Histopathological Image Analysis using attention-based MIL and Transfer Learning techniques

Authors

Nupur Yadav
Lokesh Vadlamudi
Chetan Kulkarni

December 6, 2020

Advisor

Prof. Vijay Eranti

Table of Contents

Abstract

Deep learning is being used in every field nowadays, but it has played a very crucial role in healthcare and pathology. The advanced equipment, such as specialized scanning machines and strides in storage capabilities has made it easy to store and process microscopic glass slides in digital forms on the computer. This has resulted in remote diagnosis, faster analysis, improved accuracy, and safe storage of pathology information.

However, classifying histopathological slides is challenging because of the two reasons. Firstly, the histopathological images are high-resolution images that are meant to be examined under the microscope by the pathologists. Secondly, it is extremely difficult to obtain a large dataset of WSIs with detailed pixel-wise annotations that makes it challenging to use a supervised deep learning classifier

Therefore, in this project, we aim to use weakly supervised techniques such as attention-based Multiple Instance Learning (MIL) for histopathological image classification that addresses the above problems of fully supervised learning. MIL uses the whole-slide labels instead of using pixel-wise annotations. We have also use transfer learning technique to classify histopathological images. We have performed histopathological classification on BreakHis and Breast Histopathology images (IDC) dataset. Though Attention-based Deep Multiple Instance Learning is applied in a wide range of medical imaging applications. Surprisingly, it didn't work for the breast cancer dataset. The transfer learning technique yielded the best accuracy of 94% on BreakHis dataset.

Introduction

The fully supervised techniques such as CNN have proved to be state-of-the-art for many image classification datasets such as ImageNet, automating the classification of histopathological slides is challenging because of the following reasons. Firstly, the images that CNN works well on are low-resolution images while histopathological images are high-resolution images that are meant to be examined under the microscope by the pathologists. The image can contain billion of pixels while the area of interest can be just a few thousand pixels. So, to apply a fully supervised technique, the Whole Slide Image (WSI) has to be divided into several small tiles, and the classifier is to be applied individually to each tile. The output is then aggregated to obtain a final classification for the WSI. Secondly, it is extremely difficult to obtain a large dataset of WSIs with detailed pixel-wise annotations that make it challenging to use a supervised deep learning classifier.

In this project, we have used two techniques to classify histopathological slides. First is Attention-based Deep Multiple Instance Learning which has proven to be successful in a wide range of medical imaging applications. Second is transfer learning where we have used “DenseNet201” architecture with weights pre-trained on ImageNet data. Initially we performed attention-based MIL on BreakHis dataset using PyTorch. The model yielded an accuracy of '69%' which is very low for medical domain where most precision is required. In general, this dataset is difficult due to high variability of slides and small number of cases. This is also backed by the paper “Attention-based Deep Multiple Instance Learning” (Ilse et al. 2018) which we later came across.

So, we thought of experimenting a bit and used another breast cancer (IDC) dataset from Kaggle to apply attention-based MIL using Keras. This time we also did data augmentation to increase the number of samples and hoping to improve the model accuracy. The model yielded an accuracy of '73%' this time which is better than the previous one but still low for the medical domain. Though Attention-based Deep Multiple Instance Learning is applied in a wide range of medical imaging applications. Surprisingly, it didn't work for the breast cancer dataset. So, we thought of using a pretrained model (transfer learning technique) for breast cancer classification on BreakHis dataset to see if it can achieve better accuracy. The model achieved an accuracy of '94%' which is quite good.

Related Work

We have taken reference from the paper “Attention-based Deep Multiple Instance Learning” (Ilse et al. 2018) where they stated the MIL problem as learning the Bernoulli distribution of the bag label where the bag label probability is fully parameterized by neural networks. Further they proposed a neural network-based permutation-invariant aggregation operator that corresponds to the attention mechanism. Notably, an application of the proposed attention-based operator provides insight into the contribution of each instance to the bag label. They showed empirically that their approach achieved comparable performance to the best MIL methods on benchmark MIL datasets and it outperforms other methods on a MNIST-based MIL dataset and two real-life histopathology datasets without sacrificing interpretability.

