Training and Evaluation of Deep Neural Networks

Hwa Chong Institution
Andrew Ke, Chu Yuan

Background

Deep learning is able to process vasts amounts of data swiftly and accurately and is invaluable in many industrial applications, from enabling autonomous vehicles to navigate safely, to distilling important insights from image collections, and even climate predictions

We aim to maximise the impact of deep learning and its applications by optimising its backpropogation training process through fine-tuning of hyperparameters in order to obtain high accuracy and low convergence speeds.

Methodology

Training of the neural network Googlenet was conducted using Caffe and ImageData layer. Hyperparameters are adjusted before each training session. Resultant accuracy and loss data are then plotted and smoothed with a Savitzky-Golay filter.

Kaggle's Dogs vs Cats dataset (25000 images) was chosen as a simple classification problem to compare peak accuracy and loss and spot overfitting.

MIT's Places 205 dataset was used to compare learning rate decay schedulers.

2 classes from MIT's Indoor Scene Recognition database (1150 images) were used for neural network fine tuning.

Conclusion

Based on the Cats vs Dogs classification problem, Googlenet's default values of 32 (batch size) and 0.01 (base learning rate) appear to be ideal for training, striking a balance between weight update stability and nimbleness of optimisation algorithm.

Step size however must be tweaked based on the complexity of the classification problem. In simpler problems, lower values may result in rapidly diminishing learning rates that stall training.

In the case of momentum, a value of 0.68 (default 0.9) appears to be ideal.

For weight decay, L2 regularisation is effective at reducing overfitting. Googlenet's default value of 2E-4 reduces overfitting slightly without compromising peak accuracy. Larger values of weight decay can drastically slow convergence or lower peak accuracy.

For complex classification problems that require learning rate decay, our results have shown that it is most effective to use Adam solver at a learning rate of 1E-4 until the loss/accuracy plateaus.

Finally, transfer learning has proven to be powerful in boosting neural network accuracy for small datasets.

Results

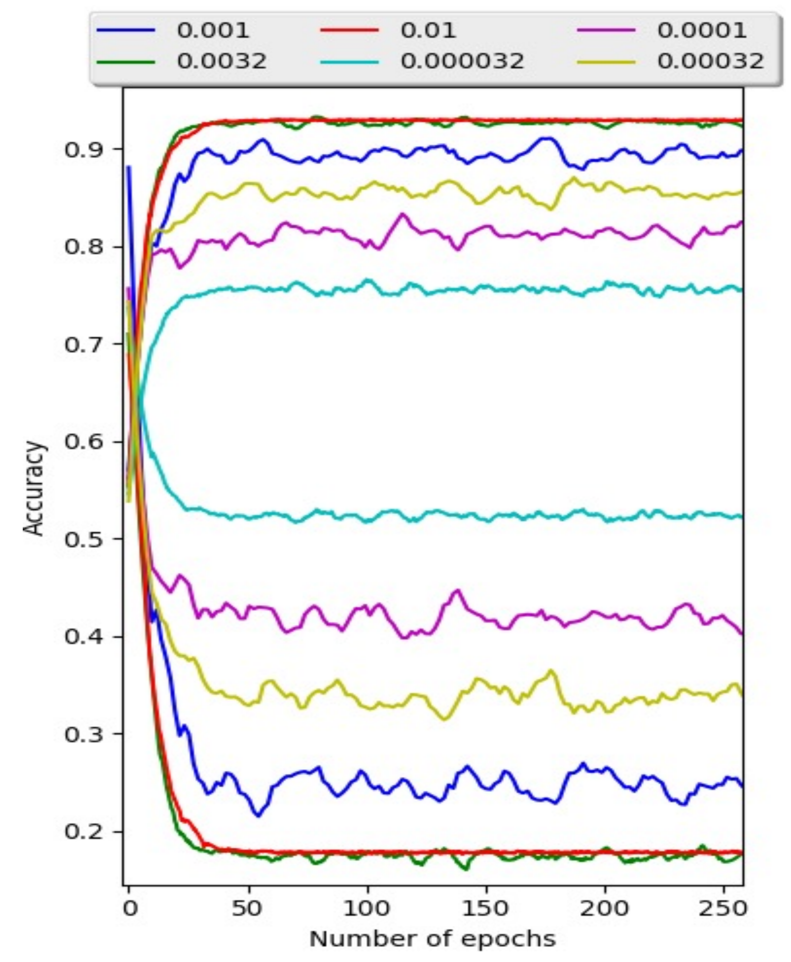

Figure 1: (Lower series) Validation loss at various learning rates; (Upper series) Validation accuracy at various learning rates

