

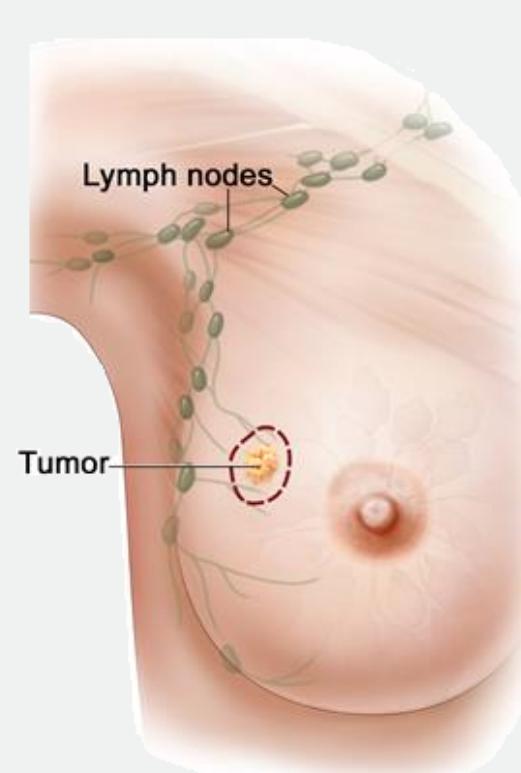
Efficient nuclei detection and classification with deep learning

Peter Grønborg, Mathematical Modelling and Computing, Technical University of Denmark

We use deep learning to automatically detect and classify breast cancer metastases, providing crucial information for the prognosis and treatment decision of breast cancer. Our method qualified for IEEE International Symposium on Biomedical Imaging 2017 Grand Challenge CAMELYON17 Melbourne, Australia. We were ranked 5th out of 23 qualifying teams of international research groups and commercial teams, with a marginal score difference to the winner of the CAMELYON17 competition.

Introduction

Breast cancer is the most common cancer disease for women worldwide (2012), where metastatic involvement of lymph nodes is one of the most important prognostic factors. Therefore, it is crucial to quickly determine if and which lymph node(s), the tumor has spread to.



Pathologists usually examine metastases in the lymph nodes by measuring the size of the tumor and/or counting the number of tumor cells. Thus, they manually categorize the type of metastases into macro metastases, micro metastases or isolated tumor cells (ITC), see Table 1. Assessment of multiple lymph nodes are then combined into a pN-stage for a single patient, see Table 2. However, this procedure is very labor intense for the examining pathologist and most importantly, small metastases are very difficult to detect and sometimes they are missed.

Figure 1: Breast cancer and lymph node localization. It is most likely to be the lymph node(s) located in the axilla which the cancer spreads to first, when the patient has breast cancer.

Methods

Preprocessing Tissue detection at low-resolution using algorithm developed in Visiopharm Integrator System (VIS) (Visiopharm A/S, Denmark)

Sampling Uniform random sampling from both classes followed by hard negative mining.

Data augmentation Spatial transformations (flip, rotations) combined with H&E-stain specific augmentation scheme, creating more stain variation than available in the data. See below.

Implementation details We implement and train a convolutional neural network based on the Inception V3 architecture [1] to discriminate between metastatic tissue and non-metastatic tissue.

WSI inference We use a sliding window analysis where each non-overlapping patch is assigned a tumor probability $p \in [0, 1]$ creating a 2D-tumor probability distribution, i.e. tumor heatmap.

Post-processing We threshold the tumor probability distribution $p > 0.5$ and find the 5 largest metastases M based on area.

Classification We train a random forest-classifier on features extracted from M to discriminate between macro, micro, ITC or normal WSIs.

