Comparative Analysis of Face Recognition: Transfer Learning with AlexNet and VGG-19 in Deep Neural Networks

Bindu Parvati, Jonnala Gadda
School of Science and Engineering
University of Missouri, Kansas City
bj35v@umkc.edu

ABSTRACT

In this work, two well-known deep neural networks—AlexNet and VGG-19—are used to compare face recognition techniques that use transfer learning [1]. This work explores the effectiveness of these architectures in the field of face recognition by utilizing advances in convolutional neural networks (CNNs) and transfer learning techniques. Extensive testing and analysis are used in this study to examine how well both models perform in terms of facial feature extraction and recognition across a variety of datasets, as well as their accuracy and computational efficiency. The results present insights into each architecture's robustness and adaptability in practical face recognition applications, in addition to highlighting its advantages and disadvantages. Understanding the implications and sensitivities of using various CNN architectures for face recognition tasks is greatly enhanced by this comparative study.

INTRODUCTION

This study embarks on an exploration of face recognition techniques by expanding upon the AT&T dataset and integrating the "cropped" iteration of the Georgia Tech face recognition dataset [2]. The dataset curation involves segregating the first 10 images of each subject for training/validation, with the remaining 5 images reserved for testing, encompassing 50 subjects, each with 15 instances.

The primary objective is to leverage a pre-trained CNN architecture, such as AlexNet, VGG, or SqueezeNet, in conjunction with MATLAB's transfer learning approach. The methodology follows MATLAB's provided example for transfer learning using AlexNet, aiming to construct a network capable of classifying subjects as unique classes while refining parameters to optimize performance.

IMPLEMENTATION

The implementation phase entails fine-tuning the network using the training dataset and utilizing the Fully Connected (FC) output as biometric templates, disregarding the classification head. By computing the cosine distance between enrollment and verification images within the testing dataset, genuine and impostor score sets are established.

Graphical representations, comprising testing score distribution histograms and Receiver Operating Characteristic (ROC) curves for genuine and impostor classes, facilitate an insightful analysis of the face recognition system's performance. Furthermore, the study quantifies the efficiency of the verification task through calculations of ROC Area Under the Curve (AUC) and d' (d-prime) metrics.

Additionally, the investigation adopts a "subject-independent" protocol by training the network on the initial 40 subjects and evaluating its performance on the last 10 subjects, providing critical insights into the model's generalization capacity across novel subjects.

RESULTS

The Figure-2 and Figure-3 summarizes diverse training configurations applied to AlexNet and VGG models, highlighting their corresponding accuracies and training times.

For AlexNet, adjustments in mini-batch sizes and pixel ranges resulted in varying accuracies. Initial settings with 8 epochs, a mini-batch size of 10, and a wider pixel range of [-40, 40] achieved a 95% accuracy within 1 minute and 11 seconds. By tweaking the mini-batch size to 8 and adjusting the pixel range to [-20, 20] while maintaining 8 epochs, accuracy improved to 96% within 1 minute and 21 seconds.

In the case of VGG, similar adjustments in training parameters demonstrated parallel trends. The initial settings with 8 epochs, a mini-batch size of 8, and a pixel range of [-20, 20] achieved a 95% accuracy within 1 minute and 31 seconds. Maintaining these parameters yielded a higher accuracy of 96% in 1 minute and 48 seconds.

These variations underscore the influence of parameter fine-tuning on accuracy and training duration, offering insights for optimizing model performance while considering computational efficiency.

CONCLUSION

In conclusion, the experimentation with diverse training settings for AlexNet and VGG models highlights the substantial impact of parameter variations on both model accuracy and training duration. These findings underscore the significance of meticulous parameter tuning in achieving higher accuracies while considering computational efficiency. The nuanced adjustments in mini-batch sizes and pixel ranges showcase the potential for optimizing face recognition models, offering valuable insights into balancing accuracy with training resource requirements for efficient model deployment.

REFERENCES

[1] https://www.mathworks.com/help/deeplearning/ug/transfer-learning-using-alexnet.html
[2] https://www.anefian.com/research/face_reco.htm

