Comparison of Deep Learning Preprocessing Algorithms of Nuclei Segmentation on Fluorescence Immunohistology Images of Cancer Cells

Xu Silun
Dept. of System Analysis and Computer Simulation
Belarusian State University
Minsk, Belarus
xusilun@hotmail.com

Victor Skakun
Dept. of System Analysis and Computer Simulation
Belarusian State University
Minsk, Belarus
skakun@bsu.by

Abstract. **Immunohistology fluorescence image analysis is an important method for cancer diagnosis. With the widespread application of convolutional neural networks in computer vision, segmentation of images of cancer cells has become an important topic in medical image analysis. Although there are many publications describing the success in application of deep learning models for segmentation of different kind of histology images, the universal algorithm is still not developed. The image preprocessing consisting in splitting images in smaller parts and normalization is important in deep learning especially when the training set is of a limited size. In this study, we compared several approaches to create the training set of a sufficient size while having a limited number of labeled whole slide immunohistology images of cancer cells. Also, we explored different normalization methods.**

Keywords: **CNN, medical image analysis, image preprocessing, image segmentation, nucleus of cancer cell, U-Net**

I. INTRODUCTION

Convolutional neural networks (CNN) have gradually begun to be applied in the field of image analysis since they achieved a huge breakthrough in the field of handwritten font recognition [1]. The outstanding performance of CNN in the ImageNet competition shows that it has great potential in image analysis fields such as image feature extraction and image classification. At present, CNN has widely been applied also in medical image segmentation. The U-Net [2] architecture, specially developed to segment objects like cell nucleus on biomedical images, is widely used in medical image segmentation. The skip connections introduced in U-Net helps to merge the features of different scales that enhance its performance. For example, Neha Todewale successfully applied U-Net to perform segmentation of mammogram images [3]. Ajinkya Jawale et al. made segmentation of the brain tumor images using U-Net [4]. Adnan Saood et al. realized COVID-19 lung CT image segmentation and comparative analysis using U-Net [5]. Other CNN models, such as VGGNet [6], ResNet [7], FCN [8], Inception [9], all based on ideas of deep learning, were developed in the past ten years for image segmentation. Most of them has been successfully applied also in a field of biomedical image segmentation and classification [10]. We selected U-Net for our study because having a relatively simple architecture it still shows very good performance.

The purpose of our research is to segment cell nuclei on the fluorescence images of cancer tissue slices. Cancer tissue slice, stained with fluorescent agents accordingly to a certain protocol, is observed by the fluorescence confocal microscopy to obtain a three-channel color picture. Then the attempt to segment cells (nuclei and cytoplasm) on these images is made for a subsequent quantitative and qualitative pathological analysis.

The segmentation of objects like nucleus or cytoplasm on biomedical images is a pixelwise binary classification problem. The goal is to assign each pixel either to the class of pixels that forms the nuclei area or not. As a step of image segmentation using neural networks, the data preprocessing has a significant impact on the segmentation results. The main purpose of image preprocessing is to eliminate irrelevant information and enhance the detectability of useful information (represented by the pixels or features in the CNN terms) to the greatest extent, thereby improving the accuracy of cell segmentation [11]. For biomedical images segmentation, preprocessing steps usually require cropping, splitting in smaller parts, regularization, intensity enhancement and normalizing to the range [0, 1] and so on [12]. Among them, due to the uneven distribution of cells in biological tissues, preprocessing steps such as the selection of cell nucleus regions also affects the ability to obtain correct segmentation results.

In the cancer pathology analysis, one usually works with samples of relatively small sizes. High diversity of cancer cases does not allow to get a large set of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

histological images of desired similarity. Another problem is a labeling of target objects on images. When the task is a segmentation of cells or nuclei on images, the person who does labeling must be an expert in both the cancer diagnosis and in the microscopy. Fortunately, each whole slide immunohistology image contains hundreds of cells (target objects for the segmentation). It opens the perspective for enhancing the training set by splitting the input images into a number of small images (patches). Simple splitting images into not overlapped patches, splitting with overlap (sliding), random extraction of patches are ways to enhance the training set [10].

