

3D CNN-BASED SOMA SEGMENTATION FROM BRAIN IMAGES AT SINGLE-NEURON RESOLUTION

Meng Dong, Dong Liu, Zhiwei Xiong, Chaoyu Yang, Xuejin Chen, Zheng-Jun Zha, Guoqiang Bi, Feng Wu

University of Science and Technology of China

ABSTRACT

Neuron segmentation is an important task for automatic analyses of brain images that are of huge volume. Previous methods for neuron segmentation rely on handcrafted image features, and have difficulty in coping with high-resolution, low signal-to-noise-ratio brain images. Convolutional neural network (CNN) has achieved remarkable success in natural image segmentation, but CNN requires accurately labeled data for training that are difficult to achieve on brain images of huge volume. In this paper, we present a weakly supervised learning strategy to deal with the inaccurate training data problem, and thus adopt 3D CNN to perform automatic soma segmentation from brain images. We test our method on our own collected mouse brain images that are of single-neuron resolution, and results show that 3D CNN-based method outperforms the traditional methods by a significant margin.

Index Terms— Brain images, convolutional neural network, soma segmentation, weakly supervised learning

1. INTRODUCTION

Recent advances in biological microscopy imaging techniques have enabled the production of brain images at single-neuron resolution, which greatly facilitates the observation of brain neural morphology and its structure. For example, as shown in Fig. 1, we are now able to collect mouse brain images at the physical resolution of $0.5 \mu\text{m}^3/\text{voxel}$, which clearly enables the identification of every neuron. However, due to their high resolution, such brain images are of huge volume, e.g. one mouse brain produces around 3 TB images under our setting. Manually observing and processing these images are not possible. Thus, it is an urgent need to develop efficient automatic analysis methods for high-resolution brain images.

Segmenting neurons from brain images is an important analysis task, as it is the enabling technology of many following tasks including neuron morphologic analysis, neuron

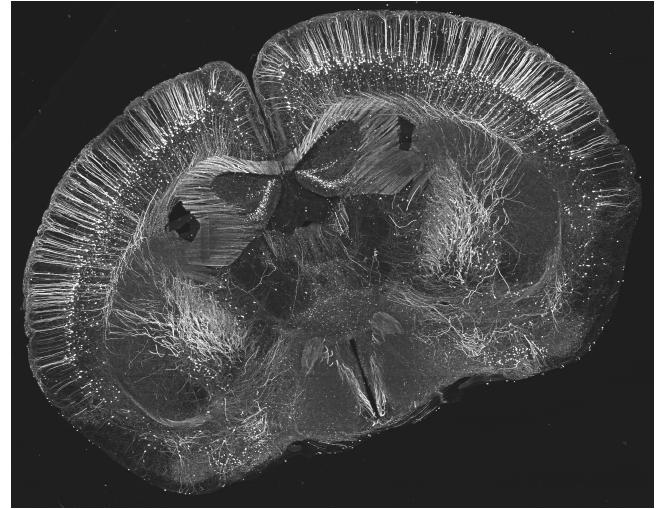


Fig. 1. An example of our own collected mouse brain images at single-neuron resolution.

counting, functional brain mapping, and so on. Although this problem has been studied for a long while, it remains a great challenge in neuroscience due to the following challenges: first, neurons in different brain functional regions vary very much in size, morphology, intensity, or density; second, the imaging process may introduce distortion or noise and produce images of low signal-to-noise ratio (SNR); third, the previous imaging techniques can only produce low-resolution images where neurons are not distinguishable. In addition, previous studies usually adopted handcrafted image features to perform segmentation, whilst such features are not robust for brain images that contain significant variance.

Recently, convolutional neural network (CNN) has achieved remarkable success in processing and understanding natural images. Also, several works have been done to adopt CNN for biomedical images and achieved certain success. When adopting CNN for brain images, there is a significant difficulty that CNN usually requires a lot of accurately labeled data for training, but manually labeling neurons from brain images is both labor and time expensive.

In this paper, we present a weakly supervised learning strategy to address the training data problem associated with

M. Dong, D. Liu, Z. Xiong, X. Chen, Z.-J. Zha, and F. Wu are with CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System. C. Yang and G. Bi are with School of Life Sciences. (*Corresponding author: Dong Liu.*)

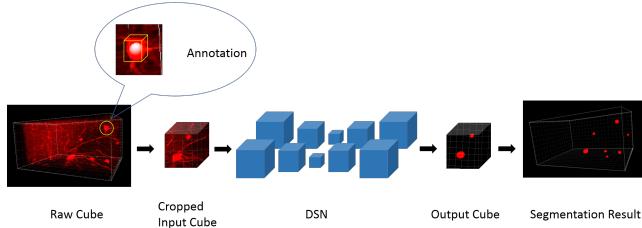


Fig. 2. Pipeline of the proposed 3D CNN-based soma segmentation.

