computer Science and Engineering, Northwestern Polytechnical University, China

## ABSTRACT

Structural analysis of neurons can provide valuable insights of brain function. Semantic segmentation of neurons thus becomes an important technique in bioinformatics. Deep learning approaches have shown promising performance in various semantic segmentation problems. However, segmentation of neurons in Electron Microscopy (EM) images has some differences compared with typical segmentation tasks due to the image noise and the disturbance of the intracellular structures. In our work, we propose a network with a ResNet encoder and densely connected decoder with large kernels, and then refinement with simple morphological post-possessing. Two main advantages of our method are: 1) the network can prevent the loss of high-resolution information and enlarge the reception field; 2) the post-processing method is simple and can be directly applied to the probability map from the network to enhance the unconfident area. Evaluated on the ISBI2012 EM membrane segmentation challenge, the proposed method achieves competitive performance.

***Index Terms***— neuronal boundary segmentation, electron microscopy image, deep neural network

## 1. INTRODUCTION

Understanding the anatomical connections of our brain plays an important role in dense circuit reconstruction [1]. However, the relationship between the structure and functionality of the nervous system is much more complicated to understand than the other organ systems [2]. Recently, the serial section Transmission Electron Microscopy (ssTEM) has become a widely used tool to learn about the relationship of structure and functionality of neuron connections. A key step in the study of EM images is to obtain the segmentation of neuron membranes as shown in Fig. 1. This generation of these boundary maps can be approached as a semantic segmentation problem.

Recently, several methods based on deep convolutional neural networks (CNN) have been proposed for semantic segmentation in general imaging. For example, SegNet [3] is a fully convolutional deep architecture. In the SegNet, the feature maps from the encoder are upsampled by the decoder with the transferred maxpooling indices to improve the segmentation resolution. However, when the image is resized larger, the small convolutional kernel in the network will produce a small receptive field which results in information loss. Therefore a global convolutional network (GCN) [4] with very large kernel size is proposed to ensure the receptive field is large enough during the whole process. However, these methods are not directly suitable for neuron segmentation due to the different imaging characteristics of EM images. First, there exist many vesciles (small bubbles) around the membranes, which make boundaries unclear as shown in Fig. 1. Secondly, due to the intracellular structures such as mitochondria (dark shadows), simple threshold and edge detection methods will become ineffective when detecting the neuron membrane [5].

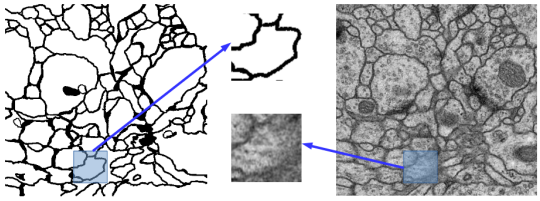

**Fig. 1**. Left: the binary segmentation for EM images (0 for membrane and 1 for the background). Right: the original EM image.

In the EM segmentation domain, some studies have shown promising performance with customized CNN architectures, such as Pyramid-LSTM [6], optree [7], DIVE [8] and M2FCN [9]. Pyramid-LSTM is based on Multi-Dimensional LSTM which changes its traditional cuboid order of computations into pyramidal style. Different from Pyramid-LSTM, optree heavily relies on a post-processing method with a tree-like structure watershed segmentation. DIVE is a model constructed with a deep neural network (DNN) pixel classifier and a post-processing method based on random forests. M2FCN is a CNN which contains several convolutional stages. All the side outputs in one stage concatenated with the original image are fed into the next stage

