NEURON SEGMENTATION USING INCOMPLETE AND NOISY LABELS VIA ADAPTIVE LEARNING WITH STRUCTURE PRIORS

Chanmin Park [2] *Kanggeun Lee* [2] *Su Yeon Kim* [3,4] *Fatma Sema Canbakis Cecen* [3,5]
Seok-Kyu Kwon [3,5] Won-Ki Jeong[1]

[1] Department of Computer Science and Engineering, Korea University, Seoul, Korea
[2] School of Electrical and Computer Engineering, UNIST, Ulsan, Korea
[3] Brain Science Institute, Korea Institute of Science and Technology, Seoul, Korea
[4] Department of Biomedical Sciences, College of Medicine, Korea University, Seoul, Korea
[5] Division of Bio-Medical Science & Technology, KIST School, UST, Seoul, Korea

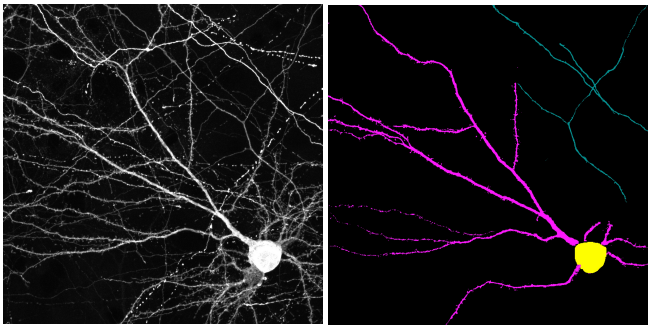
ABSTRACT

Recent advances in machine learning have shown significant success in biomedical image segmentation. Most existing high-quality segmentation algorithms rely on supervised learning with full training labels. However, such methods are more susceptible to label quality; besides, generating accurate labels in biomedical data is a labor- and time-intensive task. In this paper, we propose a novel neuron segmentation method that uses only incomplete and noisy labels. The proposed method employs a noise-tolerant adaptive loss that handles partially annotated labels. Moreover, the proposed reconstruction loss leverages prior knowledge of neuronal cell structures to reduce false segmentation near noisy labels. the proposed loss function outperforms several widely used state-of-the-art noise-tolerant losses, such as reverse cross entropy, normalized cross entropy and noise-robust dice losses.

Index Terms— Adaptive learning, Semantic segmentation, Noise label, Structure prior

1. INTRODUCTION

Neurons show highly polarized structures that typically consist of multiple dendrites and a single axon from a cell body. The morphology of a neuron is critical for various neuronal properties, including synaptic integration and neuronal excitability [1]. Moreover, abnormalities of neuronal shapes have been observed in multiple neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease [2]. Recent advances in imaging techniques led to the emergence of large scale image datasets of complex neuronal processes which require futuer analysis of automated and precise analysis tools. Currently, commercially available software and open-source tools are broadly applied for segmenting and annotating neurons [3, 4]; however, there is still a room for improvement in labeling accuracy and automation of the process.



(a) Input (b) Partial label

Fig. 1. Our neuron data are composed of (a) input fluorescence image showing complex structures, and (b) partially given labels (magenta: dendrites, teal green: axons).

Many semantic segmentation methods rely on supervised learning with clean labels [5, 6, 7]. However, our data (fluorescence microscopy images of neurons) are different from those in conventional semantic segmentation problems; neurons are narrow and thin, and they are distributed sparsely (i.e., a large portion of the image contains empty background). Moreover, multiple neurons are intertwined and cross each other. Therefore, making pixel-level accurate ground-truth labels from neuron images is laborious and difficult. The input to our method is partially labeled data (see Figure 1), that is some axons and dendrites are marked, while the others are left unmarked; in other words, weak labels. We also consider these noisy labels because unlabeled pixels work negatively during training when the loss function is the a conventional Mean Squared Error (MSE). Recently, many noise-robust loss functions have been proposed, such as Mean Absolute Error (MAE) [8], Symmetric cross entropy Learning(SL), Reverse Cross-Entropy (RCE) [9], Normalized Cross-Entropy (NCE), Active and Passive Loss (APL) [10], and Noise-Robust Dice (NR-Dice) [11]. However, none of them has shown satisfactory performance regrading our target problem.

In this paper, we propose two novel loss functions specifically designed to handle incomplete and noisy labels for neuron segmentation. To deal with noisy labels, we develop an adaptive mean squared error (ADMSE) loss, which employs spatially varying weights to prevent learning from noisy labels. Moreover, we propose a structure prior (STPR) loss to promote the correct assignment of labels when multiple classes complete with each other. We demonstrate the performance of our method by comparing it with other noise-robust losses using various segmentation accuracy metrics. We also show how our method improves the performance of neuron segmentation using the ClDice metric [12], which measures the connectivity of elongated structures.

