Deep Learning for Metastatic Breast Cancer Detection

Weitian Ding, Ziyi Cui, Yuting Xiao
Carnegie Mellon University
Pittsburgh, PA 15213
weitiand@andrew.cmu.edu, ziyic@andrew.cmu.edu, yutingx1@andrew.cmu.edu

Abstract

The rapid development of machine learning techniques has made automated cancer identification on medical images achievable. This task is receiving an increasing amount of attention recently, and many algorithms have been proposed to provide a better diagnosis. In this work,we would like to explore the cutting-edge application of deep learning on medical image classification. More specifically, we will compare the results of metastatic breast cancer identification among a diverse array of deep learning models.

1 Introduction

Modern pathological analysis of cancer depends on the microscopic images on tissues from patients. Currently, it still relies on visual analysis of each slide by pathologists, which is prone to errors due to the heavy workload of pathologists and the difficulties of standardization in the manual evaluation of the microscopic images[7]. Increasing efforts have been put on developing the computational methods to assist the histopathological analysis, which is promising in the reduction of both the workload of the pathologists and the subjectivity in cancer diagnosis.

We chose the Camelyon Grand Challenge 2016 (Camelyon16, https://camelyon16.grand-challenge.org) on cancer metastasis detection on lymph node, held by the International Symposium on Biomedical Imaging (ISBI) challenge as our project to study the performance of several deep learning algorithms on the automated detection of cancer metastasis. This project aims to predict the metastasis of breast cancer based on the whole-slide images (WSIs) of sentinel lymph node biopsy, in both micro and macro scale. The sentinel nodes are the first few lymph nodes that cancer is likely to spread into the lymphatic system, and it works as an indicator of metastasis, especially in the breast cancer evaluation[3].

A positive breast sentinel lymph node suggests the invasion of cancer cells and it is important to guide the following therapies against breast cancer, such as removing the axillary lymph node. The sensitivity of breast sentinel lymph node biopsy is improved by immunohistochemistry (IHC) with pancytokeratins[2], since it is difficult to differentiate non-cohesive cells from lymphocytes that are of a similar size on slides purely stained by Haematoxylin and eosin (HE)[1]. However, the manual diagnostic procedure for pathologists remains tedious, time-assuming and is commonly accompanied by misinterpretation. Therefore, there is an increasing demand on computational methods that can efficiently pinpoint the small metastases from pancytokeratins IHC slides. This challenge of fully automated metastasis detection is the first one to use WSI as the data in histopathology, and it is meaningful in term of both the reduction of the workload for pathologists and the improvement of diagnosis for breast cancer patients.

The input data includes whole-slide images (WSIs) of sentinel lymph node biopsy collected independently from Radbound University Medical Center, and the University Medical Center Utrecht. We

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

are going to develop deep learning algorithms that can automatically localize the tumor in each WSI and discriminate metastasis slides from normal slides.

First, we start with our baseline method. We downsample and pre-process slide images and train a three-layer convolutional neural network (CNN) to produce a binary classification for each slide, which is normal or metastatic. We show that the baseline trained on downsampled image slides overfits the training dataset and does not extract useful information from the slides. In future work, we plan to train deep neural networks on sampled image patches and combine the predictions on image patches to produce predictions for the whole slide images.

2 Related Work

Recent work has seen promising results in Camelyon16. Wang et al. propose a "patch-framework" [10]. They train a deep neural network to classify whether an 256x256 image patch is a tumor, and during test time, they combine predictions on image patches to construct a tumor heatmap, which is used in the slide-level lesion localization task. For the tumor classification task, they train a random forest classifier using hand-engineered features of the heatmap. Wang et al. won Camelyon16 in both tasks. A more recent state of the art is presented by Liu et al [6]. Their approach is similar to the patch-framework proposed by Wang et al. except that they use the maximum value of the tumor heatmap as the tumor probability of the slide image instead of training a random forest. Also, since positive tumor samples are rare, Liu et al. downsample negative image patches and they perform data augmentation to address class imbalance.

Both Wang et al. and Liu et al. uses GoogleNet to classify image patches. However, GoogleNet contains 22 layers and thus is difficult to train without a GPU cluster. Other state of arts classification CNN include AlexNet[5], VGG[9], and ResNet [4]. Image segmentation CNNs is another class of networks widely used in medical image processing. Ronneberger et al. propose U-Net which consists of a contracting path and an up-sampling path [8]. They show UNet is very effective in bio-medical image segmentation tasks with small sample size.

3 Methods

3.1 Preprocessing

The training dataset contains 270 slide images where 160 are normal and the rest contain metastases. The test dataset has 130 slide images. The input WSIs are stored in a multi-resolution pyramid, including 9 different downsampling levels where level 0 is approximately 200k $\times$100k in resolution. Each downsampling level reduces the slide to a quarter of the previous level. Thus, to reduce the training time, we choose an image at a single downsampling level for each sample. We choose the level 6 for training a baseline since we can not fit larger images in the memory. Although each image in this pyramid is stored as a series of tiles for easier retrieval of subregions, preprocessing is still needed for focusing analysis and improved computational time. We apply background eliminations and RGB color filtering to exclude the potential markers on the image.

