Tomasz STRZODA[1], Seweryn KALISZ[1], Piotr JÓŹWIK-WABIK[1],
Dawid MACHA[1], Marek HERMANSA[1], Bartłomiej GŁADYS[1],
Adam POPOWICZ[2], Paweł FOSZNER[3], Michał MARCZYK[4]

Chapter 10. TUMOR DETECTION ON HE-STAINED HISTOPATHOLOGICAL IMAGES USING DEEP LEARNING METHODS: CHALLENGES AND SOLUTIONS

10.1. Introduction

Each year millions of people are diagnosed with cancer around the world. Therefore, research in many disciplines is constantly being done to understand the disease and ultimately increase the survival. One of the areas that provides research opportunities is digital pathology, which is the presentation of pathological images on a computer screen instead of a microscope and glass slides [1]. This opens up further possibilities for analyses based on processing these high-resolution images and using them, e.g. for tumor area detection tasks [2]. In the classical approach, the pathologist looks at the whole-slide image (WSI), analyses it carefully and marks the tumor region with a marker. This solution has one major drawback, namely the long time needed to carry out the assessment. Then, there is the human factor, where fatigue occurs after looking at images for a long time, directly affecting the expert's judgement. One possible solution is to use deep learning methods that can handle such task by operating on a scanned slide image [3]. In supervised deep learning, where data training samples are associated with class labels, we identify three major tasks in digital pathology: classification, regression and segmentation, and the main purpose is to train a model which is able to work on unknown test image based on optimizing loss function [3].

[1] Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology.
[2] Department of Electronics, Electrical Engineering and Microelectronics, Silesian University of Technology.
[3] Department of Graphics, Computer Vision and Digital Systems, Silesian University of Technology.
[4] Department of Data Science and Engineering, Silesian University of Technology.

This manuscript presents the difficulties that must be taken into account when starting the task of detecting tumor regions in hematoxylin and eosin (HE) stained histopathological images. In section 2, we described the data that were used and the deep learning model created for detection task. In section 3, we presented issues devoted to usage of deep learning methods in HE stained images analysis that we found during our research, such as differences in tissue staining between the slides, magnification used during slide scanning, image patching to prepare data for model building, and diffferent type of prediction methods.

10.2. Materials and methods

10.2.1. Data

Four breast cancer datasets were used in this study. The IDC_regular_ps50_idx5 dataset [2] consist 162 whole slides prepared in magnification 40x. From it, 277 524 patches in size 50 x 50 were extracted and downsampled to the 2.5x magnification (198,738 images without cancer, and 78 786 images containing cancer). The file name consists of the patient identifier and the X and Y coordinates from which the patch was extracted. A 0 indicates no cancer and a 1 indicates the presence of cancer. BreastPathQ dataset [4] contains 96 images scanned at 20x magnification where 64 images contain tumour tissue. 2 579 patches in size 512 x 512 were extracted from the entire WSI images and each include expert opinion about cancer cellularity, i.e. proportion of cancer region to tissue region (from 0 to 1). Another dataset used in the research is TCGA-BRCA [5], which includes WSI for 1098 patients. Images have been prepared at 20x or 40x magnification. All images have been lossy compressed using JPEG with values between 30 and 70. Data contained in TCGA were collected from many sites all over the world in order to reach their accrual targets. The last collection used is BreCaHAD [6] which consist of 162 histopathological images of breast tissue from cancer patients. Each image has size 1360 x 1024 and was prepared in 40x magnification. The dataset contains various cases of tumor malignancy obtained from archived surgical pathology cases. All datasets used in this work were derived from publicly available sources. Example image from each dataset is shown in Figure 10.1.

