Multi-Head Classification

Programming Assignment 2-2 : Report

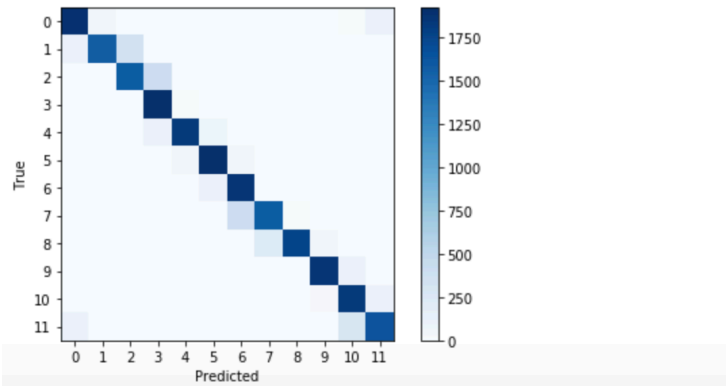
Submitted by : Group 3

Experiment 1 : Feature Extractor : 2 convolutional layers, first layer has 16 filters of size 5x5, max pooling is done to downsample image from 28x28x3 to 14x14x3, next layer has 36 filters of size 5x5,resulting images are again downsampled to 7x7x3. This is followed by flattening. The 4 heads are as specified in question. Loss of all heads are given equal weightage before training. Adam optimiser is used.

Accuracies (Len,Wid,Col,Ang)	F scores (Len,Wid,Col,Ang)
95.5,98.5,98.6,88.8	95.25,98.48,98.6,96.02

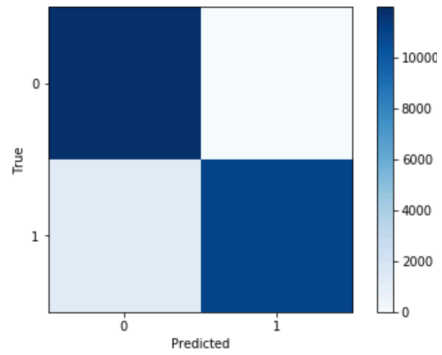
Figures :

```
[[1887   26    2    0    0    0    1    0    0    0    9  102]
 [ 105 1581  367    0    0    0    0    0    0    0    0    2]
 [   0    7 1591  408    0    0    0    0    0    0    0    0]
 [   0    0    6 1924   14    0    0    0    0    0    0    0]
 [   0    0    0  110 1839   74    0    0    0    0    0    0]
 [   0    0    0    0   33 1911   41    0    0    0    0    0]
 [   0    0    0    0    3   92 1865    6    0    0    1    0]
 [   0    0    0    0    0    0  395 1587   10    3    0    0]
 [   0    0    0    0    0    0    0  236 1772   27    0    0]
 [   0    0    0    0    0    0    0    0    5 1869  100    0]
 [   2    0    0    0    0    0    0    0    0   22 1844   91]
 [ 117    0    0    0    0    0    0    0    0    0  277 1636]]
```

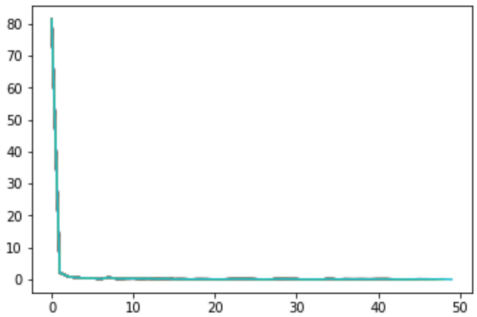


Confusion matrix for angle

```
[[11965     0]
 [ 1092 10943]]
```



Confusion matrix for angle



Loss curve

Experiment 2 : Feature extractor architecture same as experiment 1, the 4 heads are as specified in question. Loss of all heads are given equal weightage except angle head which is multiplied by 1.2 before training. Adam optimiser is used.

Accuracies (Len,Wid,Col,Ang)	F scores (Len,Wid,Col,Ang)
94.4,96.9,99.3,92.7	

Inferences :

1. As we can see from the confusion matrices, mostly errors occur for closer classes. For eg in first experiment, errors are mostly 0 for farther classes whereas for nearby classes the errors exist.
2. Giving weightage to the different losses makes significant differences in the learning, which is intuitive. For eg, in first experiment the accuracy for angle is a bit low, so in second experiment we multiply its loss by 1.2, which helps us get better accuracy for angle whilst the accuracies for others are a bit reduced though all are above 90.