CS 498 AML - Homework 9
Graham D Chester, 04-May-2018

MNIST Tutorial

TensorFlow 1.7 CPU version was installed on an iMac, and the MNIST TensorBoard tutorial downloaded as required. This tutorial code was then modified to use a batch size of 100 and to train for 10,000 batches. The results for the tutorial are shown in the TensorBoard screen print below, with an accuracy of 98.2% (orange line). The best improved model below has an accuracy of 99.45% (grey line).

MNIST Improvement

Various modifications were then applied to the tutorial code in order to improve accuracy, as follows:

Model augm: The images were augmented using various combinations of parameters. The optimal were found to be rotation 6 degrees, width shift 0.06, sheer range 0.27, height shift 0.06, zoom range 0.06. The model architecture was unchanged, with 2 hidden layers. This improved accuracy to just under 99% (green line).

Model augm_bn: Batch normalization was added to the model "*augm*" resulting in a tiny improvement in accuracy to just over 99% (light-blue line).

Model augm_bn_norm: Image normalization was added to model "*augm_bn*" resulting in a (surprising) small reduction in accuracy back to just under 99% (dark blue line).

Model augm_conv1: Model "*augm*" was modified substantially to add convolutional layers as conv(3,3,32), conv(3,3,64), max pool(2,2), conv(3,3,64), conv(3,3,64), max pool(2,2) prior to the original 2 hidden layers with dropout. This resulted in a significant improvement in accuracy to 99.4% (pink line).

Model augm_bn_conv: Batch normalization was added to Model "augm_conv1" and resulted in a tiny but consistent improvement in accuracy to 99.45% (grey line). This was the best model.

MNIST TensorBoard Test Set Accuracy
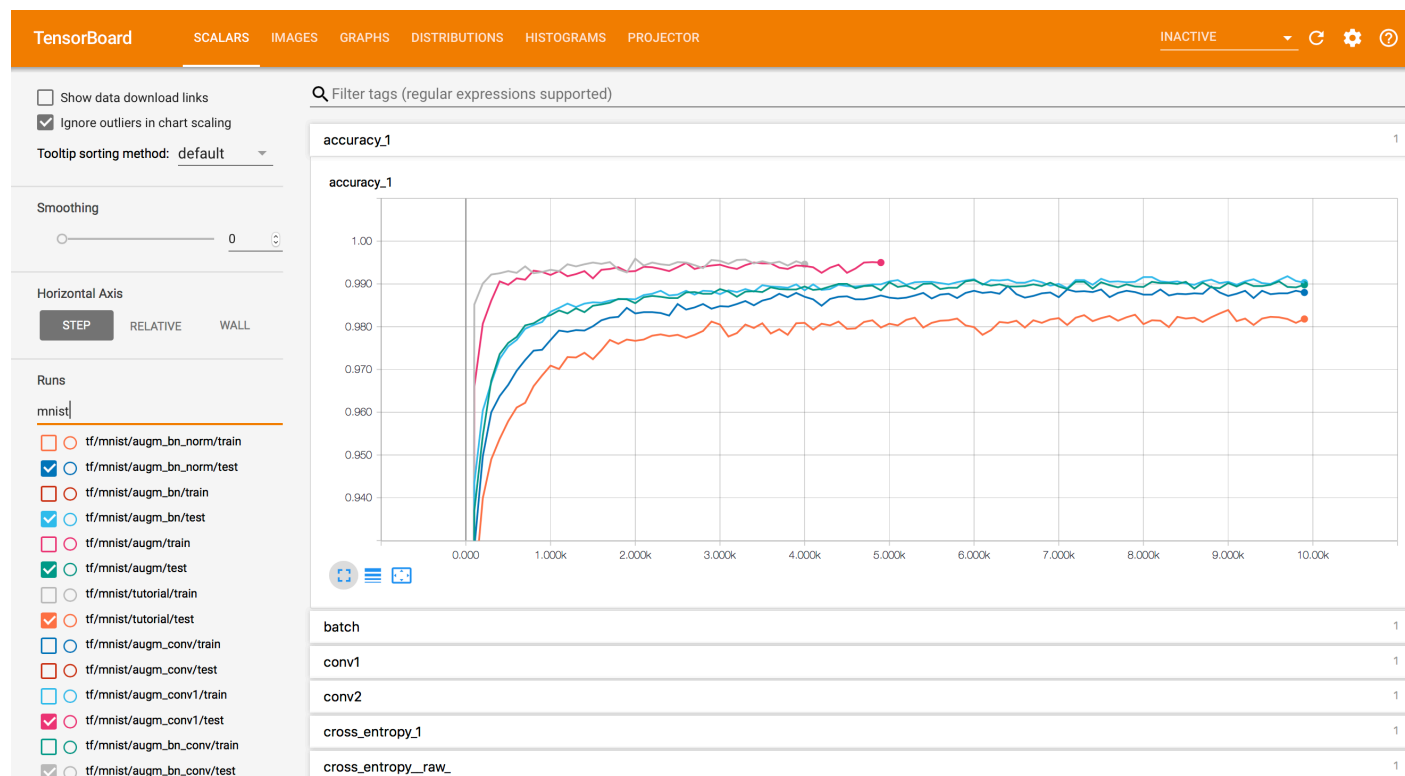
There were many other modifications tried that are not described above, such as replacing the Adam optimizer with Momentum or SGD, reducing and increasing the batch size, changing keep probability etc., however they all resulted in either no improvement or worsening of accuracy.