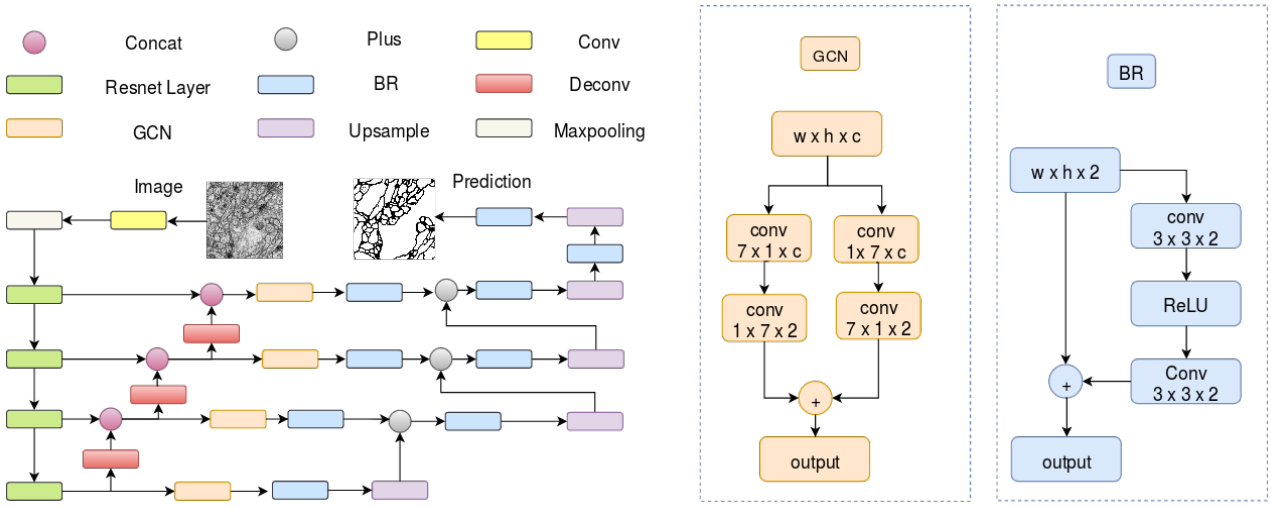


**Fig. 2**. The overall framework of our method. The kernel size of the first Conv layer after the input is 7. We use ResNet152 as the ResNet layers encoders. A Deconv layer contains a Transpose Convolutional layer and a ReLU layer. The upsampling part represents the nearest upsampling layers.

as input. While these methods achieve high performances in ISBI2012 EM segmentation challenge [10], some problems such as the receptive field being too small and high-resolution information loss still need to be solved.

In this study, we propose to incorporate densely connected parts on the global convolutional network. The new network can process the high resolution feature together with the low resolution one with a large kernel. It can not only minimize high resolution information loss, but also ensure the receptive field is large enough when the original image is resized larger. Besides the network, we also design a post-processing method which combines the final probability map with its corresponding binary mask to refine the boundaries. Our post-processing method is easy to operate because it can be directly applied to the probability map from the network and its computation load is small. Our proposed method achieves a competitive ranking on the ISBI2012 EM segmentation challenge.

## 2. METHODS

### 2.1. Densely Connected Global Convolutional Network

In this part, we will present a densely connected global convolutional network for this EM membrane segmentation task. First we consider that if the size of the receptive field is too small, and the input image is large, some information in the receptive field will be lost and this will reduce the segmentation accuracy. Therefore the receptive field should be large enough and the large kernel matter method is necessary in the EM segmentation task. However, as the structure shown in Fig. 2, the feature map from the high resolution layer will be handled by several convolution models, some of the information will be lost during this process when coming up. Thus we need to make sure there is enough high-resolution information remaining after the sampling transformation. One way to solve this problem is to combine the high-resolution feature map with the low-resolution one. In this way, the information in the high-resolution feature map can remain at the very last layers due to the network structure. In our model, we choose to combine the high-resolution feature with the lower-resolution directly after the encoder. If we do the combination after the convolution model, as some information may be lost after the transformation of the convolutional kernels.

The model we proposed is shown in Fig. 2. In the encoder part, we choose the ResNet152 [11] layers without pre-training. Although some articles [12, 4] prefer to use the ResNet model which is pre-trained on ImageNet, such methods did not work well in this particular task. As mentioned above, there are noises which blur the boundary between the neuronal membrane and the intracellular part. The pre-trained ResNet increased the precision of the boundary detection, but introduces misclassification of the intracellular part at the same time.