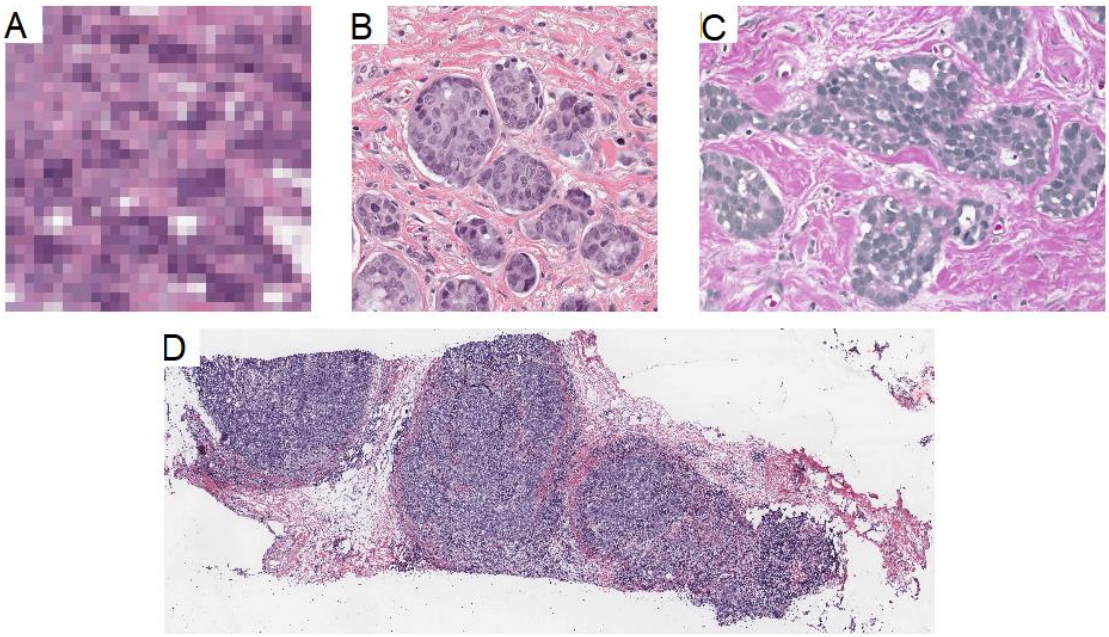

Fig. 10.1. Sample image from each dataset: (A) IDC_regular_ps50_idx5, (B) BreastPathQ, (C) BreCaHAD, (D) TCGA-BRCA
Rys. 10.1. Przykładowy obraz z każdego zbioru danych: (A) IDC_regular_ps50_idx5, (B) BreastPathQ, (C) BreCaHAD, (D) TCGA-BRCA

10.2.2. Models

The results presented in this manuscript were obtained using AlexNet deep convolutional neural network architecture, which is commonly used in image classification task [7]. AlexNet is composed of 3 convolutional layers, one max pooling layer, two average pooling layers, and 3 fully connected layers. The last layer uses softmax as an activation function. As deep learning framework, we selected open source framework Caffe, which was initially developed for deep learning vision solutions and allows rapid prototyping [8]. The models were trained on the datasets IDC_reguls_ps50_idx5 and BreasthPathQ due to the class labels held by the experienced pathologists and tested on the two other datasets. We note that the training time for each model is about 30 h on a NVIDIA Tesla C2075 graphic card using CUDA 9.0.

10.3. Results and discussion

10.3.1. Magnification of scanning

When looking at WSI from different sources, it can be notice that such images often differ in resolution. This is due to hardware aspects such as the magnification of the microscope used by the scanner, the numerical aperture of the objective and the sensor used (size and number of individual pixels) [9, 10]. A model trained on a WSI dataset of a certain magnification will not perform well with images scanned on a different magnification, because the desired regions will have other sizes and properties sought by the neural network. To overcome this problem, the apparent magnification difference (factor *K*) can be calculated and downsampling could be applied. There are several ways, but here we decided to calculate the average values for each *K* x *K* window and insert them instead of the pixel squares. However, this approach also has its drawbacks. While there is no problem when the height and width of the WSI are multiples of the number representing the apparent magnification, it arises in the opposite situation. If *K* is an integer, it is possible to crop the image by removing excessive pixels, but when doing so we lose a part of the WSI that may be valuable. If *K* is a floating point number, such a solution may not suffice (if accuracy is to be achieved) because the WSI size divisible by a factor of *K* will never be reached.

