Theory and Practices of Deep Learning Project

Chang Jun Qing 1002088, Shaun Toh Shou'en 1002012

March 2019

1 Introduction

The question posed by pascal VOC2012 classification challenge is that of multi-label classification. In order to compute loss, we use the Sigmoid function instead of the Softmax function as each class should have its probability of being present computed separately, as opposed to the Softmax function that computes the probability of the class being present in the image weighed against the probabilities of all other classes being present in the image. This results in the the final choice of binary cross entropy loss.

1.1 Chosen Metrics

Instead of using accuracy, we used the suggested tail accuracy, validation loss and mean average precision to select the best model. This is because accuracy is susceptible to imbalances in labels and fails to indicate if too many false positives or negatives are being produced by the model. As an example, if the output vector in a multilabel problem is a vector of (1,500), and most of the testing, training and validation samples only have a maximum of 2 or 3 classes present in a single input, the model will always be able to achieve a high accuracy simply by outputting zeroes for all spaces in the vector, failing to indicate that the model is simply failing to predict anything relevant.

2 Model Description

2.1 Model

Our first model is a pretrained resnet 18 model is used. The second model is a modified pretrained resnet 18 model. The average pooling layer has been changed to an adaptive pooling layer, allowing it to take in images of different size.

2.2 Hyper parameters

The following hyper parameters are used.

- For our training, we used a Learning rate of 0.01 and momentum of 0.1.
- We also set a learning rate decay of 0.1 every 10 epochs, causing the learning rate to reduce by 90% every 10 epochs.
- Our thresholds are recorded at intervals of 0.05 from 0 to 0.7.
- The inputs are normalised across the recommended pytorch mean of [0.485,0.456,0.406] and standard deviation of 0.229, 0.224, 0.225.

2.3 Data Augmentation

The following data augmentation was applied for each neural net

- During training, the unmodified resnet 18 model has its input resized to 280 before going through a random rotation up to twenty degrees. Photometric distortion with brightness, saturation, and contrast up to a factor of 2, with a hue factor of up to 0.5 is set on the image. It then undergoes a five crop process of size 224 and normalisation before being fed into the neural net. The output arrays are then averaged, and the final output placed into a sigmoid function to determine the individual probabilities for each class.
- During testing, the unmodified resnet 18 model has input images resized to 280 before five cropping them to a size of 224, and normalisation. The resultant output is again averaged across all 5 input images before going through a sigmoid function to determine probabilities.
- During training, the modified resnet 18 has its input resized to 350, and it undergoes random rotation up to a value of 20 degrees, a five crop of random size from 224 to 330. Afterwards, each individual crop undergoes photometric distortion with brightness, saturation, and contrast up to a factor of 2, with a hue factor of up to 0.5 is set on the image. Random Horizontal and vertical flips with probability 0.5 are also applied before normalisation.
- During testing, the modified resnet 18 under goes the same data augmentation procedure as that of the unmodified resnet 18 model.

2.4 Results

2.4.1 Mean Average Precision

Flip Model mean average precision: 0.0703854524832117 Color Model mean average precision: 0.7158235941950959

2.4.2 Unmodified resnet 18

The general result is that training and validation loss drop sharply after the first few epochs.
For the unmodified resnet, The tail accuracy across the top 50 images for each of the classes is displayed in the appendix. All classes are highly accurate except 10 and 16, which only achieve maximum tail accuracy at a very high threshold (close to 1), and 11,17, and 18, which suffer from drops in tail accuracy at high threshold. (although 17 only suffers a minor loss in accuracy from 0.98 to 0.96, and fails to achieve 100% accuracy.
The relevant graphs can be found in the appendices, and the precisions of the model with the best performance is listed below.

