BESNet: Boundary-Enhanced Segmentation of Cells in Histopathological Images

Hirohisa Oda[1(✉)], Holger R. Roth[2], Kosuke Chiba[3], Jure Sokolić[4], Takayuki Kitasaka[5], Masahiro Oda[3], Akinari Hinoki[3], Hiroo Uchida[3], Julia A. Schnabel[4], and Kensaku Mori[2,6,7]

[1] Graduate School of Information Science, Nagoya University, Nagoya, Japan
hoda@mori.m.is.nagoya-u.ac.jp
[2] Graduate School of Informatics, Nagoya University, Nagoya, Japan
[3] Nagoya University Graduate School of Medicine, Nagoya, Japan
[4] Division of Imaging Sciences and Biomedical Engineering, King's College London, London, UK
[5] School of Information Science, Aichi Institute of Technology, Toyota, Japan
[6] Information Technology Center, Nagoya University, Nagoya, Japan
[7] Research Center for Medical Bigdata, National Institute of Informatics, Tokyo, Japan

Abstract. We propose a novel deep learning method called Boundary-Enhanced Segmentation Network (BESNet) for the detection and semantic segmentation of cells on histopathological images. The semantic segmentation of small regions using fully convolutional networks typically suffers from inaccuracies around the boundaries of small structures, like cells, because the probabilities often become blurred. In this work, we propose a new network structure that encodes input images to feature maps similar to U-net but utilizes two decoding paths that restore the original image resolution. One decoding path enhances the boundaries of cells, which can be used to improve the quality of the entire cell segmentation achieved in the other decoding path. We explore two strategies for enhancing the boundaries of cells: (1) skip connections of feature maps, and (2) adaptive weighting of loss functions. In (1), the feature maps from the boundary decoding path are concatenated with the decoding path for entire cell segmentation. In (2), an adaptive weighting of the loss for entire cell segmentation is performed when boundaries are not enhanced strongly, because detecting such parts is difficult. The detection rate of ganglion cells was 80.0% with 1.0 false positives per histopathology slice. The mean Dice index representing segmentation accuracy was 74.0%. BESNet produced a similar detection performance and higher segmentation accuracy than comparable U-net architectures without our modifications.

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11071, pp. 228–236, 2018.
https://doi.org/10.1007/978-3-030-00934-2_26

1 Introduction

The detection or the semantic segmentation of cells in histopathological images using fully convolutional networks has been explored [1,2] for many diagnostic or medical research purposes. Our focus in this work is the ganglion cell detection of the HE-stained images of pediatric intestine specimens of Hirschsprung's disease [3]. To quicken and increase the accuracy of its pathologic diagnosis during surgery, an automatic segmentation method of ganglion cells is required. There may be several ganglion cells on HE-stained images, which have variations of color, size, shape, and contrast. Many cells or tissues also resemble ganglion cells on HE-stained images.

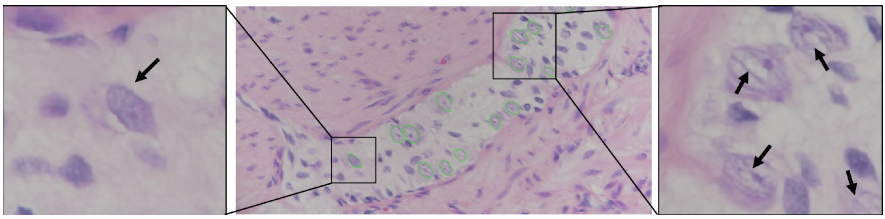


Fig. 1. Ganglion cells on HE-stained images: Green circles or black arrows represent ganglion cells, which have variations of color, shape, and contrast. Cells or tissues also exist that resemble ganglion cells.