In the decoder part, we choose the global convolutional network modules (GCN) and the boundary refinement modules (BR) [4]. In order to ensure a large receptive field, a large size kernel is required instead of the ordinary small one directly after each encoder. However, if we directly use a $K \times K$ kernel, the number of parameters in the network will become extremely large. GCN addresses the parameters problem by using one $1 \times K$ kernel with one $K \times 1$ kernel instead of a $K \times K$ kernel. Two branches which combines the two 1D kernels permuted in different orders are summed together to

produce the output of the GCN module. The feature map from each encoder will be sent to the GCN with kernel sizes of 31, 15, 9, 7 respectively. Then the output of GCN will be fed into the BR modules. Each BR module shown in Fig. 2 is a residual structure which contains two $3 \times 3$ kernels with a ReLU activation function between them. As for the upscale part, we use upsampling kernels instead of traditional deconvolutional kernels.

In the connection between encoder and decoder, we propose a densely connected style to combine the high-resolution information together with the low-resolution information. This idea is inspired by the densely connected network [13], which is constructed by several small dense blocks. The $(n+1)th$ layer of a dense block combines all the preceding n feature maps together. Denoting the preceding layers as $x_1, ..., x_n$, the $(n+1)th$ layer will be:

$$x_{n+1} = F_n[x_1, ..., x_n] \tag{1}$$

where the $[x_1, ..., x_n]$ means that the preceding features are concatenated together. In our network, we use a deconvolution layer to upscale the high-resolution feature map from the ResNet layer encoder. Then we connect them in the same style as dense block for the decoder in Fig. 2. In this way the information from high-resolution feature map of the encoder remains after the up-scaling layer of the decoder.

### 2.2. Morphological Post-processing

Although the probability map from the network reveals many details, there are flaws and noises around some boundaries, which may reduce the segmentation accuracy. Therefore, we propose a post-processing method which can reduce such noises to make the boundary clearer.

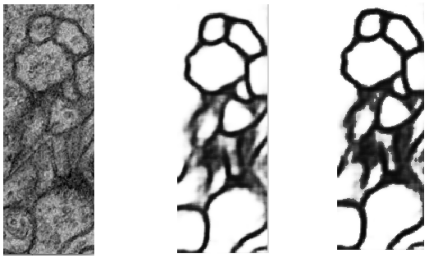

**Fig. 3**. An example of the effect of our post-processing technique. From left to right: part of the original image from test set; corresponding segmentation probability map; probability map after post-processing.

First we use a threshold value of 0.5 to get the binary image of the original probability map $P_o$. Second, we compute the exact Euclidean distance transform of the binary image to get a distance map. Based on this, we set a distance value of 1 as a threshold to get another binary image $P_b$. Then, we perform a linear combination between the binary image $P_b$ and the original probability map $P_o$ to generate the final result $P_f$:

$$P_f = kP_o + (1-k)P_b \tag{2}$$

where $k$ is a parameter which controls the ratio of the original probability map to the binary image in the final result. In our experiment, we set $k$ as 0.62 based on our empirical studies. The effect of this post-processing is shown in Fig. 3.

## 3. EXPERIMENTAL RESULTS AND DISCUSSIONS

In our experiment, we use the dataset from EM segmentation challenge. The dataset of this challenge is from the first instar larva ventral nerve cord of the Drosophila.

The training data with ground truth provided and testing data each contain 30 consecutive slices, with a size of $512 \times 512$ pixels and a resolution of $4 \times 4 \times 50$ nm/pixel [10]. Each pixel in the image can either belong to the boundary or the background. In addition, the resolution in the z-axis is lower than that of the xy-plane. So the common method is to obtain the boundary probability map of each slice separately.