Additionally the early models were susceptible to a catastrophic 'collapse' in accuracy down to 10% after several thousand training iterations – this was traced to 0log(0) in the cross-entropy calculation and adding a tiny number (1e-10) to this resolved the issue with no impact on training or accuracy.

CIFAR-10 Tutorial

The TensorBoard tutorial was downloaded and the code run as a test on an iMac as per MNIST above. The training time was found to be extremely long, so an AWS p2.xlarge GPU machine was used with Amazon Deep Learning Image, and TensorFlow was updated to V1.8 to enable the use of save checkpoint steps on the training session.

The tutorial was then modified to save the model and report the accuracy on the testing set every 100 batches as required. The requirement was to train for at least 2000 batches so 15,000 batches was chosen as a balance between training time and best accuracy obtainable from the models. The original tutorial code was modified for each model below rather than replicating code, hence only one code version is supplied with comments for each modification.

The Tutorial result is shown as the *"eval"* model in the TensorBoard screen shot below (dark blue line), with an accuracy of 83.06%, whereas the best improved model below has an accuracy of 84.61% (red line).

CIFAR-10 Improvement

Various modifications were then applied to the tutorial code in order to improve accuracy, as follows:

Model eval1: Batch size changed to 256 which improved accuracy to 83.90% (light-blue line).

Model eval2: Model *"eval1"* was updated with neurons/units in first hidden layer changed from 384 to 512 which improved accuracy slightly to 84.06% (orange line).

Model eval3: Model *"eval2"* was updated with batch size of 512 which improved accuracy slightly to 84.32% (red line).

Model eval4: Model *"eval3"* was updated to use the Adam Optimizer which resulted in a faster initial accuracy curve, but the final accuracy of 82.53% was slightly worse than the tutorial *"eval"* model (green line).

Model eval5: Model "eval3" was updated to reduce image random contrast and brightness adjustments which resulted in an accuracy of 84.61% (pink line). This was the best model.

There were many other modifications tried such as removing image standardization (very poor accuracy), using a batch size of 1024 (slow training and no accuracy improvement), adding normal noise to the images (worse accuracy for high standard deviations, and no improvement for low standard deviations), removing the random crop (slightly worse accuracy), and trying Momentum/Adadelta optimizers (slow training and worse accuracy).

Additionally a model developed for MNIST above was also modified to work on CIFAR-10, but with a poor accuracy of around 75%. To avoid clutter, this is not shown on TensorBoard below, however the code is included in the zip file.

CIFAR-10 TensorBoard Test Set Accuracy