Tissue Type Recognition in Whole Slide Histological Images

Alexander Khvostikov[1,2], Andrey Krylov[1,2], Ilya Mikhailov[3], Pavel Malkov[4] and Natalya Danilova[4]

[1]*Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskie Gory, 1, building 52, Moscow, 119991, Russia*

[2]*Moscow Center of Fundamental and Applied Mathematics, Moscow, Russia*

[3]*Faculty of Medicine, Lomonosov Moscow State University, Lomonosovskiy prospekt, 27, building 1, Moscow, 119991, Russia*

[4]*Medical Research and Educational Center, Lomonosov Moscow State University, Lomonosovskiy prospekt, 27, building 10, Moscow, 119991, Russia*

Abstract

Automatic layers recognition of the wall of the stomach and colon on whole slide images is an extremely urgent task in digital pathology as it can be used for automatic determining the depth of invasion of the digestive tract tumors. In this paper we propose a new CNN-based method of automatic tissue type recognition on whole slide histological images. We also describe an effective pipeline of training that uses 2 different training datasets. The proposed method of automatic tissue type recognition achieved 0.929 accuracy and 0.903 balanced accuracy on CRC-VAL-HE-7K dataset for 9-types classification and 0.98 accuracy and 0.926 balanced accuracy on the test subset of whole slide images from PATH-DT-MSU dataset for 5-types classification. The developed method makes it possible to classify the areas corresponding to the gastric own mucous glands in the lamina propria and also to distinguish the tubular structures of a highly differentiated gastric adenocarcinoma with normal glands.

Keywords

Deep Learning, Image Segmentation, Histology, Pathology, Whole Slide Images, Tissue Recognition

1. Introduction

Determining the depth of invasion (T) of the digestive tract tumors (stomach and colon tumors) is one of the most important tasks of surgical pathology since the depth of invasion is a reliable highly significant negative prognostic factor. If determining the depth of invasion of the digestive tract tumors at advanced stages is a relatively simple task for a pathologist, then the determination of foci of microinvasion of adenocarcinoma in polyps with low and high grade dysplasia is a rather difficult task, which can be tried to be solved using of deep learning based methods.

The key task for developing these algorithms is learning to recognize the layers of the wall of the stomach and colon on whole slide images (WSI), namely the lamina propria, muscularis mucosa, submucosa, own muscle layer, subserous layer, serous membrane and adjacent areas adipose tissue. Recognition of tumors and layers of the stomach wall in gastric cancer is an example of one of the most urgent problems in the digital pathology of the digestive tract tumors. The papers available in this area are mainly devoted to the recognition and classification tasks for endoscopic images obtained during gastroscopy. For example, it has been shown that the accuracy of deep learning-based algorithms for predicting the depth of invasion of early gastric cancer is 71.43%, which is slightly higher compared to the opinion of endoscopists of 64.41% [1].

Whole slide histological images recognition in gastric cancer using deep learning algorithms is used to estimate the density of infiltration by various types of immune cells [2]. Deep learning algorithms are also being developed to count the number of lymph nodes with metastases in gastric cancer. Already existing algorithms make it possible to identify lymphoid tissue and tumor area, and then determine the ratio of tumor area to lymphoid tissue area in order to determine lymphogenous metastases in gastric cancer. After training the tumor detection performance of the model was comparable to that of experienced pathologists and achieved similar performance in two independent cohorts of gastric cancer screening [3].

Thus, the existing algorithms based on deep learning are so far aimed only to solve specific narrow problems in relation to histological images of gastric cancer. While there are already quite effective algorithms for recognizing the depth of invasion and the layers of the intestinal wall in colorectal cancer [4], then there are no similar algorithms for gastric cancer, which can be explained by the higher frequency of diffuse type tumors and tumors with a discohesive component among gastric tumors.

2. Used data

In this work we use three different image datasets that are developed for the purpose of WSI segmentation and tissue type recognition.

