

Comparison of UNet, ENet, and BoxENet for Segmentation of Mast Cells in Scans of Histological Slices

Alexander Karimov

*Engineering School of ITTCS
Ural Federal University
Ekaterinburg, Russia
sanya1992AK@gmail.com*

Artem Razumov

*Institute of Natural Sciences
Ural Federal University
Ekaterinburg, Russia
airplaneless@yandex.ru*

Ruslana Manbatchurina

*Engineering School of ITTCS
Ural Federal University
Ekaterinburg, Russia
manbatchurina99@gmail.com*

Ksenia Simonova

*Engineering School of ITTCS
Ural Federal University
Ekaterinburg, Russia
ksansi@yandex.ru*

Irina Donets

*Engineering School of ITTCS
Ural Federal University
Ekaterinburg, Russia
ira-don1225@gmail.com*

Anastasia Vlasova

*Laboratory of Translational Medicine and Bioinformatics
Institute of Immunology and Physiology
Ekaterinburg, Russia
vlasova.9@mail.ru*

Yulia Khramtsova

*Laboratory of Immunophysiology and Immunopharmacology
Institute of Immunology and Physiology
Ekaterinburg, Russia
hramtsova15@mail.ru*

Konstantin Ushenin

*Institute of Natural Sciences
Ural Federal University
Ekaterinburg, Russia
konstantin.ushenin@urfu.ru*

Abstract—Deep neural networks show high accuracy in the problem of semantic and instance segmentation of biomedical data. However, this approach is computationally expensive. The computational cost may be reduced with network simplification after training or choosing the proper architecture, which provides segmentation with less accuracy but does it much faster. In the present study, we analyzed the accuracy and performance of UNet and ENet architectures for the problem of semantic image segmentation. In addition, we investigated the ENet architecture by replacing of some convolution layers with box-convolution layers. The analysis performed on the original dataset consisted of histology slices with mast cells. These cells provide a region for segmentation with different types of borders, which vary from clearly visible to ragged. ENet was less accurate than UNet by only about 1-2%, but ENet performance was 8-15 times faster than UNet one.

Index Terms—biomedical segmentation, semantic segmentation, neural network performance, ENet, UNet, box convolution layer, mast cells

I. INTRODUCTION

Deep neural networks show high accuracy in the semantic and instance segmentation of biomedical data [1]. However, deep neural networks are a computationally expensive approach in comparison with classical pattern recognition methods. Usually, high-quality automatic segmentation with deep

neural networks requires a computational accelerator or a CPU with good computational performance. Thus, software engineers and biomedical engineers are forced to use cloud computing or add advanced computational power in the device for offline processing of data. This leads to an increase in the infrastructure annual cost or an increase in the main cost of developing biomedical devices.

Reducing the computational cost of deep learning may be achieved with two approaches. In the first approach, a deep neural network may be simplified after training. For example, this type of method is described in work [2] and presented by pruning, trained quantization, and Huffman coding.

In another approach, several works propose neural network architectures that do not show the best accuracy of segmentation but consist of a significantly smaller number of trainable parameters than the more precise approaches. ENet [3] is this type of architecture that has been proposed for real-time semantic segmentation.

In present work, we compare the accuracy and performance of three neural networks (UNet [4], ENet [4], and BoxENet) for semantic segmentation of histological slice scans. UNet was chosen for the current study as the gold standard for biomedical segmentation problems. ENet was chosen as the architecture that shows the moderate quality in segmentation but should show a high performance in image processing. In addition, we assume before the study that replacing the usual convolution with box convolution ([5]) may improve the

The work was performed as part of the state assignment of IIF UB RAS (subject No. AAAA-A18-118020590108-7) and supported by the RF Government Act No. 211, dated March 16, 2013 (agreement 02.A03.21.0006).

quality of the segmentation. To test this idea, we implement BoxENet, which is ENet model from the original paper with a replacement of some bottleneck block to bottleneck blocks with box-convolutional layers [5].

For analysis, we chose an original dataset that included toluidine blue-stained histological slices with mast cells at different stages of degranulation. This dataset was manually segmented by an expert. The main reason for this choice is the ground truth segmentation with a region that had a different type of border: clear border, ragged border, and a cloud of points with no observable border. In particular, the last case is important because the neural network cannot learn on the cell membrane pattern [6]. Let us briefly describe the dataset.

