Visual Learning and Recognition (16-824)
Homework 1

Jennifer Isaza, jennifei

February 27, 2019

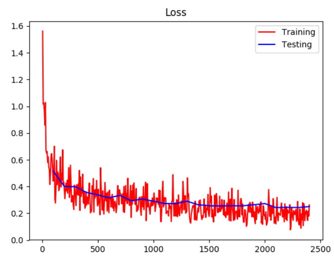
Collaborated with: Kevin Christensen Hunter Goforth

Task 0: Fashion MNIST classification in TensorFlow (5 points)

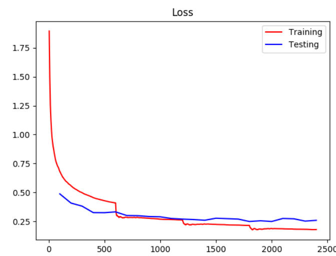
Q 0.1: Both scripts use the same neural network model, how many trainable parameters does each layer have?

First conv2d layer: 832
Max pooling: 0
Seconds conv2d layer: 51264
Max pooling: 0
Flatten: 0
Dense layer: 3212288
Dropout: 0
Dense layer: 10250
Total parameters: 3,274,634

Q 0.2: Show the loss and accuracy curves for both scripts with the default hyperparameters.

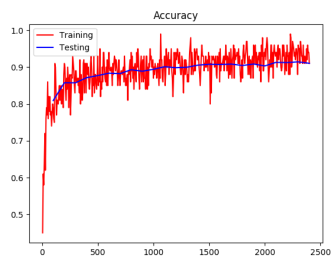


(a) 00 fashion mnist

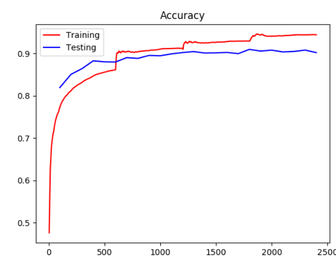


(b) 01 fashion mnist

Figure 1: Loss Curves



(a) 00 fashion mnist



(b) 01 fashion mnist

Figure 2: Accuracy Curves

Q 0.3: Why do the plots from two scripts look different? Why does the second script show smoother loss? Why are there three jumps in the training curves?

The second script uses eager execution instead of the graphical tensorflow method, which is implemented for the first script. Eager execution uses the metric Mean and Accuracy functions for the loss and accuracy results, respectively. Eager execution uses a running average over the batch norm, that the normal graphical tensorflow model does not use. The three jumps in the training curves occur after training each epoch. There were four epochs by default, therefore, three jumps.

Q 0.4: What happens if you train the network for 10 epochs?

When training the network for 10 epochs, there are nine jumps in the training curves. Both scripts have an increased loss and decreased accuracy when testing, most likely, due to overfitting to the training data.

Task 1: Simple CNN network for PASCAL multi-label classification (20 points)

(a) SimpleCNN Test mAP

(b) SimpleCNN Training Loss

Figure 3: SimpleCNN Results

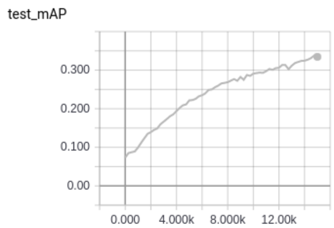
Task 2: Lets go deeper! CaffeNet for PASCAL classification (20 points)

(a) CaffeNet Test mAP

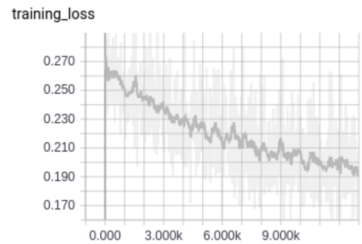
(b) CaffeNet Training Loss

Figure 4: CaffeNet Results

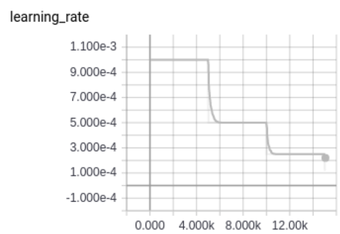
Task 3: Even deeper! VGG-16 for PASCAL classification (15 points)



