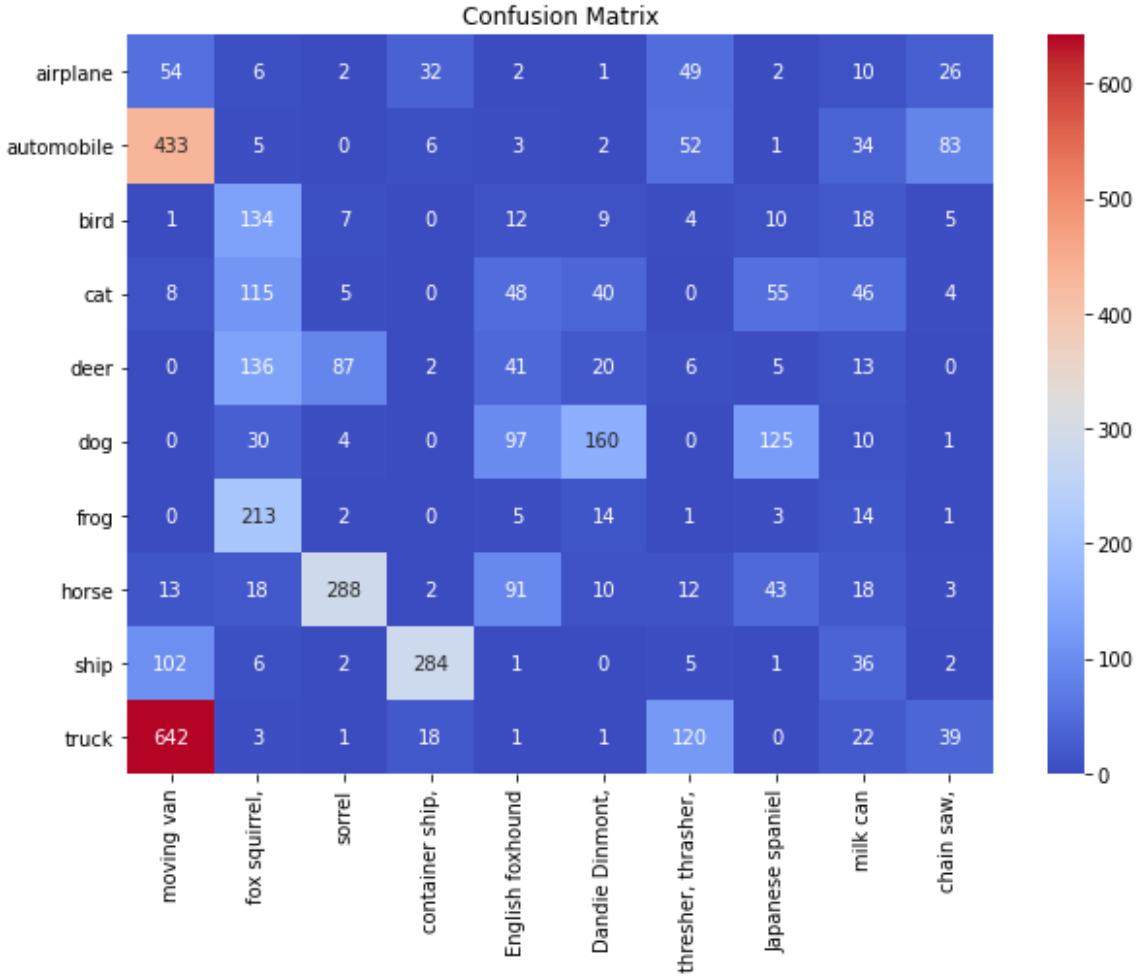
B) Classify Images

Top 10 Images in CIPHAR according to Pretrained Alexnet

```
{'moving van': 1253,
 'fox squirrel, eastern fox squirrel, Sciurus niger': 666,
 'sorrel': 398,
 'container ship, containership, container vessel': 344,
 'English foxhound': 301,
 'Dandie Dinmont, Dandie Dinmont terrier': 257,
 'thresher, thrasher, threshing machine': 249,
 'Japanese spaniel': 245,
 'milk can': 221,
 'chain saw, chainsaw': 164}
```