2. METHOD

2.1. Definition

Let $\mathcal{X}$ and $\mathcal{Y}$ be the distributions of observed images and neuron labels, respectively. Our goal of neuron segmentation is to predict a response map $\mathbf{Y} \sim \mathcal{Y}$ for every class representing neuronal structures from a sample image $\mathbf{X} \sim \mathcal{X}$ by feeding a deep neural network $f(\cdot;\theta)$ parameterized by θ using a paired random variable $(\mathbf{X}, \mathbf{Y})$ from the distribution $\mathcal{D}$. In our problem, we use four classes, such as background, cell body, dendrite, and axon, which are denoted as labels 0, 1, 2, and 3 respectively, as shown in Figure 2. Since the labels are partially given, we cannot treat all unlabeled pixels as background. Instead, we apply an intensity-based thresholding to the input image to generate the background image $\mathbf{Y}^{(0)}$. We define K as the set of label indices (in our case, $K = \{0, 1, 2, 3\}$). Note that the cell body and background labels are considered as clean labels (i.e., not partial labels). Therefore, we define the set of clean labels $S_1 = \{0, 1\}$ and that of noisy labels $S_2 = \{2, 3\}$, and thus, $K = S_1 \cup S_2$.

2.2. ADMSE loss

The MSE loss is widely used to minimize the distance between the prediction $f(\mathbf{X};\theta)$ and the ground-truth label $\mathbf{Y}$ as follows:

$$\mathcal{L}_{mse} = \min_{\theta} \mathbb{E}_{(\mathbf{X},\mathbf{Y})\sim\mathcal{D}} ||f(\mathbf{X};\theta) - \mathbf{Y}||^2 \tag{1}$$

Even though the MSE loss effectively penalizes large errors in L^2 sense, it is not resilient to noise in training data as in our problem. To address this issue, we propose an adaptive weight based on the prediction of the network to control the pixel-level backpropagation. Let $\mathbf{Y}^{(k)}$ be the label map for the k-th class and $Y_i^{(k)}$ be the i-th pixel value (i.e., per-pixel label) in $\mathbf{Y}^{(k)}$. Then, we can define L_{admse} using a per-pixel

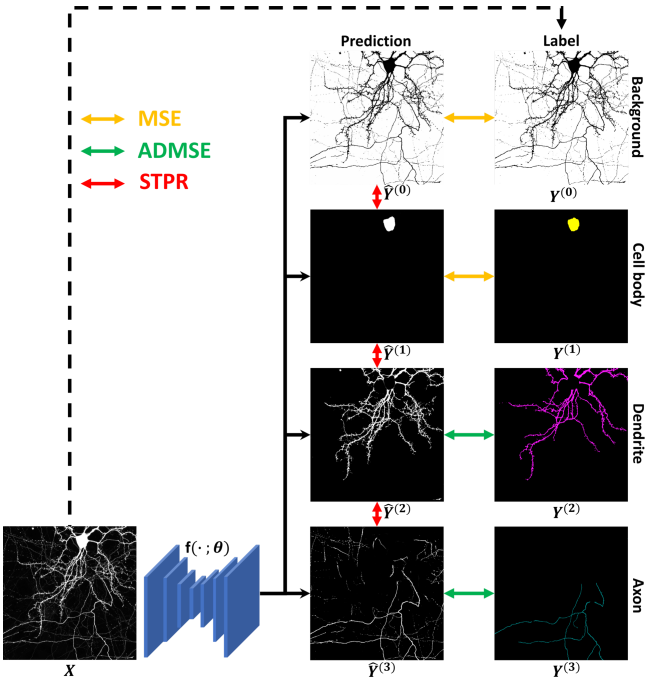


Fig. 2. Overview of the proposed method. There are two sets of training data; one is the clean label (background and cell body) and the other is the partial label (dendrite and axon), to which MSE and ADMSE losses are applied, respectively (yellow and green arrows). Predictions from the model is self-regularized using the STPR loss (red arrows).

adaptive weight function $w(\cdot,\cdot)$ as follows:

$$\mathcal{L}_{admse} = \mathbb{E}_{(\mathbf{X},\mathbf{Y})\sim\mathcal{D}}[\sum_{k}^{K}\sum_{i}^{I} w(\hat{Y}_i^{(k)}, Y_i^{(k)})(\hat{Y}_i^{(k)} - Y_i^{(k)})^2] \tag{2}$$

$$w(\hat{Y}_i^{(k)}, Y_i^{(k)}) = \begin{cases} e^{-(\frac{\hat{Y}_i^{(k)}}{\alpha})^2} + \beta & \text{if } Y_i^{(k)} = 0, k \in S_2 \\ 1 & \text{otherwise} \end{cases} \tag{3}$$

