

AUTOMATIC BREAST CANCER GRADING IN LYMPH NODES USING A DEEP NEURAL NETWORK

T. Wollmann, K. Rohr

University of Heidelberg, BioQuant, IPMB, and DKFZ Heidelberg,
Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group,
Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

ABSTRACT

The progression of breast cancer can be quantified in lymph node whole-slide images (WSIs). We describe a novel method for effectively performing classification of whole-slide images and patient level breast cancer grading. Our method utilises a deep neural network. The method performs classification on small patches and uses model averaging for boosting. In the first step, region of interest patches are determined and cropped automatically by color thresholding and then classified by the deep neural network. The classification results are used to determine a slide level class and for further aggregation to predict a patient level grade. Fast processing speed of our method enables high throughput image analysis.

Index Terms— Deep Learning, Densely Connected Convolutional Networks, Microscopy, Machine Learning, Histology

1. INTRODUCTION

Breast cancer is the most mortal cancer disease for women, worldwide [1]. The progression of the disease is quantified by pathologists using whole-slide images (WSIs) of lymph nodes, which are stained with hematoxylin and eosin (H&E). For grading the progression of cancer diseases the TNM system is used [2]. In the TNM classification system, the parameter N describes the degree of cancer spread to regional lymph nodes. Currently, pathologists perform the pathologic N-stage grading manually, which is tedious, time-consuming, and error prone. Automation of this process could lead to the reduction of manual work and errors.

In the past years several challenges for whole-slide image analysis were conducted (e.g., [3, 4]). However, most methods use dense classification of the whole-slide images, which is slow and requires processing of many non-meaningful image regions [5, 6, 7]. Sparse classification methods speed up the classification process [8, 9]. Thus, such methods could run on the workstation of a pathologist and support the decision process. However, sparse classification generally reduces the performance which, however, can be alleviated by using model averaging.

In this work, we describe a method which uses a deep neural network (DNN) for sparse classification of WSIs. Our approach determines regions of interest (ROIs) within the WSIs using color thresholding and morphological operations followed by sparse classification using a deep neural network. The network is trained and tested using the same image resolution. Since the WSIs have varying downsampling factors, we employ a normalization.

2. METHOD

Our method uses a region of interest selection method and a deep neural network to perform sparse classification of the WSI. The individual classification results are aggregated to a slide wide class. Additional decision rules, provided by the CAMELYON17 team, are used for patient level classification.

Figure 1 illustrates the overall workflow of our method. The deep neural network is implemented in Tensorflow [10].

2.1. Region of Interest Selection

A region of interest (ROI) is determined by color thresholding similarly to the method in [8]. Colour thresholding is performed using a ratio of the intensities of 0.9 in the green and red channel. Afterwards, a median filter with a 50 pixels disk shaped structuring element is applied. Based on the whole slide with a downsampling factor of 64 as input, a region of interest map is calculated. Figure 2 shows examples of regions of interest determined by the region of interest selection method.

For tissue classification after ROI selection, 20 region centroids with a size of 512×512 pixels with a downsampling factor of 64 are sampled from the WSI using the region of interest map.

2.2. Tissue Classification

Tissue classification is performed on image regions determined by the region of interest selection mechanism. Computed centroid coordinates are used to extract 512×512 pixels image patches with a downsampling factor of 64. A sample

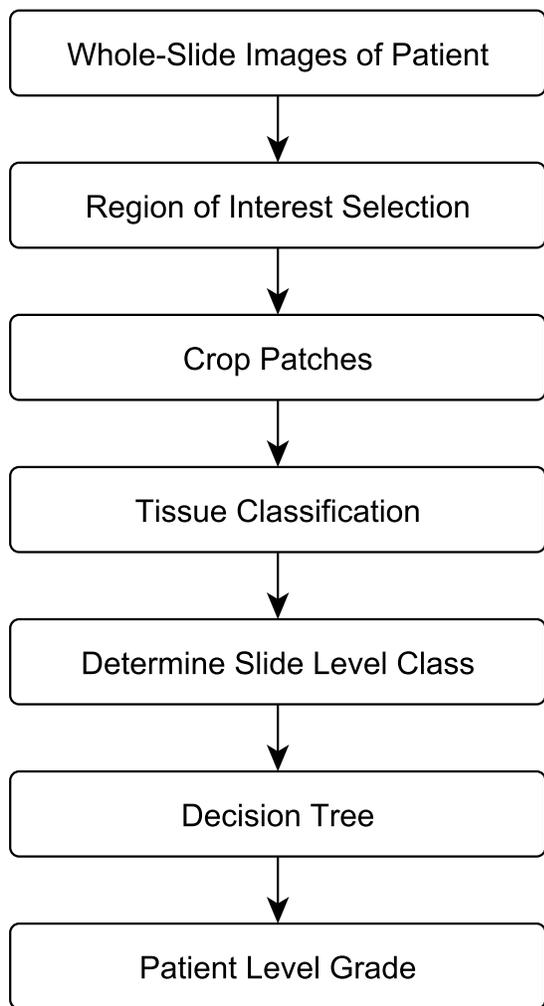


Fig. 1: Overall workflow of our method

patch is show in Figure 3. The patches are augmented using random rotations and flipping, and used for model averaging. All patches are classified into a one-hot encoding of the four classes: contains isolated tumour cells (ITC), contains a macro metastasis, contains a micro metastasis or is negative.