CNN. We thus adopt a 3D CNN structure, known as Deeply Supervised Network (DSN), to perform soma segmentation from brain images. We build a weakly annotated dataset from our own collected mouse brain images, and test our method on the dataset in comparison with the traditional methods. Results show that our method outperforms the previous one [1, 2] by a significant margin, and ablation study verifies the effectiveness of our weakly supervised learning strategy.

2. RELATED WORK

There are many methods proposed for neuron detection or segmentation, most of which are based on traditional image processing algorithms, such as watershed [3, 4], energy functional approach [1, 2, 5], mean-shift clustering [6], and so on. Traditional algorithms depend on handcrafted features and carefully tuned parameters, and usually perform well only in the case of high signal-to-noise ratio (SNR).

In recent years, CNN has achieved remarkable success in processing and understanding natural images, such as classification [7, 8], object detection or recognition [9], semantic segmentation [10], and so on. CNN is advantageous in automatically learning the most discriminative features, which outperform handcrafted features in many tasks, especially in the case of low SNR.

There have been some works devoted to adopt CNN for biological/biomedical image analysis, and demonstrated the superiority of deep learning [11, 12]. For 3D images, a series of deep learning models have been proposed, such as 3D CNN-based cerebral microbleed detection [13], neuron segmentation based on CNN [14], joint sequence learning combined with LSTM for segmentation [15], 3D auto-encoder for segmenting cellular organization [16]. Recently, there are two 3D fully convolutional structures known as Deeply Supervised Network (DSN) [17] and DenseVoxNet[18], both achieve comparable or better performance than the state-of-the-art approaches in the challenge held in conjunction with MICCAI 2017. In this paper, we adopt the network structure of DSN to solve soma segmentation problem, but with a new design of an effective training strategy.

3. PROPOSED METHOD

The pipeline of our proposed 3D CNN-based soma segmentation is shown in Figure 2, which will be described in detail in the following.

3.1. Training Data

In previous works of CNN-based image segmentation, pixel-wise annotation is assumed for supervised learning. However such annotation is very laborious to obtain especially in our task, as neurons have diverse sizes, shapes, and densities. We simplify the annotation by marking each soma as a sphere, i.e. each soma is associated with a label of its center (3D coordinates) as well as its diameter. Such annotation strategy can save considerable labor and time.

Since the annotation is not accurate, if we naively set all the voxels inside spheres as foreground (soma) and those outside as background (not soma), then the CNN may be misled during training. Especially, the voxels at boundary of somas cannot be identified correctly by the CNN. To address this problem, we convert the data label into voxel-wise annotation in the following manner: if a voxel is inside a sphere, it belongs to foreground; or if it is distant enough from any sphere, it belongs to background; otherwise, it is ambiguous. For the ambiguous voxels, we “do not care” whether their segmentation result is correct or not during training, i.e. these voxels are excluded when calculating loss. Figure 2 shows an example 3D patch, where the voxels inside the white sphere are foreground, and the others inside the yellow cube are regarded ambiguous.

3.2. Network Structure Analysis

For volumetric images like our data, it is beneficial to extract features across three dimensions. Previous researches like [14, 15] also investigated biomedical segmentation problems with CNN on volumetric data, but their models are based on 2D CNN that extracts features from every image individually, and performs fusion later. 2D CNN may not be able to make full use of the 3D information and relation among image sequences. Since somas vary in size, morphology, and intensity, per our experience, we need to distinguish somas in different views of 3D images. Hence we choose a 3D CNN for our usage. The chosen DSN network includes convolutional layers, pooling layers and deconvolutional layers, all in 3D fashion. Convolutional layers and pooling layers act as feature extractor, while deconvolutional layers followed by softmax layer aim to up-sample the feature cubes to the same size as input. In the DSN, there are two more branches at the shallower layers in order to avoid gradient vanishing. Please refer to [17] for more details.

3.3. Training Strategy

One problem of training 3D CNN is memory limitation because the volumetric feature cubes are huge with respect to the input size. Therefore, we **crop the 3D brain images into smaller cubes with appropriate overlaps between cubes**. We try to use maximal possible cubes to avoid splitting a soma into two cubes, so we set batch size to 1 during training.