U-net [1] is one of the most popular and widely used fully convolutional architectures, which segment biomedical images well. However, for small objects in large images, blurring of the probability response maps occurs around the boundaries of these small objects. This problem is caused by lack of consideration of difficulties around the target object borders. Hand-crafted weighting schemes of the loss for outside objects have been introduced for improving the prediction in these regions [1,2]. However, adaptive weighting schemes for improving the responses around the border based on the difficulty of training have not yet been considered. We tackle these problems by proposing a (1) new network architecture called Boundary-Enhanced Segmentation Network (BESNet) and a (2) Boundary-Enhanced Cross Entropy (BECE) loss. BESNet consists of a network with two decoding paths. One is trained for boundary prediction but suffers from inaccuracies because detecting the boundaries is difficult. This information on the degree of difficulty of detecting a network's boundaries can be fused with the decoder path for entire cell segmentation by skip connections and (2) Boundary-Enhanced Cross Entropy loss. Accessing and modifying feature maps in the layers that haven't been decoded yet is inspired by deep supervision, which is especially useful for edge enhancement [4]. In this work, we enhance the segmentation of the entire cell by utilizing the feature maps of boundaries.

The BESNet performance is shown by the detection and the segmentation of ganglion cells from the HE-stained images of the histopathological samples of

pediatric intestine. As shown in Fig. 1, ganglion cells are scattered on HE-stained images, and many similar regions surround them. A computer-aided diagnosis system that detects and measures the size of the ganglion cells is required for assisting the rapid pathologic diagnosis during surgery, which finds ganglion cells from HE-stained images. To the best of our knowledge, no other work has addressed the detection or segmentation of ganglion cells apart from our preliminary work [5].

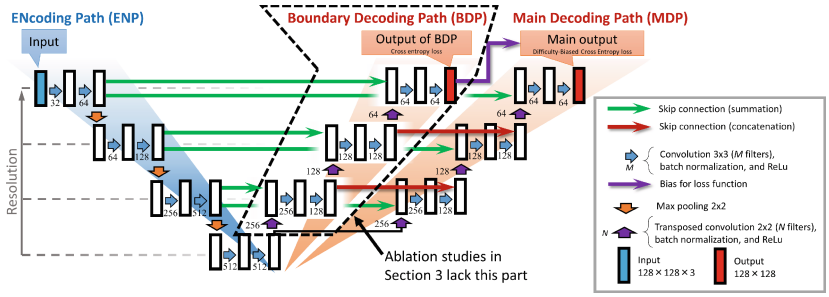


Fig. 2. Network structure of BESNet: While encoding part resembles standard U-net, BESNet has two decoding parts, Boundary Decoding Path (BDP) and Main Decoding Path (MDP). Feature maps in BDP are concatenated with MDP. Loss function for MDP is weighted by BDP output.

2 Method

2.1 Boundary-Enhanced Segmentation Network (BESNet)

BESNet is a novel, fully convolutional network for semantic segmentation. Its concept is to train the boundaries of the targeted cells and use their responses to adaptively weight the training loss for entire cell segmentation. This allows us to apply a stronger weight in the more difficult part of the targeted cell during training. Our proposed network structure is shown in Fig. 2. Any input patch is encoded into feature maps in a similar way to U-net [1] on the ENcoding Path (ENP). Unlike U-net, BESNet has two decoding paths. A Boundary Decoding Path (BDP) is trained using the boundary labels of the annotated cells. Feature maps in this path are concatenated with Main Decoding Path (MDP), which is trained on all of the cell labels. After two layers of 3×3 convolution (CV), batch normalization (BN), and ReLU activation functions (RA), 2×2 max pooling (MP) decreases the resolution at each level of the ENP. After repeating these layers (CV, BN, RA, CV, BN, RA, and MP) three times and this sequence twice (CV, BN, RA), we obtain feature maps whose resolution is the lowest but has the highest level of abstraction for effective semantic segmentation. Here, the network is branched into BDP and MDP. The resolution is restored by 2×2 transposed convolutions (TC) at each resolution level. Both BDP and MDP have three times of the sequence (RA, TC, CV, BN) with a final CV layer with 1×1

convolution kernels and *sigmoid* activations. For each TC on BDP and MDP, feature maps after the last RA in the same resolution in ENP are summed by skip connections. Moreover, for each RA on MDP, feature maps after the last RA in the same resolution in BDP is concatenated using skip connections.

