Task 3: Classification using VGG-16 Architecture

Sanjeev Narayanan

22nd February 2018

This task aims to understand the architecture of VGG16 model[1] of convolution neural network. In this experiment, instead of using the pre-trained weights, the model itself is constructed and trained. The data set used for the training is CIFAR-10[2]. The CIFAR-10 dataset (Canadian Institute For Advanced Research) is a collection of images that are commonly used to train machine learning and computer vision algorithms. It is one of the most widely used datasets for machine learning research. The CIFAR-10 dataset contains 60,000 32×32 color images in 10 different classes.The 10 different classes represent airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. There are 6000 images of each class.
The topology of the VGG16 model is constructed as described in the paper:

- Two convolution layers with 64 filters with feature map of sizes 3×3 and 30% dropout and Rectified Linear Unit Activation.
- A max pool layer of size 2×2
- Two convolution layers with 128 filters with feature map of sizes 3×3 and 40% dropout and Rectified Linear Unit Activation.
- A max pool layer of size 2×2
- Three convolution layers with 256 filters with feature map of sizes 3×3 and 40% dropout and Rectified Linear Unit Activation.
- A max pool layer of size 2×2
- Six convolution layers with 512 filters with feature map of sizes 3×3 and 40% dropout and Rectified Linear Unit Activation.
- A max pool layer of size 2×2
- A layer with dropout of 50%
- A flatten layer
- A fully connected Dense layer with Rectified Linear Unit Activation.

- A fully connected Dense layer with Softmax Activation.

The model was initially trained using 10 epochs, but did not give good results. The training epochs were increased to 25, 50, 100, 150, 200 and 250 epochs where the accuracy baseline error saturated at 6.78%.

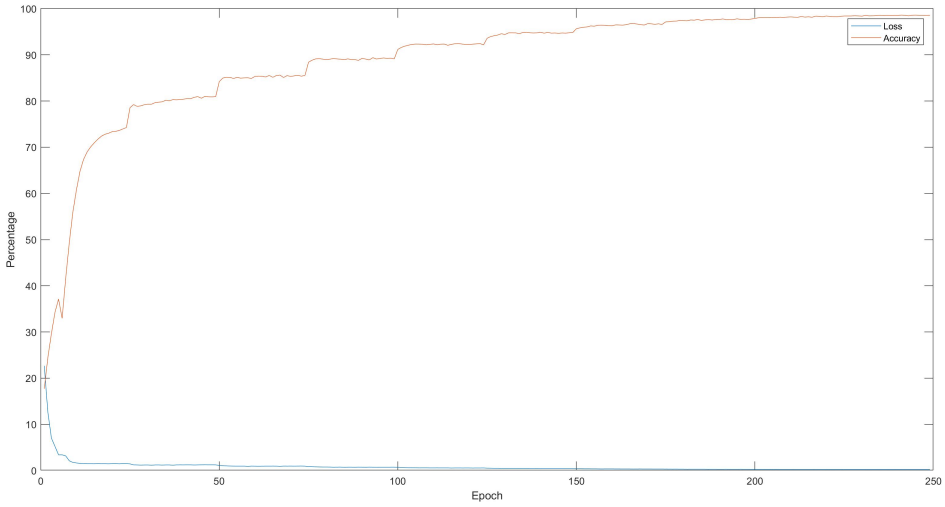

Figure 1: Plot of Loss percentage and Accuracy versus epoch

Training each epoch on an Intel i7 processor with 16GB RAM took roughly 1 hour and 27 minutes. In order to speed up the training, Colabaoratry was used. Colaboratory is a Google research project created to help disseminate machine learning education and research. It's a Jupyter notebook environment that requires no setup to use and runs entirely in the cloud. Colab supports running TensorFlow computations on a GPU. The plot of percentage of losses and accuracy achieved at each epoch are shown in the figure.

Acknowledgement

Special thanks to Dr.Anxiao (Andrew) Jiang for sharing the Colab resources for faster training of the VGG16 model.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] A. Krizhevsky, V. Nair, and G. Hinton, "The cifar-10 dataset," *online: http://www. cs. toronto. edu/kriz/cifar. html*, 2014.