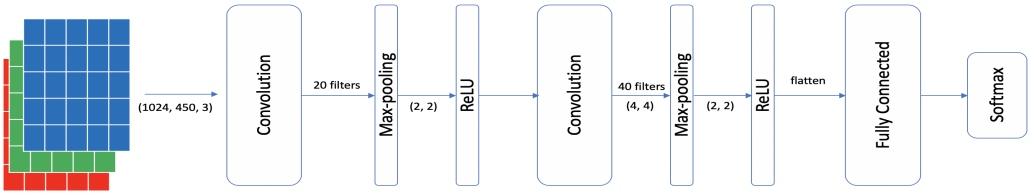


Figure 1: Architecture for baseline CNN

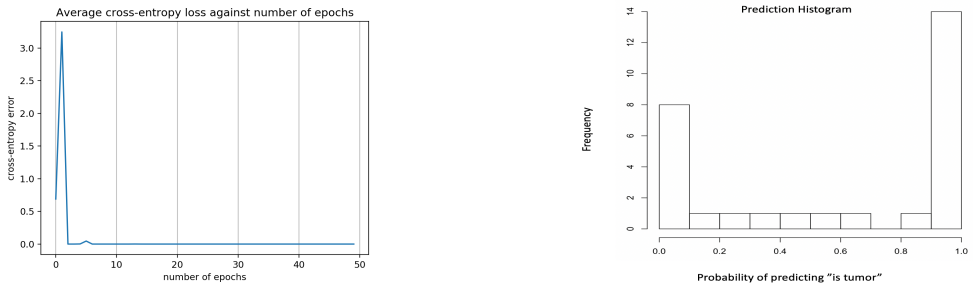
3.2 Baseline CNN

We build a three-layer CNN as the baseline for this project. The structure of the baseline method is shown in figure 1. All input images are reshaped and cropped to (1024, 450).

4 Preliminary Results

We train the CNN using 120 downsampled slide images for 50 epochs, with a learning rate of 0.001. We supervise the training with binary cross-entropy loss. The training data is rebalanced between the two labels. The trained network is tested on 29 test cases. We obtain perfect training accuracy but 0.48 test accuracy, which is close to random prediction. In the incorrect predictions, we have 10 false positives and 5 false negatives. To reduce overfitting, we reduce the training to 10 epochs, but a similar accuracy result was obtained.

The plots of the binary-cross-entropy loss and the histogram of predicted probabilities of 'tumor' are shown in figure 2. The training loss (figure 2a) shows the model converges and reaches zero training loss after 15 epochs, which is due to the small sample size. Also, the low test accuracy we obtain shows our model is significantly over-fitted on the training set. The prediction histogram(figure 2b) shows the model tends to make extremely positive or negative predictions, possibly due to overfitting. There are two explanations for the poor performance. One is that the training sample is too small for training neural networks from scratch. We can either use a pre-trained model or reduce the number of parameters. The second explanation which we think is more likely is that the level 6 downsampled slide images do not contain features related to tumor prediction and we should use a more magnified slide image. Level 6 is $\frac{1}{4^6}$ of the level 0 slide image. We show a solution in the future work section.



(a) The average cross-entropy loss after 50 epochs (b) The probability histogram of predicting 'tumor'

Figure 2: Plot results for baseline CNN

5 Future Work and Plan

The preliminary results show that naive machine learning pipelines which assume training data has been pre-processed do not work on Camelyon16. In future work, we adopt the patching-framework proposed by Wang et al [10]. We sample 256×256 patches from the level 0 image of each WSIand create 256x256 binary mask for each image patch using the annotated contours. We also effectively solve the small sample size problem by sampling image patches from the level 0. Weitian Ding and Ziyi Cui are responsible for this data engineering task, which should be completed by mid-March.

Next, we train deep neural networks on the image patches instead of the whole slide images. Both Wang et al. and Liu et al. use GoogleNet [10, 6]. We plan to experiment with an array of state of art deep neural networks, and different methods to supervise the training. For example, we can reduce the binary-mask to a probability that an image patch is a tumor and train classification CNNs such as AlexNet, VGG, and ResNet [5, 9, 4]. Another option is training image segmentation CNNs such as UNet that can be supervised by binary masks directly [8]. Also, we can apply multi-task training that is supervised by both the tumor probability and the binary mask to improve model interpret-ability and generalization. Yuting Xiao is responsible for evaluating the deep learning models. Weitian and Ziyi will also participate. We plan to complete this by early April.

Once we achieve promising prediction accuracy on image patches, we can combine the predictions on image patches to produce a prediction of an entire slide image. The combining step can either achieved by statistical methods or machine learning methods such as another deep neural network. This task should be completed by mid-April so that we have enough time for writing the final report.

References

[1] Gábor Cserni, Simonetta Bianchi, Vania Vezzosi, Hans Peterse, Anna Sapino, Riccardo Arisio, Angelika Reiner-Concin, Peter Regitnig, Jean-Pierre Bellocq, Cristi Marin, et al. The value of cytokeratin immunohistochemistry in the evaluation of axillary sentinel lymph nodes in patients with lobular breast carcinoma. *Journal of clinical pathology*, 59(5):518–522, 2006.

[2] Brian J Czerniecki, Alice M Scheff, Linda S Callans, Francis R Spitz, Isabelle Bedrosian, Emily F Conant, Susan G Orel, Jesse Berlin, Cynthia Helsabeck, Douglas L Fraker, et al. Immunohistochemistry with pancytokeratins improves the sensitivity of sentinel lymph node biopsy in patients with breast carcinoma. *Cancer*, 85(5):1098–1103, 1999.

[3] Stephen B Edge and Carolyn 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*, 17(6):1471–1474, 2010.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[6] 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.

[7] Stephen S Raab, Dana Marie Grzybicki, Janine E Janosky, Richard J Zarbo, Frederick A Meier, Chris Jensen, and Stanley J Geyer. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer*, 104(10):2205–2213, 2005.

[8] 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.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[10] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016.