
Melanocyte Detection in Skin Whole-Slide Histopathological Images

Meredith Wu

Department of Biomedical and Health Informatics
University of Washington
Seattle, WA, 98195
wenjunw@uw.edu

Yiran Zhao

Department of Biomedical and Health Informatics
University of Washington
Seattle, WA 98195
yzhao362@uw.edu

Abstract

Skin cancer is one of the most malignant types of cancer and melanoma is the most aggressive type among skin cancers. In melanoma diagnosis, analysis of images in cellular is important and melanocytic nuclei is the most important cellular-level indicators of skin melanoma. The segmentation of melanocytes could lead to more accurate diagnosis of melanoma. A previous study developed an unsupervised method to segment melanocytes by estimating the area of the "halo" region around a nuclei and using a area threshold to determine if the given nuclei is a melanocyte. However, according to our pathologist in contact, the assumption of the existence of the "halo" region around a melanocyte may not be valid. In this project, we aimed to develop a more accurate supervised classifier for detecting melanocytes from histopathological whole slide images (WSI). We used a nuclei detector based on AdaBoost, which have been tested on histopathological WSI dataset with successful performance. We used this classifier to extract nuclei from WSIs. Then, we obtained nuclei patches which has a nuclei's centroid at the center of the patch. Finally, we trained a convolutional neural network (CNN) and a Dense CNN to classify whether a nuclei in the center of a patch is a melanocyte or not. We examined the effect of learning rate on model performance and compared the performance between the proposed CNN and Dense CNN. We were able to achieve classification accuracy of 75% and F1 score of 0.72, which proved the feasibility of our approach.

1 Introduction

The incidence of melanoma is rising faster than any other cancer, and about 1 in 50 U.S. adults will be diagnosed with melanoma this year alone. The gold standard for diagnosis of melanoma is the histopathological examination which the specimen is examined by pathologist manually [Cireřan et al., 2013]. However, the whole slide image (WSI) of one tissue sample usually has the size of 2.2×10^9 pixels, which imposes large difficulty for the pathologist to analyze the image completely. Thus, manual diagnosis only based on a small portion of the image is often subjective and prone to variability and automated diagnosis based on the entire slide is more desired [Lu and Mandal, 2012].

Melanocytic nuclei is the most important cellular-level indicators of skin melanoma [Lu et al., 2013]. If we can segment melanocytes accurately, cellular features and architectural features developed from

melanocytes can later be used for diagnosis of the melanocytic lesions. As the only former study worked on this issue, Lu et al. derived an algorithm based on the observation that melanocytes were surrounded by a pink "halo" region [Lu et al., 2013]. They used gradient magnitude and radial vectors to estimate the area of the halo region of given nuclei. An area ratio between the halo region and the nuclei was used to separate the melanocytes. However, based on our pathologist in contact and our observation of the data, the "halo" region was not an accurate measure of the melanocyte. To solve this issue, the goal of this project was to detect melanocyte from WSIs more accurately using a supervised method.

2 Dataset

We had access to a set of skin biopsy images from the MPATH study (R01 CA151306) conducted by University of Washington School of Medicine. This dataset contains two images with resolution of approximately 40000×25000 pixels. One of the images is a H&E Staining for all cells (Figure 1a), while the other is a Sox-10 staining for melanocytes (Figure 1b).

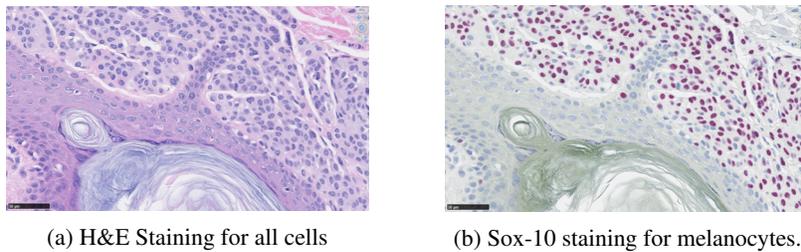


Figure 1: Sample WSI images.