Regardless of the method, the use of downsampling itself involves a significant loss of information. It is especially visible when the factor *K* has a large value (Figure 10.2). The larger the *K*, the more distant and less detailed the image is.

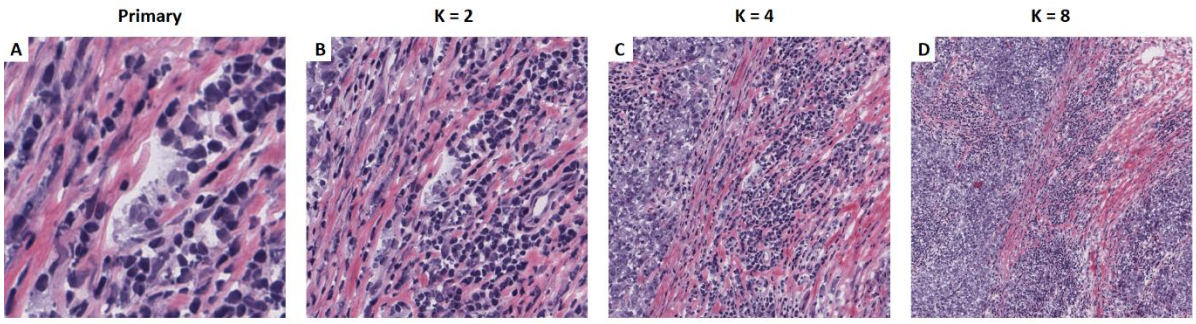

Fig. 10.2. WSI fragment (500 x 500 px) showing the effect of downsampling on the primary image. From left, magnification: (A) 40x, (B) 20x, (C) 10x, (D) 5x

Rys. 10.2. Fragment WSI (500 x 500 px) pokazujący efekt próbkowania w dół na obrazie bazowym. Od lewej, powiększenia: (A) 40x, (B) 20x, (C) 10x, (D) 5x

The effect of different magnification is also seen in a tissue region detection task (Figure 10.3). All elements from the 20x magnification dataset were downsampled (K = 8) to obtain 2.5x. The neural network model was then trained and the 2.5x (64 x 64) and 20x (512 x 512) image predictions were made (Figure 10.3 B, C). As can be seen, the prediction of the image before downsampling (Figure 10.3 C), although it appears more detailed (consisting of more patches), contains much more noise, indicating a large number of false positives.

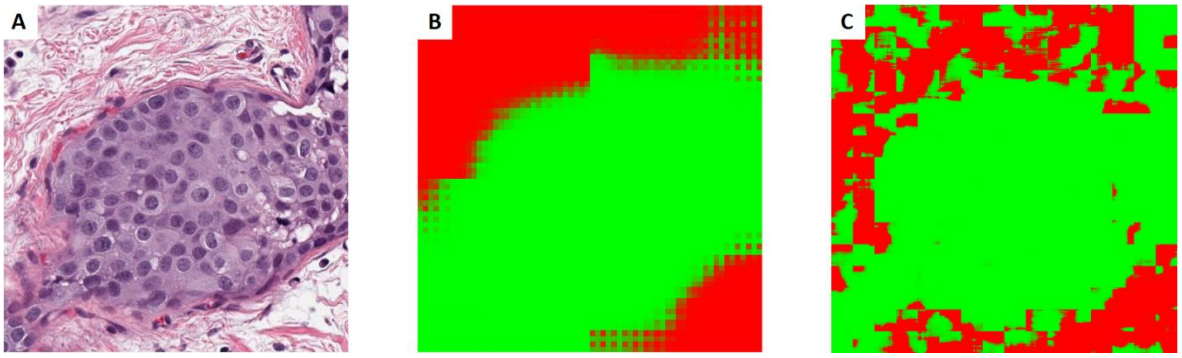

Fig. 10.3. From left: (A) input image, (B) model prediction, (C) model prediction for the image with inappropriate magnification. The green color indicates that the network has classified the region as a cancer, while the red color indicates no cancer