In this dataset, there are only 30 slices of 2D images in the training data. Data augmentation is thus necessary for training. We tried flip, rotation, Gaussian blur, elastic transformation, random drop out and affine transformation. Each augmentation methods is operated on a randomly cropped image with a size of $256 \times 256$. Also, in order to enhance the robustness of the training data, online data augmentation at the beginning of each epoch is performed. After the experiment, we find out that Gaussian blur, affine transformation and elastic transformation are beneficial to the final result while image dropout is harmful. This is because the image dropout causes too much loss of image details. As for the optimiser, we use SGD and ADAM. In this experiment, ADAM outperforms the SGD optimiser. Then we use the Adam optimiser with a learning rate schedule, in which the learning rate of the current epoch is:

$$lr_{epoch} = \begin{cases} 0.0005 & \text{if } 0 < epoch < 101 \\ lr_{epoch-1} \times (1 - \frac{epoch-100}{epoch_{max}})^{power} & \text{otherwise} \end{cases} \tag{3}$$

where we set the $epoch_{max}$ as 50 and the $power$ as 0.9. When using a lower learning rate, the model is able to learn more details in the images and avoid local-minimum. In addition, from the ground truth of the training dataset, we find that the ratio of the boundary and the background is 1:4, which means there exists class imbalance which will make the segmentation less effective. Thus we assign a weight for each class in the cross-entropy loss.

In order to evaluate the effect of the methods, the ISBI2012 challenge proposed foreground-restricted Rand Scoring after border thinning (${V^{rand}}_{thinned}$) and Information Theoretic

Scoring after border thinning ($V^{Info}{}_{thinned}$) [10]. From the experiment in [10], $V^{rand}{}_{thinned}$ seems to be more robust than $V^{Info}{}_{thinned}$. In this way, the leader board is decided by the $V^{rand}{}_{thinned}$ of the result. In this challenge, all the results are evaluated by submitting them to the official website.

**Table 1**. The $V^{rand}{}_{thinned}$ evaluation of our model when there are no densely connected decoders or large kernels.

| large kernels? | ✗ | ✓ | ✗ | ✓ |
|---|---|---|---|---|
| dense connection? | ✗ | ✗ | ✓ | ✓ |
| | 0.9610 | 0.9639 | 0.9649 | 0.9739 |

Table 1 shows the evaluation of our model when there are no dense connected decoders or large kernels. We can see the model without the dense connected decoders and large kernels has the worst performance. By adding large kernels, the model can get a higher score because the larger receptive field in the network prevents details loss when the original images are enlarged. The dense connections also make the results better by keeping the high-resolution information remaining. If the model contains both dense connections and large kernels, it outperforms all the other methods. Thus the large kernels and dense connection are necessary in our method. Fig. 4 shows one slice of the test prediction. From the result we can see that intracellular mitochondria can be removed and some boundaries surrounded with vesicles can be clearly distinguished.

**Table 2**. Evalution on ISBI2012 challenge.

| Method | $V_{Rand}$ | $V_{Infor}$ |
|---|---|---|
| M2FCN [9] | 0.9780 | 0.9901 |
| Ours (with post-processing) | 0.9764 | 0.9858 |
| DIVE-SCI [8] | 0.9762 | 0.9874 |
| Ours (without post-processing) | 0.9739 | 0.9866 |
| IDSIA [14] | 0.9730 | 0.9874 |
| RotEqNet [15] | 0.9712 | 0.9790 |
| optree [7] | 0.9712 | 0.9849 |
| PolyMtl [16] | 0.9690 | 0.9860 |
| Pyramid-LSTM [6] | 0.9676 | 0.9829 |

