Face Recognition Using Deep Transfer Learning: A Study on the Georgia Tech Dataset

Luke Miller

December 18, 2023

Attestation of Originality:
I attest that this work is solely a product of my efforts, and that all quotations, explicit and implicit, have been appropriately attributed.

Abstract

This study delves into the efficacy of transfer learning for facial recognition on the Georgia Tech face recognition data set, utilizing the Inception V3 Convolutional Neural Network (CNN) as a base model. The project's cornerstone is the comparison of four models: subject-independent, subject-dependent without dropout and regularization, and subject-dependent with dropout, and subject-dependent with dropout and regularization. Furthermore, a test case with dimensionality reduction through Principal Component Analysis (PCA) was employed on the subject-dependent model to compare its effects on computational efficiency versus accuracy. The models were assessed using cosine distance to generate genuine and impostor score sets, with the performance quantified by Receiver Operating Characteristic (ROC) Area Under Curve (AUC) and d-prime statistics. The results indicated that the subject-dependent models achieved an ROC AUC of 0.97 with a d-prime of 2.66. The application of PCA yielded comparable AUC of 0.96 and a slightly higher d-prime of 2.72. The subject-independent model demonstrated robust generalization with an AUC of 0.96 and a d-prime of 2.56. An ensemble approach, combining the three subject-dependent models, resulted in a superior ROC AUC of 1.00 and a d-prime of 3.71. These results underscore the potential of ensemble learning and the strategic use of PCA in facial recognition tasks.

1 Introduction

Facial recognition systems are integral to numerous applications, ranging from security to user interface design. The advent of deep learning has revolutionized these systems, providing unprecedented accuracy and robustness. This paper presents a comprehensive study using transfer learning using the InceptionV3 Convolutional Neural Network to recognize faces within the Georgia Tech face recognition dataset. The significance of this work lies in its exploration of transfer learning techniques, dimensionality reduction via Principal Component Analysis (PCA), and ensemble methods to improve identification accuracy across various subjects.

2 Methods

2.1 Data Preparation

The Georgia Tech facial recognition data set[1], consisting of 50 subjects each with 15 cropped images, was utilized. In the subject-dependent protocol, the first 10 images from each subject were used as the training and validation set, while the last five were used for testing. In the subject-independent protocol, 10 subjects were set aside for testing, and the remaining 40 subjects were used for training and validation of the model. Test images were further separated into enrollment, genuine and impostor images for use in evaluating the models. The pre-processing pipeline included resizing to match the size of the Inception V3 network and augmentation via random translation, scaling by up to 10 percent, and minor color shifts as seen in Figure 1. These augmentation steps were only conducted on the training set. While useful for other image recognition tasks, horizontal flips were avoided

due to their adverse effect on face recognition[2, page 147]. This augmentation was vital due to the limited size of the data set and the regularizing effects it has upon the data set[2, page 237].

2.2 Network Architecture

The InceptionV3 architecture[3] was employed as the backbone for this study. The layers of the InceptionV3 Model include convolutional layers(equation 4), average pooling layers (equation 6), batch normalization layers (equation 8), various concatenation layers, and activation layers using the RELU (equation 9) activation function. The classification layers of Inception were not imported. All layers of the imported pretrained models were frozen. A global average pooling layer and two dense layers were then added to the network for classification of the 50 subjects for the subject-dependent models and 40 subjects for the subject-indpendent model. The Adam optimizer(Figure 6) was used for adaptive gradient descent and accuracy(equation 3) was the loss metric. The network underwent five stages of fine-tuning involving the sequential unfreezing of layers and adjustment of the learning rate to refine feature extraction[2, page 240]. Finally, to create the ensemble model, two variations of the subject-dependent model were added: one with a dropout of 0.50, and another with dropout of 0.50 and l2 regularization (equation 2).