The first dataset is NCT-CRC-HE-100K [5]. It consists of 100,000 non-overlapping image patches from hematoxylin & eosin (H&E) stained histological images of human colorectal cancer and normal tissue (data available at http://dx.doi.org/10.5281/zenodo.1214456). Each patch has a resolution of 224×224 pixels and is matched with one of the 9 class labels according to the specific tissue type or background: adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), colorectal adenocarcinoma epithelium (TUM). These images were manually extracted from 86 H&E stained human cancer tissue slides from formalin-fixed paraffin-embedded (FFPE) samples from the NCT Biobank (National Center for Tumor Diseases, Heidelberg, Germany) and the UMM pathology archive (University Medical Center Mannheim, Mannheim, Germany). Tissue samples contained CRC primary tumor slides and tumor tissue from CRC liver metastases; normal tissue classes were augmented with non-tumorous regions from gastrectomy specimen to increase variability.

The second dataset is CRC-VAL-HE-7K [5]. It consists of 7180 image patches from 50 patients with colorectal adenocarcinoma (no overlap with patients in NCT-CRC-HE-100K). The authors

of the dataset recommend to use this dataset as a validation set for models trained on the larger NCT-CRC-HE-100K. Like in the NCT-CRC-HE-100K dataset, image patches are 224×224 pixels at 0.5 MPP and correspond to the same 9 classes. All tissue samples were provided by the NCT tissue bank.

The third dataset is the subset of PATH-DT-MSU dataset [6, 7], containing 7 H&E whole slide histological images of digestive tract tumors. Each WSI image is a full-thickness fragment of the stomach wall, cut from the surgical material, and includes areas of adenocarcinoma, adjacent areas of visually unchanged lamina propria and the underlying layers of the stomach wall (muscularis mucosae, submucosa, own muscle layer, subserous areas). The images are collected with 40X scanning magnification and have a resolution up to 111.552×90.473 pixels. These whole slide images contain annotation of 5 classes (tissue types): areas of gastric adenocarcinoma (TUM), unchanged areas of the lamina propria (LP), unchanged areas of the muscularis mucosae (MM), areas of the submucosa, the own muscle layer of the stomach and subserous areas in one class (AT), background (BG). The annotations represent the polygons, all pixels inside them belong to the same class (tissue type or background). An example of WSI image from PATH-DT-MSU with its annotation is shown in Fig.1. Images were captured using a scanning microscope Leica Aperio AT2 (Leica Microsystems Inc., Germany), the annotations were made using the Aperio ImageScope 12.3.3 (Leica Microsystems Inc., Germany).

It is also worth noting that area of annotated regions of whole slide images in PATH-DT-MSU dataset is relatively small compared to the area of WSI images. The main reason of this is the necessity of choosing only the regions with "clear" texture that is most characteristic to the one of the corresponding 5 classes. Also an objective reason of this fact is the complexity and laboriousness of WSI annotation process.

3. Proposed methodology

At the moment, the direct implementation of segmentation models for whole slide image processing using fully-supervised deep learning methods is almost impossible due to the necessity of preparing accurate full pixel-wise annotations of whole slide images. The common way of getting the approximate result of automatic tissue segmentation is splitting WSI image into small patches and predicting the class label for each patch using the classification model. Since the impressive resolution of WSI images this way provides a tissue segmentation map with the resolution which is appropriate from the medical point of view [4].

In this paper we propose a CNN-based model for automatic tissue type recognition of whole slide histological images based on the patch classification approach. We also propose a special pipeline for training the model using several different datasets to maximize the efficiency of the model.