3 Method

There are four major stages in the proposed workflow (Figure 2). From the original WSIs, we first performed image registration, which included utilizing Scale Invariant Feature Transformation (SIFT) and homography to transform the two images onto the same coordinate system. Next, we performed nuclei detection to obtain the nuclei masks. More specifically, we developed a Adaboost classifier based on pixel-wise and structural features. After obtaining the nuclei masks, we cropped the original image into nuclei patches of 100 pixels with the center of each patch being the centroid of a nuclei. From the nuclei patches, we performed data augmentation on the training set samples. Finally, we trained a convolutional neural network (CNN) and a DenseNet for melanocytes vs. non-melanocytes binary classification.

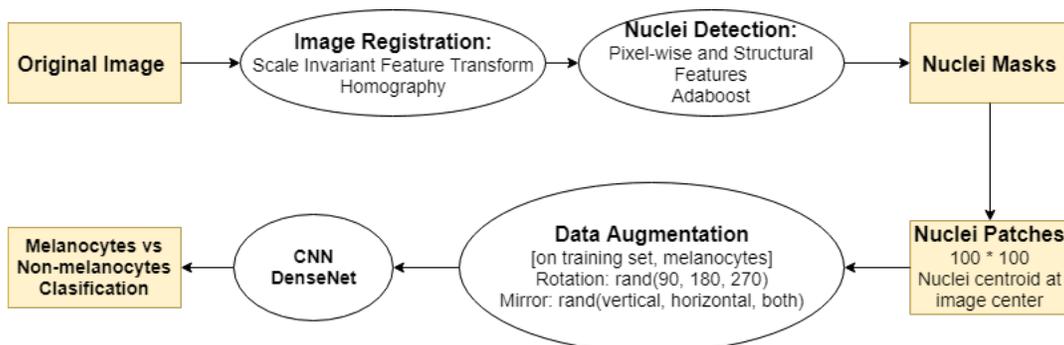
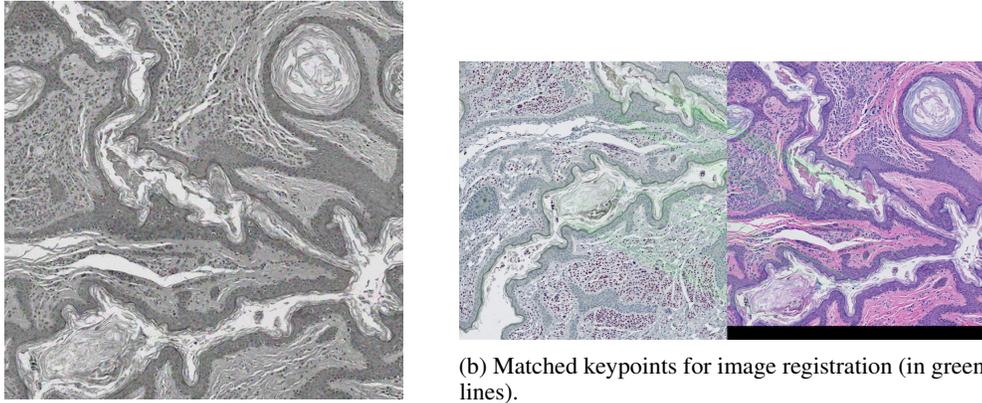


Figure 2: An overview of proposed method.

3.1 Image Preparation Registration

In order for the H&E staining to be matched with the Sox-10 melanocyte staining, we performed image registration with four steps: 1) extract distinctive image features from Scale-Invariant keypoints,

2) match the descriptors between the two images, 3) use the RANSAC algorithm to estimate a homography matrix from the matched keypoints and 4) use the homography matrix to perform a perspective transformation. Figure 3a shows an example of keypoints detected in an grayscale image and Figure 3b shows the matched important keypoints between two images. With the homography matrix computed from matched keypoints, we were able to match the H&E staining and Sox-10 staining together. A matched perspective view can be seen in Figure 1.



(a) Keypoints detected (as circles) from H&E staining in grayscale.