The deep neural network is inspired by densely connected deep neural networks [11]. In the first block a 3×3 convolution is performed yielding 32 feature maps. Since it was empirically observed that the first layers benefit from negative parts of activation functions we replaced the original ReLU with PReLU [12, 13]. For downsampling and feature disentangling we alternate densely connected blocks with downsampling blocks as described in [11]. Within the densely connected blocks we use six densely connected feature map extractors. Each downsampling block doubles the number of feature maps. The predicted class is calculated using global average pooling [14], a fully connected layer with 128 hid-

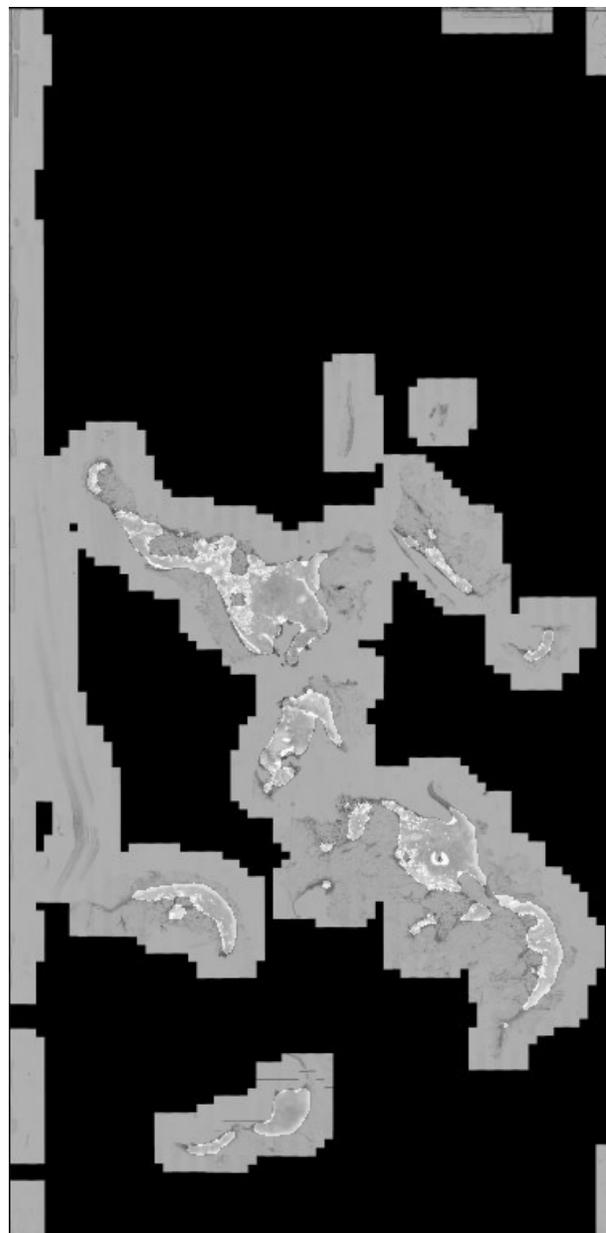


Fig. 2: Example demonstrating the region of interest selection mechanism

den units, and the softmax operator. All non-linear layers are initialised using MSRA initialisation and we used ReLU activation functions [12].

2.3. Patient Level Grading

A slide level class is calculated by adding the activations of the class activations of the last layer and calculating the maximum over all 20 patches. By selecting the maximum after adding up the activations of multiple predictions we perform model averaging. Instead of using the maximum, we also

