Histopathological Image Classification for Cancer Detection using Convolutional Neural Networks

Mikolaj Jaworski
Collegium of Economic Analysis
SGH Warsaw School of Economics

Abstract

Medical screening is one of the main applications of image recognition systems that are supporting doctors in patient's condition evaluation. Different models are being developed simultaneously for photos coming from histopathological examinations, mammographics, blood smears, X-rays, ultrasounds, MRIs. Conclusions coming from the literature states that many models are preforming better or close to dostor's classification accuracy and are much more time efficient. Such a support could be useful considering continuously increasing volume of medical data. Development and analysis of such a solution was the goal of the thesis. Binary classification of histopathologic images of lymph nodes was performed in order to detect presence of metastatic cells. Data was published on Kaggle and it included over 275 thousands of images. Convolutional neural networks with transfer learning and regularization methods were applied. The best performing architecture turned out to be DenseNet 201, which led to the result of 97.56% AUC on 51% of the test set and 97.08% on the rest of it. It places the model among first 100 of the competition's leaderboard. Taking into consideration simplicity of the program's code and relative short time spent on training (almost 8 hours) the model could be beneficial in supporting oncologists' decicion processes.

1 Introduction

This paper is shortened version of author's master thesis. The purpose of this paper is twofold. Firstly, is to highlight some most effective applications of image analysis methods using convolutional neural networks in medical screening and diagnosis. Secondly, is to present prototype of classifier detecting metastatic cells in histopathological images. In the following sections literature, methods, and experiments will be described.

2 Literature

First reviewed paper is based on Camelyeon16 dataset [1], which is initial version of the dataset used in this paper. Researchers presented a model based on GoogLeNet CNN architecture [2]. To evaluate their model they used FROC (free-response receiver operating characteristic) [3], which was metric used by radiologists and provides information on how accurately suspected areas on the images are detected. Comparing to radiologist who had 30 hours to examine the test set, model outperformed him by 0.9 (97.5%) and 14.7 (88%) percent points on AUC and FROC respectively. Second paper focuses on mammography images [4]. The main challenges in this case were data availability, multiclass problem and fact that one person could have multiple images from different perspectives. Metric combining AUC for multiple classes was used (macAUC) which reached level of 0.688,which was close to board of oncologists with score of 0.704. Cooperation of model and human returned 0.735 macAUC. Third paper presents results of classification of blood smear images to detect malaria [5]. Ensemble of two CNN architectures: VGG-19 and SqueezeNet returned the result of 99.92% AUC and 99.51% accuracy [6]. Model presented in fourth paper [7] was detecting tuberculosis on X-ray images. Again, ensemble of two architectures - AlexNet and GoogLeNet returned the best result, which was 99% AUC. Fifth paper tackles segmentation problem on MRI brain images for cancer detection using BRATS 2015 dataset [8]. The challenges involved data variability dependent on patients and medical personnel conducting examination. The best segmentation models based on CNN returned Dice scores: 0.78, 0.65, 0.75 for three versions of the cancer images respectively: full, main parts, only active cells [9].

3 Experiments and results

The goal of experimental part was to create as accurate CNN model as possible for detecting metastatic cells in histopathological images. Dataset is a modified version of Camelyeon16, which was prepared for online data science competition [10][11]. It consists of 220 020 training samples and 57 458 test samples. As test set was not publicly available, more metrics and information could be gained from the training set. It is imbalanced as only 89 117 images belongs to positive class and 130908 to negative. Conducted experiments focused on manipulating size of train and validation split, type of pretrained architecture, resizing of the images, applying augmentations and modifying hyperparameters. Computations were programmed using PyTorch (FastAI) Python package and performed on instance with NVIDIA Tesla P4 GPU, on which one epoch of training took roughly 55 minutes. After 50 experiments with various combinations with parameters, the best results were returned using below configuration of parameters:

- 20% of training set was used for validation
- Images were resized to 224px and transformed (with probability of 75%) using horizontal and vertical flips, rotations, zoom, contrast, brightness, perspective warp and jitter.
- ADAM optimization, weight decay equal to 0.01 and 0.5 dropout rate.
- Learning rate was being changed through learning process (from 1e-03 in the beginning and up to 1e-07 in the end) which length was 6 epochs.

The most significant differences were noted between different pretrained architecture types and their depths: ResNet [12] and DenseNet [13]. Table 1. allows comparison of AUC score per network.

Network	51% of test set	49% of test set	Validation set
ResNet50	97.18%	96.83%	99.27%
ResNet101	97.11%	96.92%	99.21%
DenseNet169	97.15%	97.01%	99.16%
DenseNet201	97.56%	97.08%	99.08%

Table 1: AUC scores per network type and dataset part.

It can be noticed, than the deeper the network, the better is AUC score. Also when metric is gaining score on test set, it losses on Validation set, which could be related to overfitting. The result of Densenet 201 could be placed among 100 best submissions in the competition.

4 References

[1] Ehteshami Bejnordi. et. al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience. 2018.

[2] Liu, Yun. et. al. Detecting Cancer Metastases on Gigapixel Pathology Images. Google AI. 2017.

[3] Bandos, Andriy. et. al. Area under the Free-Response ROC Curve (FROC) and a Related Summary Index. PubMed-NCBI. 2009.

[4] Geras, Krzysztof, et. al. High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks. arXiv:1703.07047. 2018.

[5] Rajaraman, Sivaramakrishnan. et. al. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PubMed-NCBI. 2018.

[6] Rajaraman, Sivaramakrishnan. et. al. Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. PubMed-NCBI. 2019.

[7] Lakhani, Paras. Sundaram, Baskaran. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. RSNA Radiology. 2017.

[8] Menze, Bjoern. et. al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging, 34(10). 2015.

[9] Pereira, Sergio. et. al. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Transactions on Medical Imaging, 35(5). 2016.

[10] Ehteshami Bejnordi et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA, 318(22). 2017.

[11] Veeling, Basitan. et. al. "Rotation Equivariant CNNs for Digital Pathology". arXiv:1806.03962. 2018.

[12] He, Kaiming. et. al. Deep Residual Learning for Image Recognition. Microsoft Research. 2015.

[13] Huang, Gao. et. al. Densely Connected Convolutional Networks. arXiv:1608.06993. 2018.