FAST CLASSIFICATION OF WHOLE SLIDE HISTOPATHOLOGY IMAGES FOR BREAST CANCER DETECTION

Larisa Chervony and Simon Polak

Mechanomind

ABSTRACT

In this paper we present a framework for whole slide histopathology image analysis using modern computer vision and machine learning methods. Our approach is based on a deep learning architecture and our aim is to show that a viable and robust solution can be achieved even with limited computational resources. We use two convolutional neural networks, applied in a cascade, followed by local maxima extraction and SVM classification of local maxima regions. The resulting system's accuracy is comparable to the best results achieved in Camelyon17 challenge, while maintaining low computational cost.

Index Terms— Histopathology, convolutional neural networks, Camelyon17, breast cancer detection

1. INTRODUCTION

Analysis of histopathological images is one of the main methods of cancer diagnosis. It requires skillful pathologists with many years of training, and is a very time consuming and tedious procedure, and therefore prone to human error. Computer aided diagnostics is the new frontier of modern medicine. While there is a lot of research in the automated analysis of the radiology images, the field of Digital Pathology and the analysis of whole slide histopathology images is relatively new and poses unique challenges.

Last decade have brought huge advancements in the Computer Vision and Machine Learning fields, mainly due to the introduction of the convolutional neural networks (CNNs) [1], which showed great results in such areas as image classification [1] and object localization [2]. However, application of CNNs to the task of whole slide images (WSI) analysis is not straightforward due to a number of factors. First of all a typical WSI is huge, compared to a typical natural image - it can be 200,000 x 100,000 pixels in size. Second, the tumor classification and localization problem is highly unbalanced - the area of tumor cells is usually very small, compared to the area of the normal tissue. Third, there are significant color variations between slides due to image acquisition equipment.

In this paper we will present a method for fast and robust analysis of whole slide histopathology images. Our method is designed to overcome the problems described above and consists of 3 stages. The first stage (section 2.2) is the removal of the non-tissue regions. The second stage is the patch-level classification (section 2.3) done on two magnification levels in a cascaded manner. The third stage is the slide-level classification (section 2.3), followed by the pN-stage rule-based decision.

2. METHODOLOGY

When designing an architecture for WSI analysis, a number of issues play a major role. One is the size of the WSI - in order to achieve a feasible, in terms of computational costs, solution we have to reduce the number of patches considered for classification. Another issue, is the fact that decision for a specific WSI location have to be based not only on the data at this location, but also rely on its surroundings. Third is the unbalanced nature of the task - the number of tumor regions is small, compared to overall tissue regions.

We address the size problem in a number of ways. First is the preprocessing step, which removes the unnecessary computations in the areas of WSI not containing the tissue. Second is the cascaded patch-level classification approach, similar to the one used in [5], where we first apply the classifier on a lower resolution (10x) and only suspicious areas (above some threshold) are passed to the next cascade level, where these areas are classified at a higher resolution (20x). This approach is also very helpful for dealing with inherently unbalanced problems, as it was shown in [5].

The issue of the increased field of view is addressed by the design of the patch-level classifier. Our neural network architecture is inspired by the GoogLeNet architecture [6] with a number of changes (see Figure 1). Specifically, the input is a 416x416 image, which is divided to a grid of 13x13 cells (each cell 32x32 pixels) and each cell receives a classification score. The first part

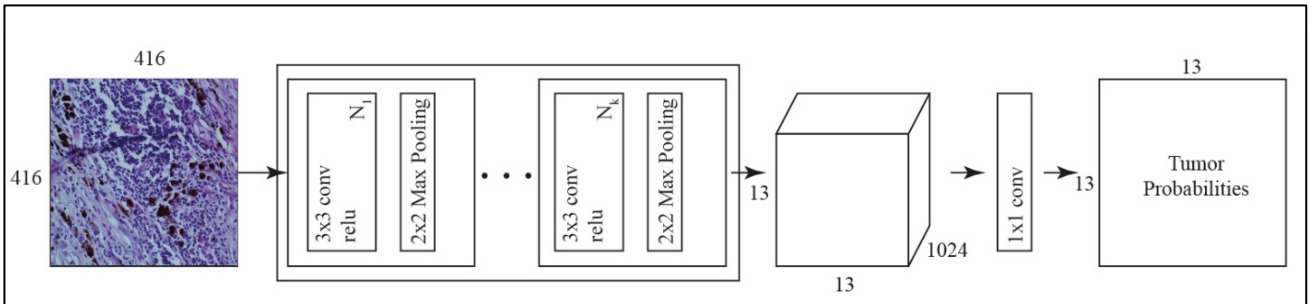

Figure 1: Patch-level convolutional neural network classifier architecture. The architecture is inspired by the GoogLeNet network. The input image of size 416x416 is passed through a series of convolutional, relu and max pooling layers (the feature extraction part) and the final layer performs classification of each of the 13x13 grid cells, so that each 32x32 pixels square in the original image receives a separate classification score.