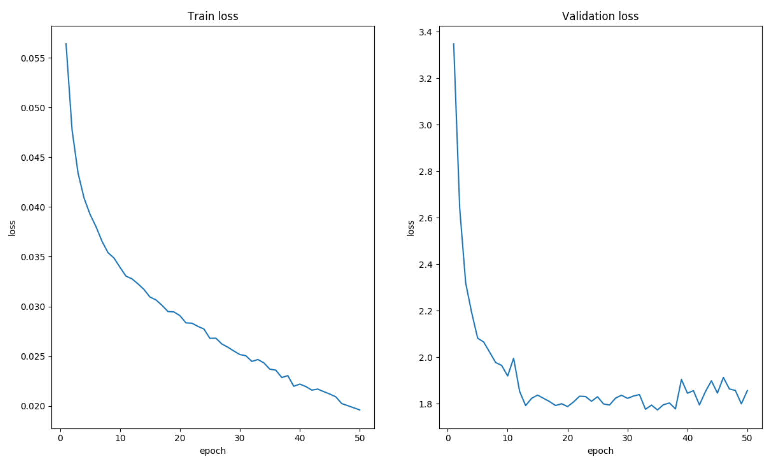
- Class 0 Precision: 0.8368434440748926
- Class 1 Precision: 0.7970547891504961
- Class 2 Precision: 0.8041453282255399
- Class 3 Precision: 0.6087456834362974
- Class 4 Precision: 0.5482012797737804
- Class 5 Precision: 0.8476616266014083
- Class 6 Precision: 0.8247062679451339
- Class 7 Precision: 0.86486952158056
- Class 8 Precision: 0.7019448783770822
- Class 9 Precision: 0.6563594321713416
- Class 10 Precision: 0.5302081263018066
- Class 11 Precision: 0.7989165626424751
- Class 12 Precision: 0.7744321635544146
- Class 13 Precision: 0.7735043303951861
- Class 14 Precision: 0.8569503268802684
- Class 15 Precision: 0.4636866144592823
- Class 16 Precision: 0.5313918977194275
- Class 17 Precision: 0.5568057414046912
- Class 18 Precision: 0.7907201038992737
- Class 19 Precision: 0.7493237653085617
- Model precision: 0.7158235941950959

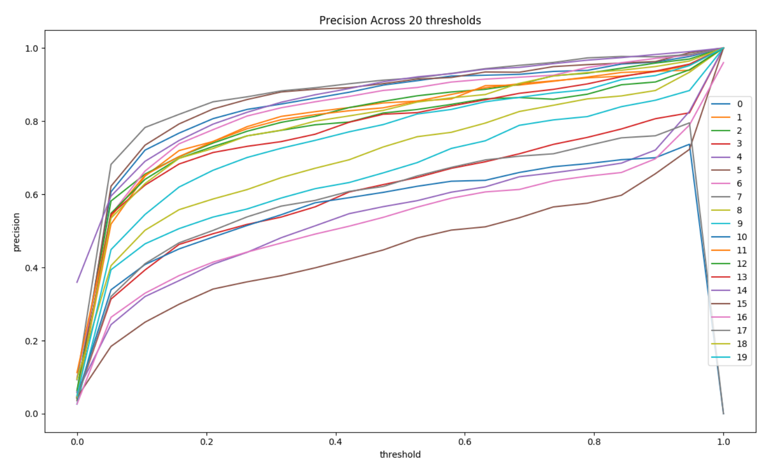
2.4.3 Modified resnet 18

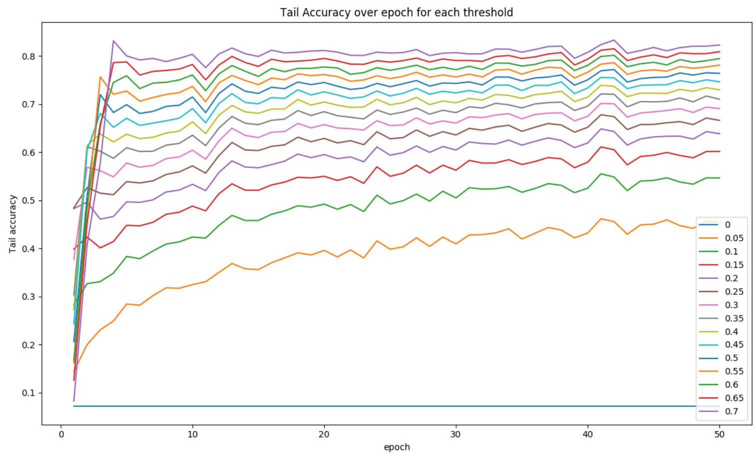
The modified resnet 18 does not perform as well as it's unmodified counterpart. There are several possible reasons for this. The first is that more epochs might be required before proper convergence due to the large number of transformations that are being applied on the dataset. The next is that it is possible that vertical flip might be causing additional difficulty in the model learning the proper weights.