The rationale behind the proposed loss is as follows; if an unlabeled pixel ($Y_i^{(k)} = 0$) in the noisy label map ($k \in S_2$) has a higher prediction value $\hat{Y}_i^{(k)}$, then we consider it as noise because it is not a true unlabeled pixel (i.e., false negative). ADMSE suppress learning from those pixels by assigning small weights during backpropagation. α is a user-defined parameter to control the degree of adaptive weight applied to the loss function. β is a small constant value added to the loss function to ensure stable convergence at the early stage of training (before convergence to correct $\hat{Y}_i^{(k)}$).

2.3. STPR loss

In addition to supervised training using partial labels, we leverage additional constraints based on prior information of

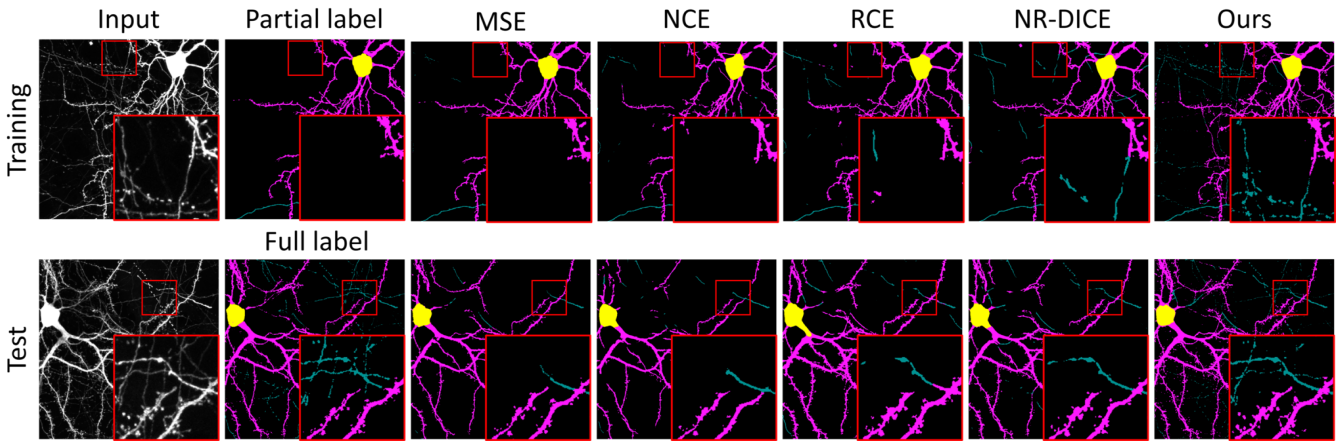


Fig. 3. Results of using robust-noise loss functions, such as NCE, RCE, and NR-DICE, as well as the proposed method. Our approach led to detailed annotated results when performing partial label training and demonstrate similar quality training with full label.

neuronal structures. Because the proposed model is based on multi-task learning using multi-class labels, the model may output strong prediction on more than one class; this is due to the fact that the data and label are noisy and different neuronal structures are overlapped and closely located. The rationale behind the proposed loss is that, for a given pixel location, there should be only one correct class assigned (e.g., a pixel cannot belong to axon and dendrite simultaneously). This constraint can be expressed using conditional probability with random variables $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$P(\mathbf{Y}^{(0)} \cup \mathbf{Y}^{(2)}|\mathbf{X};\theta) \simeq P(\neg\mathbf{Y}^{(1)} \cap \neg\mathbf{Y}^{(3)}|\mathbf{X};\theta)$$
$$P(\mathbf{Y}^{(0)} \cup \mathbf{Y}^{(3)}|\mathbf{X};\theta) \simeq P(\neg\mathbf{Y}^{(1)} \cap \neg\mathbf{Y}^{(2)}|\mathbf{X};\theta) \quad (4)$$

Intuitively, the above relationship implies that a pixel should not be classified as the cell body or axon if such a pixel has high background or dendrite probability (vice versa for axon as well). Based on this, we propose the STPR loss as follows:

$$\mathcal{L}_{stpr} = \mathbb{E}_{\mathbf{X}\sim\mathcal{X}} \sum_{k\in S_2} ||\hat{\mathbf{Y}}^{(0)} + \hat{\mathbf{Y}}^{(k)} - \prod_{j\in S(k)} (1 - \hat{\mathbf{Y}}^{(j)})||^2 \quad (5)$$

where $S(k) = \{x|x \in K \setminus \{0, k\}\}$. This loss is applied to noisy labels (i.e., S_2) to serve as a regularizer.