Rys. 10.3. Od lewej: (A) obraz wejściowy, (B) predykcja modelu, (C) predykcja modelu dla obrazu z nieodpowiednim powiększeniem. Kolor zielony oznacza, że sieć sklasyfikowała region jako nowotwór, natomiast kolor czerwony oznacza brak nowotworu

10.3.2. Variation in tissue staining

The next problem concerns the colors observed on the slides (Figure 10.4). The visible variations in the tissue staining stage are due to several factors. One of them is the lab protocols applied by individual laboratories. Also, parameters such as tissue thickness, staining concentration and staining time affect the final color. The second aspect is the variety of whole slide scanners that are on the market. The production of digital slides consists of several stages that largely depend on certain hardware or software properties. Unfortunately, manufacturers do not have commonly used standards, so there are visible differences in the final color of each structure [11].

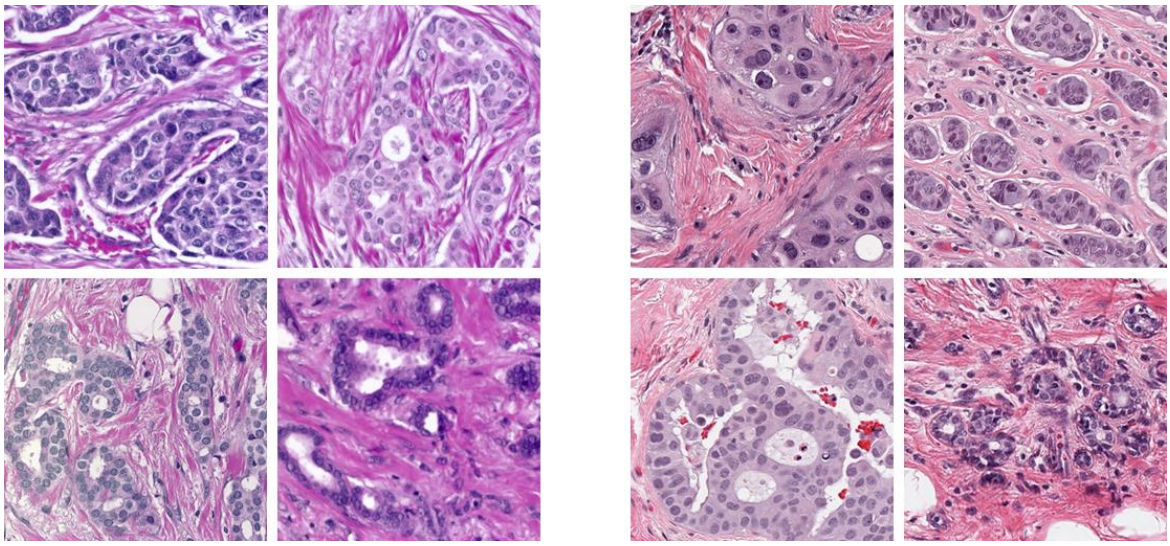

Fig. 10.4. Comparison of the WSI staining between slides
Rys. 10.4. Porównanie barwienia WSI pomiędzy szkiełkami

A trained neural network may interpret staining variation as analysing images from a different cohort. It is enough for the input slides to have more intense coloring to cause improper prediction of the well-trained deep learning model (Figure 10.5).

