## Assignment 4

1. Data overview

Divide dataset into test set, training set, validation set

Test set:

Artifacts: 100 images

Cancer regions: 100 images

Normal regions: 100 images

Other: 100 images

Training set:

Artifacts: 1824 images

Cancer regions: 2223 images

Normal regions: 308 images

Other: 487 images

Validation set:

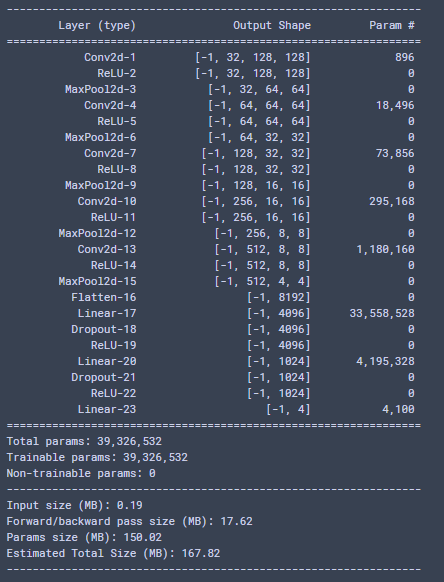
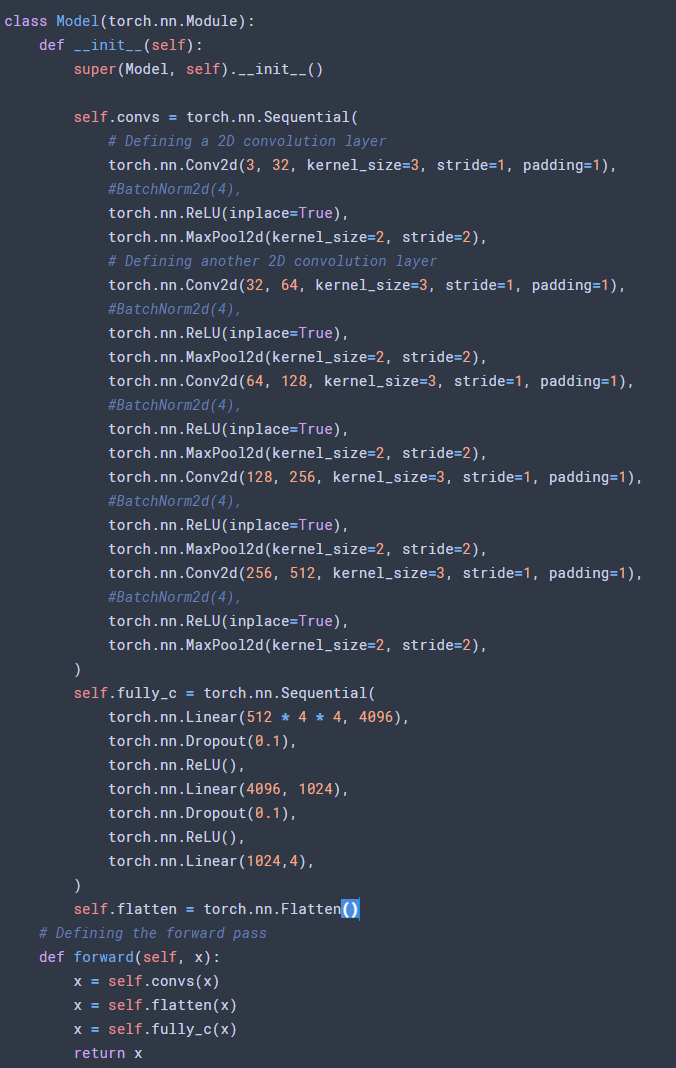
Artifacts: 456 images

Cancer regions: 555 images

Normal regions: 217 images

Other: 76 images

1. Deep learning model



The input shape is (3, 128, 128), which is 3 channel, 128 pixel by 128 pixel

After the convolutional layers to extract features, the feature map shape is (512, 4, 4). 512 feature maps, 4 pixel by 4 pixel.

Then flatten the features to be a 512\*4\*4 dimensional vector.

Put the high dimensional vector into fully connected layers, and output the (4, 1) result representing the 4 class.

Hyperparameters:

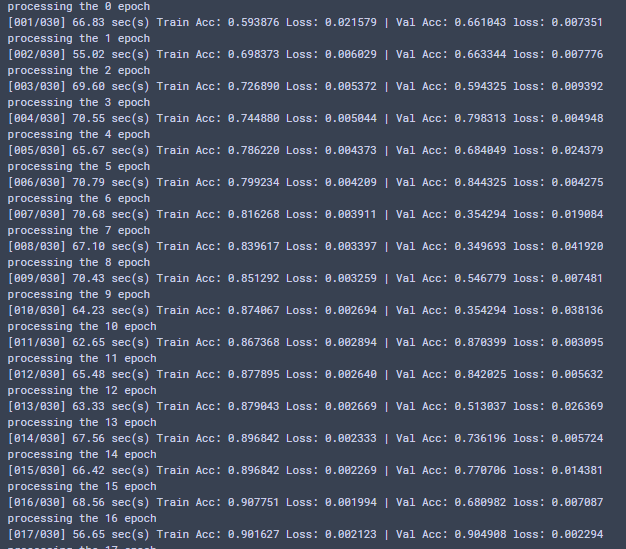
Optimizer: Adam optimizer

Learning rate: 0.01

Learning rate decay: none

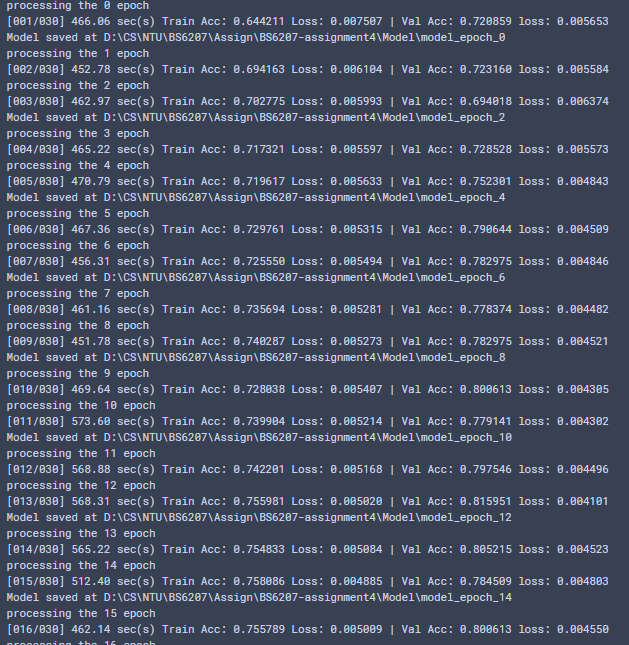
Dropout: none

Training result on all training set(4000+):



After 17 epoches of training, the training acc reach 0.90, and the validation acc reach 0.90

Training result on 400 images:



After 15 epoches of training, the training acc reach 0.80, and the validation acc reach 0.80

But the test acc is very low



Training result on 400 images:



Result:

Training on 40 images: acc 0.57

Training on 400 images: acc 0.78

Training on 4000+ images: acc 0.85