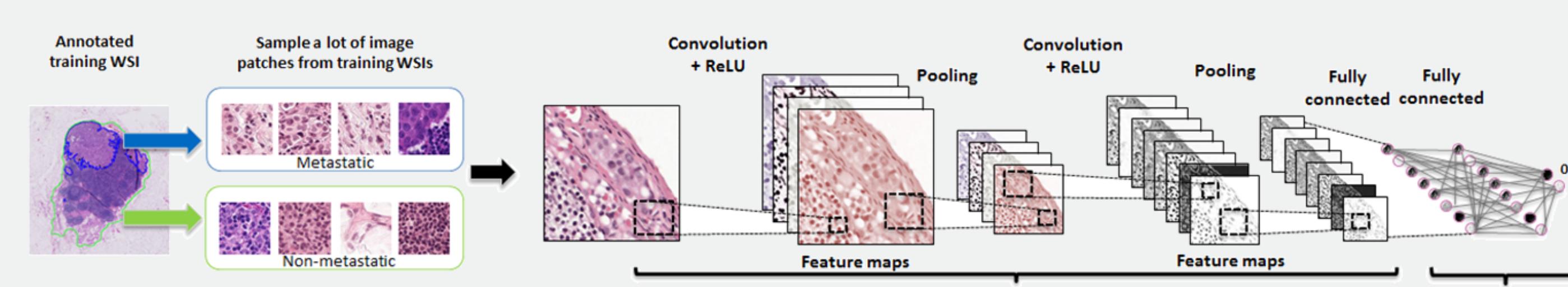
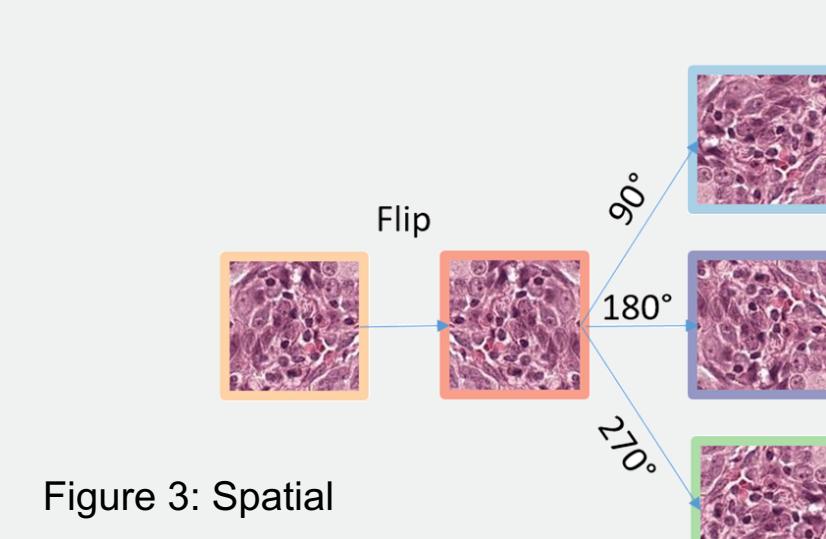
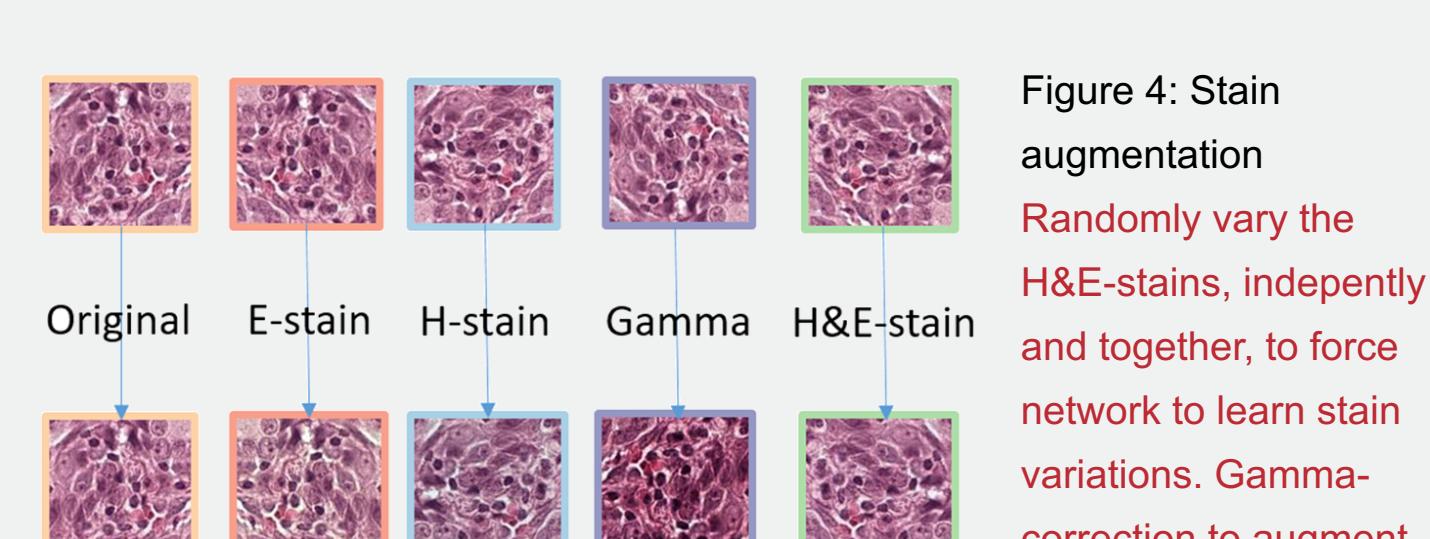