Moreover, as the somas are quite sparse in brain images, non-soma area is much larger. This brings the data imbalance problem and can influence the performance of CNN-based training. We use a **data balancing technique** to address the problem. Specifically, for each cube, we randomly select a certain percent of background voxels as non-soma, while the rest background voxels are labeled as “do not care,” i.e. not considered in computing loss. Further, only those cubes that contain more than 2,000 soma voxels are kept as training data.

Last but not the least, due to the imaging process, the 3D images in our dataset are not uniform across different dimensions. Especially, there was interpolated voxels along one dimension but not along the others. To eliminate this effect, we **perform random transposing of the data cubes as data augmentation**.

4. EXPERIMENTS

4.1. Implementation

In this experiment, the raw volumetric data are of $512 \times 869 \times 1024$ in 16-bit depth. We manually annotate the somas in these data, and there are more than 7,000 somas, whose radius ranges from less than 4 μm to around 30 μm . We crop them into 191 cubes of $160 \times 160 \times 160$ with more than 30 μm overlaps between adjacent cubes. 52 cubes are randomly selected for testing and the others are kept for training. To balance the ratio between soma and non-soma voxels, we randomly keep non-soma voxels to make their amount is ten times that of soma voxels, while the rest background voxels are not considered during training. In addition, we perform a common pre-processing, to **subtract mean and divide by the standard deviation on the images**.

The DSN model is implemented by Caffe and its MATLAB interface. The model is trained from random initialization with Gaussian distribution whose mean is 0 and variance is 0.01. The optimization is finished with stochastic gradient descent. The base learning rate is 0.001 and descended with “poly” policy, which means at “iter” iteration the learning rate multiplies $(1 - \frac{\text{iter}}{\text{max_iter}})^{\text{power}}$ where max_iter and power equal 15,000 and 0.9, respectively. All the experiments are completed on a GeForce GTX 1080Ti.

4.2. Test Protocol

To the best of our knowledge, the most relevant work to ours is **NeuroGPS (NGPS)** [1, 2], whose software is available to

Table 1. Quantitative comparison results of soma segmentation on our test data

Method	Jaccard	Dice	Precision	Recall
NGPS [1]	0.6103	0.7505	0.7332	0.7983
Ours	0.6811	0.7910	0.7870	0.8329

us, so we compare our method with NGPS.

We test DSN with the final model that is trained after 15,000 iterations. Each test image is also cropped to the same size as in the training procedure. The segmentation results of cubes are summed up by simple averaging in the overlapped area. And we perform a simple post-processing to **remove those too small detected regions as they are unlikely somas**.

When testing NGPS, we use the author provided software. All the test images are tested with the same parameters: binarization threshold equals to 6.0, and minimum radius equals to 4.0. These parameters had been carefully tuned to ensure the best achievable result of NGPS on our data. Please note that NGPS detects somas as spheres, so it outputs the center and radius of each detected soma.

We employ four commonly used metrics to quantitatively evaluate the performance of our method and NGPS: Jaccard, Dice, Precision, and Recall. Jaccard is also known as IoU, i.e. the *Intersection* between prediction result and ground truth *Overing the Union* of them. The four metrics are defined as follows:

$$\text{Jaccard}(R, G) = \frac{|R \cap G|}{|R \cup G|}, \text{Dice}(R, G) = \frac{2|R \cap G|}{|R| + |G|} \quad (1)$$

$$\text{Precision}(R, G) = \frac{|R \cap G|}{|R|}, \text{Recall}(R, G) = \frac{|R \cap G|}{|G|} \quad (2)$$

where R denotes segmentation result and G denotes ground truth. It is worth noting that the ground truth is also labeled with the sphere approximation, i.e. each soma is recorded as a sphere.

4.3. Results

The quantitative results of both methods are shown in Table 1. It can be observed that our method is significantly superior to NGPS.

We perform visual inspection of the segmentation results. Some examples are shown in Figure 3. We can see that our method outperforms NGPS in both sparse and dense cases. For example, in (a) and (b), the somas are too dense to identify, but we still can observe that the segmented somas by our model are more similar to the ground truth. In (a), there are some false positives in the result of NGPS. In (c), (d), and (e), somas are much sparse, where NGPS has difficulty in detecting somas at brain boundary or at regions of low SNR.