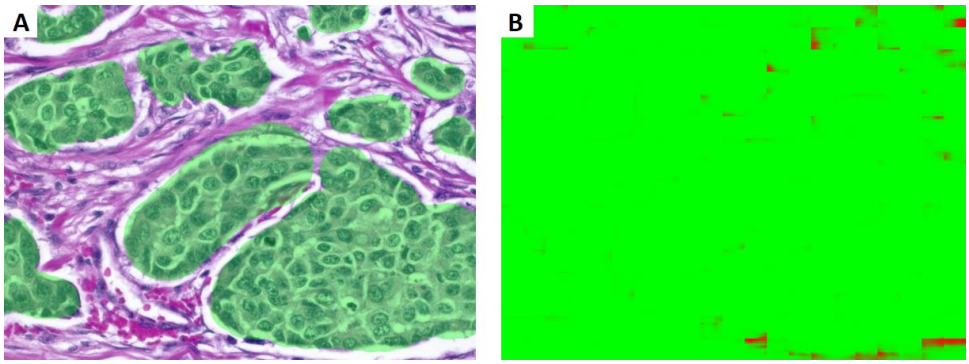

Fig. 10.5. Output image of the deep network, when given a part of the slide with more intense staining. From left: (A) expert opinion (in green), (B) prediction
Rys. 10.5. Obraz wyjściowy sieci głębokiej, po podaniu fragmentu szkiełka z bardziej intensywnym barwieniem. Od lewej: (A) opinia eksperta (na zielono), (B) predykcja

One solution to this problem is tissue stain normalization. There are several algorithms available, but in our experiments we used the method introduced by Azevedo, et al. [12]. The effects of tissue stain normalization are shown in Figure 10.6.

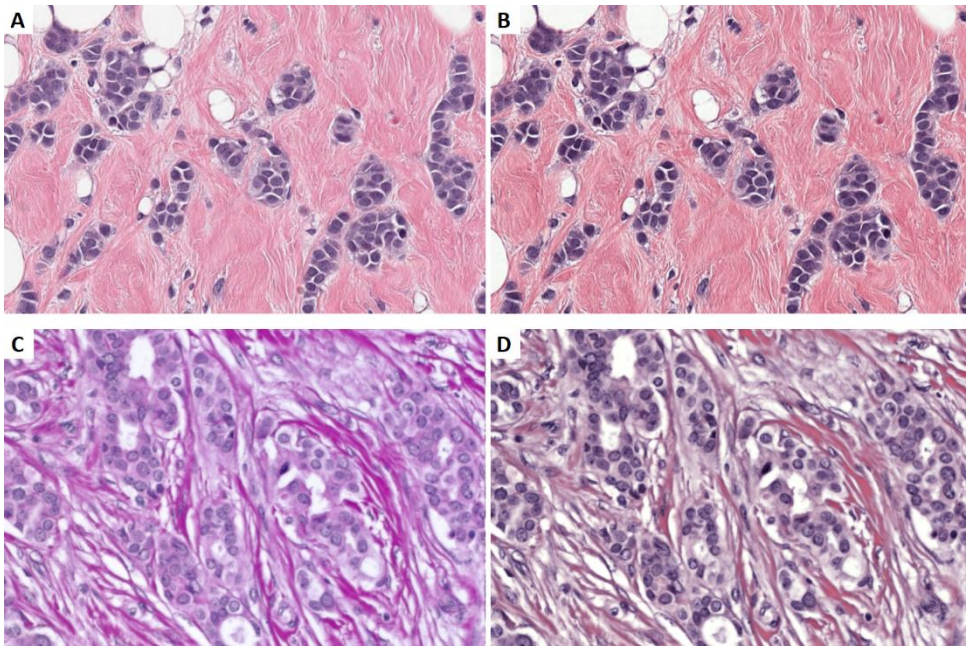

Fig. 10.6. Example of tissue stain normalisation. Left: original images (A, C), right: normalised images (B, D)
Rys. 10.6. Przykład normalizacji barwienia tkanek. Po lewej: obrazy oryginalne (A, C), po prawej: obrazy znormalizowane (B, D)