Figure 2: Training Framework We sample millions of patches (128x128 pixels) from whole slide images (WSI) to train a convolutional neural network to distinguish between metastatic and normal tissue, i.e. a binary classification problem. We use a slightly modified version of the Inception V3 [1] (the architecture shown are only for illustration purposes).



Macro: metastases greater than 2.0 mm.
Micro: metastases greater than 0.2 mm or more than 200 cells, but smaller than 2.0 mm.
ITC: Single tumor cells or a cluster of tumor cells smaller than 0.2 mm or less than 200 cells

Table 1 Metastases subtypes and WSI classification.



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.

Table 2 pN-stage definitions

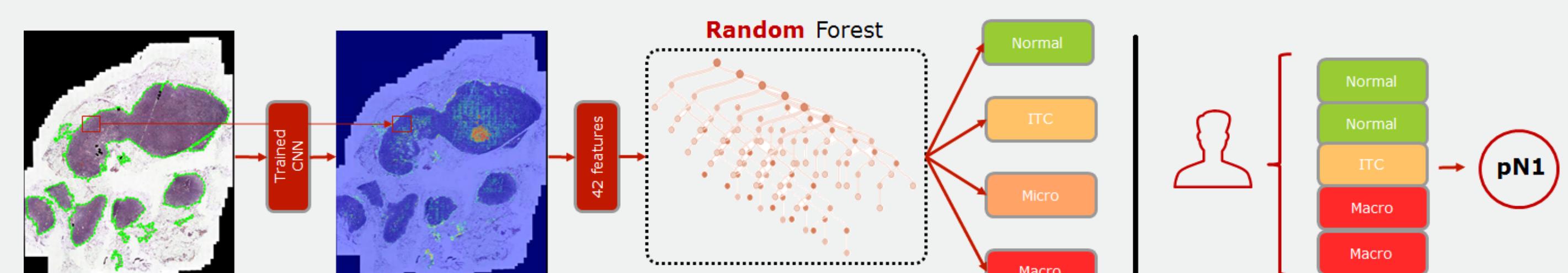


Figure 5: Whole-slide image inference and classification. Inference of histology gigapixel images is followed by post-processing and feature extraction for WSI classification.