of the network is a series of convolutional, relu and max pooling layers (as in the original network), where the last convolutional layer

outputs 13x13x1024 feature map. The last part is the classification layer, where for each cell in the 13x13 grid a fully connected classifier is applied, which produces a 13x13 heat map of classification results. This architecture has a number of advantages - wide field of view, due to the large support area for features in each grid cell, each location in the image is classified and the classification step is small enough in order not to miss ITC areas.

Patch-level classification produces a WSI tumor probabilities map. To perform slide-level classification, we extract patches of tumor probabilities around the highest N local maxima of the WSI probability map and apply one-vs-all support vector machine classification on the histograms of these patches.

Classification is followed by a simple decision rule - if "macro" patch was detected, the slide is assigned the "macro" label, otherwise, if "micro" patch was detected, the slide is assigned the "micro" label, otherwise, if "itc" was detected, the slide is assigned the "itc" label, and otherwise the slide is assigned the "normal" label. The block-diagram of the classification process is depicted in Figure 2.

2.1 Data Set

This work is based on the data provided by the Camelyon17 challenge [3]. The data consists of fully annotated 270 images from the Camelyon16 challenge [4] and 500 images (5 slides per patient) with slide-level annotations to one of 4 classes - normal, macro, micro and itc (a small portion of these slides were also fully annotated). In addition another 500 images were set aside to serve as a test set.

In our work we have divided the second (pN-stage annotated) set to training set (60 patients or 300 slides) and validation set (40 patients or 200 slides).

2.2. Image Preprocessing

The goal of the preprocessing step is to reduce unnecessary computation by removing non-tissue slide regions. We have used a very simple procedure based on thresholding image's grayscale values to remove completely black or completely white patches. Any patch where 90% of its pixels were below (black) or above (white) the threshold was removed. After this simple procedure an average of 8000 patches containing tissue are left. These patches are used for training the patch-level classifier or passed to patch-level classification during testing.

2.3 Training

The proposed WSI classification system consists of three classifiers - two patch-level CNN classifiers (working at different magnifications) organized in a classification cascade and one slide-level SVM classifier.

During training of the first cascade level classifier we randomly select slides and extract tissue patches at magnification 10x from the selected slide. We have used large batches of 256 images and kept constant ration of ¼ between tumor and normal patches in each batch. Our choice of the loss function was focal loss [7] which is particularly suitable for unbalanced problems.
In order to account for color and cell size variations in slides we add noise to the extracted patches' hue, saturation and brightness and rescale the patch by a random factor varying between 0.9 and 1.1.