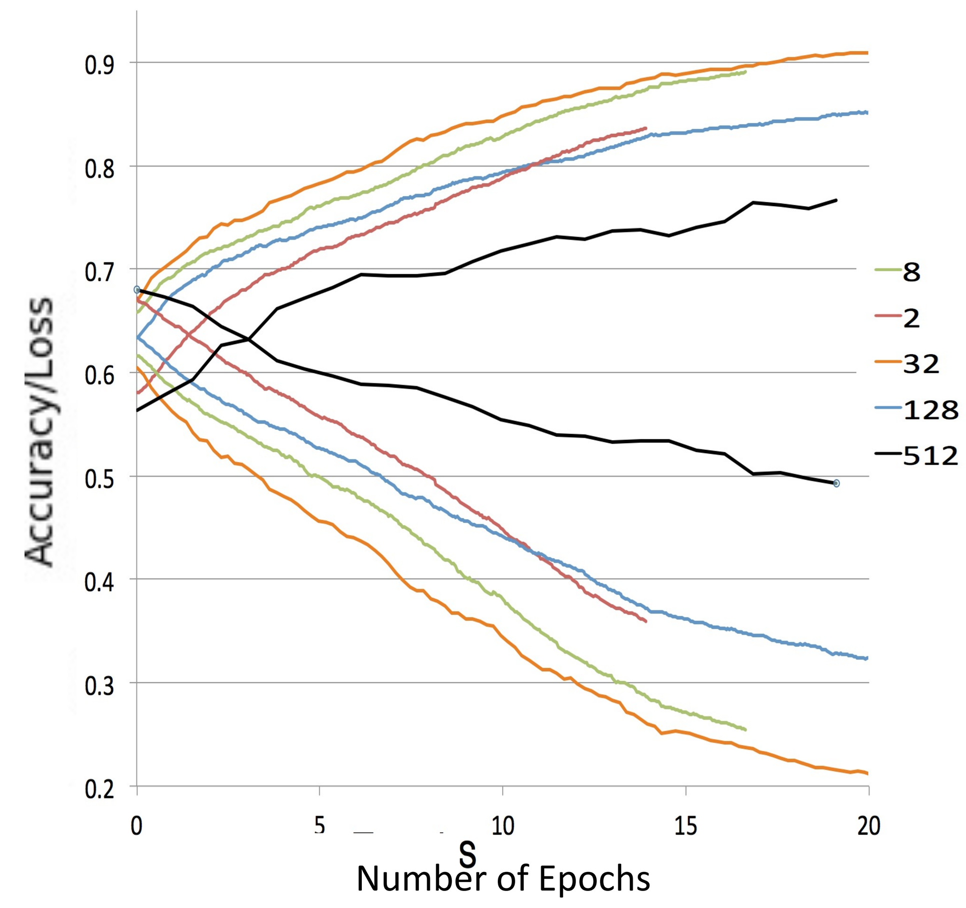

Figure 2: (Lower series) Validation loss at different batch sizes; (Upper series) Validation accuracy at different batch sizes