Figure 6: Patient classification 5 WSI analysis determines the pN-stage

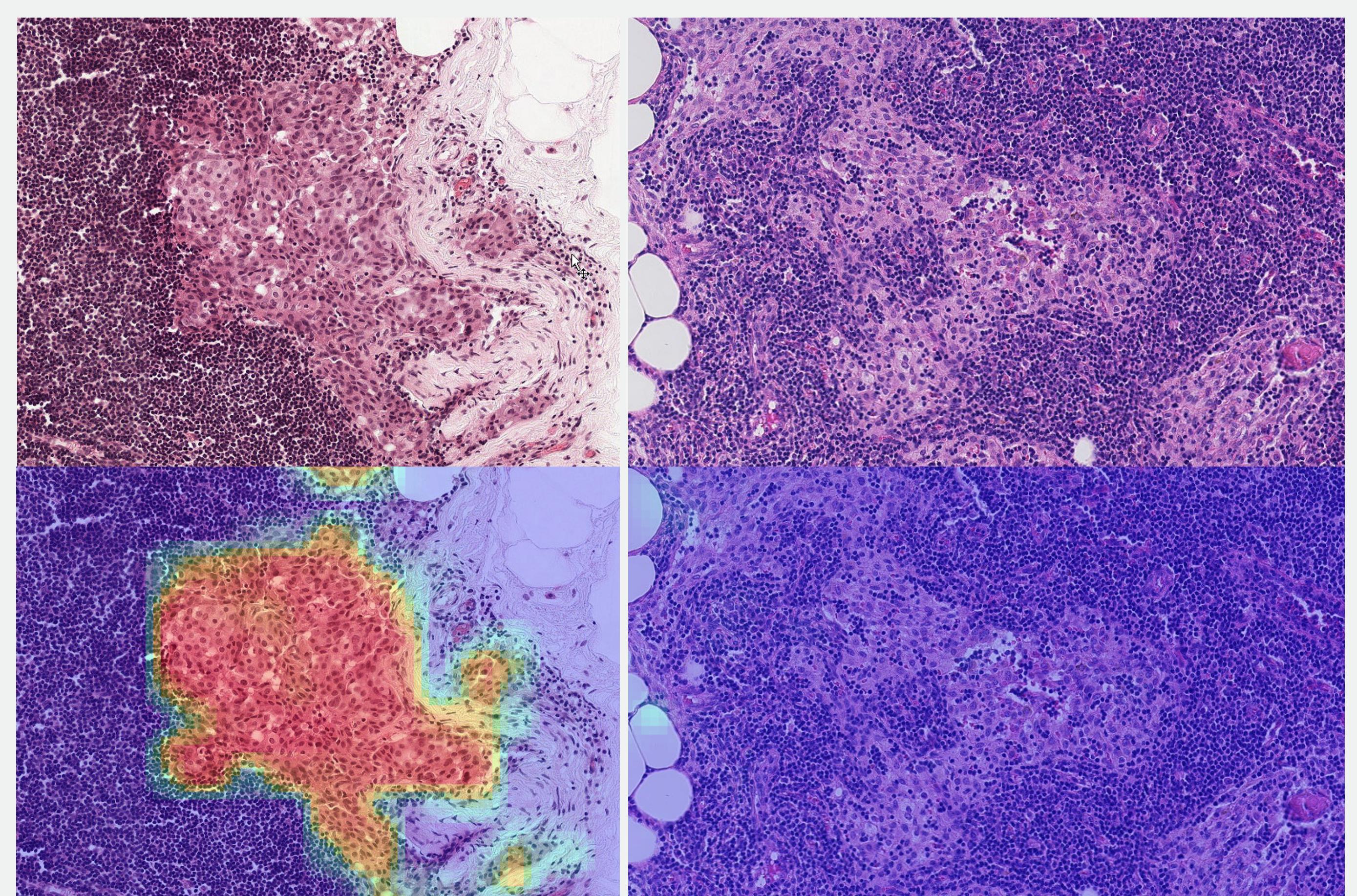


Figure 7: True positive tumor region Top: Shows metastatic tissue with cancer cells surrounded by lymphatic tissue. Bottom: Probability heatmap with high probability of tumor (red).