3 Results and Discussions

3.1 Subject Dependent Model

Cosine distance(equation 10) was chosen for its effectiveness in high-dimensional spaces and its common use in face recognition tasks[2, page 269] to test the model's capabilities in distinguishing impostor subjects from genuine subjects. After training, the subject dependent model was evaluated and found to have an Area Under the Curve2 (AUC) of its Reciever Operating Characteristic (ROC) of 0.97. This demonstrates outstanding predictive capability as is exemplified by its easily separable similarity scores between genuine and impostor pairings. Additionally, the model achieved a d-prime of 2.66 which indicates a strong ability to differentiate faces. These metrics affirm the model's capability in distinguishing individual faces with high confidence[2, page 147].

3.2 PCA Dimensionality Reduction

Principle Component Analysis (PCA) was then applied, retaining 95 percent of the variance in the data, resulting in a compressed yet informative feature set. The PCA-enhanced model3 sustained an AUC of 0.97 and its d-prime actually increased to 2.72, indicating that dimensionality reduction preserved discriminative power while enhancing computational efficiency[2, page 240].

3.3 Rank Identification Rates

Features were extracted from the the layer of the model prior to the classification layers by passing an enrollment image and comparing the cosine distance between it and genuine and impostor images. These served as unique biometric templates for each subject[2, page 321]. Both rank 1 and rank 5 identification rates2 were perfect (1.0), which remained unchanged after finding the optimal threshold for identification of 0.20. This demonstrated

the model's precision in correctly identifying subjects[2, page 321]. This same procedure was repeated for the subject-independent model covered in the next section, and it also achieve rank 1 and rank 5 identification rates of 1.0.

3.4 Subject-Independent Protocol

The subject-independent model, was trained on a 40 subjects and was tested on the remaining 10 unseen subjects. Despite testing on unseen subjects, the subject-independent model achieved a ROC AUC4 of 0.96 and a d-prime of 2.56 by comparing cosine distance of subjects. This slight performance dip compared to the subject-dependent model was anticipated due to the challenges inherent in generalizing to unseen data[2, page 269].

3.5 Committee of Models

The ensemble approach, involving a committee of three models with variations in dropout, regularization, and pooling. Adding a dropout factor of 0.5 allowed the model to simulate training multiple models without incurring the computational expense[2, page 255][4, slide 9]. While data augmentation provides a form of regularization[2, page 237] L2 regularization(equation 2) was added which encourages smaller weights for those vectors that contribute least[2, 230]. Finally, pooling was changed from average pooling(equation 6 to max pooling (equation 5). The reason for this change was to add diversity to the ensemble models as different pooling methods may better capture different features[2, page 335-339]. The ensemble models capitalized on capturing diverse patterns in the data. The fusion of model outputs led to a ROC AUC of 1.00 and a d-prime of 3.71, showcasing the power of ensemble methods in improving prediction accuracy[2, page 250]. The results indicated that the ensemble model, which combines the strengths of individual models, could significantly outperform single-model architectures1. This finding underscores the potential of ensemble learning in the field of facial recognition.

4 Conclusion

This paper presented a transfer learning approach to facial recognition using the Georgia Tech dataset and the InceptionV3 CNN. The exploration of different model configurations and the incorporation of PCA for dimensionality reduction confirmed that such methodologies could significantly enhance the accuracy and efficiency of facial recognition systems. The ensemble of diverse models demonstrated a remarkable improvement over individual models, achieving perfect classification metrics. These findings reinforce the notion that transfer learning, when combined with ensemble techniques and PCA, can be a potent strategy in the domain of facial recognition. The study's success paves the way for future research to explore similar approaches in more complex and varied datasets, further advancing the field of computer vision and deep learning.

References

[1] Georgia Tech College of Computing, "Georgia tech facial recognition database," Accessed 2023, accessed: 2023-12-18. [Online]. Available: https://www.anefian.com/research/face_reco.htm

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[4] "Chapter 6-8 overview," Lecture note, 2023.

[5] "Object detection and segmentation in deep learning part 2," Lecture note, 2023.

[6] "Autoencoders and generative adversarial networks," Lecture note, 2023.

[7] "Classification metrics part 1," Lecture note, 2023.

[8] "Nonlinear systems part 1," Lecture note, 2023.

[9] "Miniproject2," Lecture note, 2023.