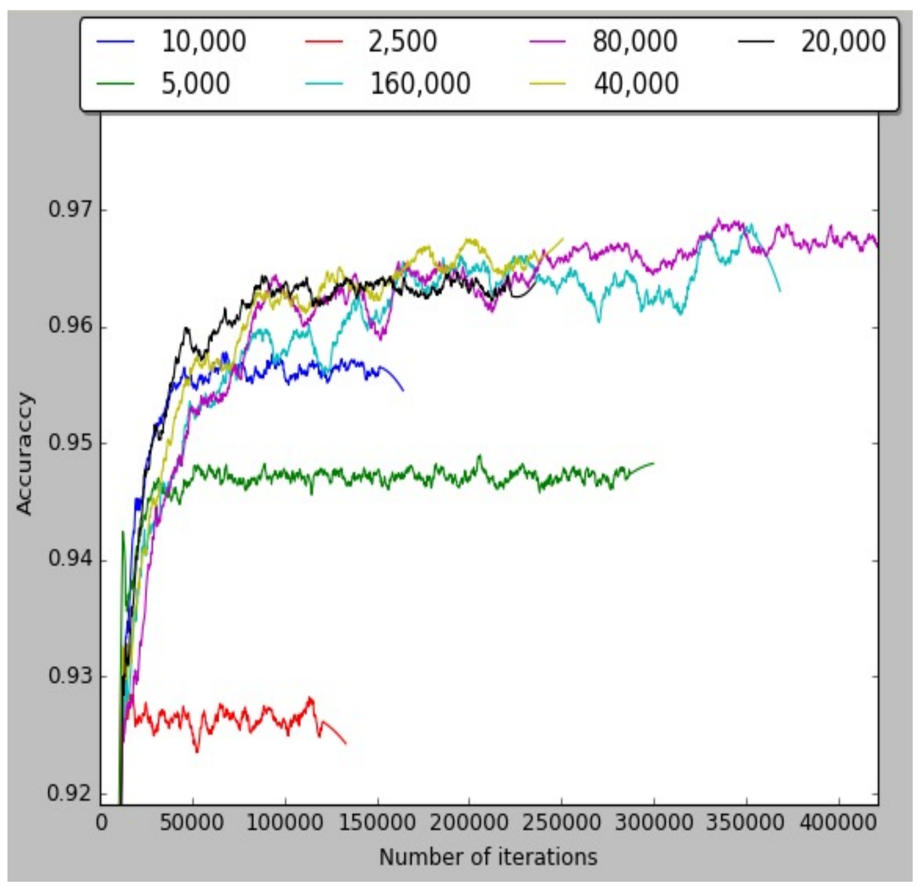

Figure 3: Validation accuracy at various decay step size values.

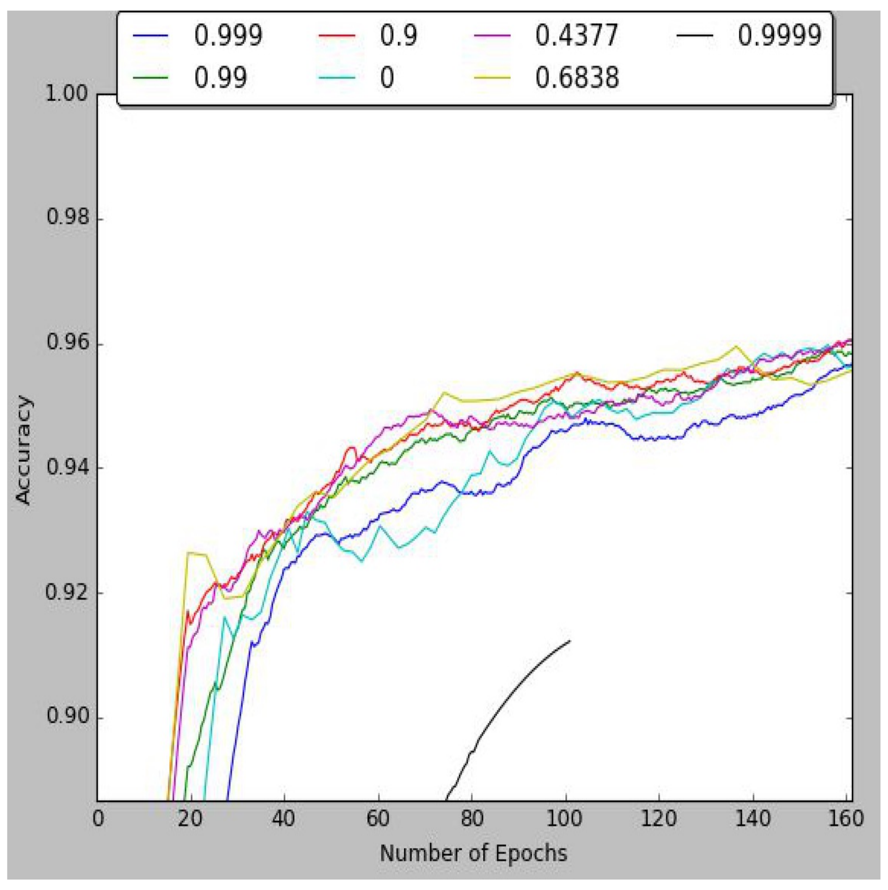

Figure 4: Validation accuracy at various momentum values.