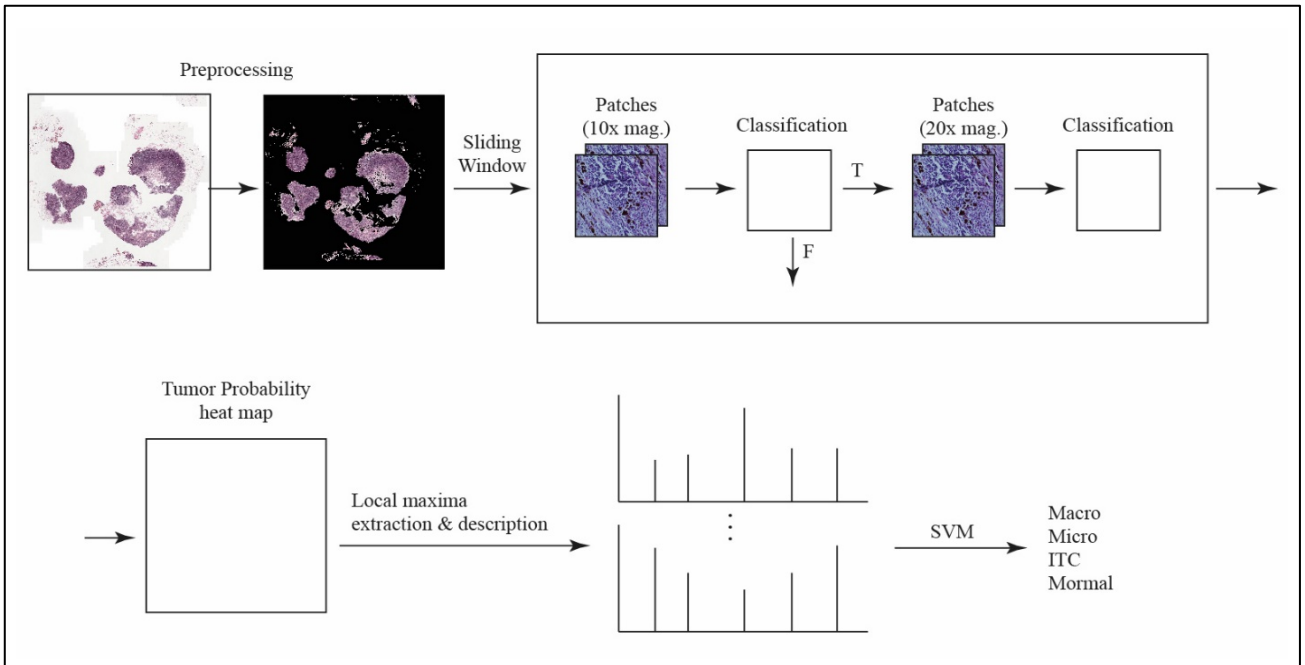

Figure 2: WSI classification system overview. An input WSI is passed through a preprocessing step, where tissue areas are extracted. Next we extract patches from tissue regions at magnification 10x, which are classified by first cascade level classifier. Patches having score above the threshold of 0.14 are passed (at magnification level of 20x) to the second cascade level classifier. In the resulting tumor probabilities heat-map N highest local maxima are extracted and a descriptor for each extracted local maxima is calculated and classified by an SVM classifier. Later on a simple decision rule is used to deduce slide-level classification.

In order to train the second cascade level classifier, we have set the threshold so that 95% of the tumor patches will be correctly classified. Next we randomly select patches at magnification 20x with classification score above the calculated threshold and train the classifier. Training of the slide-level classifier was done on the subset of the training data where only slide-level labels were available. For each

such slide we apply the classifier cascade and select N (=100 in our final setup) highest local maxima in the resulting tumor probabilities map. For each selected local maxima we have extracted a square area of the tumor probability map around it and transformed it to L1 normalized histogram of 50 uniformly distributed bins between 0 and 1. Next we have trained a one-vs-all SVM classifier, where the label of each histogram was given based on the label of the slide.

2.4 Classification

When classifying a WSI the image was passed through the preprocessing step and patches of size 416x416 with step 140 (which is approximately ⅓ of the patch size) were extracted for classification. This resulted in a mean of 8K patches per slide at 10x magnification for the first level classifier. Patches with at least 5 cells having probabilities above the threshold of 0.14 were passed to the second level classifier (a mean of 2.5K patches per slide). Later on the global probabilities map was assembled and histograms of probabilities surrounding local maxima were passed to

SVM classifier, followed by slide-level decision rule. Mean classification time of a slide on a single NVIDIA TITAN X GPU was approximately 200s.

3. Results

We have evaluated the proposed WSI analysis system on the validation set of 200 slides selected from the 500 training examples of Camelyon17 data. Our slide-level classification model and patient-level pN-stage prediction achieved 0.92 accuracy and 0.88 quadratic weighted kappa score, respectively.

4. Conclusions

We have presented a whole slide histopathology image analysis framework. The presented approach is able to perform highly accurate classification of breast cancer metastases in 3 minutes per slide. Future work can reduce the computation cost even more by exploring lighter

network architectures and deeper classification cascade, starting from lower magnification WSI levels. Inn addition, accuracy of the system can be improved by applying more appropriate for histopathology images color scheme normalization techniques, as was done by some of the participants in the Camelyon17 challenge [3]

5. REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[2] R. B. Girshick. Fast R-CNN. In ICCV, pages 1440–1448, 2015

[3] Oscar Geessink, Peter B ´ andi, Geert Litjens, and Jeroen ´ van der Laak, "Camelyon17: Grand challenge on cancer metastasis detection and classification in lymph nodes," 2017.

[4] Babak Ehteshami Bejnordi and Jeroen van der Laak, "Camelyon16: Grand challenge on cancer metastasis detection in lymph nodes," 2016.

[5] P.A. Viola, M.J. Jones, Rapid object detection using a boosted cascade of simple features, in: CVPR, issue 1, 2001, pp. 511–518.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. CoRR, abs/1409.4842, 2014

[7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. ´ Focal loss for dense object detection. arXiv preprint arXiv:1708.02002, 2017.