Mast cells are part of the human immune system. They are related to allergy reactions, and several diseases. Mast cells look substantially different at a different stage of their physiological role. Before the release of granules with histamine, these cells look like compact objects with a clearly visible border. Under some factor, the cells start to spread capsules with histamine. This process is named *degranulation*. During degranulation, the cells' borders become unclear, and the cell surrounded by histamine capsules that look like a cloud of points. At the end of the degranulation process, the original location of the cell is indistinguishable, and the cell is fully replaced by the cloud of points. At the final stage of the degranulation process, it is difficult to separate one cell from another.

II. METHODS

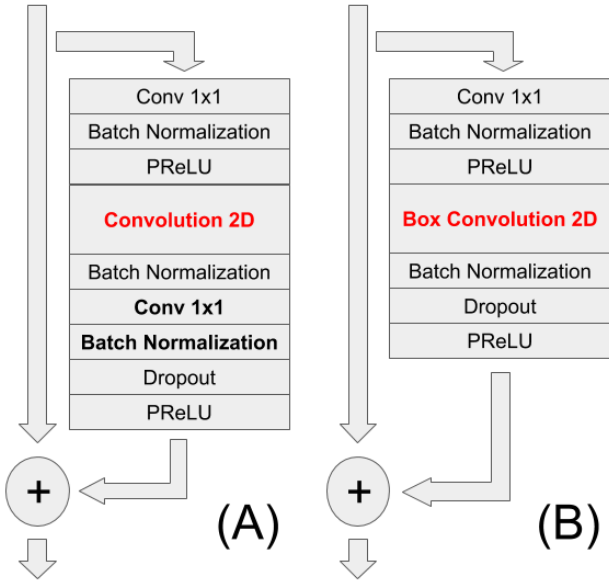


Fig. 1. Bottleneck blocks that are used for the current work in ENet (A) and BoxENet (B) neural networks. Red and bold fonts indicate the differences.

A. Dataset

The mast cell dataset was acquired from histology preparations from previous studies, and laboratory animals were not

ENet	BoxENet
Downsampler(3→16)	Downsampler(3→16)
Bottleneck(16→64, downsample)	Bottleneck(16→64, downsample)
Bottleneck(64→64)	Bottleneck(64→64)
Bottleneck(64→64)	BottleneckBoxConv(64→64)
Bottleneck(64→64)	Bottleneck(64→64)
Bottleneck(64→64)	BottleneckBoxConv(64→64)
Bottleneck(64→128, downsample)	Bottleneck(64→128, downsample)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=2)	BottleneckBoxConv(128→128)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=4)	BottleneckBoxConv(128→128)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=8)	BottleneckBoxConv(128→128)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=16)	BottleneckBoxConv(128→128)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=2)	BottleneckBoxConv(128→128)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=4)	BottleneckBoxConv(128→128)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=8)	BottleneckBoxConv(128→128)
Bottleneck(128→128)	Bottleneck(128→128)
Bottleneck(128→128, dilation=16)	BottleneckBoxConv(128→128)
Upsampler(128→64)	Upsampler(128→64)
Bottleneck(64→64)	Bottleneck(64→64)
Bottleneck(64→64)	Bottleneck(64→64)
Upsampler(64→16)	Upsampler(64→16)
Bottleneck(16→16)	Bottleneck(16→16)
ConvTranspose2d(16→2)	ConvTranspose2d(16→2)

TABLE I

THE IMPLEMENTED ENET AND BOXENET ARCHITECTURES. THE SECOND ARCHITECTURE REPEATS THE FIRST, BUT BOTTLENECK BLOCKS WITH THE DILATED CONVOLUTIONS ARE REPLACED TO BOTTLENECK BLOCKS WITH BOX-CONVOLUTION LAYERS.

purposefully processed for this study. The conditions of the housing and treatment of the animals during the experiment complied with the Directive of the European Parliament and the Council of September 22, 2010, on the protection of animals used for scientific purposes (2010/63/EU). Moreover, the study was approved by the local ethical committee of the Institute of Immunology and Physiology of the Ural Branch of the Russian Academy of Sciences (IIF UB RAS) (protocol No. 10, dated April 3, 2017).

