Histopathological Image Analysis Using Deep Learning

Faye Hall
March 2021

Domain Background

I. Histopathology

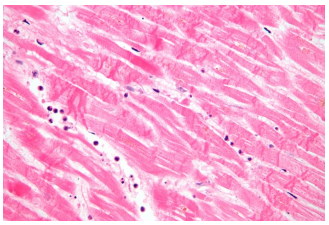
Histopathology refers to the microscopic examination of tissue in order to study the causes and effects of disease [1]. Histopathology process involves these following producers:

1. Collection of tissues

    During surgery, biopsy, or autopsy, the tissue is removed from the body, and then, often following expert dissection in the fresh state, placed in a fixative such as formalin, which stabilizes the tissue to prevent decay [1].

2. Preparation for histology

    The tissue is processed with either chemical fixation or frozen section process:



    Chemical Fixation

    The tissue is placed in a fixative, mostly formalin, then dehydrated using alcohol and then xylene, afterwards, wax is used to infiltrate the tissue. The wax infiltrated tissue is then transferred to an individual embedding container. Finally molten wax is introduced around the tissue in the container and cooled to solidification so as to embed it in the wax block. Once the wax embedded block is finished, sections will be cut from it and placed on slides. Afterwards, the thin section mounted slide is stained and a protective cover slip is mounted [1].

    Frozen Section

    The tissue is frozen and sliced thinly using a microtome mounted in a below-freezing refrigeration device called the cryostat. It is then mounted on a glass slide, fixed immediately and briefly in liquid fixative, and stained [1].

    Comparing to chemical fixation, frozen section process is rapid, require less equipment, and less ventilation in the lab. The disadvantage is the poor quality of the final slide.

Both chemical fixation and frozen section use similar stain. The most commonly used stain is a combination of hematoxylin and eosin (H&E). Hematoxylin is used to stain uncle blue, while eosin stains cytoplasm and the extracellular connective tissue matrix pink [1].

3. Interpretation

The resulting slides, i.e. the tissue histopathology slides, are then examined by a pathologist under a microscope by a pathologist to formulate medical diagnosis.

II. Digital Pathology

With the advent of whole slide digital scanners, tissue histopathology slides are digitized and stored in digital image form. Large amount of image-rich pathology data available online, gives rises to mining features that not visually discernible by a pathologist, thus improve prediction of disease [2][3].

III. Deep Learning

Deep Learning paradigms represent end-to-end unsupervised feature generation methods that take advantage of large amounts of training data in conjunction with multi-layered neural network architectures. Histopathology images, given their data complexity and density are ideally suited for interrogation via deep learning. Deep learning also makes less demand on the understanding the nuances of domain complexity. Specifically, it can be applied in the following areas of pathology:

1. Segmentation and detection of histologic primitives

One application of Deep Learning is detection and classification of histologic primitives (e.g. nuclei, glands). One approach proposed by Sirinukunwattana et al. [4] is to detect nucleus using a Spatially Constrained Convolutional Neural Network (SC-CNN). SC-CNN regresses the likelihood of a pixel being the center of a nucleus, where high probability values are constrained to locate the vicinity of the centers of unclei. For classification of nuclei, he proposes a novel Neighboring Ensemble Predictor (NEP) coupled with CNN to more accurately predict the class label of detected cell nuclei.

Deep learning approaches can also be used for mitoses detection of breast cancer from routine H&E stained tissue images.

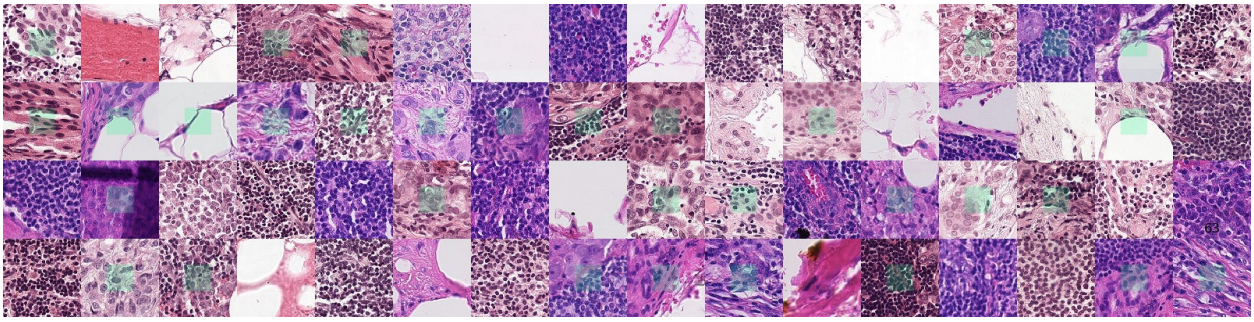
2. Tissue classification, grading and precision medicine

Problem Statement

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

Datasets and inputs

The PatchCamelyon (PCam) dataset is used here. It consists of 327 color images (96x96) extracted from histopathologic scans of lymph node sections. Each image is annoyed with a binary label indicating presence of metastatic tissue. (https://github.com/basveeling/pcam) [7]



The dataset is divided into a training set of 262 examples, and a validation and test set each of 32 examples. All sets have 50/50 balance between positive and negative examples. [7]

A positive label indicates the center 32x32px region of a patch contains at least one pixel of tumor tissue. Tumor tissue in the outer region of the patch does not influence the label. This outer region is provided to enable the design of fully-convolutional models that do not use any zero-padding, to ensure consistent behavior when applied to a whole-slide image. [7]

PCam is derived from the Camelyon16 Challenge, which contains 400 H&E stained WSIs (Whole Slide Images) of sentinel lymph node sections. [7]

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): measures the proportion of positives that are correctly identified, i.e. the proportion of those who have some condition who are correctly identified as having the condition. [6]

2. specificity (True Negative rate): measures the proportion of negatives that are correctly identified, i.e. the proportion of those who do not have the condition who are correctly identified as not having the condition. [6]

A benchmark model

A panel of 11 pathologists were given the 129 slides, flexible 2-hour time limit, to classify each slide as with or without metastases . The group has average age 47.7, 10 practicing pathologists, and 1 resident pathologist. Three of these pathologists had breast pathology as a special interest area. The group achieved a mean sensitivity of 62.8% and a mean specificity of 98.5%.

Characteristics of the Whole-Slide Images and Glass Slides in the Data Sets Used in the CAMELYON16 Challenge

Data Set (N = 399 Slides and Images)[a]	Hospital Providing the Slides and Images	Primary Tumor Histotype[b]		Slides Containing Metastases, No.			No. of Lesions per Slide or Image, Median (Range)	Total Slides or Images
		IDC	Non-IDC	None	Macro	Micro		
Training (n = 270 images)	RUMC	54	16	100	35	35	2 (1-20)	170
	UMCU	30	10	60	26	14	3 (1-27)	100
Test (n=129 slides and images)	RUMC	23	6	50	14	15	2 (1-14)	79
	UMCU	15	5	30	8	12	3 (1-25)	50

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: modifying the color appearance of WSIs so that they resemble some reference sample with the goal of reducing the appearance variability within a dataset.
    (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 CrossEntryopyLoss 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