To fulfil this task one question is naturally arising, what size of a patch is optimal for the segmentation of nuclei of cancer cells of a certain average size? Also, it is interesting to examine different strategies of the patch extraction in order to increase the size of the training set. One has to take into account that it is impossible to increase the size of the training set infinitively by a simple extraction more and more overlapped patches. Even in a case of using abundant data augmentation the segmentation will suffer from overfitting. In addition, it is interesting to know how the image preprocessing, like normalization and standardization, influences the segmentation results.

Therefore, the goal of this study is to find the most effective algorithm of preprocessing of immunohistology fluorescence images in the task of segmentation of nuclei of cancer cells.

II. Materials and Methods

In this study we are performing the segmentation of fluorescence images of the breast tumor tissue slices. The images where obtained using Nikon TE200 epi-fluorescent inverted microscope equipped with the Photometrics 300 series CCD camera at 10x magnification and stored in RGB color system. The size of the images is 2048 × 2048 pixels in each of the three channels, the resolution is 0.2 μm / pixel, or 5 μm.

The protein estrogen receptor was used as an cancer indicator [13]. In contrast to healthy cells, a protein cytokeratin appears in the cytoplasm of cancer cells. This protein is labeled with cyanine dye Cy3 and registered in the green color channel of the image. To label all nuclei, the 4,6-diamidino-2-phenylindole dihydrochloride (DAPI) dye was used. Its fluorescence was recorded in the blue channel. The cyanine dye Cy5 (recorded in the red channel of the image) was used to label the estrogen receptor, which is located primarily in the nuclei of cancer cells. Accordingly, two dyes, Cy5 and Cy3, are markers of cancer cells.

Nine experimental images were labeled by experts initially semi automatically using CellProfiler (https://cellprofiler.org/) then manually. Labeled images (ground truth) are the binary images, where pixels of nucleus of cancer cells were set to 1.

TABLE I. Layer Structure of U-Net

Encoder layers	**Decoder layers**
3x3 Conv+ELU, F=16	2x2 ConvTranspose(S=2), F=128 (V)
Dropout=0.1	3x3 Conv+ELU, F=128
3x3 Conv+ELU, F=16 (I)	Dropout=0.2
MaxPool (S= 2)	3x3 Conv+ELU, F=128
3x3 Conv+ELU, F=32	2x2 ConvTranspose(S=2), F=64 (VI)
Dropout=0.1	3x3 Conv+ELU, F=64
3x3 Conv+ELU, F=32 (II)	Dropout=0.2
MaxPool (S= 2)	3x3 Conv+ELU, F=64
3x3 Conv+ELU, F=64	2x2 ConvTranspose(S=2), F=32 (VII)
Dropout=0.2	3x3 Conv+ELU, F=32
3x3 Conv+ELU, F=64 (III)	Dropout=0.2
MaxPool (S= 2)	3x3 Conv+ELU, F=32
3x3 Conv+ELU, F=128	2x2 ConvTranspose(S=2), F=16 (VIII)
Dropout=0.2	3x3 Conv+ELU, F=16
3x3 Conv+ELU, F=128 (IV)	Dropout=0.1
MaxPool (S= 2)	3x3 Conv+ELU, F=16
Middle layers	1x1 Conv+Sigmoid, F=1
3x3 Conv+ELU, F=256	
Dropout=0.3	
3x3 Conv+ELU, F=256	

Abbreviations in the table: S – stride, F – number of filters. The following layers were concatenated: (I)-(VIII), (II)-(VII), (III)-(VI), (IV)-(V).

The U-Net architecture, specially developed for a segmentation of biomedical images and showed good performance, was selected in our study. It is symmetrical neural net and has five sets of convolution/deconvolution layers, see Table I (in TensorFlow notation).

