DD2424 - Assignment 3 Bonus

Timo Nicolai

April 29, 2019

I attempted to incorporate the following four techniques: increasing the depth of the network, increasing the size of its hidden layers, applying dropout and augmenting the dataset by mirroring all training images. I will describe each of these in a separate section below.

1 Increasing Depth

As a simple sanity check I tried appending a single additional hidden layer to the best three layer network found in the non-bonus part of this assignment. Figure 1 shows the resulting validation set accuracies achieved by varying the size of this addtional hidden layer (hyperparameters are as before with the same λ found via grid search and batch normalization enabled, I trained for two cycles in each case).

Hidden Node in 3rd Hidden Layer	Accuracy
30	0.5498
40	0.5504
50	0.5528
60	0.5446
70	0.5464

Figure 1: Validation set accuracy achieved by appending a third hidden layer of varying size.

Unfortunately, adding another hidden layer does not seem to improve validation set accuracy, it is possible that training a deeper network with good performance would require larger hidden layers in combination with more elaborate regularization. I did however not have the hardware resources to perform more exhaustive tests here so I decided to stick with a three layer network.

2 Larger Hidden Layers

Figure 2 shows the accuracies achieved by training a three layer network with larger hidden layers. Increasing the size of the hidden layers seems to significantly improve validation set accuracy, but this effect strongly diminishes after

around 300 nodes in each hidden layer. To retain a reasonable training time I therefore decided to use 300 node in both hidden layers from here on.

Hidden Layer Nodes	Accuracy
100, 100	0.5614
200, 200	0.5846
300, 300	0.604
400, 400	0.6016
500, 500	0.6048

Figure 2: Validation set accuracies achieved by increasing the size of the hidden layers.

3 Dropout

Applying inverted dropout with a dropout probability of 50% during training dropped validation set accuracy down to less than 0.5, so I did not use it when training the final network.

4 Data Augmentation

As a data augmentation measure I doubled the size of the training set by mirroring each image horizontally. To actually make use of the enlarged dataset I also trained for more cycles (six to be exact). Figure 3 shows the training curves from this final training run and Figure 4 shows the test set confusion matrix for the trained network.

Unfortunately it seems like neither training set augmentation nor longer training time significantly improve the networks performance. The final network does however perform slightly better than the baseline network before attempting any optimizations. The only real improvement was due to the increase in hidden layer sizes.

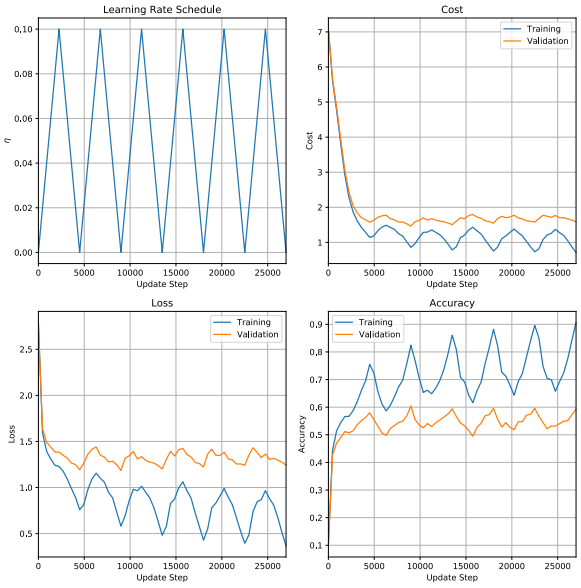

Figure 3: Training curves for the final network trained on augmented data.

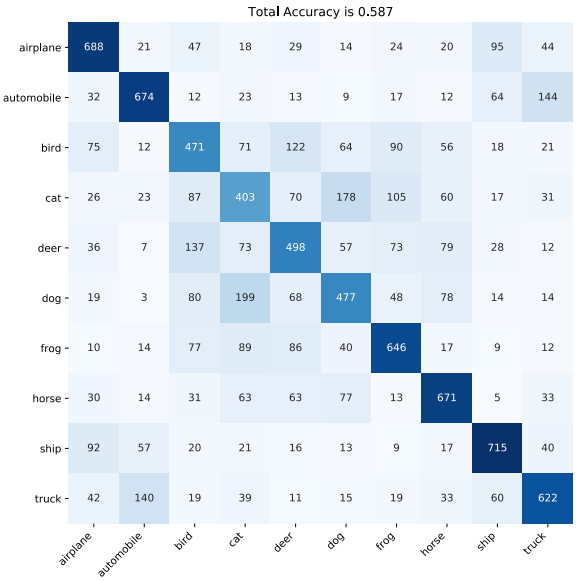

Figure 4: Test set confusion matrix for the final network trained on augmented data.