In summary, the total loss for our model is defined as the sum of the ADMSE and STPR losses as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{admse} + \mathcal{L}_{stpr} \quad (6)$$

3. EXPERIMENT

3.1. Implementation Details

Our neuron image data consists of paired training data with partial annotations from a total of 23 single cells and 3 test sets with full labels. Each image (1024$\times$ 1024 in size) was split into small patches (128 $\times$ 128 in size). We discarded totally empty (background) patches, and applied data augmentation using rotation and flipping. Since the number of pixel labels of dendrite was about five times higher than that of axons, we applied oversampling on axon labels to prevent class imbalanced problem. The proposed network is based on a conventional U-Net with the ResNet-34 backbone encoder. We trained the model using the Adam optimizer with an initial learning rate of 3e-3, with a random batch size of 300, and gradually decreased the learning rate of 3e-4 with the cosine annealing method. We generated a binary image $\mathbf{Y}^{(0)}$ by setting a threshold of 0.3 for the input image (see Fig. 2). We used $\alpha = 0.1$ to apply the optimal weight map for noisy labels, and $\beta = 0.03$ to converge the learning at initial training, both of which were empirically found via 10-fold cross-validation.

3.2. Evaluation

To demonstrate the efficacy on partial labels, we compared the proposed method with other noise-robust losses, such as SCE, APL, and NR-Dice loss. For the parameters, we used $\alpha = 0.1$ and $\beta = 1$ for SCE, $\alpha = 1$ and $\beta = 1$ for APL, and $\gamma = 1$ for NR-Dice,which were chosen empirically to obtain the best performance.

To quantitatively assess the performance of our method, we used three error metrics which are commonly used in semantic segmentation, such as Intersection over Union (IoU), F1 score, and ClDice [12]. In particular, our data are composed of a tubular structure in each neuron; thus, we employed ClDice which can evaluate the connectivity between structures. To obtain ClDice to the evaluation score from our structures, we can obtain the score by inserting the skeletonized ground truth into the prediction label.

	IoU			Precision			Recall			F1			ClDice		
	body	**Dend**	**Axon**	**body**	**Dend**	**Axon**	**body**	**Dend**	**Axon**	**body**	**Dend**	**Axon**	**body**	**Dend**	**Axon**
MSE	0.8509	**0.7094**	0.0889	0.9138	0.8121	0.6975	0.9251	0.849	0.0954	0.9194	**0.8299**	0.1608	0.8771	0.8105	0.2486
RCE[9]	0.7335	0.6476	0.1719	0.7447	0.7188	0.7661	**0.9799**	0.8677	0.184	0.8461	0.7859	0.2924	0.7321	0.7082	0.4873
APL[10]	0.855	0.6448	0.1304	0.9116	**0.9004**	0.772	0.9325	0.6947	0.1375	0.9218	0.784	0.2293	0.8066	0.7635	0.2851
NR-Dice[11]	0.8374	0.7007	0.1523	0.8707	0.8065	0.6469	0.9566	0.8425	0.1676	0.9115	0.8239	0.2628	0.8215	0.8245	0.3536
$\mathcal{L}_{admse}$(ours)	0.8349	0.6986	0.2871	0.9312	0.7241	0.6084	0.8899	**0.9531**	0.3564	0.9101	0.8223	0.4447	0.8871	0.8101	0.3965
$\mathcal{L}_{mse} + \mathcal{L}_{stpr}$(ours)	0.8626	0.6729	0.2961	0.8848	0.7822	0.7735	0.972	0.8277	0.3241	0.9262	0.8043	0.4568	0.8814	0.7561	0.1683
$\mathcal{L}_{admse} + \mathcal{L}_{stpr}$(ours)	**0.8892**	0.7086	**0.5183**	**0.9676**	0.7435	**0.8439**	0.9165	0.9375	**0.5731**	**0.9413**	0.8292	**0.6826**	**0.9043**	**0.8461**	**0.4948**

Table 1. Comparison of segmentation performance over different loss functions measured using various segmentation quality metrics. The top two results in each case are marked in bold (1st) and underlined (2nd), respectively.