3.1. Data preparation

Image patches from NCT-CRC-HE-100K dataset are available in two versions: with and without color-normalization using Macenko's method [8], image patches from CRC-VAL-HE-7K are available only in color-normalized version. Since that we use only color-normalized version of
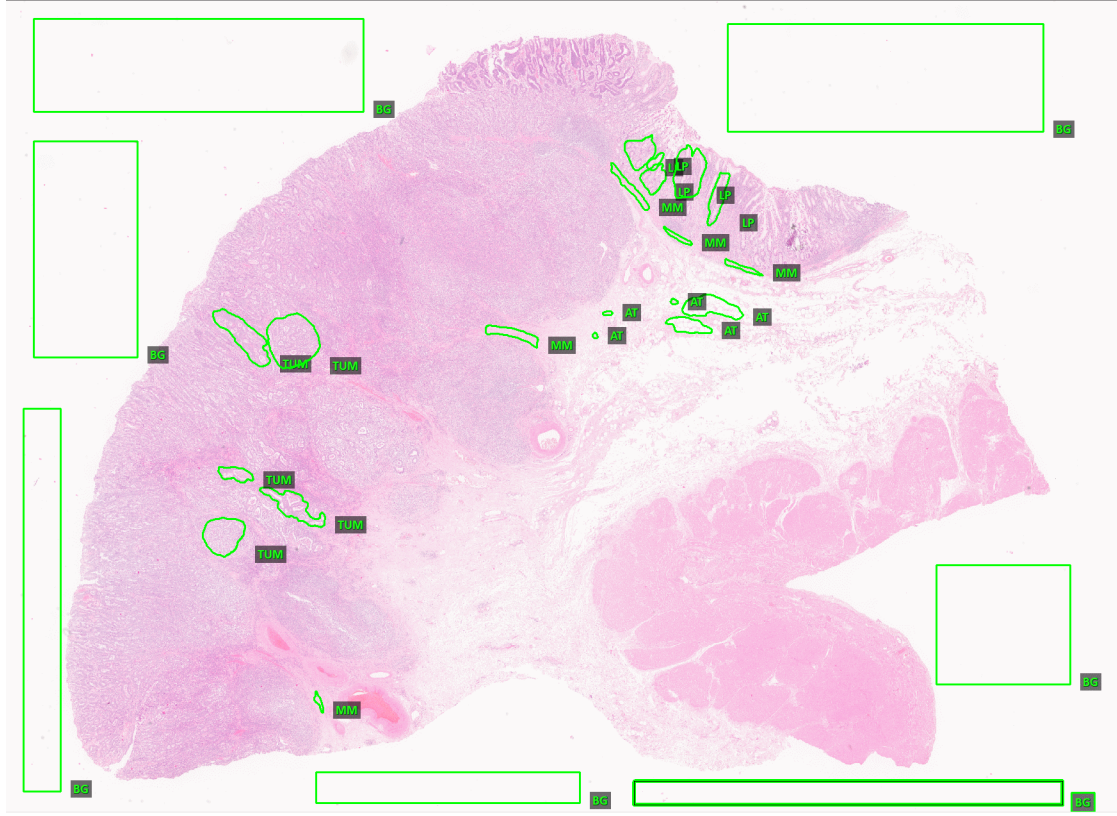
Figure 1: An example of annotation of H&E stained whole slide image from PATH-DT-MSU dataset representing gastric cancer. Image resolution is 73704×56317 pixels. The inside content of each green polygons correspond to 5 classes: AT, BG, LP, MM, TUM.

NCT-CRC-HE-100K dataset. This allows us to use CRC-VAL-HE-7K dataset for validation of the model trained on NCT-CRC-HE-100K without any modifications.

The whole slide images from the PATH-DT-MSU dataset are also split into train and test subsets. The train subset contains 5 annotated images, the test subset contains 2 annotated images.

The annotations of WSI images from PATH-DT-MSU dataset made with Aperio ImageScope software can easily be exported in the XML markup format. In order to train the patch classification model and make our PATH-DT-MSU dataset compatible with NCT-CRC-HE-100K and CRC-VAL-HE-7K we extract patches from the WSI images in accordance with annotations.

For the train subset we use the window of 320×320 pixels, which is moved over each WSI image with stride of 160 pixels. At each position we calculate whether the window intersects any of the annotation polygons and if the intersection happens the area of intersection is calculated. If the area of intersection is larger than 0.75 of the window area, the patch corresponding to the current position of the window is extracted and saved.

For the test subset the process of patch extraction is the same except that the window size is 224×224 pixels and the stride is 112 pixels. The different size for the windows is chosen in order to augment the training set of extracted patches with rotation. The stride of the moving window is always half of the window size.

Moreover, due to the 2 times difference in the magnification level of images from NCT-CRC-HE-100K and PATH-DT-MSU datasets, all the described above patch extraction procedure from PATH-DT-MSU WSI images is performed on the half of the largest scale (at 20x magnification level).

