Fine-tuning and Evaluation of Alexnet and VGG-16 Pre-trained Networks by Face Recognition

KangKang Li kl689

Abstract

This project fine-tunes and evaluate pre-trained Networks by face recognition. The performance of AlexNet and VGG-16 before and after fine-tuning by Euclidean distance is discussed in receiver operating characteristic curve. VGG-16 is slightly better than AlexNet, and improvement of both networks was observed with fine-tuning.

1 Introduction

Neural network is the most efficient and promising tool for image recognition. Various convolutional neutral networks (ConvNet) were develop and showed excellent accuracy for image recognition. Among these networks, AlexNet and VGG-16 are milestones in this area and fundamental models to begin with. This study applied fine-tuning on AlexNet[1] and VGG-16[2] models with LFW dataset.

1.1 AlexNet

AlexNet is the first prevalent ConvNet model for image recognition, developed by Alex Krizhevsky, Ilya Sutskever and Geoff Hinton[1]. AlexNet achieved top-1 and top-5 error rates of 37.5% and 17.0% which was the state-of-art result in 2012. It has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax, as shown in Figure 1. Non-saturating neurons and dropout were used to accelerate training and reduce overfitting. In the ILSVRC-2012 competition, AlexNet achieved a winning top-5 test error rate of 15.3%, compared to 26.2% of the second one.

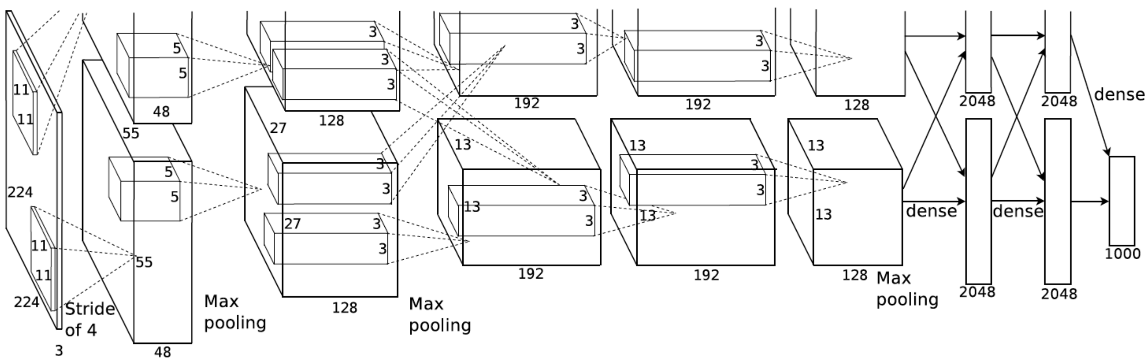

Figure 1: Network structure of AlexNet

1.2 VGG-16

VGG is introduced by Ken Chatfield et al. in 2014, and VGG-16 is one model of VGG that has 16 layers[2]. Improvement over AlexNet was made in building of VGG model, by replacing large filters with multiple smaller filters, as shown in Figure 2. In many situations, multiple stacked smaller filters is better than the one with a larger one. The reason may be that multiple non-linear layers increases the depth of the network so that it learn more complex features with lower cost. In ImageNet Challenge 2014, VGG won first and the second places in the localization and classification tracks respectively. VGG also generalize well to other datasets.