A Appendices

A.1 Figures

Fig. 1. Illustration of Data Augmentation. Original images in left column.

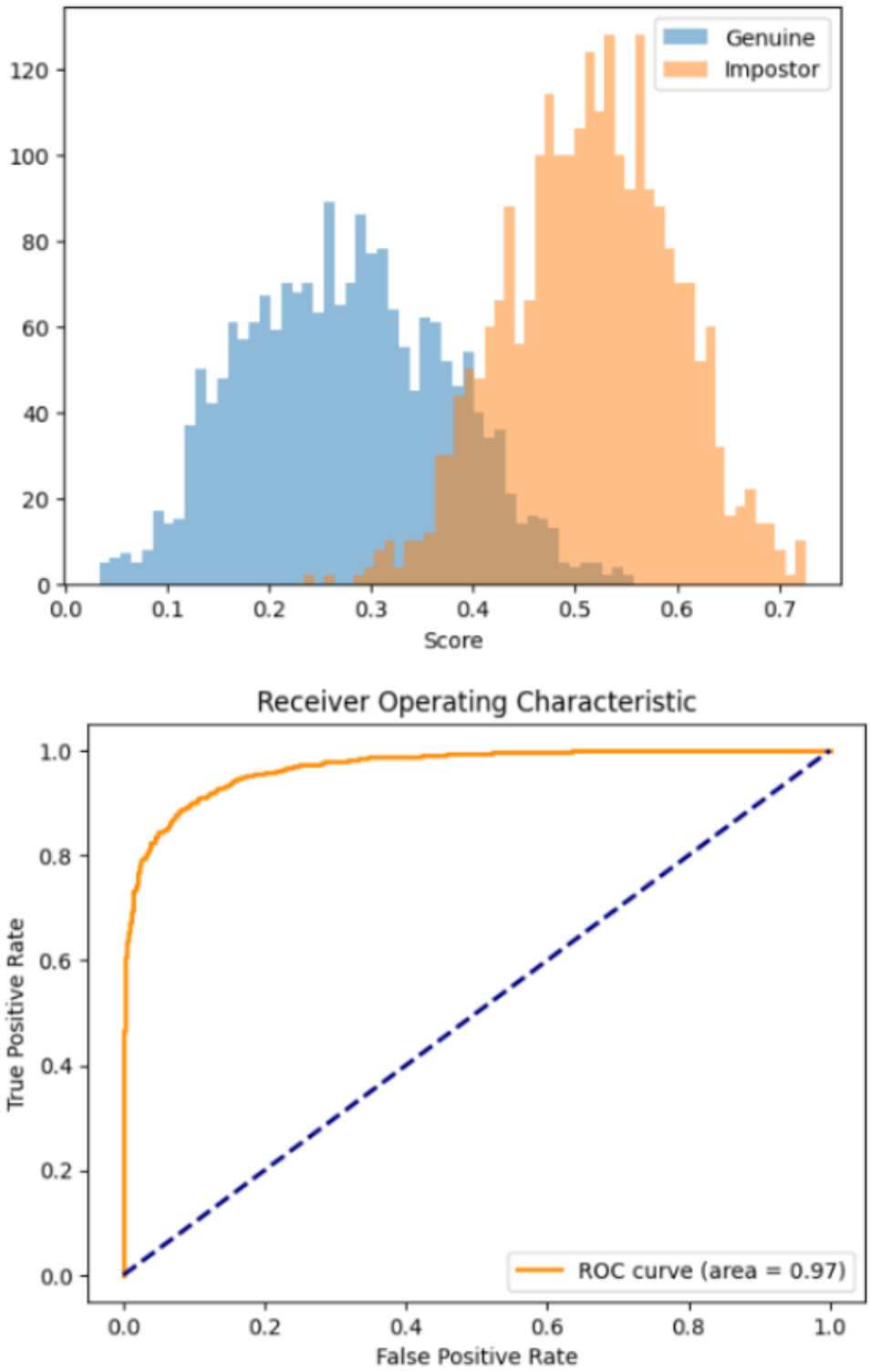

Fig. 2. Score Distribution and ROC for Subject-Dependent Model without dropout or regularization