The described patch extraction procedure with the current annotations of WSI images from PATH-DT-MSU dataset allowed us to extract 70871 patch from training subset and 14462 patches from test subset. Some examples from the extracted set of patches are shown in Fig.2.

It is also necessary to mention that all three datasets used in this work are imbalanced (the number of patches per class for all used datasets is given below):

- NCT-CRC-HE-100K. ADI - 10407, BACK - 10566, DEB - 11512, LYM - 11557, MUC - 8896, MUS - 13536, N0RM - 8763, STR - 10446, TUM - 14317,
- CRC-VAL-HE-7K. ADI - 1338, BACK - 847, DEB - 339, LYM - 634, MUC - 1035, MUS - 592, NORM - 741, STR - 421, TUM - 1233,
- Train subset of PATH-DT-MSU. AT - 15207, BG - 45977, LP - 3911, MM - 675, TUM - 5101,
- Test subset of PATH-DT-MSU. AT - 1491, BG - 11252, LP - 590, MM - 211, TUM - 918.

To overcome the class imbalance we use the simple oversampling technique that guarantees that the number of patches of each class fed into the CNN model during training is equal.

3.2. CNN model and training strategy

In this work we use a CNN-based model for automatic tissue type recognition of whole slide histological images based on the patch classification approach. Since the amount of data in the obtained datasets is limited, we chose the DenseNet architecture [9] as it tends to perform well in the case of relatively small number of training samples.

DenseNet is the further development of the ResNet architecture [10] with additional direct connections between any two layers within the same feature-map size. The DenseNet consists of several dense blocks, each of them corresponding to fixed feature-map size, connected with pooling layers. Visualization of single dense block of the DenseNet architecture is shown in Fig.3, visualization of the full DenseNet architecture is shown in Fig.4. As a direct consequence of the input concatenation, the feature-maps learned by any of the DenseNet layers can be accessed by all subsequent layers. This encourages feature reuse throughout the network, and leads to more compact model.

To get the most efficient model of tissue type recognition in this work we use the transfer learning principle gradually turning the DenseNet model to be able to classify patches from the target PATH-DT-MSU whole slide images.

Thus, the training of the proposed model consists of two main steps:

- take the DenseNet model pretrained on ImageNet dataset and fine tune it on the NCT-CRC-HE-100K dataset,
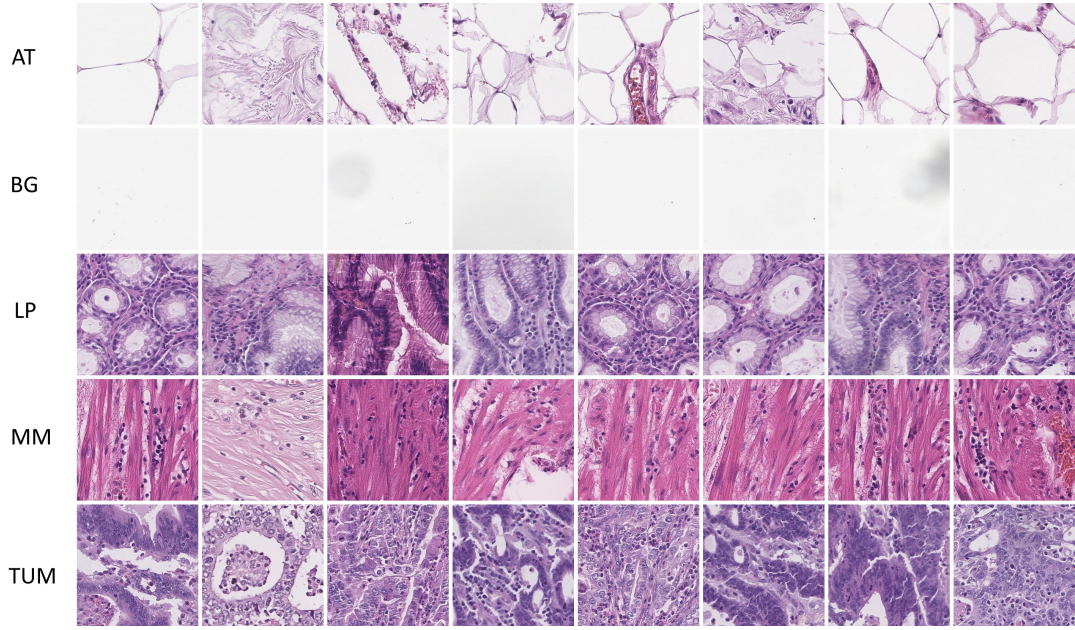
Figure 2: Example images for each of the five tissue classes extracted from whole slide images from PATH-DT-MSU dataset. Areas of gastric adenocarcinoma (TUM), unchanged areas of the lamina propria (LP), unchanged areas of the muscularis mucosae (MM), areas of the submucosa, the own muscle layer of the stomach and subserous areas in one class (AT), background (BG).
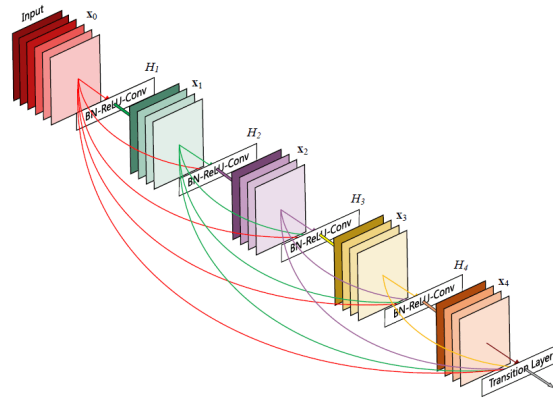