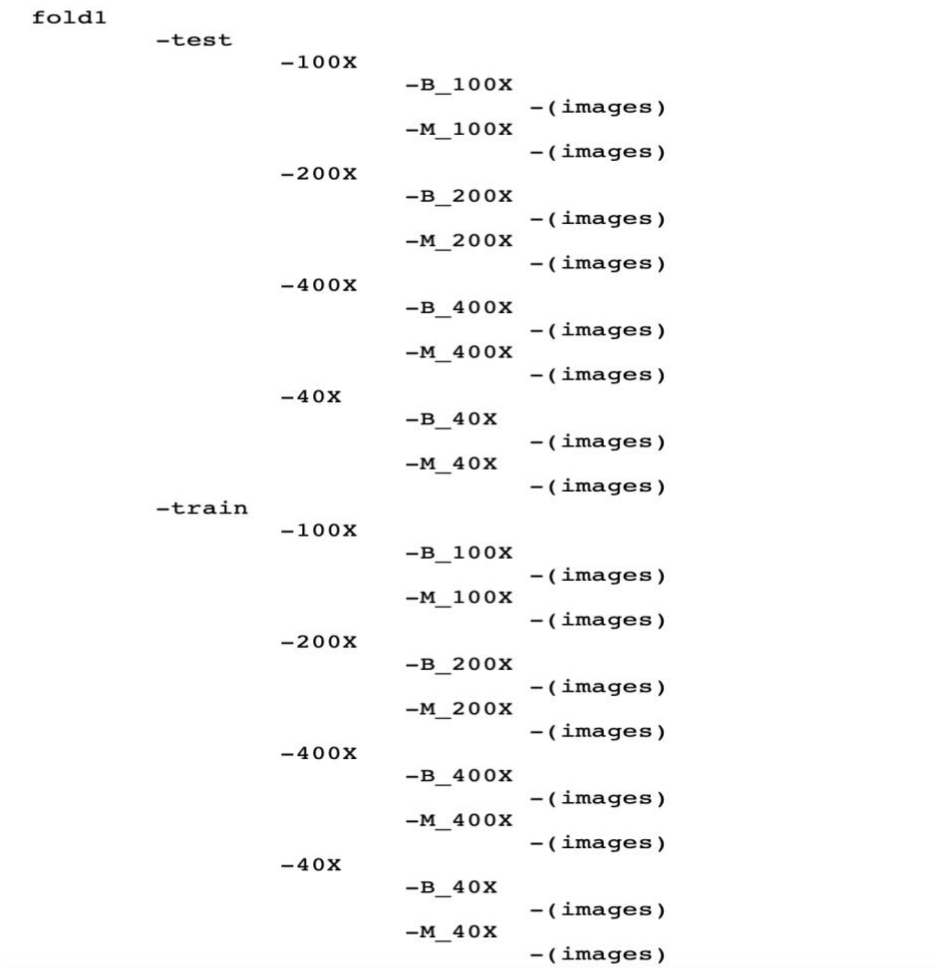
So, in our project we used breast cancer histopathology images (BreakHis) dataset to perform attention-based MIL using the same approach in PyTorch but surprisingly our model gave a very low accuracy of just 69%. Later on, we came to know that breast cancer dataset is difficult due to high variability of slides and small number of cases. It was also backed by the above-mentioned paper. Upon knowing this we thought of experimenting a bit and got another breast cancer histopathology (IDC) dataset from Kaggle. This time we used Keras and also performed data augmentation to increase the number of samples for training our model. This time we got an accuracy of 74%, somewhat better than the previous model. But still, it is low and not acceptable for breast cancer classification.
Then we performed transfer learning on BreakHis dataset. In this approach, we used “DensetNet201” architecture with weights pre-trained on ImageNet data. We also used ReduceLROnPlateau callback which Reduce learning rate when a metric has stopped improving. Model often benefits from this. This time our model yielded an accuracy of 94% which is quite good.

Data

We have used two datasets in our project. Both are taken from Kaggle.

1. [BreakHis (Breast Cancer Histopathological Database) :]()This dataset is divided into two main groups: benign tumors and malignant tumors. It is composed of 9,109 microscopic images of breast tumor tissue collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). To date, it contains 2,480 benign and 5,429 malignant samples (700X460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format).

