# MULTI STAGE CONVOLUTION NEURAL NETWORK FOR BREAST CANCER STAGE CLASSIFICATION IN HISTOPATHOLOGY IMAGES

*Xiaoyu Shi*

## ABSTRACT

The TNM stage evaluation system is well accepted in classifying the spreading degree of cancers in patient with a solid tumor. TNM staging takes into account three different types of stages to make the assessment of breast cancer, the size of the tumor (T-stage), the spread of the regional lymph nodes (N-stage), and the metastasis (M-stage). The pathologic N-stage (pN-stage) is commonly used by clinical pathologist. However, this staging evaluation method requires a diagnosis through the scanned histopathological image in high resolution, which would be time-consuming and leaves heavy manual work. In this paper, we present a multi-stage framework with two convolutional neural networks to predict the pN-stage category. The performance is evaluated on Camelyon 17 Challenges [1].

***Index Terms***— Concolutional neural network, Deep Learning, hierarchical clustering, Metastasis detection, Camelyon 17.

## 1. INTRODUCTION

The TNM stage system [2] is internationally used to classify the extent of cancer spread and a crucial reference value in clinical treatment. The decision of whether the breast cancer has spread to the regional lymph nodes (N-stage) requires an accurate detection of metastasis of the cancer in breast lymph nodes. However, the detection of the metastasis relies on careful diagnosis in histopathological image slide in huge resolution, which brings heavy manual work and enormous time consumption. Automated metastasis detection and histopathological slide classification shows great value in reducing workload of pathologist.

In the past few years, convolutional neural network (CNN) has shown significant improvement in computer vision and medical image understanding tasks. CNN based framework has been proposed in different pathological image analysis work [3, 4]. [3] presents a unsupervised framework to analysis the component of basal cell carcinoma. [4] demonstrates the effectiveness of CNN based architecture in detecting lymph node metastasis on Camelyon 16 dataset.

In this paper, we propose an automated framework to predict the pN-stage using the whole slide image (WSI). The proposed framework consists of two image processing module, a CNN based metastasis detection module and a CNN based slide classification module. Specifically, a ROI extraction module is proposed to extract the valid tissue regions from WSI, after which a CNN is trained to locate the metastasis within the ROI. A hierarchical clustering is used on the heatmap generated by the CNN to filter out the noise responses on the heatmap. Finally, a small CNN is designed to classify the pN-stage categories.

## 2. METHODOLOGY

Figure 1 introduces the overall framework of our system. The details are described in this section respectively. The Whole Slide Image (WSI) usually consist of 8 different resolution level and the highest resolution contains approximately 200,000 x 100,000 pixels. The tissue only takes a certain area of the slide and leaves a large proportion of the slide as background or invalid area. Selecting the entire slide as input for CNN could be computational wasted and would generate a large number of useless responses during the computation within the network. As a result, we employ a image patch classification to detect the metastasis area in slide and a heatmap classification to determine the pN-stage categories on each slide.

Section 2.1 explains the image preprocess procedure, especially the ROI extraction module, which erase the slide background and improve the efficiency of sampling in patch classification. Section 2.2 introduces the CNN based patch classification and probability heatmap generation. Section 2.3 analyses the noises in heatmap and introduces a hierarchical clustering framework based on DBSCAN [5] to filter out some of the false responses. Section 2.4 represents a small neural network that we designed to bring out pN-stage classification in slide level.

### 2.1. ROI extraction

The hematoxylin and eosin (H&E) stain is the most widely used staining method in histological slide preparation, which gives sharp blue/pick contrast across various cell structures. Comparing to RGB colour space, the HSV space better described color contrast and the saturation difference among various structures. Otsu [6] threshold binaryzation is employed on the saturation channel to extract the tissue area.