2.2 Boundary-Enhanced Cross-Entropy (BECE) Loss

The basic idea of cross-entropy, which is one of the most commonly used loss functions, is to penalize the loss more when the network's output is more different than the ground-truth. We utilize cross-entropy loss $\mathcal{L}_{\mathrm{C}}$ for BDP. Since this is a binary problem but we are only interested in how difficult it is to learn the foreground pixels of the boundary, $\mathcal{L}_{\mathrm{C}}$ is defined by

$$\mathcal{L}_{\mathrm{C}} = -\sum_{\boldsymbol{x}\in M} B(\boldsymbol{x}) \log\big(p_{\mathrm{B}}(\boldsymbol{x})\big) \tag{1}$$

where $\boldsymbol{x}$ represents a pixel in mini-batch M, $B(\boldsymbol{x}) \in \{0, 1\}$ represents the boundary label of the ground-truth at $\boldsymbol{x}$, and $p_{\mathrm{B}}(\boldsymbol{x}) \in \{0, \cdots, 1\}^{\mathbb{R}}$ represents the BDP output at $\boldsymbol{x}$.

BDP output $p_{\mathrm{B}}(\boldsymbol{x})$ usually performs well at the boundaries, but it may become low at the boundary parts that are less clear or have rare types of appearances. This means that the features of the boundaries with low output of BDP probability are difficult to train by the network. Therefore, these parts should be trained more strongly by MDP by adaptively weighting the loss function for the MDP branch. For MDP, we newly define a Boundary-Enhanced Cross-Entropy (BECE) loss:

$$\mathcal{L}_{\mathrm{D}} = -\sum_{\boldsymbol{x}\in M} \big\{\big[1 + b(\boldsymbol{x})\big] G(\boldsymbol{x}) \log\big(p_{\mathrm{M}}(\boldsymbol{x})\big) + w\big[1 - G(\boldsymbol{x})\big] \log\big(1 - p_{\mathrm{M}}(\boldsymbol{x})\big)\big\} \tag{2}$$

$$b(\boldsymbol{x}) = \alpha \ \max\big(\beta - p_{\mathrm{G}}(\boldsymbol{x}), 0\big) \tag{3}$$

where $G(\boldsymbol{x}) \in \{0, 1\}$ and $p_{\mathrm{M}}(\boldsymbol{x}) \in \{0, \cdots, 1\}^{\mathbb{R}}$ represent the ground-truth label and the MDP output at $\boldsymbol{x}$, respectively. $b(\boldsymbol{x})$ is a function that represents the training difficulty of the boundary at $\boldsymbol{x}$. $\alpha \in \{0, \cdots, 1\}^{\mathbb{R}}$ and $\beta \in \{0, \cdots, 1\}^{\mathbb{R}}$ are coefficients for the strength of boundary-enhanced weighting and minimum value of p_{B} that are enhanced well, respectively. w is weight for background pixels, which is the ratio between numbers of positive and negative pixels. This loss definition is partly inspired by Focal Loss [6], but it adjusts the weighting just from the probabilities of the same output of the network.

2.3 Training and Testing

Input and Output: Our method detects and segments cells from histopathological images. For training, a set of images and their ground-truth labels G_n

are required. Detection and segmentation of cells are performed on the images for testing. The output is a set of ganglion cell regions.

Training: Histopathological images, which are usually scanned in high resolution, are much bigger (e.g., 1636×1088 pixels) than what we can fit on GPU memory as input to BESNet, (see Sect. 2.1 for more details). Hence, we first perform d-times downsampling of the images and the ground-truth. Then patches ($s_{\mathrm{x}} \times s_{\mathrm{y}}$ pixels) are cropped randomly, but at least one positive pixel must exist in the ground-truth. We employ a data augmentation process during training that consists of random rotation, translation, and elastic deformations by B-spline splitting. We collect m images as a mini-batch for training at each iteration.