The slices were obtained during a histological examination of the seminal vesicles and appendages of male Wistar rats of different ages. The preparations were fixed in 10% formalin solution for 1 day. After standard tissue processing on a Shandon Excelsior closed-type automatic machine (MICROM International GmbH, Germany), the material was embedded in paraffin using an EG 1160 paraffin embedding station for biological tissue (Leica, Germany). Next, the paraffin blocks were cut on a semi-automatic microtome Thermo Microm HM 450 (MICROM International GmbH, Germany). The slice thickness was 4 μm . Then, the preparations were scanned with a light optical microscope (Leica, Germany).

The full dataset includes 168 images with a resolution of 1024 x 1280 px and standard RGB channels. Each image consists of 5 to 7 mast cells. About 40% of all cells appear close to each other and have joined borders. Images were cut on 244 tiles with 256 x 256 tiles accordingly to input the size of the neural networks. Ground truth segmentation of the

dataset was performed manually by an expert. Separation of the full dataset on test and validation part was performed in 74:51 ratio and a random shuffling of the samples.

B. Artificial Neural Networks

For the present purposes, we compared three architectures: UNet [4], ENet [4], and an ENet modification (BoxENet), where bottleneck block with the standard convolutional layer replaced to the bottleneck block with box-convolution layers [5].

UNet [4] has become a gold standard for comparison of biomedical segmentation accuracy and performance. UNet [4] is a convolutional autoencoder with additional connections between the encoder and the decoder parts. This type of neural network also is named "an hourglass architecture". UNet and neural networks that are used the same approach show highly accurate results in a wide area of biomedical applications. In addition, this family of architectures works faster than approaches that are based on the pixel-wise classification [4].

ENet architecture [3] was developed for the real-time semantic segmentation problem. In total, the ENet from the original paper contains twice less trainable parameters than UNet. This neural network uses a convolutional block with short skipped connections, which is also named as a bottleneck block (see. Fig. 1). Also, this architecture used dilated convolutions, asymmetric convolutions, and other approaches that affect the perceptive field and the distribution of weights in the deep layers.

Besides that, we implement ENet architecture with the replacement of some bottleneck blocks with usual convolution layers to bottlenecks block with box-convolution layers proposed in [5]. From our assumption, this replacement may improve the quality of the segmentation. Variation of ENet architectures and replaced layers for BoxENet are presented in Fig. I. Fig. 1 shows the differences between the usual bottleneck block and the bottleneck block with box convolution layer.

All neural networks from the original paper were implemented with minor changes. The training was performed using dice similarity coefficient (DSC) loss and ADAM optimizer with a learning rate of 0.0001. We did not apply the overlapping strategy and weight map that is usually used for the Unet architecture [4]. This unification of training approaches was performed to focus on the primary goal of the broader study because a more wide study is required to analyze the effect of other losses and weight maps on the accuracy of each neural network.

III. RESULTS

As shown in Fig. 2, the ENet, and BoxENet accuracy increase faster than the UNet accuracy during the neural network training process. The best neural network accuracy is reached before 300 epochs for each of the three architectures.

Fig. 3 and Table II present a detailed comparison of the accuracy metrics on the validation dataset for the three neural networks. Accordingly to the table, UNet shows higher values

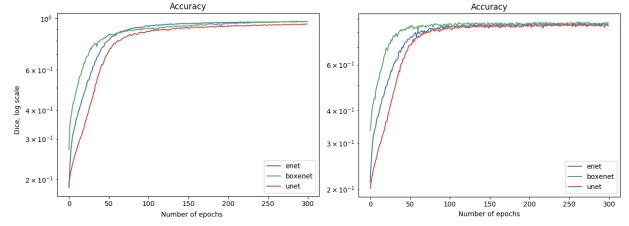


Fig. 2. Accuracy by epochs of the neural network training in terms of the DSC. The left image is for the training dataset, and the right image is for the validation dataset.