Figure 3: A 5-layer dense block. Each layer takes all preceding feature-maps as input. BN stands for batch normalization, Conv stands for convolution.

- fine tune the obtained DenseNet model on the PATH-DT-MSU dataset.

These steps will be split into three phases of training and will be described in detail in the next section of the paper.
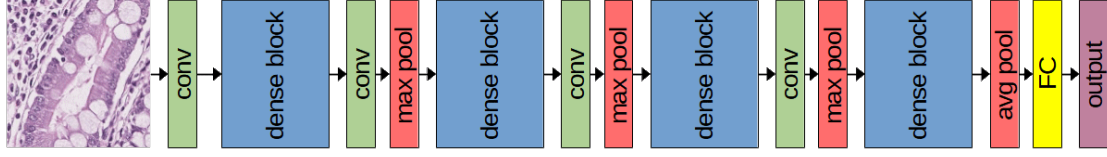
Figure 4: A DenseNet architecture [9] with 4 dense blocks. Max pool stands for max pooling operation with stride 2, avg pool stands for average pooling, conv stands for convolution, FC stands for fully-connected layer.

4. Experiments and results

DenseNet-121 version from [9] was chosen as the base model. All described below training procedures were done with Adam [11] optimizer with automatic decrease of learning rate on plateau. The batch size was chosen to be 16.

Since all used in this work histological datasets are imbalanced concerning the number of patches per class (tissue type), we use the oversampling technique while training CNN-based model. Particularly the data batches that are fed into CNN are created by generator in the following way:

- randomly select one of the classes,
- randomly select a patch from the list existing patches corresponding to the selected class,
- augment the patch (different augmentation methods are used at different phases of training),
- repeat until the batch is collected.

This technique equalizes the amount of patches for each class passed through the CNN during training thus making the classification more effective and robust.

During the validation step all image patches from the validation set are passed through the CNN, accuracy, balanced accuracy and confusion matrix are calculated based on the retrieved prediction.

4.1. First phase of training

The first phase of training consists in fine tuning of the DenseNet-121 model that is pretrained on the ImageNet dataset. Due to the enormous amount of different types of images in ImageNet dataset the head of the pretrained DenseNet-121 can be used as a universal image feature extractor. We replace the last fully-connected layer with the new one with 9 outputs in correspondence with the number of classes in NCT-CRC-HE-100K dataset, freeze all head layers and train the obtained modified DenseNet for 20 epochs with the initial learning rate of $2 \cdot 10^{-5}$. At this step the augmentation with random flip and 90 degrees turn is applied. The validation is performed on CRC-VAL-HE-7K dataset. We reached the accuracy value of 0.919 and balanced accuracy of 0.8816 at this phase.