Our realization of U-Net was based on the architecture that shown nice results in the 2018 Data Science Bowl Kaggle competition and available at GitHub (https://github.com/dubeyakshat07/Cell-Nuclei-Image-Segmentation-using-U-Net). The model has 1,941,105 learning parameters. We used binary cross entropy as the loss function and intersection over union (IOU, known also as Jaccard similarity coefficient) calculated at the threshold 0.5, as the segmentation metric. The smaller the loss value, the better the learning effect. The larger the IOU metric – the better the segmentation (closer to the ground truth).

Taking into account the wide range and scalability of the application, we used a Python language

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

environment to build the U-Net model, and selected the Jupyter Notebook to run and test the code. The main model components were implemented through the Keras library (TensorFlow kernel). Augmentation was performed using the Albumentations library (https://albumentations.ai/) and calculation of the loss function and the evaluation metric was done using the Segmentation models library (https://segmentation-models.readthedocs.io).

III. Results and Discussions

As we have only 9 labeled images (of size 2048 × 2048 px.) the main goal at deep learning is to increase the size of the training set. The rough estimation of the nuclei size of 1250 ± 800 pixels corresponds approximately to the circle with radius of 20 px. It allows us to split the input images into a set of much smaller sizes – patches, thus defining a training set of optimal size. The dimension of the patch size is not trivial to guess because the segmentation results are obviously depended on how close to the ground truth is the predicted border of a nucleus. If we select small patch window the learning process will definitely suffer from the fact that probably most of nuclei will be located not in the center of the window, but over their borders. In opposite case we can get just rough estimation of the nuclei area. The answer to the question how many times an overlapped part of the image can be present in the training set is also not evident. The overfitting makes useless all our work otherwise.

Therefore, to answer these questions we initially performed the segmentation of nuclei at different patch sizes and extraction strategies. At the second step we have studied several methods of preprocessing the patches such as image histogram equalization, standardization and normalization to the range [0, 1].

The U-Net model used in our study is described in details in the Materials and Methods section. We selected a simple augmentation including random 90-degree rotation, horizontal flipping and transposing. Although we utilized simple augmentation, it can effectively prevent overfitting. The intersection over union (at level 0.5) metric was set as a measure of accuracy of the segmentation. Seven images were used for training, one image for validation and one for testing. The same algorithm of preprocessing was applied to both training and validation images. The empty patches (the number of pixels less than 150) were removed from the training set. Number of epochs was set to 30. Earlier stopping callback with patience = 8 was used in the training process.

The first task is to obtain a sufficiently large training set and to avoid overfitting. We started with a patch size of 512 × 512 and decreased it to 64 × 64. The latter case gave much worse results, probably because the most nuclei crossed borders of patches, therefore we did not include it into the analysis. More specifically we studied the following cases: 1) simple splitting into patches of the size 128 × 128 without overlap; 2) splitting into patches of the size 128 × 128 with overlap in 64 px. (1/2 of the patch size); 3) splitting into patches of the size 256 × 256 with overlap in 128 px. (1/2 of the patch size); 4) random cropping 128 × 128 patches 8 times after intermediate splitting into 256 × 256 subimages; 5) splitting into patches of the size 512 × 512 with overlap in 128 px. (1/4 of the patch size). The side lengths of patches are therefore 128, 256, and 512 pixels. At the case 1) we got the training set of 1576 images. A at the case 4) we got the training set of 12064 images.

The values of the loss function and the IOU metric on the test set are summarized in the TABLE II. Other relevant information about the number of patches and the total number of pixels in the training set as well as the batch size, number of steps per epoch is also summarized there. The values of the loss function and IOU metric for 30 CNN epochs, obtained for both training and validation data, are shown in Fig. 1.