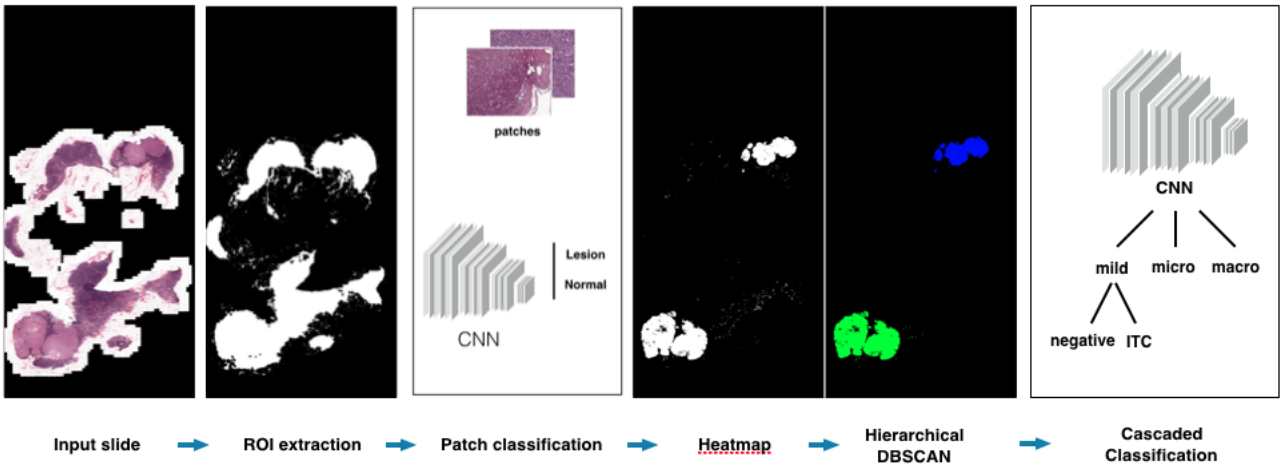


**Fig. 1**. The overall framework procedure

We also found that a certain gray [7] threshold binaryzation on RGB channel could be efficient to filter our some blank or black background and improve the entire ROI extraction procedure.The extraction is employed on the level 3 resolution of WSI.

### 2.2. metastasis detection

The annotated metastasis pixel mask usually include some fat tissue or some other non-metastasis area or a certain number of lymph nodes. It's extremely difficult to annotate the pixel mask with 100 percent accuracy. As a result, we build a large scale dataset using image patches extracted from WSI. We labeled the patch with positive label (tumor) if it has more than 50% pixels being annotated as tumor and labeled the patch with negative label (normal) if it has no overlap with the annotated tumor area. The patches are extracted from the ROI area we generated as described in section 2.1.

GoogleNet is employed as the basic CNN architecture for tumor and non-tumor binary classification. To deal with the imbalance among the number of tumor, non-tumor and normal patches(the non-tumor patch is extracted from normal area within the slide with tumor area, normal patch is extracted from normal slide, both of these patches are regarded as negative sample), we sample the normal and non-tumor patches with uniform distribution and make the ratio of positive/negative patches number as 1:1.5. Data augment, such as rotating the patches and random mirror flipping are also involved. Hard negative mining is employed to improve the performance of the network.

### 2.3. Hierarchical Clustering

Before conducting the slide classification, we involve a hierarchical DBSCAN clustering to clean the scattered responses at the edge of the heatmap cluster. The cleaning procedure improves the feature of the heatmap responses, making the detected area much compact. We also find that most of ITC slides are with scattered responses, which makes it extremely difficult to be distinguished from the negative slides with tiny false responses. The clustering erase the difference between negative and itc at the first place, which enhance the classification among negative, micro and macro categories. A cascaded classification architecture will be employed as the second classification level to discriminate negative and itc.

### 2.4. slide classification

Instead of using manually designed topology features, we create a small CNN with 5 layers to extract the feature from slide heatmap and build the classification. The architecture of the network is shown in Table 1. We pick 672x672 as the input image size, since resizing heatmap into smaller size would sacrifice some image content and key heatmap area. Max pooling, which is popular in most of the network architecture, is replaced by average pooling to keep the area information. Rotation of the heatmap from 0 to 360 degree with 10 degree as interval is conducted as data augmentation.