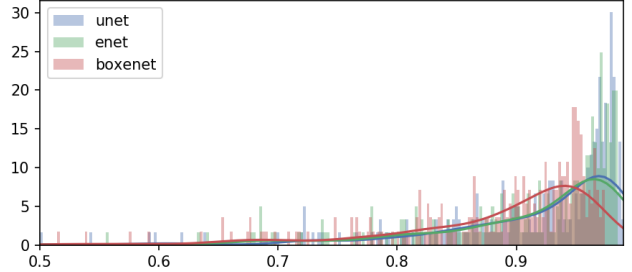


Fig. 3. Distribution of the DSC over images from the validation dataset.

in three metrics of the semantic segmentation accuracy. The median of the DSC distribution is lower for BoxENet in comparison with UNet and ENet as shown in Fig.3. Therefore, the box convolutions did not sufficiently improve the accuracy of the ENet architecture.

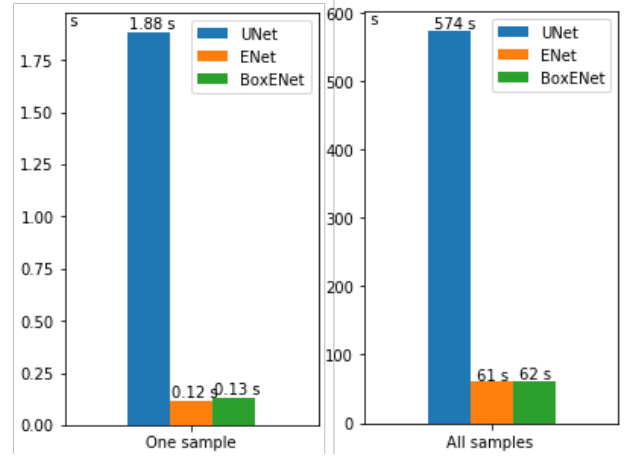


Fig. 4. Comparison of computational performance for three architectures on one sample and the full validation dataset.

As shown in Fig. 4, the ENet, and BoxENet computational performance was up to 10 to 15 times higher than the UNet performance. That analysis were performed on a computational accelerator NVIDIA Tesla K80.

Figure 5 shows segmentation results. Each of the three neural networks shows the correct segmentation for objects on the image border. All neural networks were insensitive to the stage of mast cell degranulation and correctly segmented cells with any border. However, UNet shows a better result in