Figure 3: Image registration: keypoints detection and matching

3.2 Nuclei Detection

Nuclei detection was performed using an Adaboost-based classifier we previously developed and tested. The nuclei detector has F1 score of 89.28% and accuracy of 90.52%. In brief, this classifier utilized intensity-based features (e.g. pixel values) and structure-based features (e.g. Laplacian of Gaussian and Gaussian gradient magnitude). After Adaboost classification, morphological functions including hole filling, opening and closing, were applied to have better output. Opening function is the dilation of the erosion of an image by a structuring element. Closing is the erosion of the dilation of an image by a structuring element.

In order to label individual nuclei, we first found out the connected components in the classified mask. Then, using a area threshold, we identified the large components that could contain multiple cells. Finally, we apply watershed segmentation method to separate connected nucleus and label each nuclei (Figure 4b). The detected nuclei can be visualized in Figure 4a.

3.3 Melanocyte Detection and Patch Generation

From the original Sox-10 melanocytes-specific staining, we extracted the melanocyte mask by thresholding with the RGB channels (Figure 5a) and obtaining all red pixels. After thresholding, in order to fill the holes and have a better output, we used morphological functions as a post-processing step to clean up the mask. In particular, we applied an opening function and then a closing function, which are described in section 3.2.

After obtaining the melanocyte mask, we generated the nuclei patches. For each nuclei identified, we first extracted 100×100 pixels with the centroid of the nuclei at the center (Figure 5b). Then, we used the melanocyte masks generated to determine whether the nuclei in the center of the patch is a melanocyte or not. This gave us the binary label for the classifier training. This process can be visualized in Figure 5.

3.4 Data Augmentation and Splitting

After obtaining nuclei patches, we performed data splitting and data augmentation prior to inputting the patches into the CNN and the DenseNet. We performed data augmentation only to the training

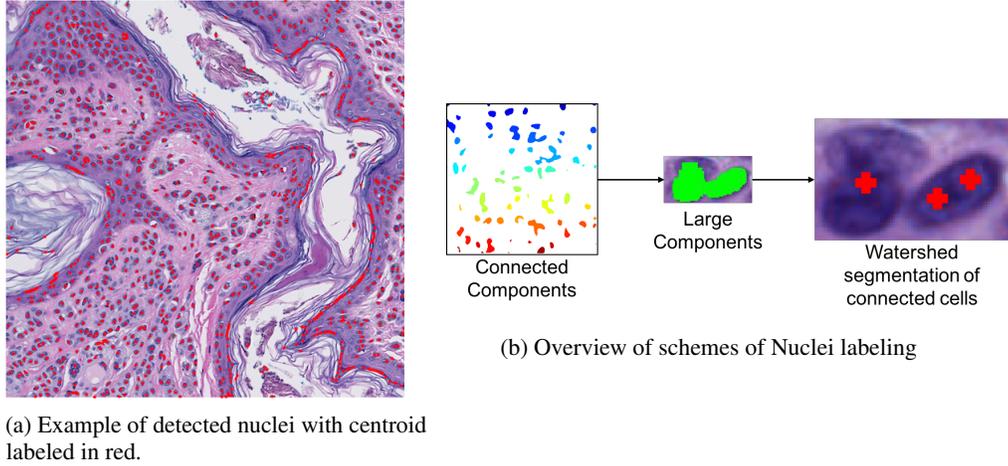
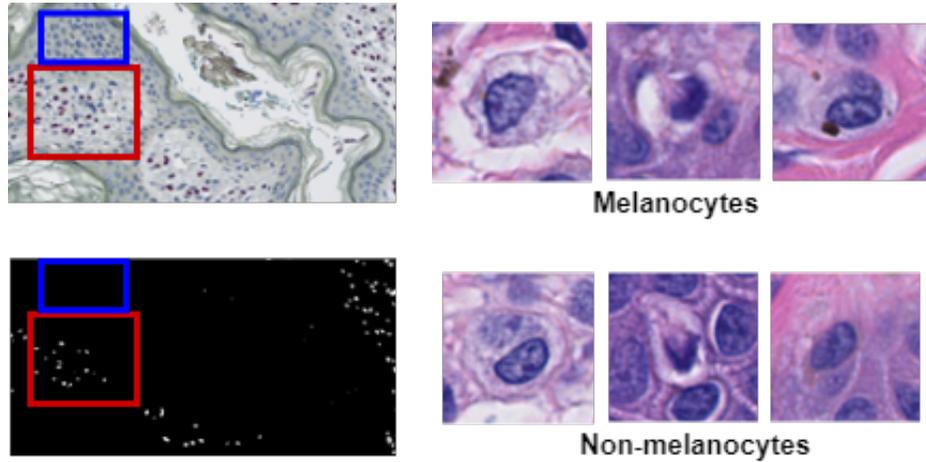