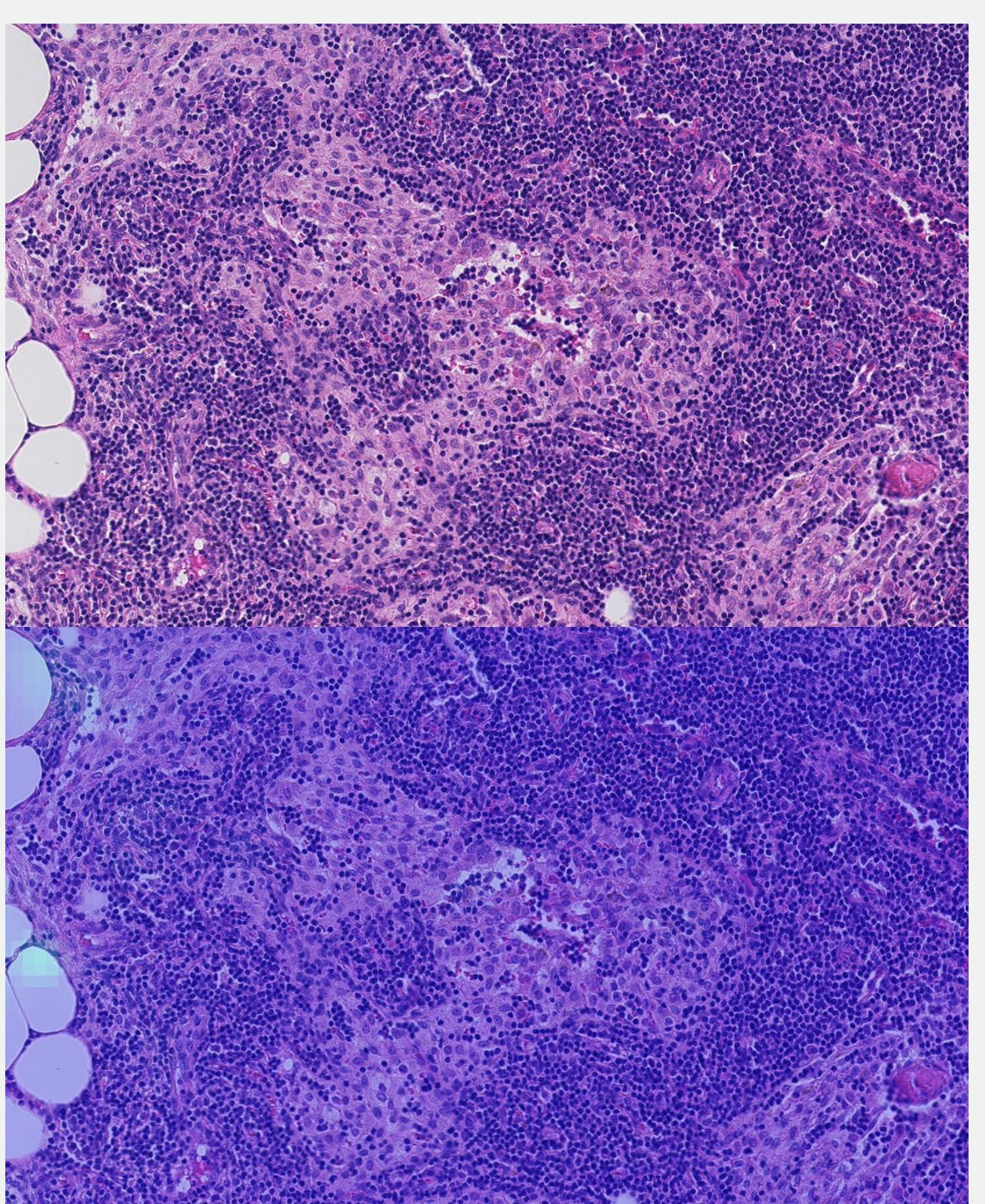


Figure 8: True negative normal region Top: Shows normal tissue (macrophages etc.) surrounded by lymphatic tissue. Bottom: Probability heatmap with low probability of tumor (blue). Illustrate difficult regions that looks similar to metastatic regions.

Results

Our method reaches a weighted kappa value of 0.8172 on the challenge test set (100 patients), showing very good agreement between our algorithm and the human pathologist.

Conclusions and future work

We demonstrate a deep learning-based approach to analyze and diagnose breast cancer metastases in whole-slide images of histological lymph node sections. Our method could also be implemented as a screening-tool that significantly decrease the clinical workload. Future work will focus on optimizing stain-augmentation and increasing sensitivity towards ITCs.

References

- [1] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818-2826)..
- [2] <https://camelyon17.grand-challenge.org/results>