4.2. Second phase of training

The DenseNet obtained at the previous phase now represent an universal feature extractor tuned for tissue type classification. In order to better adapt it for working with histological images at the second phase of training we unfreeze all layers in the model and further train it with the same NCT-CRC-HE-100K dataset for 30 epochs with the initial learning rate of 10^{-4}. The augmentation and validation are the same as at the previous phase. At this point we reached the accuracy value of 0.929 and balanced accuracy of 0.903.

4.3. Third phase of training

After the previous phase of training the DenseNet model is much more adapted to work with histological images. Finally we fine tune it on patches from whole slide images of PATH-DT-MSU dataset. We replace the last fully-connected layer with the new one with 5 outputs in correspondence with the number of classes in PATH-DT-MSU dataset, freeze all head layers and train the model for 8 epochs with the initial learning rate of 10^{-5}. Herewith the data augmentation besides the random flip includes the random rotation with subsequent crop to the 224×224 size. The validation is performed on the test set of patches extracted from PATH-DT-MSU whole slide images. At this phase we finally achieved the accuracy value of 0.98 and balanced accuracy of 0.926. The obtained confusion matrix is shown in Table.1.

4.4. Applying model to whole slide images

In order to perform the visual inspection of the effectiveness of the trained CNN model in the task of automatic tissue type recognition we apply it to the whole slide images from the test subset of PATH-DT-MSU. Each WSI image is split into patches of size 224×224 pixels without overlap, for each patch the trained model predicts the class label, after that all the predictions are combined into a matrix, which is then visualized as a semi-transparent layer imposed on the source WSI image.

The visualization of the proposed method of tissue type recognition is shown in Fig.5. The developed algorithm makes it possible to classify with sufficient accuracy the areas corresponding to the gastric own mucous glands in the lamina propria and also makes it possible to distinguish the tubular structures of a highly differentiated gastric adenocarcinoma with normal glands. Less accurate results obtained with respect to the muscularis mucosa can be explained by a similar structure with longitudinal smooth muscle fibers in the own muscle layer of the stomach.

4.5. Comparison with alternative models and training techniques

Several experiments with alternative models and training techniques were tested within this work.

Different shallow CNN models with reduced number of parameters (compared to DenseNet-121) without pretraining were only able to provide the balanced accuracy value of 0.76 on PATH-DT-MSU dataset.