Figure 4: Nuclei detector results and scheme.



(a) Melanocyte detection from original WSI image. The red rectangle denotes detected melanocytes, while the blue rectangle denotes a region without melanocytes.

(b) Nuclei patches generated. Each patch has a size of 100×100 pixels. The center of each patch matches with the centroid of a nuclei.

Figure 5: Melanocytes mask and nuclei patches generated.

set in order to ensure balanced samples from positive and negative classes. Data augmentation was performed with two methods: 1) randomly rotate the image with one angle from $\{90, 180, 270\}$ degrees, and 2) randomly mirror the image along one out of the three axis: the vertical axis, the horizontal axis or both. Thus, from one image in the training set, we can obtain up to 3 training samples with two of them being augmented samples. The data split between training, validation and testing group is summarized in Table 1.

	Melanocytes	Non-melanocytes
Training	19597	19287
Validation	4818	7848
Testing	4818	7848

Table 1: The sample sizes of positive and negative cases in training, validation and testing set.

3.5 Convolutional Neural Networks (CNN)

CNN has the capability of end-to-end training and has shown potentials in efficient feature training, with both natural and medical images [Ronneberger et al., 2015]. Following the success of CNN on image classification task [Simonyan and Zisserman, 2014], we reasoned that they would also be a good choice for melanocyte detection since there is no established explicit features. By researching similar task in breast biopsy whole slide images, we adopted an architecture that is found to be superior by a previous study using similar patch sizes [Li et al.].

3.5.1 Network architecture

The architecture of the CNN proposed can be seen in Figure 6. There are 6 Convolutional-Batch-ReLU (CBR) units. Each CBR units has three consecutive function: a convolutional layer with size indicated in the graph, a batch normalization layer and finally a rectified linear unit (ReLU). Finally, a fully connected layer computes the activations of feature maps from previous layers with dimension of 2, which represents the binary classification.

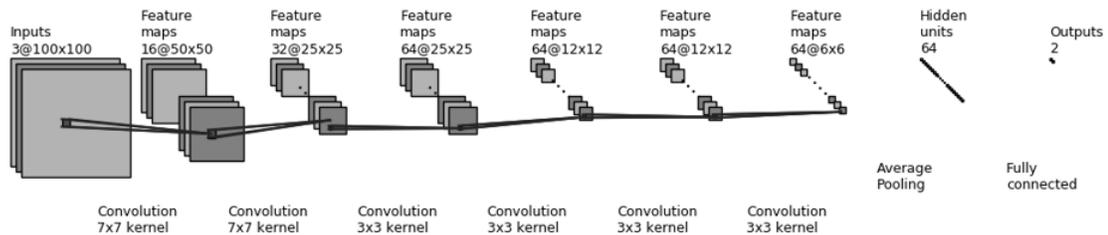


Figure 6: The architecture of the proposed CNN.

3.5.2 Training Hyperparameter optimization

We mainly investigated the effect of learning rate on performance metrics. We investigated the common learning rate used in previous literature performing similar task and selected the best out of three values: 0.1, 0.01, 0.001 [Li et al., Li]. We adopted a small and safe batch size suggested by [Li] to be 32.

To investigate the performance of the proposed CNN with different learning rate, we trained the model using stochastic gradient descent for 150 epochs and compared the accuracy and F1 scores. Then, we selected the learning rate with highest validation accuracy and F1 score combined and trained the model up to 300 epochs to get the training/validation/test performance curve vs. number of epochs trained.