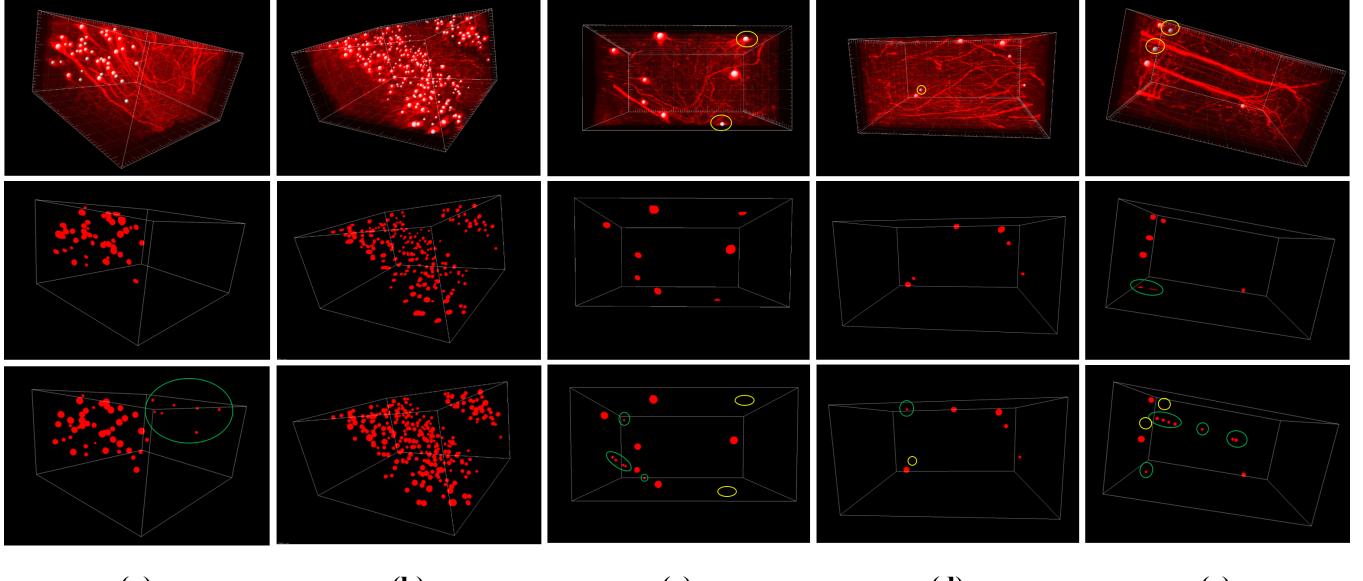


Fig. 3. Visual comparison of the segmentation results. From top to bottom: the original image with ground-truth labels, each soma is denoted by a white sphere, yellow circles indicate some difficult cases; the segmentation result by our method; the segmentation result by NGPS. In the segmentation results, green circles indicate false positive, yellow circles indicate false negative (i.e. missing).

Table 2. Ablation study results

Balancing	Transposing	Jaccard	Dice	Precision	Recall
	✓	0.5327	0.6615	0.8987	0.5621
✓		0.6738	0.7862	0.7744	0.8419
✓	✓	0.6811	0.7910	0.7870	0.8329

NGPS may mistake axons, dendrites, or other tissues as somas, which results in false positives. On the contrary, our approach can work well in such conditions.

It is interesting to observe from Fig. 3 that the segmented somas of our method have different shapes and sizes. This is not trivial because our DSN is trained with simplified labels that assume somas to be spheres. However, CNN-based learning has the capacity to extract both low-level and high-level features to identify the concrete morphology of somas. On the contrary, NGPS always uses spheres to fit somas and may easily fail in the regions of low SNR.

4.4. Ablation Study

In order to validate the efficiency of our proposed training strategy, we perform ablation study whose results are shown in Table 2. Balancing means whether to perform data balancing, i.e. to control the number of non-soma voxels during training. Transposing means whether to perform random transposing as data augmentation. From the results, we can see that both operations improve the performance of our model in terms of Jaccard and Dice. It is worth noting that,

when not using data balancing, the Precision can be very high because CNN will incline to identify voxels as non-soma and avoid false positive, but then the Recall is very low. Thus, using data balancing provides a better trade-off between Precision and Recall, resulting in superior performance.

5. CONCLUSION

In this paper, we propose a 3D CNN-based method to perform soma segmentation from brain images at high resolution. We annotate the collected dataset with simplified labels as weak supervision, but experiments verify that the trained CNN still can predict the concrete surface of somas. Such annotating manner saves considerable labor and time. Compared with the traditional method–NeuroGPS, our model performs obviously better on our dataset. The proposed method can be extended to neuron counting, brain functional mapping, neuron reconstruction, and so on. Our future work will investigate other network structures for soma segmentation, and focus on the difficult cases of dense somas.

