Computer Vision and Image Analysis
Course Project Proposal

Tonghe Zhang

Third-year Undergraduate Student From EE class 7
Department of Electronic Engineering, Tsinghua University
zhang-th21@mails.tsinghua.edu.cn

2024-05-12

1 Introduction

This proposal draws the technical blueprint for the computer vision project, which aims to train a neural network to execute image segementation task on a human stenosis dataset.

2 Preliminary Experimental Results

Training We run the simplest algorithm provided with the project description and we keep the hyperparametes unchanged, which indicates that we will train the model for 10 epochs. In each epoch 250 training samples are used. our model is trained on a single AMD A40 GPU. Empirically, running each epoch costs some 83 seconds. The loss curve in the entire training process is presented below:
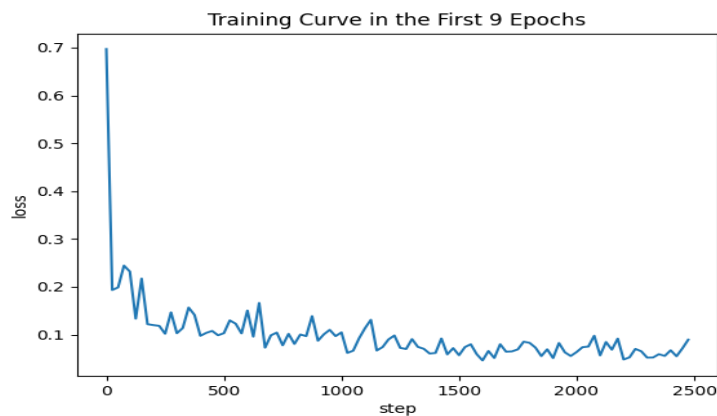

We can see that the loss curve goes down quickly.

Evaluation When evaluating the models saved in different epochs with the validation set, we obtain the following result:
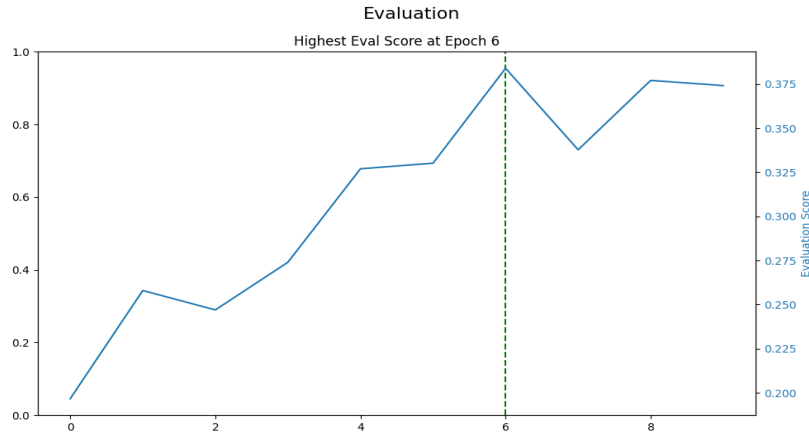
which indicates that the best model in this simplest setting is obtained in epoch 6 and further training only causes the function approximator to overfit the training set.

Testing We test the trained model on the test set and obtain the following result:
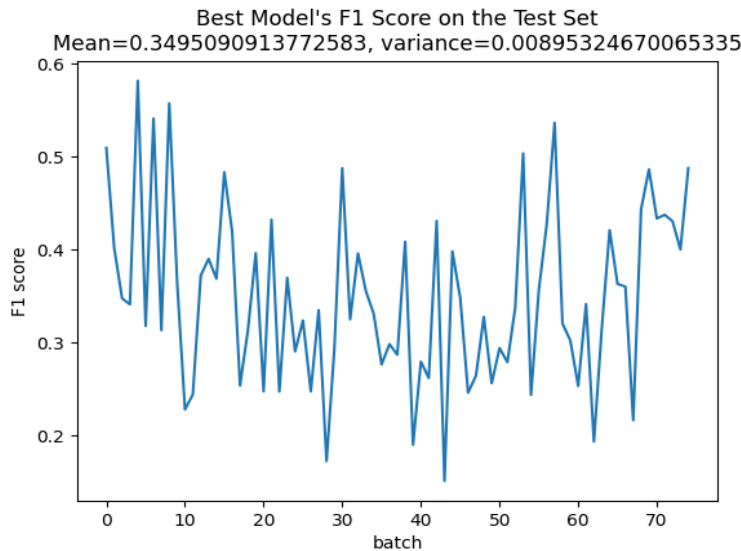
We observe that the generalization ability of the model is quite strong, in that the f1 score on the test set is close to that on the evaluation set. However, the model cannot yield an F1 score higher than 40% on the test set or the evaluation set.

Remaining Problems Although the training and evaluation curves seems natural, the F1 scores on the evaluation set are all below 0.4, which is very low. We cannot trust this model even if it is the best model obtained. We must try to improve the F1 score.

3   Initiative for Improvement

- Better loss function: Week 12-13. We will try to use loss functions other than the Cross Entropy Loss to pay specail attention on an unbalanced dataset. For example we may try to use focal loss functions and compare the results.

- Better representation and architecture: Week 12-13. We plan to draw inspiration from state-of-the-art computer vision models designs to enhance the representation and generalization ability of our model. For

example, we might switch to using Visual Transformer, Detection Transformer and other architectures. We can also try to optimize the design of the U net to adapt to the small-scale dataset.

- Data augmentation: Week 14-15. a. Synthesize more data. We can train a diffusion model on our train set and generate new data for training to increase the volume and diversity of the data set, in the hope that we can improve the generalizability. b. Enhance the quality of the data. We also play to adopt classical image augmentation techniques, like increasing the contrast and rotating the image to discourage overfitting.