10.3.3. Selection of image patch size

When preparing a deep learning model, selection of hyperparameters or patch size might have much greater impact on the outcome than selection of network architecture itself [2, 3]. Choosing proper window size is a task-specific, time-consuming and problematic process. The small patch size is a frequently used in H&E image analysis, since it is not hardware limited and can be a good solution for the small number of images available for network training. Unfortunately, the disadvantage of this solution is the lack of information about the environment outside the patch, which generates opposite information at the edges of the two patches. Two tumor detection models trained on the same dataset with different patch window sizes can give much different prediction results (Figure 10.7). For the smaller window size (Figure 10.7C) prediction is much better on the borders of each structure, however smaller tissue components, like invidual lymphocytes, are also predicted as cancer. On the other hand, when patch size is too big (Figure 10.7B), the detected cancer region is much more imprecise.

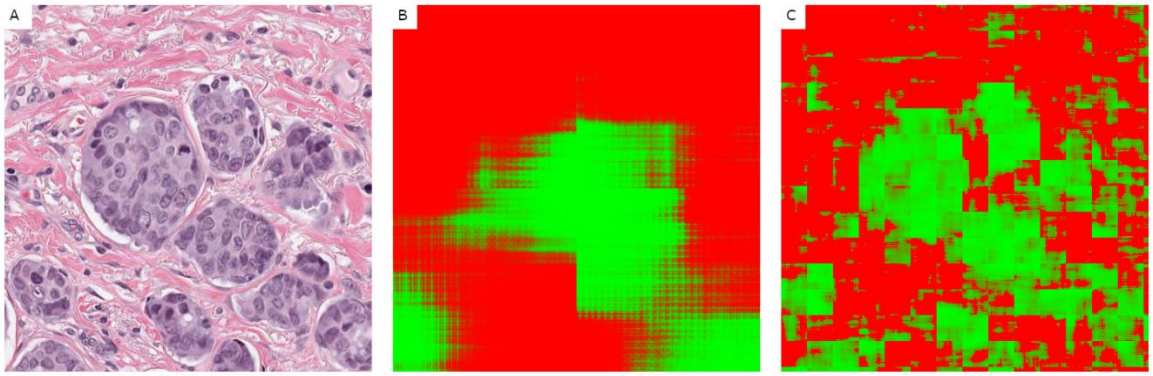

Fig. 10.7. Prediction results for models with different windows size: (A) input image, BreastPathQ dataset (B) window size 227 x 227 and (C) window size 32 x 32

Rys. 10.7. Wyniki predykcji dla modeli z różnymi rozmiarami okien: (A) obraz wejściowy (zbiór danych BreastPathQ), (B) rozmiar okna 227 x 227 oraz (C) rozmiar okna 32 x 32

The observed effect of sharp edges resulting from the merging of adjacent patches can be removed using a method of overlapping windows. To do this, a prediction must be made for the windows located at the boundaries of neighbours patches. When generating the output image for more values at a single point, an average score, min or max value can be determined (Figure 10.8).