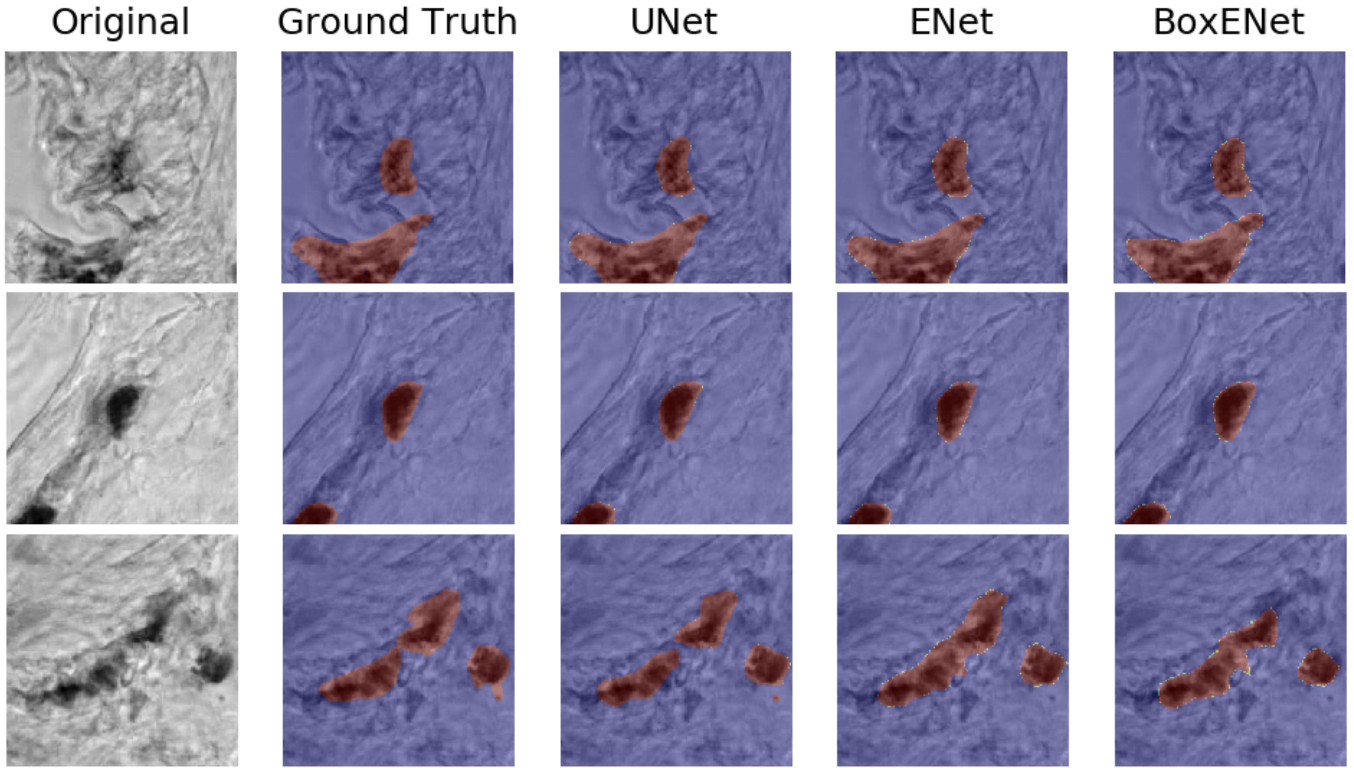


Fig. 5. Original image, ground truth, and semantic segmentation performed with three neural networks.

	UNet	ENet	BoxENet
DSC coef. (mean)	0.9159	0.9107	0.8859
DSC coef. (median)	0.9514	0.9469	0.9157
DSC coef. (max)	0.9909	0.9851	0.9741
DSC coef. (percentile 10)	0.8075	0.7966	0.7671
IoU (mean)	0.8538	0.8445	0.8092
IoU (median)	0.9079	0.8996	0.8525
IoU (max)	0.9757	0.9771	0.9571
IoU (percentile 10)	0.6744	0.6624	0.6161
F1 (mean)	0.8509	0.8442	0.8076
F1 (median)	0.8993	0.9023	0.8578
F1 (max)	0.9785	0.9738	0.9487
F1 (percentile 10)	0.6713	0.6539	0.6207

TABLE II

COMPARISON OF THE NEURAL NETWORK ACCURACY FOR THE VALIDATION DATASET. THE BOLD TEXT SHOWS THE HIGHEST VALUE IN EACH ROW.

segregating closely located cells even without the application of the weighted map in training. That may lead to issues with the counting of objects on images, and complicate the instance segmentation with ENet instead of UNet.

IV. DISCUSSION AND CONCLUSION

In the current study, we compared the accuracy and performance of UNet, ENet, and ENet with box convolutional layers (BoxENet). The comparison was performed on histological data with mast cells. The dataset had regions with different types of borders. The borders of the regions varied from clearly visible to ragged, and in some cases, looked like a cloud of points.

For the semantic segmentation problem, ENet lost about 1-2% of accuracy at the comparison with UNet in terms of the DSC and the F1-score. On the other hand, ENet computational performance is significantly higher (up to 10-15 times) that is important for many applications.

REFERENCES

- [1] G. Carneiro, Y. Zheng, F. Xing, and L. Yang, "Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis," in *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Springer, 2017, pp. 11–32.
- [2] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding," *arXiv preprint arXiv:1510.00149*, 2015.
- [3] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.
- [4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [5] E. Burkov and V. Lempitsky, "Deep neural networks with box convolutions," in *Advances in Neural Information Processing Systems*, 2018, pp. 6211–6221.
- [6] S. Dimopoulos, C. E. Mayer, F. Rudolf, and J. Stelling, "Accurate cell segmentation in microscopy images using membrane patterns," *Bioinformatics*, vol. 30, no. 18, pp. 2644–2651, 2014.