3.6 DenseNet

In a plain CNN model, as the information about the input or gradient propagates through layers and layers, the information can be "washed out" by the time it reaches the end or the beginning of the network. Also, the down sampling operations in the convolutional units could result in a loss of spatial information. Dense CNN (DenseNet) was introduced to compensate the loss of information and improve the information flow [Huang et al., 2017]. In DenseNet, all layers are connected directly with each other, as shown in Figure 7. The dense connection allows direct access to gradients from the loss function and the original input, which may yield better parameter efficiency. This is especially true in medical images since local features in those biopsy images are alike. Thus, we would like to compare the performance of DenseNet with a plain CNN. The architecture we used is very similar to the plain CNN shown in Figure 6. The DenseNet we used has 6 dense blocks, which is connected

with each other. Each blocks has the same composite function of three consecutive operations: batch normalization, ReLU unit and convolution. We used a growth rate of 12, which would yield feature maps with roughly the same dimension as the plain CNN.

3.6.1 Hyperparameter and Training

Due to time limitation, we did not perform exhaustive hyperparameter selection for this model, we trained the model for 300 epochs with the selected batch size and learning rate for the plain CNN.

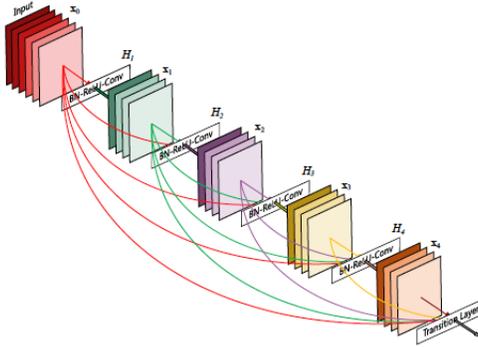


Figure 7: A 5-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input [Huang et al., 2017].

4 Results

To evaluate the proposed model, we used the standard evaluation metrics: accuracy and F1 score. The formula for each metric is as follows: $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ and $F1 = \frac{2 \cdot precision \cdot recall}{precision+recall}$, where $precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$. TP and TN is the number of true positives and true negatives; FP and FN is the number of false positives and false negatives.

In order to determine the most optimal learning rate, we examined the validation accuracy and F1 score of the CNN trained with learning rates discussed in section 3.5.2 (Figure 8b, 8a). As can be seen from the graph, the learning rate of 0.01 achieved the highest accuracy and F1 for the validation set. Thus, we chose 0.01 as the optimal learning rate.

After optimizing the learning rate, we trained the CNN for up to 300 epochs and examined the learning curve, which can be seen from Figure 8c and Figure 8d. From the plot, we could see that the model performance converges around epoch 200. Training beyond 300 epochs did not yield any significant change in performance on validation and test set.

For comparison purpose, we also trained the proposed DenseNet with learning rate of 0.01 for up to 300 epochs. The best accuracy from the plain CNN and the DenseNet is summarized in Table 2. To our surprise, DenseNet did not perform better than the plain CNN despite its maximum information flow between layers in the network. We think the reason could be that we did not have time to tune the hyperparameters (e.g. learning rate) specifically for the DenseNet, which could affect the performance of the model.

	(a) Accuracy		(b) F1 score	
	Plain CNN	DenseNet	Plain CNN	DenseNet
Training	0.8147	0.7565	0.7707	0.7623
Validation	0.7488	0.7181	0.7223	0.7065
Testing	0.7495	0.7194	0.7205	0.6998