Direct fine tuning of the pretrained DenseNet-121 model on NCT-CRC-HE-100K dataset in one step instead of the proposed two phases of training and further training the model on
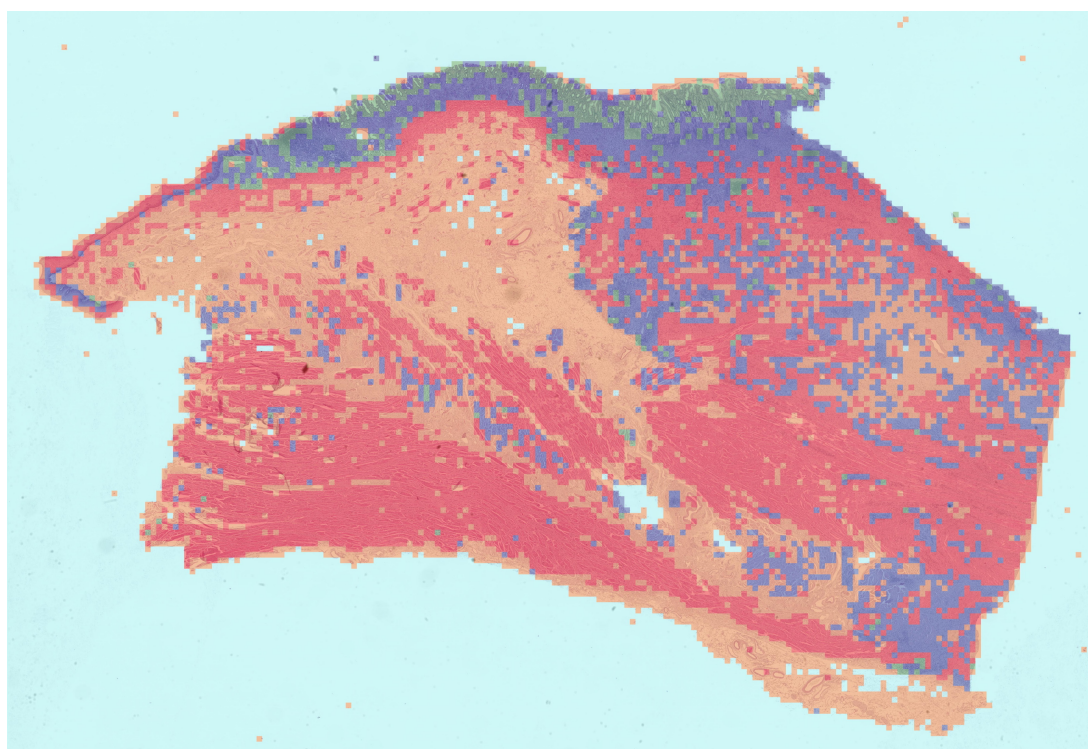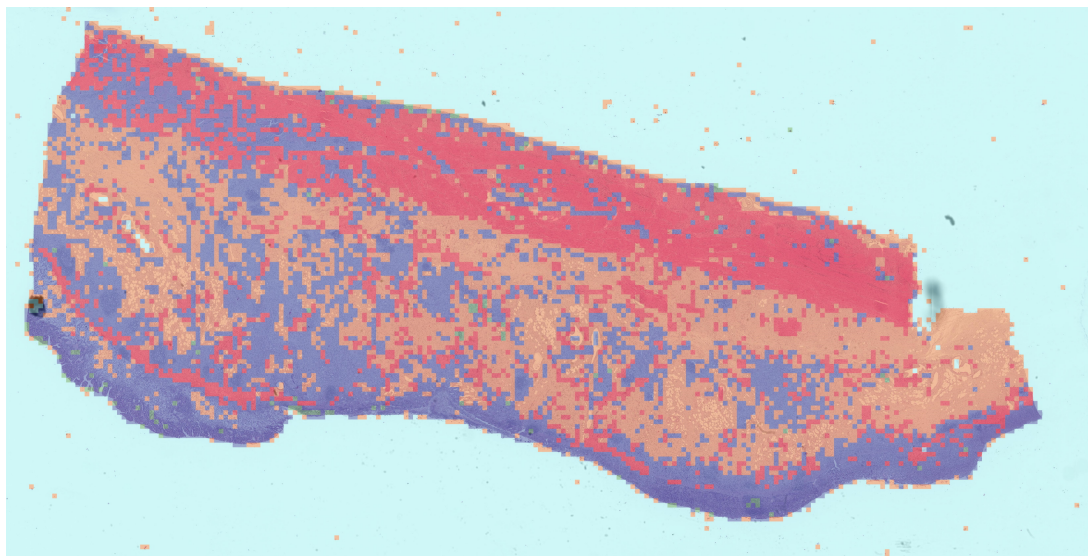
Figure 5: Visualization of the proposed method of tissue type recognition for test whole slide images from PATH-DT-MSU dataset. AT areas are highlighted with orange color, BG areas are highlighted with cyan color, LP areas are highlighted with green color, MM areas are highlighted with red color, TUM areas are highlighted with purple color.

Table 1
Confusion matrix for the test set of patches extracted from whole slide images of PATH-DT-MSU dataset. The columns correspond to predicted values, the rows correspond to ground truth values.

pred\gt	AT	BG	LP	MM	TUM
AT	1481	0	1	9	0
BG	41	11210	1	0	0
LP	2	0	492	0	96
MM	9	0	1	200	1
TUM	0	0	109	19	790

PATH-DT-MSU dataset allowed us to get the balanced accuracy value of 0.9 on PATH-DT-MSU dataset.

