Capstone Project Report

Executive Summary

**I want to apply and test the ability of state-of-art techniques connecting text with image, zero-shot classification models, to predict or annotate unlabeled medical images, MedMNIST dataset which is a benchmark dataset used to test new algorithms or new methods. The general ImageNet zero-shot models can annotate the high-level labels, like chest or breast, with a high accuracy on this medical benchmark dataset. However, it performs poorly on granular-level labels like malignant, normal, benign and so on. We then train some traditional CNN models on one of dataset by using labels annotated by zero-shot model and fine-tune it to have the best accuracy on its test dataset.**

Introduction/background

**Because there are millions of unlabeled medical images generated in hospitals waiting for annotations before they can be downstream analyzed, but manually labeling them is too time consuming. Also, sometimes, the labels or classes that we want to predict are not in the all pretrained models or some of them are not. At this time, we come across zero-shot, one-shot or few-shot transfer learning.**

Data

**MedMNIST v2: A large-scale lightweight benchmark for 2D and 3D biomedical image classification. This benchmark dataset was collected and processed by 8 researchers from six institutions including universities and hospitals. They have published a paper about this dataset:** [**https://arxiv.org/abs/2110.14795**](https://arxiv.org/abs/2110.14795) **on the arxiv and they have a GitHub repository containing the guides and instruction of how to use this dataset by installing the package they wrote for this dataset, called medmnist. You can download each 2D and 3D respectively by clicking this website:** [**https://zenodo.org/record/5208230**](https://zenodo.org/record/5208230) **and also download all dataset automatically by setting download parameter in the package equal to True. The whole dataset is very easy to access. The researchers now are still actively supporting and developing this dataset and its GitHub Repo. Although dataset is not proposed for clinical use, it has many advantages of research purposes, such as diverse, standardize, lightweight, educational. The diverse dataset provides users with different organ images, different tasks, like binary/multi-class, regression problem.**

Methodology

**My goal is to automatically annotate the millions of unlabeled medical images without too much human involvement or manually labelling. I will use zero-shot clip model and convolutional neural network together to tackle this problem. The whole process involves two procedures. I first split data which we want to annotate into two parts. The first part of data will be annotated by zero-shot clip model choosing the highest probability label among all. Then I will use convolutional neural network to train the first part of data with their predictions from zero-shot clip model as their CNN ground truth labels. Eventually, I will evaluate this trained CNN model on the second part of the data with their truly ground labels. Once I have this best tuned CNN model, I am able to leverage this CNN to annotate new unlabeled medical images. In this way, I can show that using two method together can achieve a higher accuracy on annotation than only using zero-shot clip model.**

Zero-shot clip model

**I will first introduce the state-of-the-art model which connects text and images, clip model. The clip model efficiently learns visual concepts from natural language supervision. Clip model can be applied to any visual classification task by simply giving the visual labels to be recognized. The clip model is trained on a very wide range of natural language supervisions and images on the internet. The general clip model performs better than ImageNet Resnet101 on ImageNet, ImageNet v2, ImageNet Rendition, ObjectNet, ImageNet Sketch, ImageNet Adversarial.**

**Clip is the short name for Contrastive Language Image Pretraining.**

Convolutional Neural Network

Results

Conclusion/Next Steps

**There are some pretrained medical image deep learning models, like in the monai, medical open network for artificial Integillence which provides domain-optimized foundational capabilities for developing healthcare imaging training workflows in native PyTorch language. In the future, I may plan to embed them into current zero-shot, N-shot transfer learning architecture then apply them into MedMNIST image dataset because current N-shot learning is trained on ImgaeNet which is far off medical fields.**

**Once the open AI releases their codes which are used to trained the their zero-shot classification model, I plan to use those codes to train on all medical dataset, like all MedMNIST benchmark dataset and fine-tune it. Then I can have a medical domain adaptation zero-shot models. Then I test and apply them to some other medical dataset to annotate them and evaluate the results. In this case, the results should be much better than the general ImageNet -based zero-shot’s performance.**