- Class 0 Precision: 0.7655817008004314
- Class 1 Precision: 0.6221535815152912
- Class 2 Precision: 0.8194227857107667
- Class 3 Precision: 0.6344372379082684
- Class 4 Precision: 0.4980564921427198
- Class 5 Precision: 0.7001492597973047
- Class 6 Precision: 0.6294007237447092
- Class 7 Precision: 0.823389444462463
- Class 8 Precision: 0.5905959626927356
- Class 9 Precision: 0.35600781891928446
- Class 10 Precision: 0.44376767927207855
- Class 11 Precision: 0.7359101669021507
- Class 12 Precision: 0.5534534384710782
- Class 13 Precision: 0.5676119710057084
- Class 14 Precision: 0.8530710343071736
- Class 15 Precision: 0.3827272537909598
- Class 16 Precision: 0.451392749857119
- Class 17 Precision: 0.4742520903533869
- Class 18 Precision: 0.742906086018349
- Class 19 Precision: 0.6508914068791075
- Model precision: 0.6147589442275543

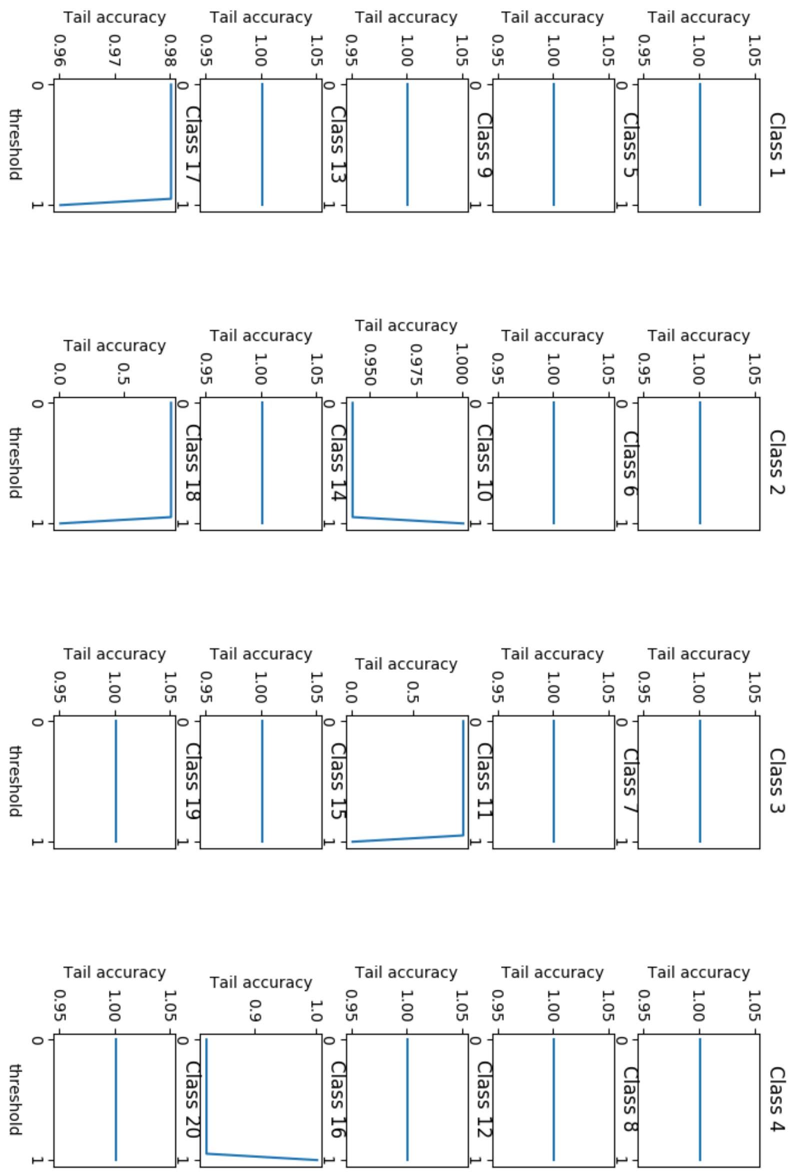
A Ummodified Resnet 18 graphs



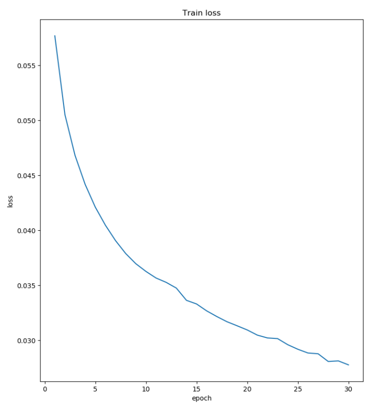


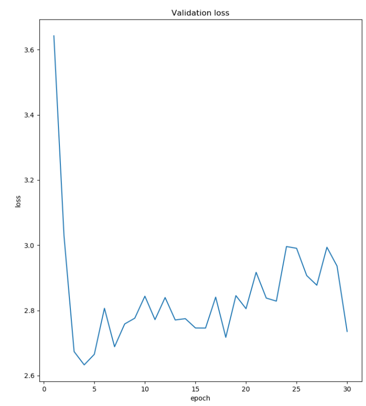


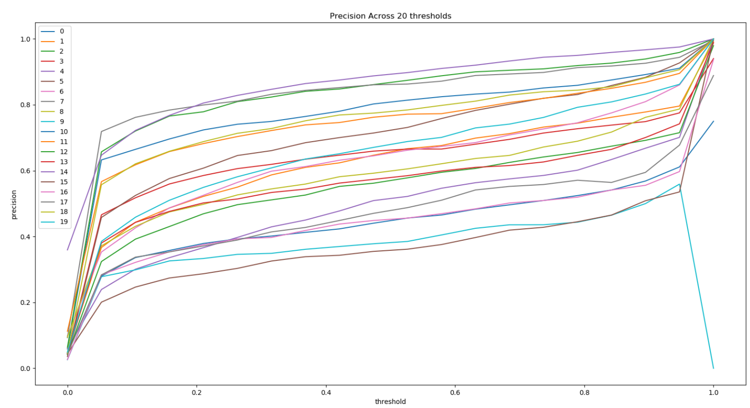
Tail Accuracy over epoch for each threshold
Tail accuracy
epoch
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
10
20
30
40
50
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7

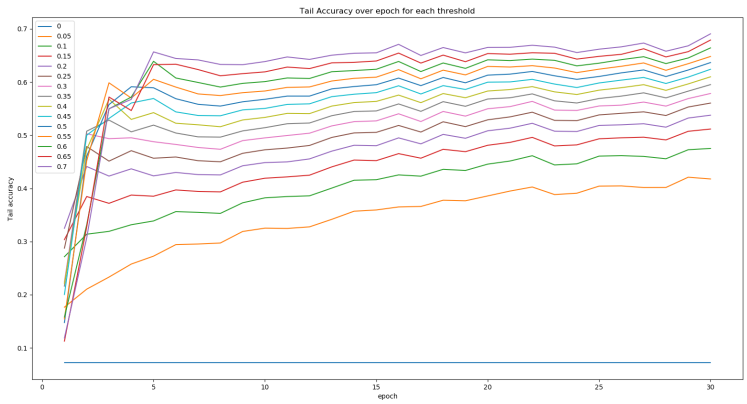

Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
Class 8
Class 9
Class 10
Class 11
Class 12
Class 13
Class 14
Class 15
Class 16
Class 17
Class 18
Class 19
Class 20
Tail accuracy
threshold