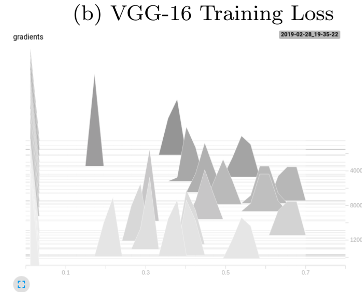
(a) VGG-16 Test mAP



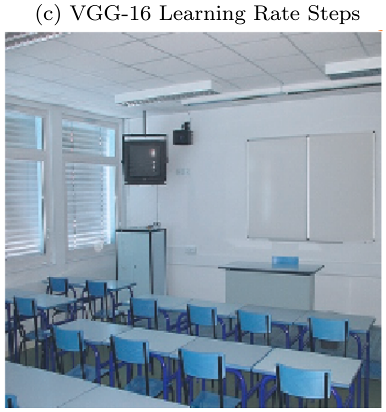
(b) VGG-16 Training Loss



(c) VGG-16 Learning Rate Steps

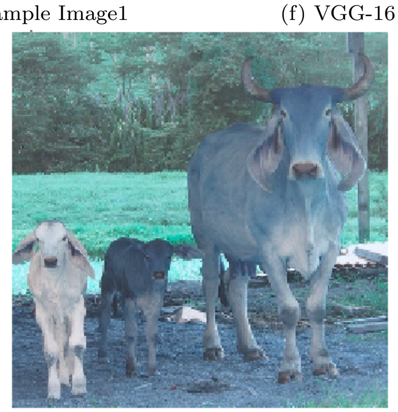


(d) VGG-16 Histogram of Gradients



(e) VGG-16 Example Image1

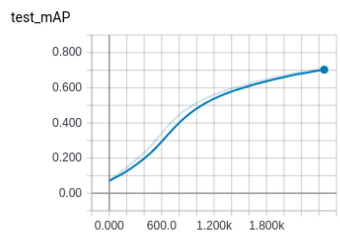
(f) VGG-16 Example Image2



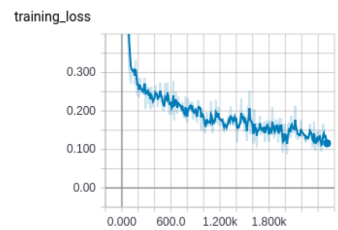
(g) VGG-16 Example Image3

Figure 5: VGG-16 Scratch Results

Task 4: Standing on the shoulder of the giants: finetuning from ImageNet (20 points)



(a) VGG-16 Finetune Test mAP

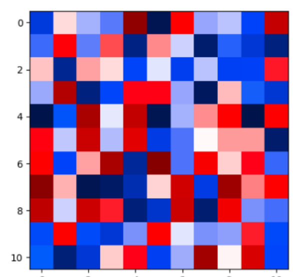


(b) VGG-16 Finetune Training Loss

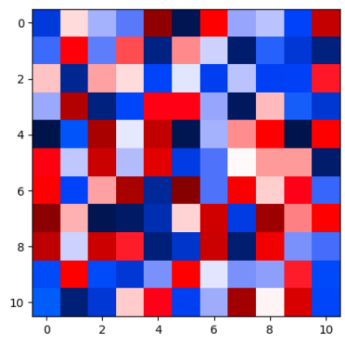
Figure 6: VGG-16 Finetune Results

Task 5: Analysis (20 points)

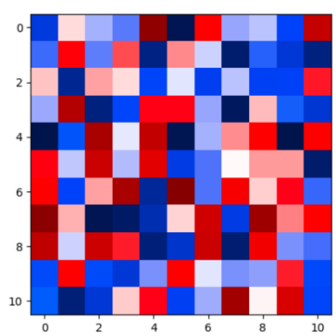
Task 5.1:



(a) Conv1 Filter Checkpoint1



(b) Conv1 Filter Checkpoint2



(c) Conv1 Filter Checkpoint3

Figure 7: Conv1 Filters from CaffeNet

Task 5.2:

Task 5.3:

Task 5.4:

Finetune:
aeroplane: 0.8561969742497731
bicycle: 0.8481625827564562
bird: 0.8760629404472989
boat: 0.8351535389059069
bottle: 0.3741916639155627
bus: 0.7396670315762395
car: 0.9030402256662998
cat: 0.8430686339608217
chair: 0.5972241315416849
cow: 0.4364611496006592
diningtable: 0.5748649248991563
dog: 0.7775302305966582
horse: 0.7964708455087699
motorbike: 0.8352819696573272
person: 0.9525325342763798
pottedplant: 0.3567856180575276
sheep: 0.572401367657898
sofa: 0.6206713930574237
train: 0.8844496444130966
tvmonitor: 0.5193042108561509

Cafe:
aeroplane: 0.024902865987539928
bicycle: 0.03947270686539033
bird: 0.05487464784384127
boat: 0.1100555884228687
bottle: 0.040772488204883 26
bus: 0.039258055496987035
car: 0.12674227906043106
cat: 0.0747620644690308
chair: 0.09517480641581859
cow: 0.03435047345791343
diningtable: 0.03315857133618677
dog: 0.09448038524887968
horse: 0.052544656254295025
motorbike: 0.06475243490305531
person: 0.4156627243279519
pottedplant: 0.05625860110303207
sheep: 0.01607770243932006
sofa: 0.053969301224213014

train: 0.04434139506208257
tvmonitor: 0.050002820089177086

Most of the objects that have a better accuracy score have more complex variations of how the object looks, is situated, what backgrounds it is in, etc. For example, cars, which had the highest accuracy, are shown in multiple different zoomed-in angles, far away, with different backgrounds, etc, while cows are generally only shown in fields at a medium distance away.