The network is first trained as a three categories classifier on the heatmap images processed by hierarchical clustering to discriminate Macro, Micro and mild (including itc and negative). A second network with the same architecture is trained as fine grained binary classifier to classify negative and itc.

## 3. EXPERIMENT

### 3.1. Dataset

The Camelyon 16 [8] dataset contains 400 WSI with pixel level annotation masks. The Camelyon 17 dataset contains 100 WSI for training and testing respectively, in which 50 slides are with pixel annotations. Since the number of posi-

| type | kernel-stride-channel | |
|---|---|---|
| conv1 | 3x3-c4 | |
| relu | | |
| pool1 | 3x3-s3 | average |
| conv2 | 3x3-c8 | |
| relu | | |
| pool2 | 3x3-s3 | average |
| conv3 | 3x3-c8 | |
| relu | | |
| pool3 | 3x3-s3 | average |
| conv4 | 3x3-c16 | |
| relu | | |
| pool4 | 3x3-s3 | |
| softmax | | |

**Table 1**. Network architecture for slide classification

tive patches varies a lot among each WSI, we extract patches from all the WSIs with mask annotation first and randomly pick 80% of the patches as training set, leaving 20% of the patches as validation. We keep the same ratio of the number of patches from negative, ITC, Micro, Macro in train and test, picking patches from 16 and 17 dataset respectively. We also make sure that the slide we extracting patches in training set is different from the slide we extracting validation set.

For slide classification, we use all the 500 slides from Camelyon 17 to generate 3 fold cross validation. During each validation, we keep the ratio of itc, negative, macro, micro as the same as the ratio they live in the 500 slides.

### 3.2. Experiment Setup

Caffe [9] framework and GoogLeNet [10] architecture are used to train our patch classification model. A self designed network is created for slide classification. We use 256x256 and 672x672 as input size for these two classification respectively. SGD with step decay is chosen as learning strategy. The GoogLeNet is initialized with model pre-trained on ImageNet dataset. The network for slide classification is trained from scratch.

### 3.3. Result

We use the average classification accuracy to evaluate the performance of patch classification and slide classification, as shown in Table 2.

| Stage | Accuracy |
|---|---|
| Patch Classification | 0.982 |
| Slide Classification | 0.872 |

**Table 2**. Classification Accuracy

## 4. CONCLUSION

In this paper, we propose a multi-stage framework to predict pN-stage in histopathology image automatically. CNN based metastasis detection and slide classification are also created. From our observation, the accuracy of the patch classification limited by the performance on the discrimination between ITC and negative cases. A single false positive response in negative cases could be recognized as ITC.

In future work, we would further improve the accuracy of patch classification, and build an end-to-end learning architecture for the prediction.

## 5. REFERENCES


[1] "Camelyon 2017," https://camelyon17. grand-challenge.org/, Accessed: 2017-07-24.

[2] http://www.uicc.org/resources/tnm

[3] Arevalo J, Cruz-Roa A, Arias V, et al. An unsupervised feature learning framework for basal cell carcinoma image analysis[J]. Artificial Intelligence in Medicine, 2015, 64(2):131-145.

[4] Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q Nelson, Greg S Corrado, et al., "Detecting cancer metastases on gigapixel pathology images," arXiv preprint arXiv:1703.02442, 2017.

[5] Martin Ester, Hans-Peter Kriegel, Jorg Sander, and Xiaowei Xu, "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996, KDD'96, pp. 226–231, AAAI Press.

[6] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[7] Andrew Janowczyk and Anant Madabhushi, "Deep learning for digital pathology image analysis: A com- prehensive tutorial with selected use cases," Journal of pathology informatics, vol. 7, 2016.

[8] "Camelyon 2016," https://camelyon16. grand-challenge.org/

[9] Jia, Yangqing, Shelhamer, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J]. 2014:675-678.

[10] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1-9.