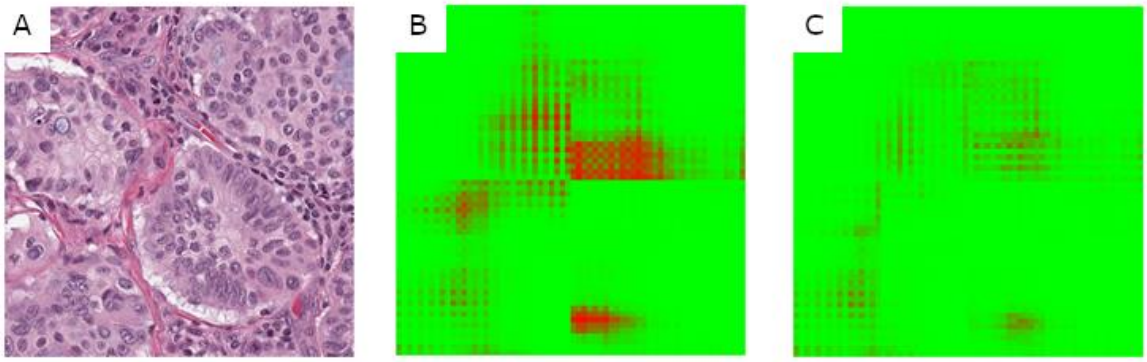

Fig. 10.8. Removal of sharp edges artifacts: (A) input image (BreastPathQ), (B) output image without overlapping, (C) output image after overlapping with max value per pixel

Rys. 10.8. Usuwanie artefaktów ostrych krawędzi: (A) obraz wejściowy (BreastPathQ), (B) obraz wyjściowy bez nakładania, (C) obraz wyjściowy po nałożeniu z wartością maksymalną na piksel

10.3.4. Prediction methods

Image classification with deep neural network is done to assign a class for the entire patch or each individual pixel. In order to achieve higher accuracy, a pixel-level classification method can be used, but it results in a significant increase in computation time. For example, tissue presented on Figure 10.9 consists of 978 patches with size

32 x 32. The time required to perform the prediction for each pixel and generate the output image was 130 minutes. Remember that WSI images can be very large, which translates into tens of thousands of small patches. If the goal is to identify the tumor region for further analysis, e.g. to count cells in this region, image classification at the patch level will allow faster computations and could lead to obtain satisfactory results (Figure 10.9D).

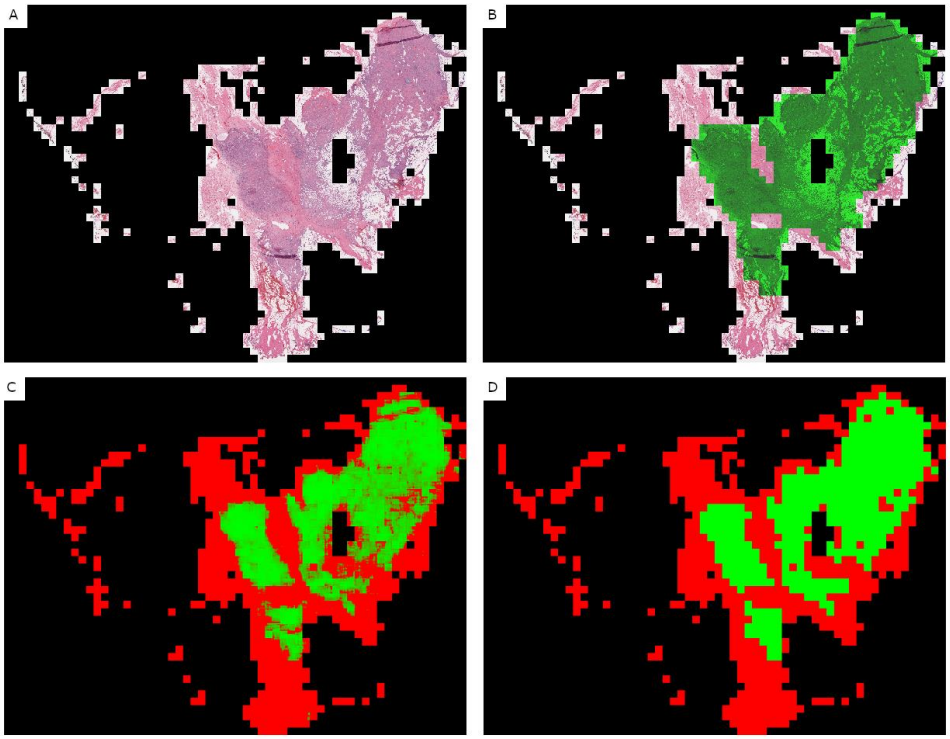

Fig. 10.9. Classification per pixel and patch results (A) input image (IDC_regular_ps50_idx5), (B) expert opinion (C) output image, prediction per pixel, (D) output image, prediction per patch