Testing: BESNet is reshaped so that the input and output sizes cover larger region $s_{\mathrm{rx}} \times s_{\mathrm{ry}}$ pixels. The testing image is divided into patches in a grid pattern with v-voxel overlap to the neighboring patches. The MDP output is computed for each patch, and the output for every histopathological image is combined from all the patch predictions. The average responses are computed on the overlapping parts of multiple patches to allow smooth transitions of the responses across patches.

3 Experiments

Overview: To evaluate the segmentation accuracy of our proposed model without decreasing the detection performance of the cells, we conducted detection and segmentation of the ganglion cells on the HE-stained images of histopathological samples. The detection performance was evaluated by the detection rate and the number of false positives (FPs) per image (FPs/image). Segmentation accuracy was evaluated by Dice index, precision, and recall. Probability threshold t was set to $0.05, 0.10, \cdots, 0.95$ for FROC evaluation.

Dataset: The HE-stained images of the intestine parts whose peristaltic movement is functioning properly were obtained from 25 patients suffering from Hirschsprung disease from whom we received ethical approval from Nagoya University Hospital (Japan). They include 741 ganglion cells from 224 images. Each specimen was imaged with an ECLIPSE Ni-U (Nikon) microscope and scanned by a DS-Ri2 (Nikon) camera as RGB-color images consisting of 1636×1088 pixels. Resolution is $250 \times 250\,\mathrm{nm}^2$. The ground-truth labels were manually created by an expert pediatric surgeon.

Condition: Three-fold cross validation was conducted by dividing the patients into three groups. The network was implemented on Keras with a TensorFlow backend. The parameters were empirically set to $d = 2, s_{\mathrm{x}} \times s_{\mathrm{y}} = 256 \times 256$, $s'_{\mathrm{x}} \times s'_{\mathrm{y}} = 768 \times 256$, $\alpha = 0.5$ and $\beta = 0.1$. DeepLearningBOX (GDEP Advance) workstations with GTX 1080 Ti (NVIDIA) GPUs, CUDA 8.0, and cuDNN 6.0 was used for the computation. We fixed the random number of seeds of NumPy and TensorFlow for reproducibility. Other training conditions were set as follows: mini-batch size to 8, iterations to 30000, and optimizer to Adam.

Ablation Studies: For a comparison with the proposed method, we conducted two ablation studies: "U-net & Cross-entropy" and "U-net & Dice", using cross-entropy loss or Dice loss, respectively. As annotated in Fig. 2, removing BDP from BESNet allows us to get a U-net-like structure. It contains BN layers, and have four levels of resolution (original one [1] has five).

4 Results and Discussions

Detection: The partial FROC curves of the three methods that were obtained by changing threshold t are shown in Fig. 3. Table 1 shows the detection performance when t is 0.20, 0.50, or 0.80. The proposed method's performance was 89.5% of the detection rate with 2.5 ± 7.1 FPs/slice with $t = 0.50$, and an example slice of the results is shown in Fig. 4. All three methods produced similar results. One difference between (c) Dice and the others is the change of the balance between the detection rate and the FPs/slice.

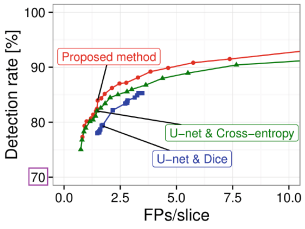


Fig. 3. Partial FROC curves: Proposed method, U-net & Cross-entropy, and U-net & Dice produced similar detection performances.

Table 1. Performances of three methods: Partial FROC curves were linearly interpolated and FPs/slice were estimated at 80.0%, 85.0%, and 90.0% of detection rates. Bold FPs/slice represent smallest average. FPs/slice of U-net & Dice at 90.0% of detection rate could not be estimated since no threshold produced detection rate of 90.0% or above.