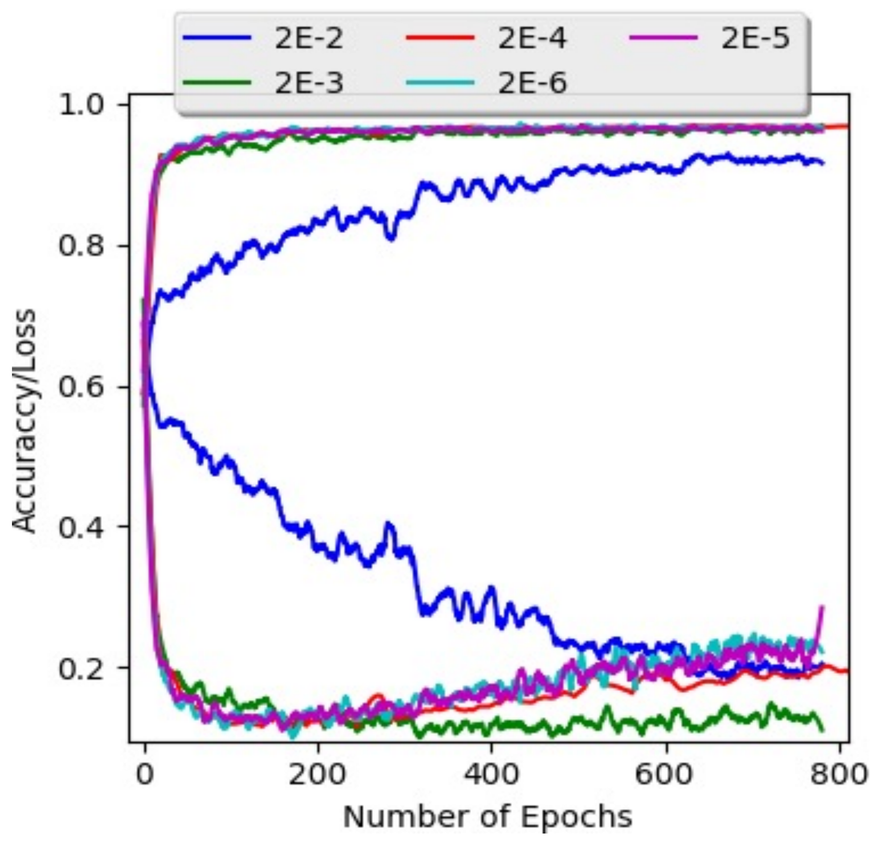

Figure 5: (Upper series) Validation accuracy at various weight decay values.
(Lower series) Validation loss at various weight decay values