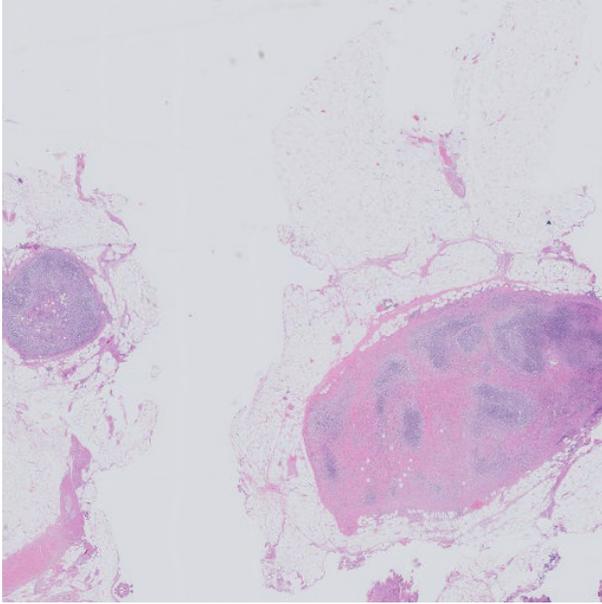


Fig. 3: Example 512×512 pixels image patch

tried to rank the classes by macro metastasis, micro metastasis, ITC, and negative, and picked the highest ranked class. However, this technique did not improve the results on the validation set.

After determining slide wide classes a patient level grade is determined using the following decision rules:

- **pN0:** No micro-metastases or macro-metastases or ITCs found.
- **pN0(i+):** Only ITCs found.
- **pN1mi:** Micro-metastases found, but no macro-metastases found.
- **pN1:** Metastases found in 1-3 lymph nodes, of which at least one is a macro-metastasis.
- **pN2:** Metastases found in 4-9 lymph nodes, of which at least one is a macro-metastasis.

2.4. Model Training

The deep neural network model is trained using the cross entropy loss. We trained the network for six epoch within eight hours on 80% of the data and kept 20% for validation of the model. In each epoch the training data class occurrence frequencies are balanced and the image patches are augmented.

The tissue classification network is trained using mini batches with 10 image patches each. The 512×512 pixels image patches are extracted from the original WSIs by a downsampling factor of 64 using our color thresholding method, augmented using random rotation, flipping, random color shifts, and elastic deformation, and passed to the deep

neural network. Data augmentation and transfer to the GPU node is optimised using multi threaded data streaming jobs running on the CPU.

The models are trained using the Adam optimizer [15] with an initial learning rate $l_{init} = 0.001$, as well as $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

2.5. Experimental

We measured the speed of our algorithm on the challenge dataset. The classification of a 512×512 pixels patch requires on average 0.04 seconds on an Intel i7-6700K workstation with a NVIDIA Geforce GTX 1070. We measured a total computation time of 0.78 seconds for a WSI and 3.90 seconds for predicting a patient level grade.

3. DISCUSSION AND CONCLUSION

We presented an automatic approach for patient level breast cancer grading for lymph node WSIs. Our approach utilises a region of interest selection mechanism and a densely connected deep neural network to perform sparse classification. The sparse classification results are aggregated using decision rules. The algorithm determines a patient level grade based on five WSIs in about 4 seconds on a standard workstation by using a region of interest selection mechanism and classifying big image crops. A trade-off between statistical power and speed can be achieved natively by changing the number of patches used for the sparse classification. Our work contributes to the field of breast cancer grading by providing a fast method, which can be trained using only slide level annotations.

Acknowledgements

This work was supported by the BMBF-funded Heidelberg Center for Human Bioinformatics (HD-HuB) within the German Network for Bioinformatics Infrastructure (de.NBI) #031A537C.

4. REFERENCES

- [1] A. McGuire, J. A. Brown, C. Malone, R. McLaughlin, and M. J. Kerin, "Effects of age on the detection and management of breast cancer," *Cancers*, vol. 7, no. 2, pp. 908–929, 2015.
- [2] S. B. Edge and C. C. Compton, "The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM," *Annals of Surgical Oncology*, vol. 17, no. 6, pp. 1471–1474, 2010.
- [3] "Tupac16 contest home page." <http://tupac.tue-image.nl/>. Accessed: 2016-11-03.
- [4] "Camelyon16 contest home page." <https://camelyon16.grand-challenge.org>. Accessed: 2016-11-03.

- [5] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, 2016.
- [6] R. Chen, Y. Jing, and H. Jackson, "Identifying metastases in sentinel lymph nodes with deep convolutional neural networks," *arXiv preprint arXiv:1608.01658*, 2016.
- [7] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, *et al.*, "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442*, 2017.
- [8] T. Wollmann and K. Rohr, "Automatic grading of breast cancer whole-slide histopathology images," in *Proc. Bildverarbeitung für die Medizin 2017*, p. 249, Springer-Verlag, 2017.
- [9] T. Wollmann and K. Rohr, "Deep residual Hough voting for cell detection in histopathology images," in *Proc. IEEE ISBI 2017*, 2017.
- [10] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.
- [11] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993*, 2016.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE ICCV 2015*, pp. 1026–1034, 2015.
- [13] 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.
- [14] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.
- [15] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.