	Detection rate	FPs/slice
Proposed method	80.0%	$\mathbf{1.0 \pm 1.7}$
U-net & Cross-entropy		1.1 ± 2.0
U-net & Dice		1.8 ± 2.9
Proposed method	85.0%	$\mathbf{1.8 \pm 2.6}$
U-net & Cross-entropy		2.4 ± 3.5
U-net & Dice		3.3 ± 4.8
Proposed method	90.0%	$\mathbf{4.8 \pm 5.4}$
U-net & Cross-entropy		7.1 ± 7.6
U-net & Dice	N/A	

Segmentation: Segmentation accuracies are shown in Table 2 and Figs. 6(a)–(c). The Dice index, precision, and recall of the proposed method were 71.4±31.9, 81.2 ± 32.9, and 67.2 ± 31.7 (mean ± std. dev.), respectively, when threshold t was 0.50. The scores of the true positives (TPs) were computed for the highest regions obtained by the methods, and the scores of all the false negatives are zero. Our proposed method produced the highest Dice index and precision. Using the Wilcoxon rank sum test, most results between the proposed method and others showed significant differences (Table 2).

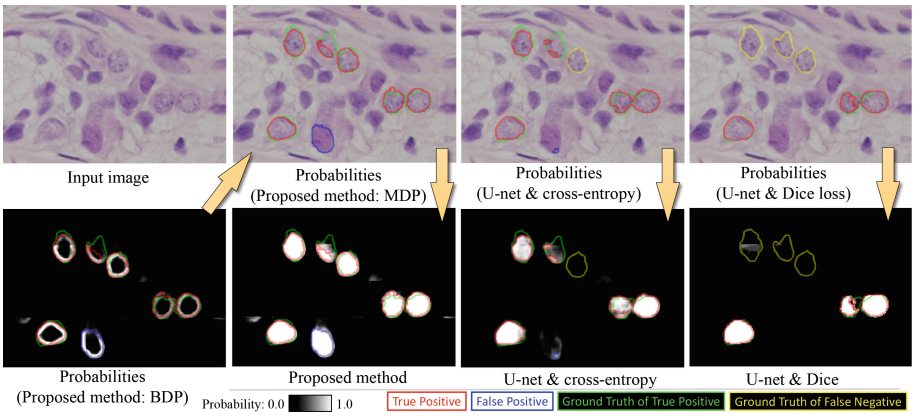


Fig. 4. Probabilities and outputs of three methods.

Table 2. Segmentation accuracy of three methods when t is 0.20, 0.50, or 0.80. Mean ± standard deviation of each measure is shown. Bold numbers show best mean of all scores among three methods with common t. (*) and (**) represent significant differences between proposed method, which has $p < 0.05$ and $p < 0.01$, respectively.

t	Method	Dice index	Precision	Recall
0.20	Proposed method	**74.0 ± 31.8**	**81.8 ± 29.9**	72.3 ± 32.8
	U-net & Cross-entropy	70.9 ± 33.6 (**)	75.7 ± 31.3 (**)	**72.6 ± 35.8** (**)
	U-net & Dice	69.7 ± 35.1 (**)	74.1 ± 34.3 (**)	70.7 ± 36.8 (**)
0.50	Proposed method	**69.1 ± 34.8**	**79.9 ± 35.2**	64.7 ± 34.1
	U-net & Cross-entropy	66.8 ± 36.7 (*)	75.8 ± 35.9 (**)	65.1 ± 37.3 (**)
	U-net & Dice	67.5 ± 36.9 (*)	73.9 ± 36.2 (**)	**67.0 ± 38.0** (**)
0.80	Proposed method	**63.5 ± 35.9**	**77.7 ± 38.9**	56.9 ± 33.8
	U-net & Cross-entropy	61.6 ± 37.9 (**)	74.2 ± 39.7 (**)	56.8 ± 36.8 (**)
	U-net & Dice	63.3 ± 38.3 (**)	71.3 ± 38.9 (**)	**61.1 ± 38.8** (**)