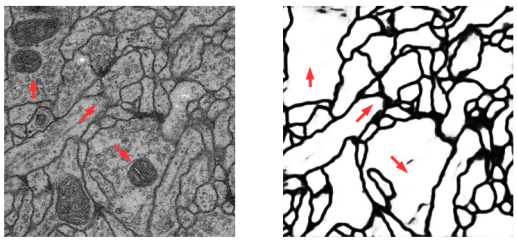

**Fig. 4**. Left: slice 27/30 of the test set. Right: corresponding prediction.

Table 2 shows the comparison of our proposed results and some other state-of-the-art results using $V^{rand}{}_{thinned}$ evaluation. There have been over 100 groups participating in this challenge and their methods vary a lot. As our model is based on the deep neural network, we choose to perform a more detailed comparison with approaches based on deep learning models. For the full leading board, please refer to the official website [1]. Comparing to the deep architecture in RotEqNet [15], IDSIA [14] and PolyMtl [16], our network with its densely connected decoders and large kernels performs better even without post-processing. As to the anisotropic property of this dataset, Pyramid-LSTM [6] which manipulated on the whole 3D dataset performs worse than some other 2D deep architecture including ours. Comparing to the complex post-processing of optree [7] and DIVE-SCI [8], our method outperforms them with a simple morphology post-processing. Although the method M2FCN [9] outperforms ours, its best structure is created by increasing the number of sub-convolutional layers in each convolutional stage in the network and increasing the number of the convolutional stages. This creates a large number of parameters which has a high cost of GPU memory. Our method contains only ResNet layers and large kernels with a few parameters, and is thus more memory efficient.

## 4. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a method for EM membrane segmentation with a densely connected large kernel global convolutional network. With the effects of large kernel of GCN and densely connected encoders, our model minimizes high-resolution information loss during training. The morphological post-processing algorithm also helps to make some blurred boundaries in the probability map clear. Our method achieves a comparative result in the ISBI2012 EM segmentation challenge. Compared to some other deep learning algorithms that require complex pre- or post-processing algorithms, our model mainly relies on the performance of the deep neural network, which shows the robustness and effectiveness of our methods. In the future work, we will implement our method on some larger ssTEM datasets to evaluate its robustness.

## 5. REFERENCES


[1] O. Sporns, G. Tononi, and R. Kötter, "The human connectome: a structural description of the human brain," *PLoS Computational Biology*, vol. 1, no. 4, pp. e42, 2005.

[2] 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.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image


[1] http://brainiac2.mit.edu

segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[4] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel mattersimprove semantic segmentation by global convolutional network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4353–4361.

[5] E. Jurrus, A. R. Paiva, S. Watanabe, J. R. Anderson, et al., "Detection of neuron membranes in electron microscopy images using a serial neural network architecture," *Medical image analysis*, vol. 14, no. 6, pp. 770–783, 2010.

[6] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," in *Advances in Neural Information Processing Systems (NIPS)*, 2015, pp. 2998–3006.

[7] M. Uzunbaş, C. Chen, and D. Metaxsas, "Optree: a learning-based adaptive watershed algorithm for neuron segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 97–105.

[8] A. Fakhry, H. Peng, and S. Ji, "Deep models for brain em image segmentation: novel insights and improved performance," *Bioinformatics*, vol. 32, no. 15, pp. 2352–2358, 2016.

[9] W. Shen, B. Wang, Y. Jiang, Y. Wang, et al., "Multi-stage multi-recursive-input fully convolutional networks for neuronal boundary detection," in *The IEEE International Conference on Computer Vision (ICCV)*, 2017.

[10] I. Arganda-Carreras, S.C. Turaga, D.R. Berger, D. Cireşan, et al., "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Frontiers in neuroanatomy*, vol. 9, 2015.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[12] S. Liu, D. Xu, S. K. Zhou, T. Mertelmeier, et al., "3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes," *arXiv preprint arXiv:1711.08580*, 2017.

[13] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[14] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems (NIPS)*, 2012, pp. 2843–2851.

[15] D. Marcos, M. Volpi, N. Komodakis, and D. Tuia, "Rotation equivariant vector field networks," in *The IEEE International Conference on Computer Vision (ICCV)*, 2017.

[16] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, et al., "The importance of skip connections in biomedical image segmentation," in *International Conference on Medical Image Comput-ing and Computer-Assisted Intervention Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, 2016, pp. 179–187.