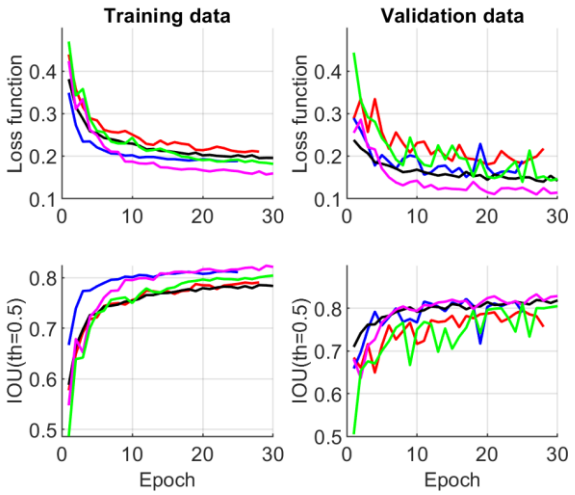

Fig. 1. Values of the loss function and IOU metric for 30 CNN epochs, after cases 1-5 of patch extraction methods. Color coding:| case 1 – red, case 2 – blue, case 3 – green, case 4 – black, case 5 – magenta

In spite of the fact that in the case 5 we got the best metric value, we conclude the case 4 with random extraction of patches of a size 128 × 128 to be the best. This conclusion came from the visual inspection of the segmentation results (data not shown). Here we use a simple but effective method to increase the size of the training set. In this case, patch overlap is useful, but when the number of patches reaches a certain level, then overfitting is occurring. When the number of patches is at the largest, we get the best results. In contrast, the maximum number of pixels is achieved in the case 5. One may come with a conclusion that if augmentation

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

is performed then as larger the training set the better the results despite of the presence of overlapping areas of the images. To prove or reject this conclusion we did an experiment (data not shown) where we selected the case 2 but increased virtually twice (because of augmentation) the training set by selection of larger number of steps per epoch to get the same number of patches in the training process. We got slightly worse value of the IOU metric (= 0.834). Thus, the random extraction of patches is better. We also tried the case with random extraction of four patches from 256 × 256 sub images. The result was worse than at the simplest case 1. Therefore, we proved that the random cropping is a powerful method but only when enough number of patches is extracted to obtain a complete coverage of the image. Also, we came to the conclusion that the patch of the size 128 × 128 px. is optimal for the segmentation of nuclei with the average radius of about 20 px.

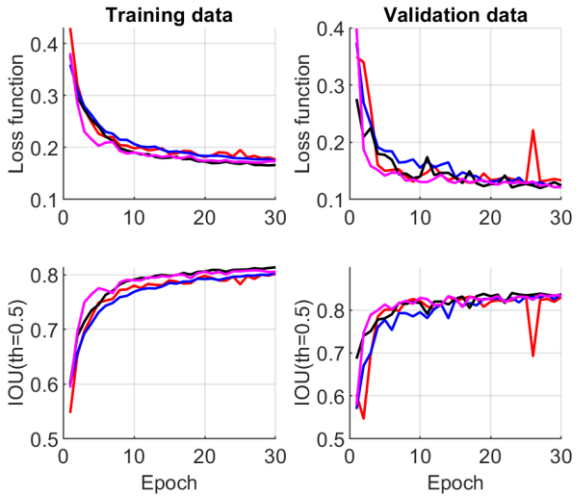

Fig. 2. Values of the loss function and IOU metric for 30 CNN epochs after different methods of patch normalization. Color coding: case 1 – black, case 2 – red, case 3 – magenta, case 4 – blue.

In the second step of our study, we picked out the case 4 of randomly selected patches with the side length of 128 pixels. The following methods of image normalization and standardization were studied: 1) raw data (without normalization); 2) normalization to the range [0, 1], where 1 corresponds to the maximum pixel intensity in the patch; 3) centering and scaling; 4) normalization to the range [0, 1] by initial patch histogram equalization then division to 255.

Surprisingly we obtained very similar results. The best method was the image histogram equalization (see Table III and Fig. 2).

The method of histogram equalization effectively increases the contrast of the cell nucleus, therefore supports its better segmentation. However, differences in the IOU metric are very small. Therefore, in our case, image normalization does not play a key role (see Table III).

IV. Conclusion

The preprocessing of input images plays important role in the deep learning. Having only 9 whole slide labeled images of the size 2048 × 2048 pixels we were able to get satisfactory results in the segmentation of nuclei on the fluorescence images of the cancer cells of the breast tumor. We concluded that the random cropping is a powerful method for increasing the training set when the number of patches is high enough to obtain a complete coverage of the input image. The simple splitting of the image with overlap may be sufficient if the augmentation is used. Also, we came to the conclusion that the patch of the size 128 × 128 px. is optimal for the segmentation of nuclei with the average radius of about 20 px.