Three cells on a slice are magnified in Fig. 5. Cell A had blurred probabilities around the boundaries from the U-net & Cross-entropy, as shown in dotted cyan squares. Predicting this part is also difficult by the BDP of our proposed method. Due to the adaptive weighting of such boundaries during training, a clearer and more accurate region segmentation was obtained by MDP. Cell B and C also had weak boundary probabilities from BDP in almost the entire cell. The MDP of our proposed method accurately produced high probabilities on entire of each cell regions, and two cells were divided well. U-net & Cross-entropy produced higher probabilities even gap between two cells, and segmentation results of two cells were connected. This is why Dice index of Cell B was only 57.4% with U-net & Cross-entropy. U-net & Dice produced high probabilities only on Cell B, and Cell C was a false negative. While just dividing two neighboring cells gives

the same advantage as other works [1] including methods specific to instance segmentation [2], BESNet also can achieve better segmentation accuracy inside cells.

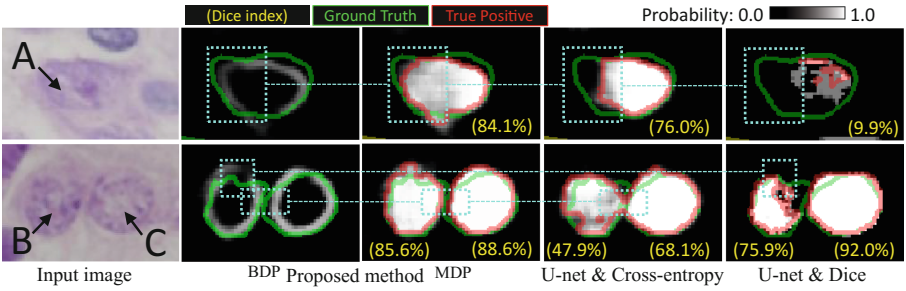
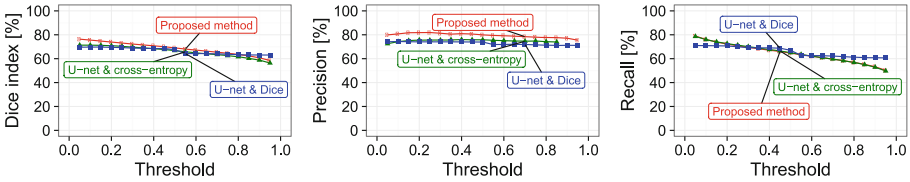


Fig. 5. Probabilities on three cells: Yellow numbers show Dice index of segmentation results where $t = 0.50$. Green circles show ground-truth. In dotted cyan squares of each cell, BDP output does not clearly show boundaries. In such regions, our proposed method produced higher probabilities inside cell and low at ones outside it, compared to U-net.



(a) Dice index (b) Precision (c) Recall

Fig. 6. Segmentation accuracy of three methods. Proposed method had higher Dice index and precision than others.

5 Conclusions

We proposed a novel deep learning method called Boundary-Enhanced Segmentation Network (BESNet) for the detection and semantic segmentation of cells on pathological images. Experimental results on ganglion cells show similar detection performances but significantly better segmentation results. One limitation is computational complexity. Ablation studies with U-net required only about 6 GB GPU memory, but BESNet required about 10 GB. More comparisons to related works are left for future work.

Acknowledgements. Parts of this research were supported by JSPS KAKENHI (26108006, 17H00867, 17K20099) and JSPS Bilateral Joint Research Project.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
2. Chen, H., Qi, X., Yu, L., Heng, P.A.: DCAN: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 2487–2496 (2016)
3. Amiel, J., Lyonnet, S.: Hirschsprung disease, associated syndromes, and genetics: a review. J. Med. Genet. **38**(11), 729–739 (2001)
4. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1395–1403 (2015)
5. Oda, H., et al.: Automated ganglion cell detection using fully convolutional networks and evaluation under different training losses. In: Computer Assisted Radiology and Surgery (CARS) 2018 (2018)
6. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: The IEEE International Conference on Computer Vision (ICCV) (2017)