

Histopathological Image Analysis Using Deep Learning

Faye Hall
March 2021

Domain Background

Histopathology is a way of diagnose disease by examining processed body tissues that were dissected during surgery or biopsy. These tissues are thinly sliced, gone through chemical fixation or frozen section process, stained commonly using hematoxylin for blue and eosin for pink (H&E), and finally placed between glass slides for pathologist to examine under microscope.

With Whole Slide Scanner, the pathology tissue slides are digitized. The availability of the huge amount of these Whole Slide Images,(WSIs) online, coupled with the rich spacial features in these WSIs, makes WSIs good candidates for deep learning.

One application of deep learning is detection of metastases in breast cancer histopathology images.

Problem Statement

Use deep learning classification algorithm to determine whether metastases are detected in a pathology tissue image.

Datasets and inputs

The PatchCamelyon (PCam) dataset, derived from Camelyon16 Challenge, will be used for this project. PCam has 327 96x96 WSIs from lymph node sections. It is divided into 262 training dataset, 32 validation dataset, and 32 testing dataset. Each dataset, half does metastases, half doesn't. (<https://github.com/basveeling/pcam>) [7]

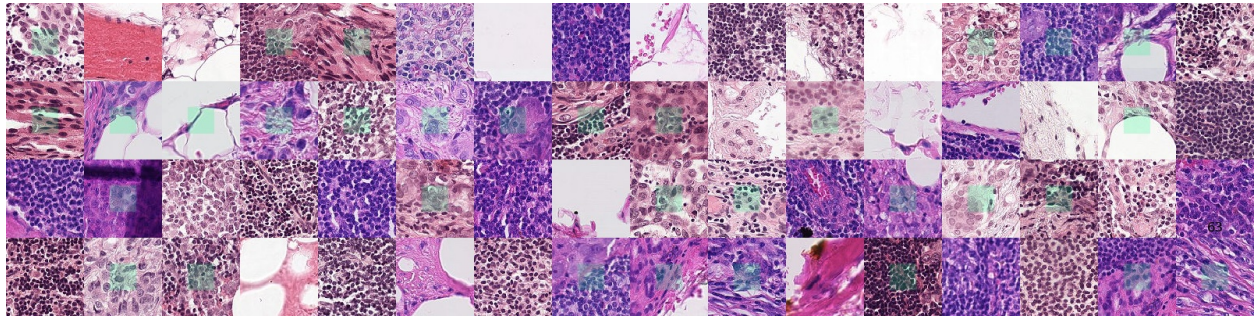


Image from <https://github.com/basveeling/pcam>

The PCam dataset size is 9GB, in HDF5 file format.

Solution statement

Use transfer learning to create a CNN model based on pre trained VGG-16 model to discriminate between WSIs with metastases vs without metastases.

Evaluation Metrics

1. **Sensitivity:** True Positive rate [6]
2. **Specificity:** True Negative rate [6]

A benchmark model

11 pathologists, with 2-hour time limit classified 129 WSIs with average:

sensitivity = 62.8%
specificity = 98.5%

Project Design

1. Preprocess the dataset
 - + data augmentation by applying transformations to increase the training set size using torchvision.transforms library
 - + normalize the data
 - + stain normalization
(https://github.com/wanghao14/Stain_Normalization/blob/master/stainNorm_Macenko.py)
2. Load the data
3. Architecture the model
 - + use pretrained VGG-16 model
 - + freeze all layers except the last layer

- + replace the last layer with linear transformation with 2 output features
 - 4. Specify Loss Function and Optimizer
 - + use CrossEntropyLoss function
 - + use SGD optimizer
 - 5. Train and Validate the Model
 - + train the model with certain learning rate, batch size and epoch
 - + validate the model using the validation dataset
 - 6. Test the model
 - + apply the model to the testing data set and calculate sensitivity and specificity rates, adjust hyper parameters in order to get comparable sensibility and specificity rates with the benchmark.
 - 7. Predict metastases detection using the model given an WSI.
-

References

- [1] <https://en.wikipedia.org/wiki/Histopathology>
- [2] <https://pubmed.ncbi.nlm.nih.gov/20671804/>
- [3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5556681/>
- [4] <https://ieeexplore.ieee.org/document/7399414>
- [5] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5820737/>
- [6] https://en.wikipedia.org/wiki/Sensitivity_and_specificity
- [7] <https://github.com/basveeling/pcam>