In our numerical tests the image normalization does not play a key role. Histogram equalization with subsequent normalization to the range [0, 1] demonstrated the best result. The method of histogram equalization effectively increases the contrast of the cell nucleus and thus supports the better segmentation.

TABLE II. Results of segmentation after different methods of patches creation

Parameters and metrics	**Methods of patches creation**				
	Splitting to 128×128 without overlap	**Splitting to 128×128 with overlap in 64 px**	**Splitting to 256×256 with overlap in 128 px**	**Random cropping 128×128 8 times after intermediate splitting into 256×256 images**	**Splitting to 512×512 with overlap in 128 px**
Total number of patches	1576	5939	1508	12064	1177
Total number of pixels in all patches	25,821,184	97,304,576	98,828,288	197,656,576	308,543,488
Batch_size	64	64	24	32	16
Steps per epoch	25	93	63	377	74

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Parameters and metrics	Methods of patches creation				
	Splitting to 128×128 without overlap	**Splitting to 128×128 with overlap in 64 px**	**Splitting to 256×256 with overlap in 128 px**	**Random cropping 128×128 8 times after intermediate splitting into 256×256 images**	**Splitting to 512×512 with overlap in 128 px**
Loss function	0.222	0.188	0.156	0.156	0.157
IOU (th=0.5)	0.786		0.837	0.837	0.839

TABLE III. RESULTS OF SEGMENTATION AFTER DIFFERENT METHODS OF PATCHES NORMALIZATION

Parameters and metrics	Methods of patches normalization			
	Without normalization	**Division by max value of pixel intensity in a patch**	**Standardization**	**Histogram equalization**
Loss function	0.159	0.154	0.160	0.154
IOU (th=0.5)	0.837	0.838	0.838	0.840

REFERENCES

[1] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," IEEE proc., vol. 86(11), pp. 2278–2324, Nov. 1998.

[2] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," arXiv:1505.04597v1 [cs.CV], May 2015.

[3] N. S. Todewale, "Lesion segmentation from mammogram imagesusing a U-Net deep learning network," IJERT, vol. 9(02), Feb. 2020.

[4] A. Jawale, P. Warole, S. Bhandare, K. Bhat, P. R. Chandre, "Jeevn-Net: Brain tumor segmentation using cascaded U-Net & overall survival prediction," IRJET, vol. 7(01), Jan. 2020.

[5] A. Saood, I. Hatem, "COVID-19 lung CT image segmentationusing deep learning methods: U-Net versus SegNet," BMC Med Imaging, 21:19, 2021.

[6] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556v6 [cs.CV], Apr. 2015.

[7] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," arXiv:1512.03385v1 [cs.CV], Dec. 2015.

[8] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation," arXiv:1411.4038v2 [cs.CV], Mar. 2015.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, et al, "Going deeper with convolutions," arXiv:1409.4842v1 [cs.CV], Sept. 2014.

[10] Q. D. Vu, S. Graham, T. Kurc, M. N. N. To, M. Shaban et al., "Methods for segmentation and classification of digital microscopy tissue images," Front. Bioeng. Biotechnol, vol. 7, article. 53, April 2019.

[11] S. Dakhare, H. Chowhan, M. B. Chandak, "Combined approach for image segmentation," IJCTT, vol. 11(03), May 2014.

[12] Y. Lisitsa, M. Yatskou, V. Skakun, A. Digris, I. Shinhariev, V. Apanasovich, "Simulation model to study denoising methods," Proc. of the 12th Intern. Conf. PRIP, Minsk [UIIP NASB, 2014, pp. 157–160].

[13] G. G. Chung, M. P. Zerkowski, S. Ghosh, R. L. Camp, D. L. Rimm, "Quantitative analysis of estrogen receptor heterogeneity in breast cancer," Lab Invest, vol. 87, no. 7, pp. 662–669, 2007.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.