   The original dataset was in the following format:



```
fold1
        -test
                -100X
                        -B_100X
                                -(images)
                        -M_100X
                                -(images)
                -200X
                        -B_200X
                                -(images)
                        -M_200X
                                -(images)
                -400X
                        -B_400X
                                -(images)
                        -M_400X
                                -(images)
                -40X
                        -B_40X
                                -(images)
                        -M_40X
                                -(images)
        -train
                -100X
                        -B_100X
                                -(images)
                        -M_100X
                                -(images)
                -200X
                        -B_200X
                                -(images)
                        -M_200X
                                -(images)
                -400X
                        -B_400X
                                -(images)
                        -M_400X
                                -(images)
                -40X
                        -B_40X
                                -(images)
                        -M_40X
                                -(images)
```

Figure 1: BreakHis dataset structure

   We converted it to the below format and saved it on google drive for ease of future use.

```
data_breakhis
        -100X
                -train
                        -0
                                -images
                        -1
                                -images
                -test
                        -0
                                -images
                        -1
                                -images
        -200X
                -train
                        -0
                                -images
                        -1
                                -images
                -test
                        -0
                                -images
                        -1
                                -images
        -400X
                -train
                        -0
                                -images
                        -1
                                -images
                -test
                        -0
                                -images
                        -1
                                -images
        -40X
                -train
                        -0
                                -images
                        -1
                                -images
                -test
                        -0
                                -images
                        -1
                                -images
```

Figure 2: BreakHis dataset formated structure

2. Breast Histopathology Images (IDC) dataset : Invasive ductal carcinoma (IDC) is the most common form of breast cancer. The original dataset consisted of 162 whole mount slide images of Breast Cancer specimens scanned at 40x. From that, 277,524 patches of size 50 x 50 were extracted (198,738 IDC negative and 78,786 IDC positive).

   The original dataset was in the following format:
   IDC dataset

```
--Patient-id

        -- 0
            --Images
        -- 1
            --Images
```

We converted it to the below format and saved it on google drive for further use.

IDC converted dataset

```
--train
    -- 0
        --Images
    -- 1
        --Images

--test
    -- 0
        --Images
    -- 1
        --Images
```

Additional preprocessing steps performed were:

1. Removed .jpeg images from the dataset and kept only .png images to train our model on.
2. Resized all images in the dataset to 50 X 50 as the dataset contained different size images.
3. Data augmentation to increase number of training samples.

Methods

In this project, we have used two techniques:

1. Attention-based Deep Multiple Instance Learning which has proven to be successful in a wide range of medical imaging applications.

2. Transfer learning technique where we have used a "DenseNet201" architecture with weights pre-trained on ImageNet.

Attention-based MIL approach

Multiple instance learning (MIL) is a variation of supervised learning where a single class label is assigned to a bag of instances. MIL uses the whole-slide labels instead of using pixel-wise annotations. In attention-based MIL approach, we have cropped an image in small square patches and made a bag of it. This bag of images will act like a batch. Passing it to the feature extractor which is basically a CNN block to generate instance level features. Then, we are passing the Instance level features to the classifier for getting Instance level attention. Here we are getting the attention weights which we are further using for attention aggregation to get the bag level features. Then we are applying Dense layer for the classification of the Benign or Malignant

Following is the diagram showing the approach step by step:

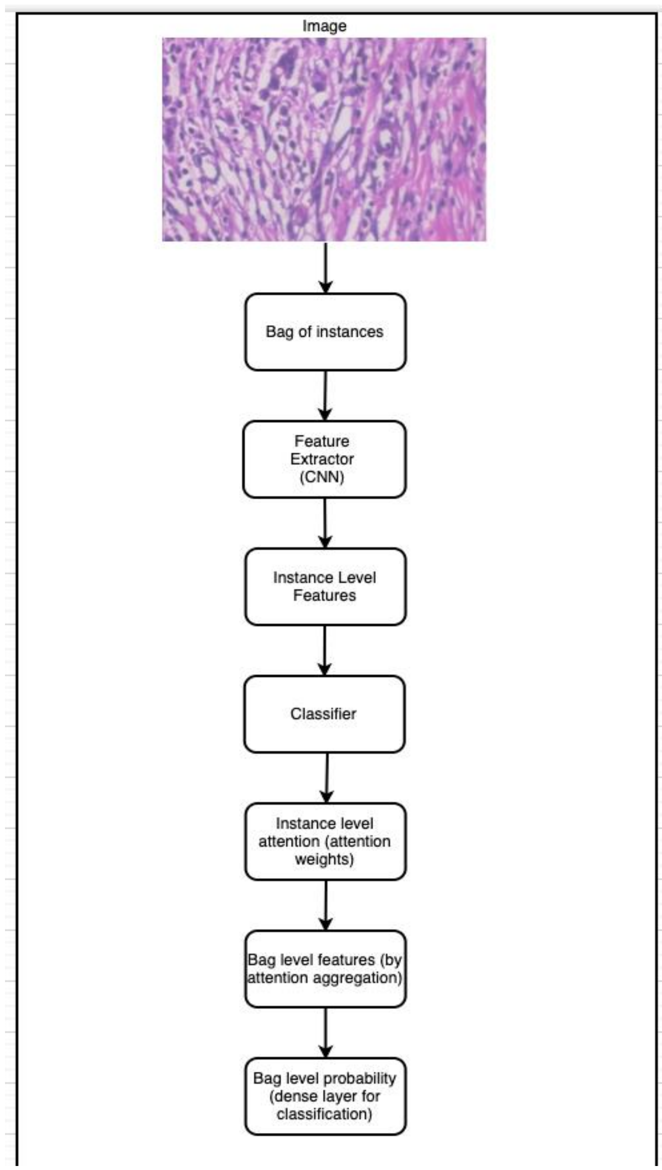


Figure 3: Attention-based MIL approach

Transfer Learning approach

Transfer learning allows us to train deep networks using significantly less data that we would need if we had to train from scratch. The idea is that the two tasks are not totally disjoint, and as such we can leverage whatever network parameters that model has learned through its extensive training, without having to do that training ourselves. Transfer learning has been consistently proven to boost model accuracy and reduce require training time. Less data, less time, more accuracy. In this approach, **we have used "DensetNet201" architecture with weights pre-trained on ImageNet data**. We have used ReduceLROnPlateau callback which Reduce learning rate when a metric has stopped improving. Models often benefit from reducing the learning rate by a factor of 2-10 once learning stagnates. This callback monitors a quantity and if no improvement is seen for a 'patience' number of epochs, the learning rate is reduced. This model achieved a better accuracy than the attention-based MIL approach on breast cancer dataset (BreakHis).

Experiments

In this project we experimented with three different techniques and used two different datasets and did an ablation study. Both datasets are taken from Kaggle. Following describes the experiments in detail.

1. Initially, we performed attention-based MIL on BreakHis dataset where we used a neural network-based permutation-invariant aggregation operator that corresponds to the attention mechanism. The model was developed using PyTorch. We used tensorboard visualizations for monitoring the training and testing of the model. The model resulted in a very low accuracy of just 69% not optimal for medical domain. In general, this dataset is this dataset is difficult due to high variability of slides and small number of cases.

2. So, we experimented a bit. We chose a different breast cancer histopathological (IDC) dataset from Kaggle. This time we performed attention-based MIL on this dataset in Keras and did data augmentation to increase the number of samples for training the model. Though model improved a bit and showed an accuracy of 73%, but still not acceptable.

3. So, we approached the problem with transfer learning technique this time, as transfer learning allows us to train deep networks using significantly less data that we would need if we had to train from scratch. We used BreakHis dataset that we used in our first experiment. We used "DensetNet201" architecture with weights pre-trained on ImageNet data. Further, we used ReduceLROnPlateau callback which Reduce learning rate when a metric has stopped improving and helps the model learn better and converge faster. Also, integrated tensorboard for monitoring model training and validation. This model achieved the best accuracy of 94%. Hence, we used this model for the deployment and making future inferences.

Conclusion

Following table summarizes all the experiments and observations.

Technique	Accuracy
Attention-based MIL on BreakHis dataset (Pytorch implementation)	69%
Attention-based MIL on Breast Cancer histopathological images (different dataset, data augmentation and Keras implementation)	73%
Transfer Learning on BreakHis dataset (using DenseNet201)	94%

Hence, we used this transfer learning model for deployment and making future inferences.

Figure 4: Models and their accuracies

Following are the learnings from this project:

4. We learned how **to apply attention-based MIL technique to classify histopathological images.**
5. We also learnt that though attention-based MIL has proved successful for various medical imaging applications such as detecting colon cancer, lung cancer, etc., but it did not work well for the breast cancer classification. In general, this dataset is difficult due to high variability of slides and small number of cases.
6. We learnt that **applying data augmentation techniques can sometimes improve model accuracy** to a great extent.
7. We learnt how transfer learning techniques can boost a model accuracy especially when we have little data and reduce the required training time.
8. We learnt using reduced **learning rate technique such as ReduceLROnPlateau** can often benefit model and help model to converge faster.
9. We learnt how to use **TFX TensorFlow serving along with Docker and Kubernetes to deploy a machine learning model in production.**

For future extensions, I think it would be interesting to see if variations of attention-based MIL technique such as gated attention or using embedding-based method with max pooling could result in improved accuracy of the model for breast cancer histopathological dataset.