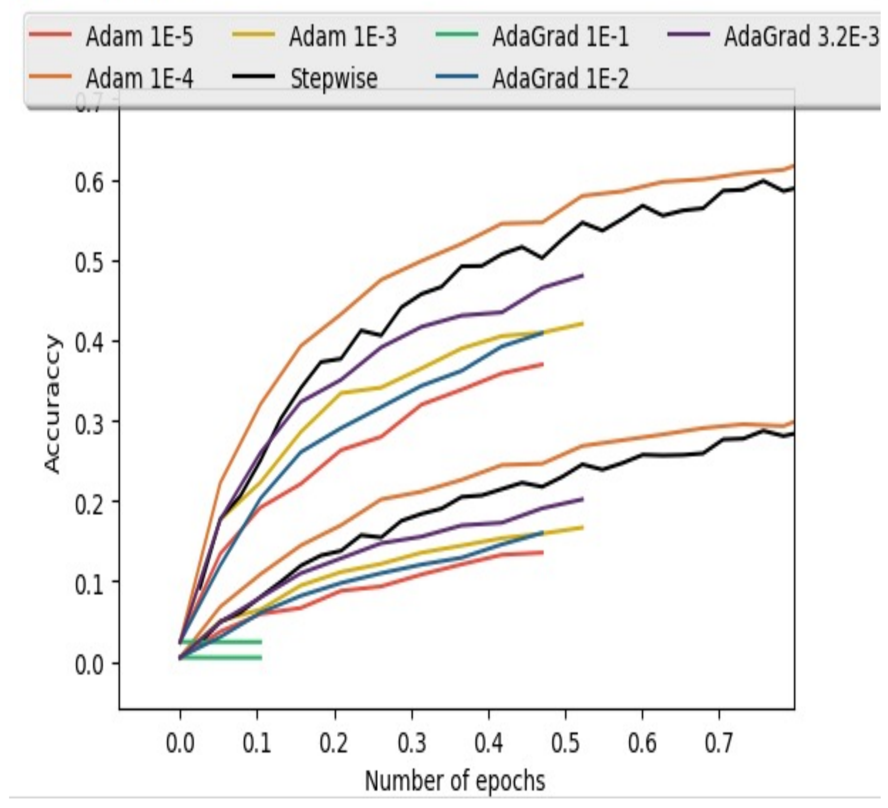

Figure 6: (Upper series) Top-5 validation accuracy with various learning rate schedulers.
(Lower series) Top-1 validation accuracy with various learning rate schedulers.

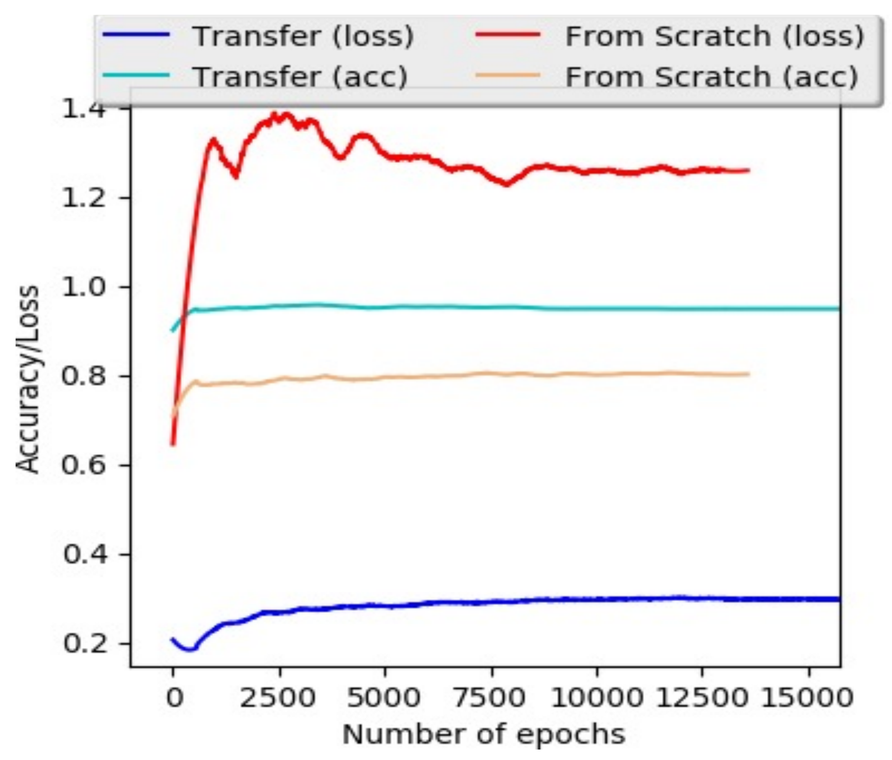

Figure 7: Accuracy and loss at when transfer learning is conducted on small dataset, Indoor Scene Recognition (ISR)

Future Work

AdaGrad could be tested at lower learning rates. RMSprop and Adadelta optimisers can be compared. Finally, data-augmentation can be attempted on small datasets to improve accuracy.

References

[1] L. Mu, Z. Tong, Y. Chen and A. J. Smola, *Efficient Mini-batch Training for Stochastic Optimization*, 1st ed. 2014, pp. 2, 5-8.
[2] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*. MIT Press, 2016, pp. 92-152, 271-316.