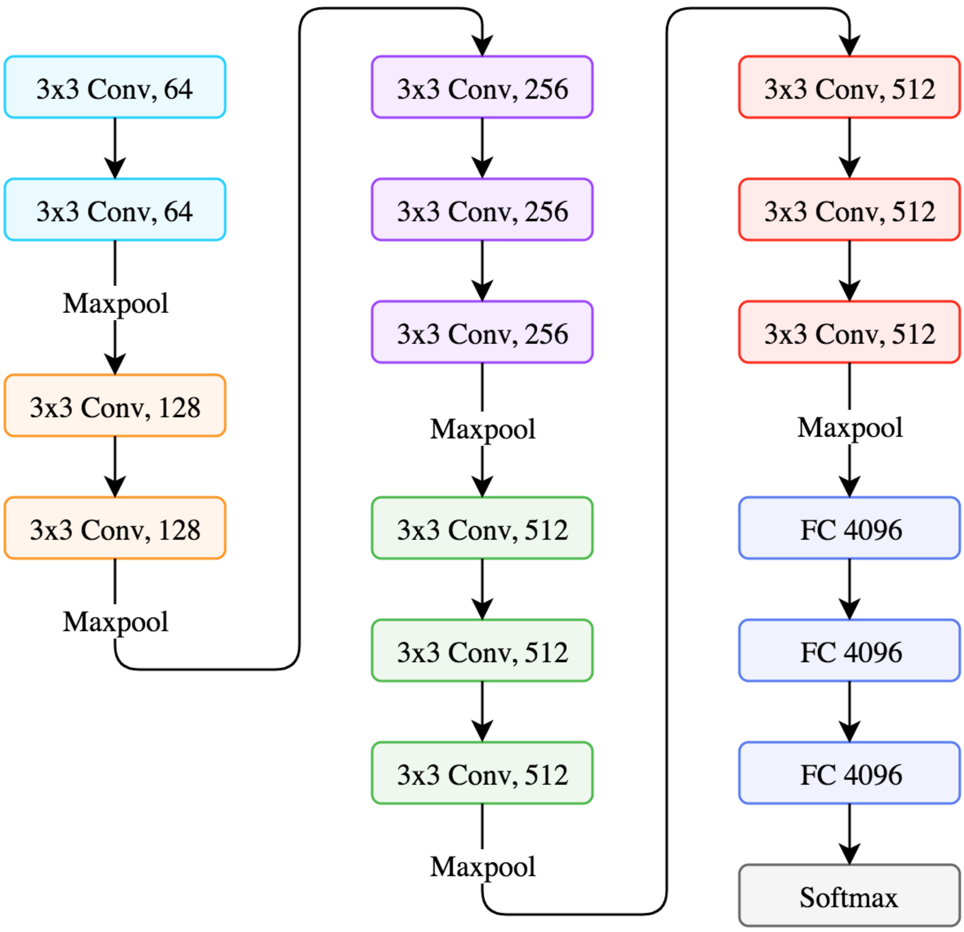

Figure 2: Network structure of VGG

2 Implementation and Discussion

The network models were trained on Tensorflow 1.8.0. ROC curves and images were generated by skliearn and matplotlib, respectively. AlexNet and VGG-16 models, and pretrained weights were obtained from kratzert and Toronto University. The dataset used for training is YaleB.

AlexNet and VGG-16 were implemented and loaded with weights for all layers except the last one. The features were obtained as the second last fc layer before and after training. Learning parameters were used as follow: learning rate at 0.1 and 0.01, epoch: 200 and 20, and batch size of 128. ROC curve was plotted with euclidean distance. Because of the lack of GPU, the training epoch is very limited to merely show the concept of fine tuning.

Because of the strong illumination bias of YaleB database, the accuracy is close to 50-50. After fine tuning AlexNet, slight improvement is shown in Figure 3. For VGG-16, the improvement after fine tuning is more evident. The epoch used for VGG is only 20 comparing to 200 for AlexNet due to much longer training time for VGG and the lack of GPU for both networks, the author assume the improvement would be more evident with sufficient training.

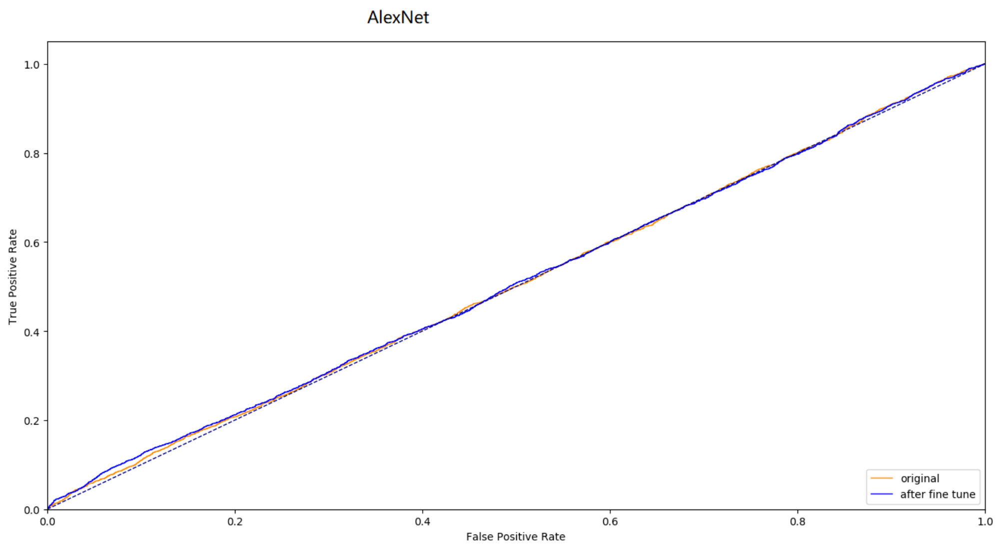

Figure 3: ROC curve of AlexNet before and after fine tuning

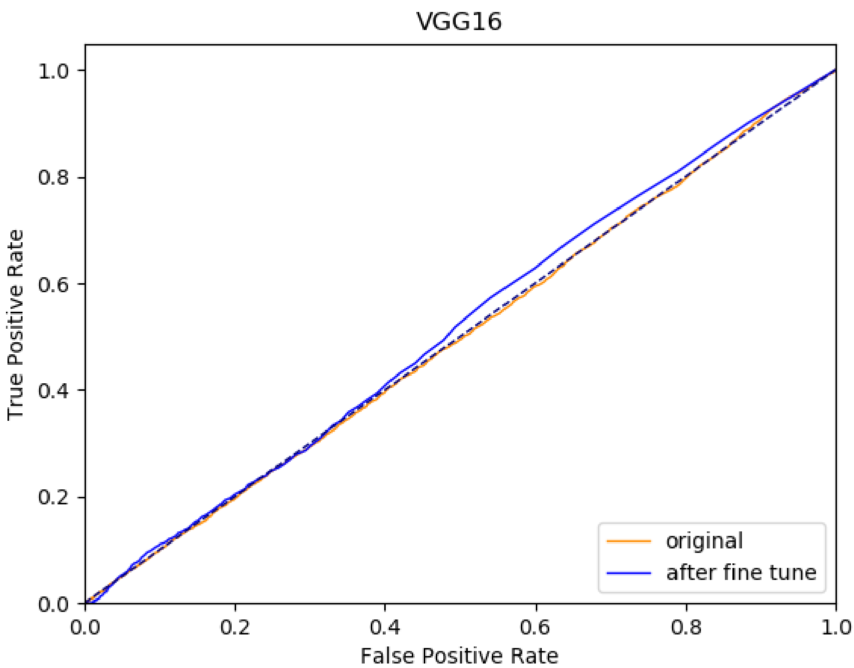

Figure 4: ROC curve of VGG before and after fine tuning

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.