CLASSIFICATION OF SUPERPIXELS FOR TUMOR SEGMENTATION ON IMMUNOHISTOCHEMISTRY IMAGES

Yauheniya Lisitsa, Victor Skakun

Faculty of Radiophysics and Computer Technologies, Belarusian State University, Republic of Belarus
lisitsa@bsu.by

The breast cancer is the most common invasive cancer in women [1]. Immunohistochemistry (IHC) is widely exploited at biological researches and medical treatments of cancer diseases. It is based specific markers, which are able to label aimed proteins and to color them [2]. Hematoxylin and eosin stain (H&E) is one of the most popular technologies for tissue staining in histology [3].

The superpixels with simple linear iterative clustering (SLIC) are used to segment different kinds of images. For example, the algorithm based on SLIC determines cellular structures in high-dimensional histopathological images of renal cell carcinoma [4]. It shows sufficient results for x-ray image segmentation [5, 6]. The random forest classification is the robust method of classification [7], which is used for different classification tasks such as non-small cell lung cancer classification [8] and dose-response prediction [9]. The aim of our work is to develop the algorithm of tumor segmentation based on superpixels and random forest classification to process images of breast cancer.

We studied the immunohistochemical image of breast cancer cells (Fig.1A), which were received by transmitted light microscope. The images are stained by H&E. The acquisition is received by Leica recording system with a DFC 420 C camera, DM5000B lens and by using LEAD Technologies Inc. V1.01 software. The image resolution is 300 pixels per inch, shutter speed is 1/12 second, image size is 2592 by 1944 pixels, color depth is 24 bits.

The first part of our algorithm is the superpixels segmentation by SLIC model. The main parameter for segmentation is average superpixels size, which default value is 50 pixels. The number of iterations for SLIC method is 10.

The second part of the algorithm is classification by random forest. The successful classification was received with the following parameters: the number of trees in the forest is 50, the nodes are expanded until all leaves are pure or until all leaves contain less than 2 samples. Each superpixel is described by 400 features, which are received as intensities (red, green and blue channels) of neighboring pixels of the superpixel center in the surrounding area at 20 pixels. The random forest was trained by the set of 29670 objects collected from three images. Each object was a pixels described by the neighboring pixels as it described above.

The fig.1D shows our results of tumor segmentation. The segmentation error is 6.3%. It is caused by misclassified superpixels, because the shape of the superpixels overcomes the linear size of cells. In order to reduce the segmentation error it is suggested to use smaller superpixel size. However, in this case it will cause incorrect edge detection of tumor edges. If the average superpixel size is 25, the segmentation error is 5.2%.

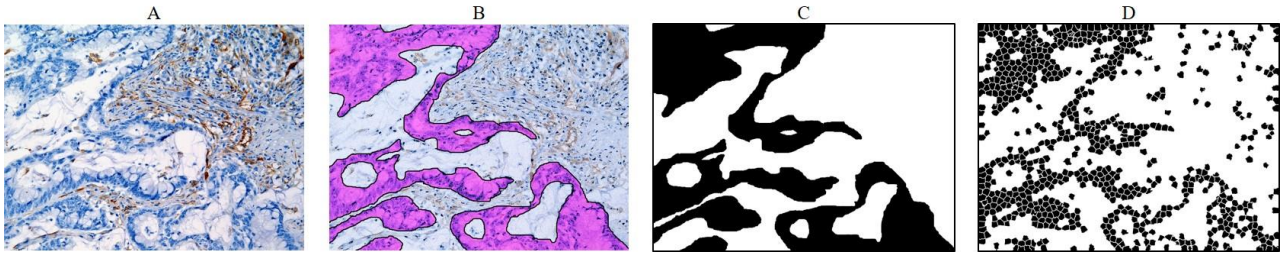

Fig. 1. Tumor segmentation on immunohistochemistry images. A) Original image B) Tumor mask received by expert C) Binary tumor mask received from C D) Segmented mask after superpixels classification

The accuracy of segmentation also depends on the quality of the random forest teaching; therefore, the method robustness can be improved by using more diverse set of data.

[1] A. Albayrak, G. Bilgin, Breast Cancer Epidemiology, Presse Med, **48(10)**, 1076-1084 (2019).
[2] N. Sukswai, J. D Khoury, Immunohistochemistry Innovations for Diagnosis and Tissue-Based Biomarker Detection, Curr Hematol Malig Rep, **14(5)**, 368-375 (2019)
[3] A. Grosset, K. Loayza-Vega, É. Grangeret et al., Hematoxylin and Eosin Counterstaining Protocol for Immunohistochemistry Interpretation and Diagnosis, Appl Immunohistochem Mol Morphol, **27 (7)**, 558-563 (2019)
[4] H. Sancho-Garnier, M. Colonna, Automatic cell segmentation in histopathological images via two-staged superpixel-based algorithms, Med Biol Eng Comput, **57(3)**, 653-665 (2019).
[5] J. Chu, H. Min, L. Liu, W. Lu, A novel computer aided breast mass detection scheme based on morphological enhancement and SLIC superpixel segmentation, Med Phys, **42(7)**, 3859-3869 (2015).
[6] L. Yao, S. Muhammad, A novel technique for analysing histogram equalized medical images using superpixels, Comput Assist Surg (Abingdon) **24(sup1)**, 53-61 (2019).
[7] L. Bierman. Random Forest. Mach. Learn., **45(1)**, 5-32 (2001)
[8] C. Guoliang Luo, Y. Zhu, R. Wang et al., Wang, Random Forest-Based Classsification and Analysis of Hemiplegia Gait Using Low-Cost Depth, Med Biol Eng Comput [Online ahead of print], (2019)
[9] R. Rahman, S. Dhruba, S. Ghosh et al., Functional Random Forest With Applications in Dose-Response Predictions, Sci Rep, **9(1)**, 1628 (2019)