Acknowledgment

This work was supported by the National Key Research and Development Plan of China under Grant 2017YFB1300201, the Natural Science Foundation of China (NSFC) under Grants 61331017, 91732304, 61622211, 61472392, and 61620106009, and the Fundamental Research Funds for the Central Universities under Grant WK2380000002.

6. REFERENCES

- [1] Tingwei Quan, Ting Zheng, Zhongqing Yang, Wenxiang Ding, Shiwei Li, Jing Li, Hang Zhou, Qingming Luo, Hui Gong, and Shaoqun Zeng, “NeuroGPS: automated localization of neurons for brain circuits using L1 minimization model,” *Scientific Reports*, vol. 3, pp. 1414, 2013.
- [2] Tingwei Quan, Hang Zhou, Jing Li, Shiwei Li, Anan Li, Yuxin Li, Xiaohua Lv, Qingming Luo, Hui Gong, and Shaoqun Zeng, “NeuroGPS-Tree: automatic reconstruction of large-scale neuronal populations with dense neurites,” *Nature Methods*, vol. 13, pp. 51, 2016.
- [3] Saket Navlakha, Parvez Ahammad, and Eugene W Myers, “Unsupervised segmentation of noisy electron microscopy images using salient watersheds and region merging,” *BMC Bioinformatics*, vol. 14, pp. 294, 2013.
- [4] Uygar Sümbül, Douglas Roossien, Dawen Cai, Fei Chen, Nicholas Barry, John P Cunningham, Edward Boyden, and Liam Paninski, “Automated scalable segmentation of neurons from multispectral images,” in *NIPS*, 2016, pp. 1912–1920.
- [5] Zhi Lu, Gustavo Carneiro, and Andrew P Bradley, “An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells,” *IEEE Transactions on Image Processing*, vol. 24, pp. 1261–1272, 2015.
- [6] Paolo Frasconi, Ludovico Silvestri, Paolo Soda, Roberto Cortini, Francesco S Pavone, and Giulio Iannello, “Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images,” *Bioinformatics*, vol. 30, pp. i587–i593, 2014.
- [7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in *NIPS*, 2012, pp. 1097–1105.
- [8] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten, “Densely connected convolutional networks,” in *CVPR*, 2017, vol. 1, p. 3.
- [9] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *NIPS*, 2015, pp. 91–99.
- [10] Jonathan Long, Evan Shelhamer, and Trevor Darrell, “Fully convolutional networks for semantic segmentation,” in *CVPR*, 2015, pp. 3431–3440.
- [11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, 2015, pp. 234–241.
- [12] Jianxu Chen, Sreya Banerjee, Abhinav Grama, Walter J Scheirer, and Danny Z Chen, “Neuron segmentation using deep complete bipartite networks,” in *MICCAI*, 2017, pp. 21–29.
- [13] Qi Dou, Lequan Yu, Hao Chen, Yueming Jin, Xin Yang, Jing Qin, and Pheng-Ann Heng, “Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1182–1195, 2016.
- [14] Kun Xu, Hang Su, Jun Zhu, Ji-Song Guan, and Bo Zhang, “Neuron segmentation based on CNN with semi-supervised regularization,” in *CVPRW*, 2016, pp. 20–28.
- [15] Kuan-Lun Tseng, Yen-Liang Lin, Winston Hsu, and Chung-Yang Huang, “Joint sequence learning and cross-modality convolution for 3D biomedical segmentation,” *arXiv preprint arXiv:1704.07754*, 2017.
- [16] Xiangrui Zeng, Miguel Ricardo Leung, Tzviya Zeev-Ben-Mordehai, and Min Xu, “A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation,” *Journal of Structural Biology*, 2017.
- [17] Qi Dou, Lequan Yu, Hao Chen, Yueming Jin, Xin Yang, Jing Qin, and Pheng-Ann Heng, “3D deeply supervised network for automated segmentation of volumetric medical images,” *Medical Image Analysis*, vol. 41, pp. 40–54, 2017.
- [18] Lequan Yu, Jie-Zhi Cheng, Qi Dou, Xin Yang, Hao Chen, Jing Qin, and Pheng-Ann Heng, “Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets,” in *MICCAI*, 2017, pp. 287–295.