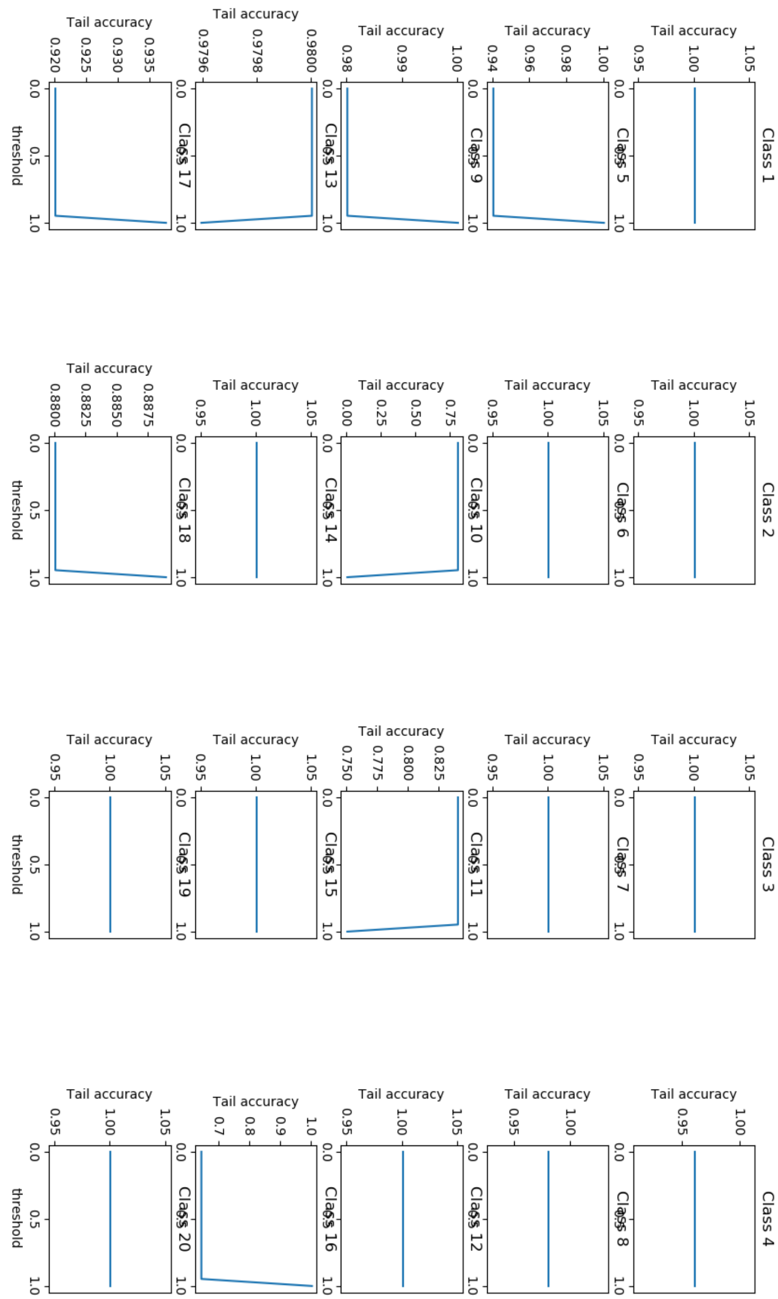
B Modified Resnet 18 graphs








Tail Accuracy over epoch for each threshold
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
Tail accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0.7
epoch
0
5
10
15
20
25
30


Tail accuracy
threshold
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
Class 8
Class 9
Class 10
Class 11
Class 12
Class 13
Class 14
Class 15
Class 16
Class 17
Class 18
Class 19
Class 20