Table 2: Performance of Plain CNN and DenseNet

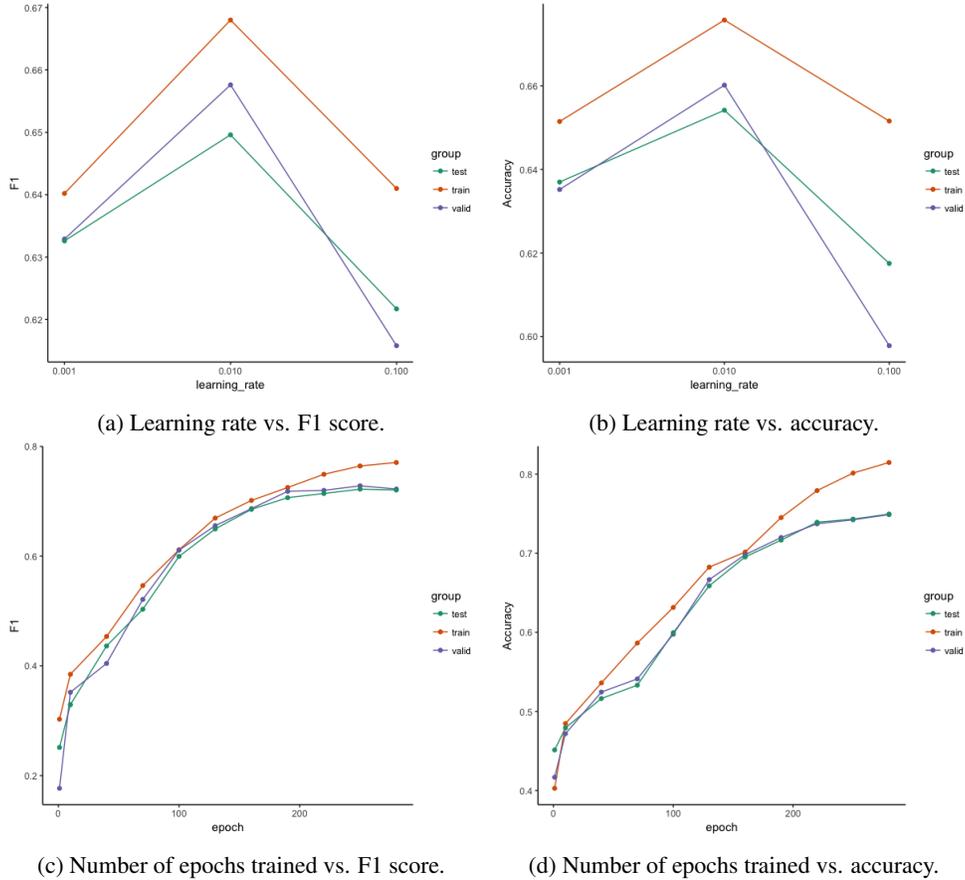


Figure 8: Results from CNN: Learning rate and Number of epochs vs. F1 and accuracy.

5 Conclusion

In this project, we were able to detect melanocytes in the nuclei patch. First, we detected the nuclei in the skin biopsy images using AdaBoost with intensity-based and structure-based features. Then, from the segmented nuclei, we were able to classify melanocytes using CNN. The classification reaches about 75% accuracy and 0.72 F1 score on test set, proving the feasibility of our approach. However, there is room for improvements. In the future, we will aim to improve the performance of the classifier by:

- Perform more comprehensive hyperparameter search for both the plain CNN and the DenseNet. Using adaptive learning rate, different batch sizes and growth rate for the DenseNet could improve the performance more.
- Try more advanced architectures like AlexNet and resNet, which reformulates the layers as learning residual functions.
- Explore various architectures of CNN: different number of layers, kernel size, strides and etc.
- Input some manually extracted features along with the CNN, which may help the model to concentrate on some features.

References

Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 411–418. Springer, 2013.

- Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, volume 1, page 3, 2017.
- Feifei Li.
- Yuguang Li, Ezgi Mercan, Stevan Knezevitch, Joann G Elmore, Linda G Shapiro 12, and Paul G Allen. Efficient and Accurate Mitosis Detection A Lightweight RCNN Approach. URL <https://homes.cs.washington.edu/~shapiro/yuguang-ICPRAM18.pdf>.
- Cheng Lu and Mrinal Mandal. Automated segmentation and analysis of the epidermis area in skin histopathological images. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 5355–5359. IEEE, 2012.
- Cheng Lu, Muhammad Mahmood, Naresh Jha, and Mrinal Mandal. Detection of melanocytes in skin histopathological images using radial line scanning. *Pattern Recognition*, 46(2):509–518, 2013.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.