Rys. 10.9. Wyniki klasyfikacji na piksel i wycinek: (A) obraz wejściowy (IDC_regular_ps50_idx5), (B) opinia eksperta, (C) obraz wyjściowy predykcja na piksel, (D) obraz wyjściowy predykcja na wycinek

10.4. Conclusions

In this paper we presented how deep learning models can be used to analyze HE-stained histopathological images and discussed what problems can be encountered during process of model development, training and prediction. Breast cancer region detection task was used here as an example. Due to the use of different WSI scanners for data acquisition and different stain preparations the input data may vary greatly and thus they require more complex pre-processing algorithms. Also, the hardware limitations

resulting in long computation time, make it necessary to choose compromise solutions of varying accuracy. Obtaining a universal solution for histopathological image analysis is a lengthy process. We hope that the issues presented in our work, together with the proposed solutions, will contribute to the development of the field of digital histopathology.

Acknowledgements

The research reported in this paper was co-financed by the European Union from the European Social Fund in the framework of the project "Silesian University of Technology as a Center of Modern Education based on research and innovation" POWR.03.05.00-00-Z098/17.

Bibliography

1. S. Nam *et al.*, Introduction to digital pathology and computer-aided pathology. *J Pathol Transl Med* **54**, 125-134 (2020).
2. A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics* **7**, 29-29 (2016).
3. C.L. Srinidhi, O. Ciga, A.L. Martel, Deep neural network models for computational histopathology: A survey. *Medical Image Analysis* **67**, 101813 (2021).
4. M. Peikari, S. Salama, S. Nofech-Mozes, A.L. Martel, Automatic cellularity assessment from post-treated breast surgical specimens. *Cytometry A* **91**, 1078-1087 (2017).
5. K. Clark *et al.*, The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging* **26**, 1045-1057 (2013).
6. A. Aksac, D.J. Demetrick, T. Ozyer, R. Alhajj, BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis. *BMC Research Notes* **12**, 82 (2019).
7. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84-90 (2017).
8. Y. Jia *et al.*, paper presented at the Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA, 2014.
9. N. Farahani, A.V. Parwani, L. Pantanowitz, Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International* **7**, 23-33 (2015).

10. T.L. Sellaro *et al.*, Relationship between magnification and resolution in digital pathology systems. *Journal of pathology informatics* **4**, (2013).
11. P.A. Bautista, N. Hashimoto, Y. Yagi, Color standardization in whole slide imaging using a color calibration slide. *Journal of pathology informatics* **5**, (2014).
12. T.A. Azevedo Tosta, L.A. Neves, M.Z. do Nascimento, Segmentation methods of H&E-stained histological images of lymphoma: A review. *Informatics in Medicine Unlocked* **9**, 35-43 (2017).

TUMOR DETECTION ON HE-STAINED HISTOPATHOLOGICAL IMAGES USING DEEP LEARNING METHODS: CHALLENGES AND SOLUTIONS

Abstract

The advent of digital histopathology has opened up new possibilities for the processing and analysis of high-resolution tissue images using deep learning methods. While working to develop a model to detect region of breast cancer on HE-stained tissue slide, we encountered several major complications. The aim of this work is to describe these challenges and propose several solutions. We addresses the following issues and showed their impact on the performance of the deep learning model: (i) variation in tissue staining, (ii) different magnification of scanning, (iii) selection of size of the image patch, (iv) usage of patch-based or pixel-based model prediction method. The analysis were done using deep learning model consisting of convolutional layers that was constructed on several independent datasets. We believe that this publication is a good starting point for image analysis practicioners who wants to construct deep learning models on histopathological images.

Keywords: digital pathology, histopathological images, HE-staining, deep learning, tumor detection.