3.3. Result and Discussion

The qualitative comparison results of different noise-robust losses with our approach are illustrated in Figure 3. The upper row represents the result of a training image (showing an example of a partial label), while the bottom row indicates the result of a testing image. As shown in this figure, the MSE loss generated incomplete results due to noisy labels. While other losses exhibited some improvement over MSE, none of them was able to generate accurate result close to the full label as in our method (see the red box in Figure 3). Note also that, due to partial labels, other losses failed to learn axons during training (because unlabeled pixels act as a negative label) while our loss successfully reconstructed most of the axons correctly (Figure 3, upper row).

Table 1 shows the quantitative comparison of different loss functions with the proposed on, which were obtained by measuring the segmentation accuracy using IoU, precision, recall, F1 score, and ClDice metrics. The last three rows in this table illustrate the ablation study of the proposed loss function. In this result, we observed that the axon class is more vulnerable to noisy partial labels than other classes, which resulted in extremely low recall values for the MSE and noise-robust loss functions(between 0.0954 and 0.1676). However, our ADMSE and STPR losses are more resilient to label noise and achieved up to 0.5731 for the recall. We also observed that dendrites were less affected by noisy labels due to their large and thick structure as compared to thin axons. Therefore, the MSE achieved the best result for IoU (which measures the overlap between two regions) and F1 scores. Although other noise-robust losses demonstrate inferior performance, our proposed loss achieve comparable results to the full label. Moreover, our approach significantly outperformed the other methods in the ClDice metric, which assesses the linearity of the structure, demonstrating that our loss fits better to the neuron segmentation problem than the other noise-robust losses.

4. CONCLUSION

In this paper, We introduced two novel losses, ADMSE and STPR, for neuron structure segmentation from noisy and incomplete training labels. The result showed that the proposed losses outperformed state-of-the-art noise robust losses in various segmentation quality metrics. Especially, the proposed method achieved best ClDice scores, which demonstrates that the method is effective to the neuron segmentation problem. In the future, we plan to extend the proposed method to the 3D neuron segmentation problem.

5. COMPLIANCE WITH ETHICAL STANDARDS

This work is about developing a computational algorithm for which no ethical approval was required.

6. ACKNOWLEDGMENTS

This work was partially supported by the National Research Foundation of Korea (NRF-2019M3E5D2A01063819) and the Institute for Information communications Technology Planning & Evaluation (IITP-2020-0-01819) funded by the Ministry of Science and ICT, and the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health Welfare (HI18C0316).

7. REFERENCES

[1] Michael London and Michael Häusser, "Dendritic computation," *Annu. Rev. Neurosci.*, vol. 28, pp. 503–532, 2005.

[2] Jaime Grutzendler, Kathryn Helmin, Julia Tsai, and Wen-Biao Gan, "Various dendritic abnormalities are associated with fibrillar amyloid deposits in Alzheimer's disease," *Annals of the New York Academy of Sciences*, vol. 1097, no. 1, pp. 30–39, 2007.

[3] Linqing Feng, Ting Zhao, and Jinhyun Kim, "neuTube 1.0: a new design for efficient neuron reconstruction software based on the SWC format," *eneuro*, vol. 2, no. 1, 2015.

[4] Hang Zhou, Shiwei Li, Anan Li, Qing Huang, Feng Xiong, Ning Li, Jiacheng Han, Hongtao Kang, Yijun Chen, Yun Li, et al., "GTree: an Open-source Tool for Dense Reconstruction of Brain-wide Neuronal Population," *Neuroinformatics*, pp. 1–13, 2020.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[6] Yuhui Yuan, Xilin Chen, and Jingdong Wang, "Object-contextual representations for semantic segmentation," *arXiv preprint arXiv:1909.11065*, 2019.

[7] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

[8] Aritra Ghosh, Himanshu Kumar, and PS Sastry, "Robust loss functions under label noise for deep neural networks," *arXiv preprint arXiv:1712.09482*, 2017.

[9] Yisen Wang, Xingjun Ma, Zaiyi Chen, Y. Luo, Jinfeng Yi, and J. Bailey, "Symmetric Cross Entropy for Robust Learning With Noisy Labels," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 322–330, 2019.

[10] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey, "Normalized Loss Functions for Deep Learning with Noisy Labels," *arXiv preprint arXiv:2006.13554*, 2020.

[11] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, "A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions From CT Images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2653–2663, 2020.

[12] Johannes C Paetzold, Suprosanna Shit, Ivan Ezhov, Giles Tetteh, Ali Ertürk, Helmholtz Zentrum Munich, and Bjoern Menze, "ClDice—a novel connectivity-preserving loss function for vessel segmentation," in *Medical Imaging Meets NeurIPS 2019 Workshop*, 2019.