Appendices

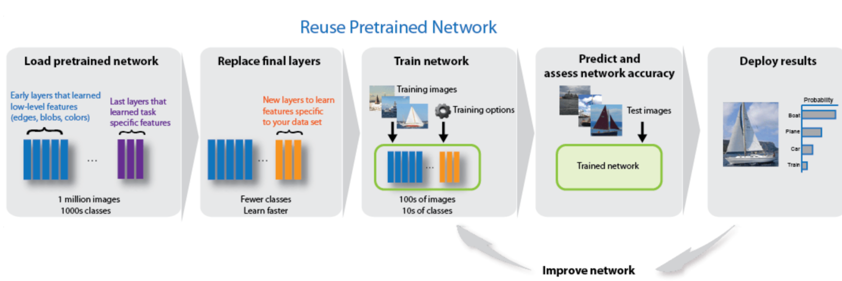

Figure-1: Reuse Pretrained Network

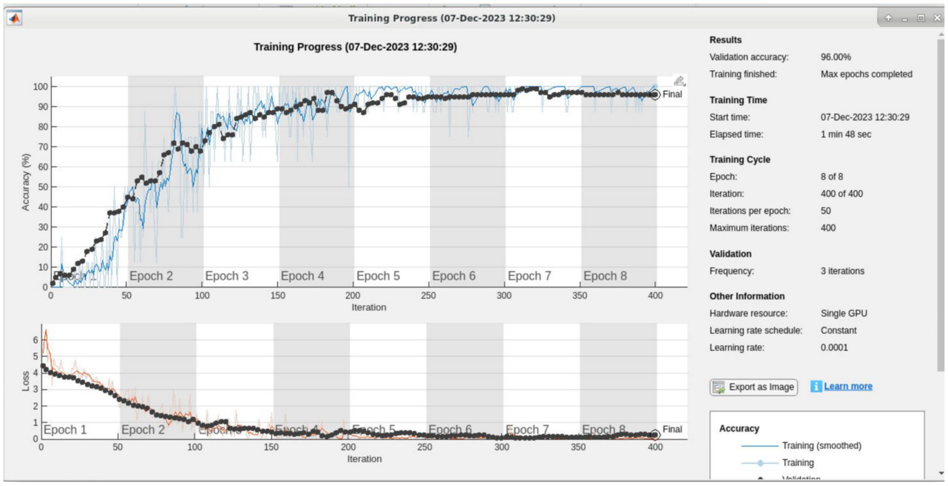

Figure-2: Fine-tuned AlexNet Training

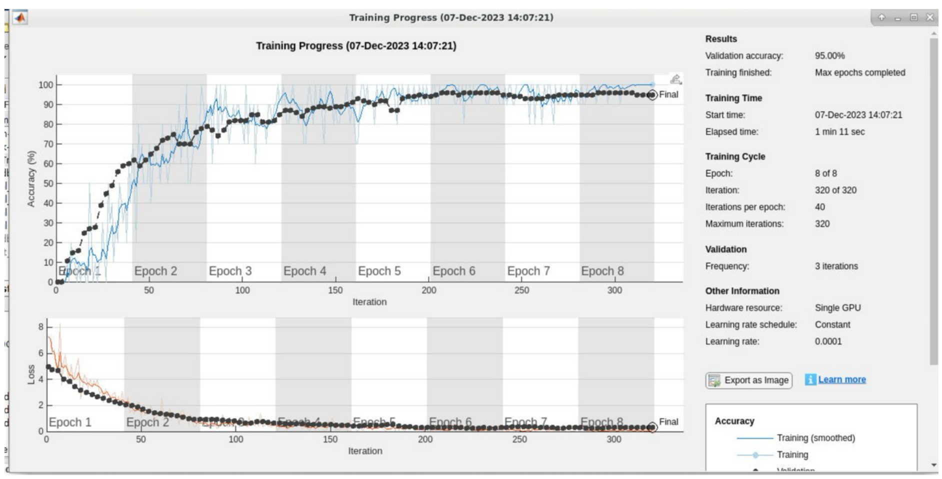

Figure-3: VGG-19 Training

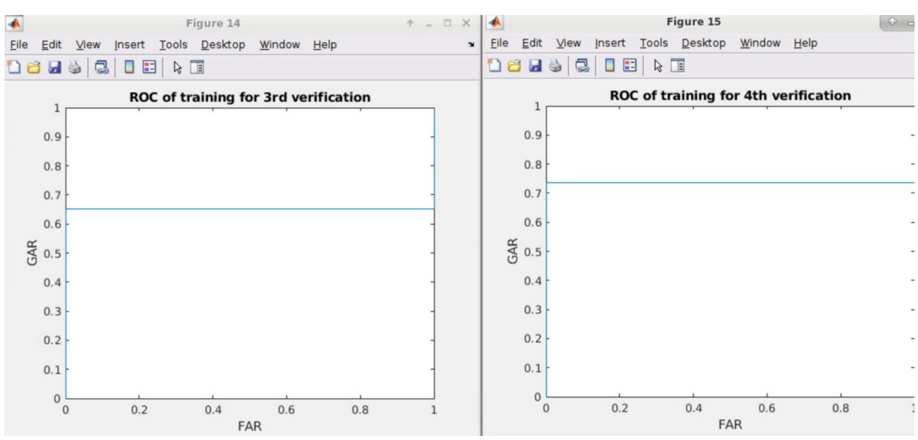

Figure-4: ROC for 3rd and 4th verification for finetuned AlexNet

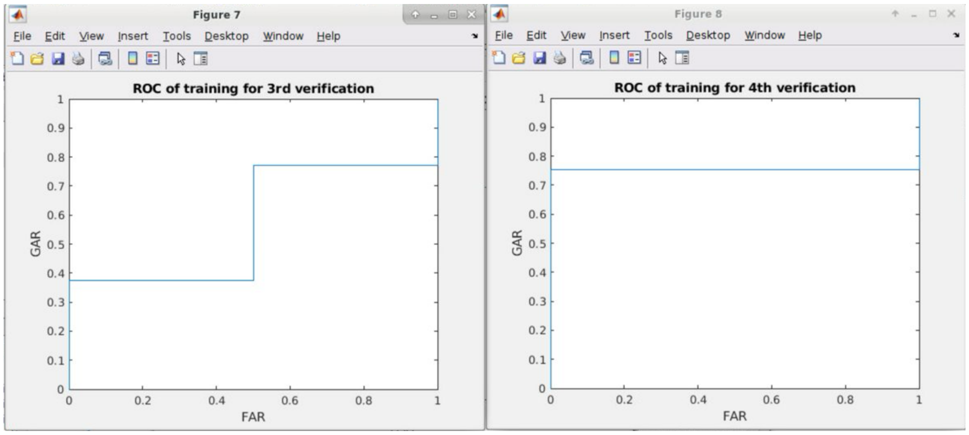

Figure-5: ROC for 3rd and 4th verification for finetuned VGG-19