Replacing the third phase of training with two phases similar with the first two phases (first training fully connected layer, then training whole model) did not improve the achieved level of accuracy. This can be explained by the small size of the training set extracted from whole slide images of PATH-DT-MSU dataset.

5. Implementation details

The proposed CNN-based method for automatic tissue type recognition in whole slide histological images was implemented using Python 3 programming language and open-source software library for machine learning Tensorflow 2 [12]. The training of CNN model was performed on a personal computer with AMD Ryzen 9 5900HS with 32 Gb RAM and Nvidia GeForce 3080 GPU.

Processing whole slide images was done using slideio library [13], shapely library [14] was used for geometric calculations during patch extraction.

After preliminary translation and saving the WSI images from PATH-DT-MSU dataset as numpy arrays, the inference time of applying the proposed model of automatic tissue type recognition with full visualization of the results takes about 1.5 minutes on the mentioned above PC configuration, which is good enough for practical use by pathologists.

6. Conclusion

In this paper we proposed a new CNN-based method of automatic tissue type recognition in whole slide histological images and a special pipeline of training CNN model within three phases using different histological datasets. The proposed model achieved accuracy value of 0.929 and balanced accuracy of 0.903 on CRC-VAL-HE-7K dataset and accuracy value of 0.98 and balanced accuracy of 0.926 on the test subset of patches from WSI images of PATH-DT-MSU dataset.

The developed method makes it possible to classify the areas corresponding to the gastric own mucous glands in the lamina propria and also to distinguish the tubular structures of a highly differentiated gastric adenocarcinoma with normal glands.

Further improvements of current work will be focused on the enlarging the number of whole slide images in PATH-DT-MSU dataset, adding more detailed annotations and increasing the number of supported types of tissues. The performance of the proposed method is also planed to be improved by pretraining the model with autoencoders on WSI images that are not annotated.

Acknowledgments

References

[1] L. Wu, J. Wang, X. He, Y. Zhu, X. Jiang, Y. Chen, Y. Wang, L. Huang, R. Shang, Z. Dong, et al., Deep learning system compared with expert endoscopists in predicting early gastric cancer and its invasion depth and differentiation status (with videos), Gastrointestinal Endoscopy (2021).

[2] Y. Chen, Z. Sun, W. Chen, C. Liu, R. Chai, J. Ding, W. Liu, X. Feng, J. Zhou, X. Shen, et al., The immune subtypes and landscape of gastric cancer and to predict based on the whole-slide images using deep learning, Frontiers in Immunology 12 (2021).

[3] X. Wang, Y. Chen, Y. Gao, H. Zhang, Z. Guan, Z. Dong, Y. Zheng, J. Jiang, H. Yang, L. Wang, et al., Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning, Nature communications 12 (2021) 1–13.

[4] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, et al., Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study, PLoS medicine 16 (2019) e1002730.

[5] J. N. Kather, N. Halama, A. Marx, 100,000 histological images of human colorectal cancer and healthy tissue, 2018. URL: https://doi.org/10.5281/zenodo.1214456. doi:10.5281/zenodo.1214456.

[6] A. Khvostikov, A. Krylov, I. Mikhailov, O. Kharlova, N. Oleynikova, P. Malkov, Automatic mucous glands segmentation in histological images., International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences (2019).

[7] N. Oleynikova, A. Khvostikov, A. Krylov, I. Mikhailov, O. Kharlova, N. Danilova, P. Malkov, N. Ageykina, E. Fedorov, Automatic glands segmentation in histological images obtained by endoscopic biopsy from various parts of the colon, Endoscopy 51 (2019) OP9.

[8] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, N. E. Thomas, A method for normalizing histology slides for quantitative analysis, in: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2009, pp. 1107–1110.

[9] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Pro-

ceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[11] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[12] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning, in: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), 2016, pp. 265–283.

[13] S. Melnikov, V. Popov, Slideio: a new python library for reading medical images, 2020–. URL: https://gitlab.com/bioslide/slideio.

[14] S. Gillies, et al., Shapely: manipulation and analysis of geometric objects, 2007–. URL: https://github.com/Toblerity/Shapely.