C) fc6 layer

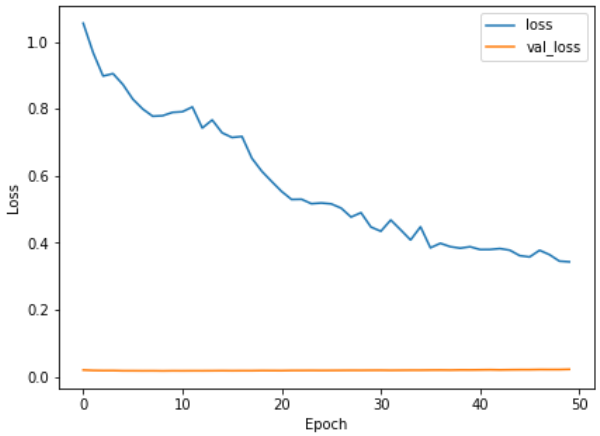
Model

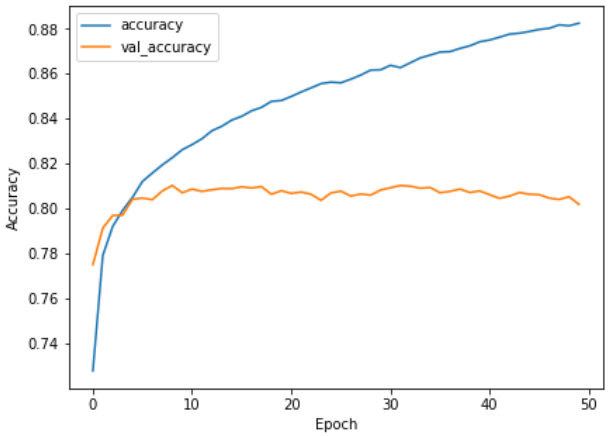
```
AlexNet(
  (features): Sequential(
    (0): Conv2d(3, 64, kernel_size=(11, 11), stride=(4, 4), padding=(2, 2))
    (1): ReLU(inplace=True)
    (2): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)
    (3): Conv2d(64, 192, kernel_size=(5, 5), stride=(1, 1), padding=(2, 2))
    (4): ReLU(inplace=True)
    (5): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)
    (6): Conv2d(192, 384, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (7): ReLU(inplace=True)
    (8): Conv2d(384, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (9): ReLU(inplace=True)
    (10): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (11): ReLU(inplace=True)
    (12): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (avgpool): AdaptiveAvgPool2d(output_size=(6, 6))
  (classifier): Sequential(
    (0): Dropout(p=0.5, inplace=False)
    (1): Linear(in_features=9216, out_features=4096, bias=True)
  )
)
```

Linear Model

```
Linear(
  (lin1): Linear(in_features=4096, out_features=64, bias=True)
  (lin2): Linear(in_features=64, out_features=10, bias=True)
)
```

Training






```
Test Dataset Loss: 0.01806300396323204, Accuracy: 0.7979
```

D) fc7 layer

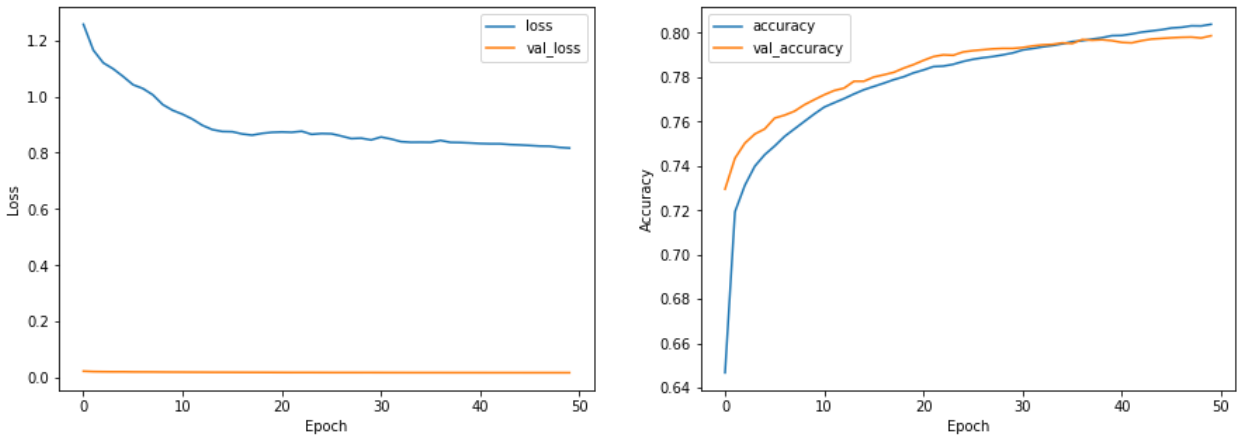
Model

```
AlexNet(
  (features): Sequential(
    (0): Conv2d(3, 64, kernel_size=(11, 11), stride=(4, 4), padding=(2, 2))
    (1): ReLU(inplace=True)
    (2): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)
    (3): Conv2d(64, 192, kernel_size=(5, 5), stride=(1, 1), padding=(2, 2))
    (4): ReLU(inplace=True)
    (5): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)
    (6): Conv2d(192, 384, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (7): ReLU(inplace=True)
    (8): Conv2d(384, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (9): ReLU(inplace=True)
    (10): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (11): ReLU(inplace=True)
    (12): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (avgpool): AdaptiveAvgPool2d(output_size=(6, 6))
  (classifier): Sequential(
    (0): Dropout(p=0.5, inplace=False)
    (1): Linear(in_features=9216, out_features=4096, bias=True)
    (2): ReLU(inplace=True)
    (3): Dropout(p=0.5, inplace=False)
    (4): Linear(in_features=4096, out_features=4096, bias=True)
  )
)
```

Linear Model

```
Linear(
  (lin1): Linear(in_features=4096, out_features=64, bias=True)
  (lin2): Linear(in_features=64, out_features=10, bias=True)
)
```

Training




```
Test Dataset Loss: 0.018707045084238053, Accuracy: 0.7926
```

Compare and Discuss

Training Accuracy for fc6 layer shows that there is higher training accuracy than fc7 layer. However, fc6 loss does not converge as smoothly in terms of loss as fc7 when training on 50 epochs. We can see that both accuracies are similar despite fc6's training accuracy being larger due to the model overfitting on the training data. This is evidenced on the large difference indicated between val_accuracy and accuracy in the accuracy graph for fc6.