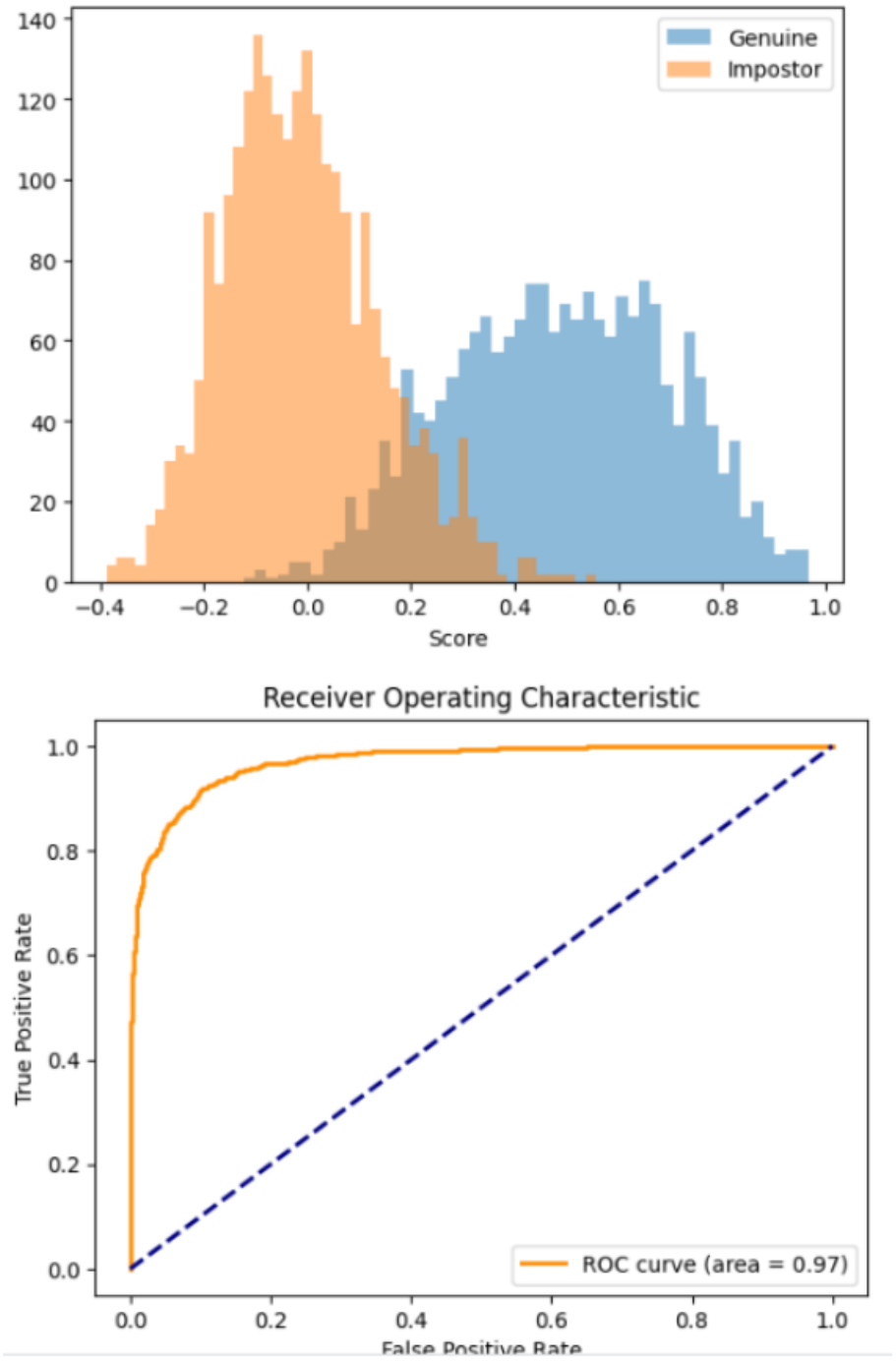

Fig. 3. Score Distribution and ROC for Subject-Dependent Model with PCA dimensionality reduction

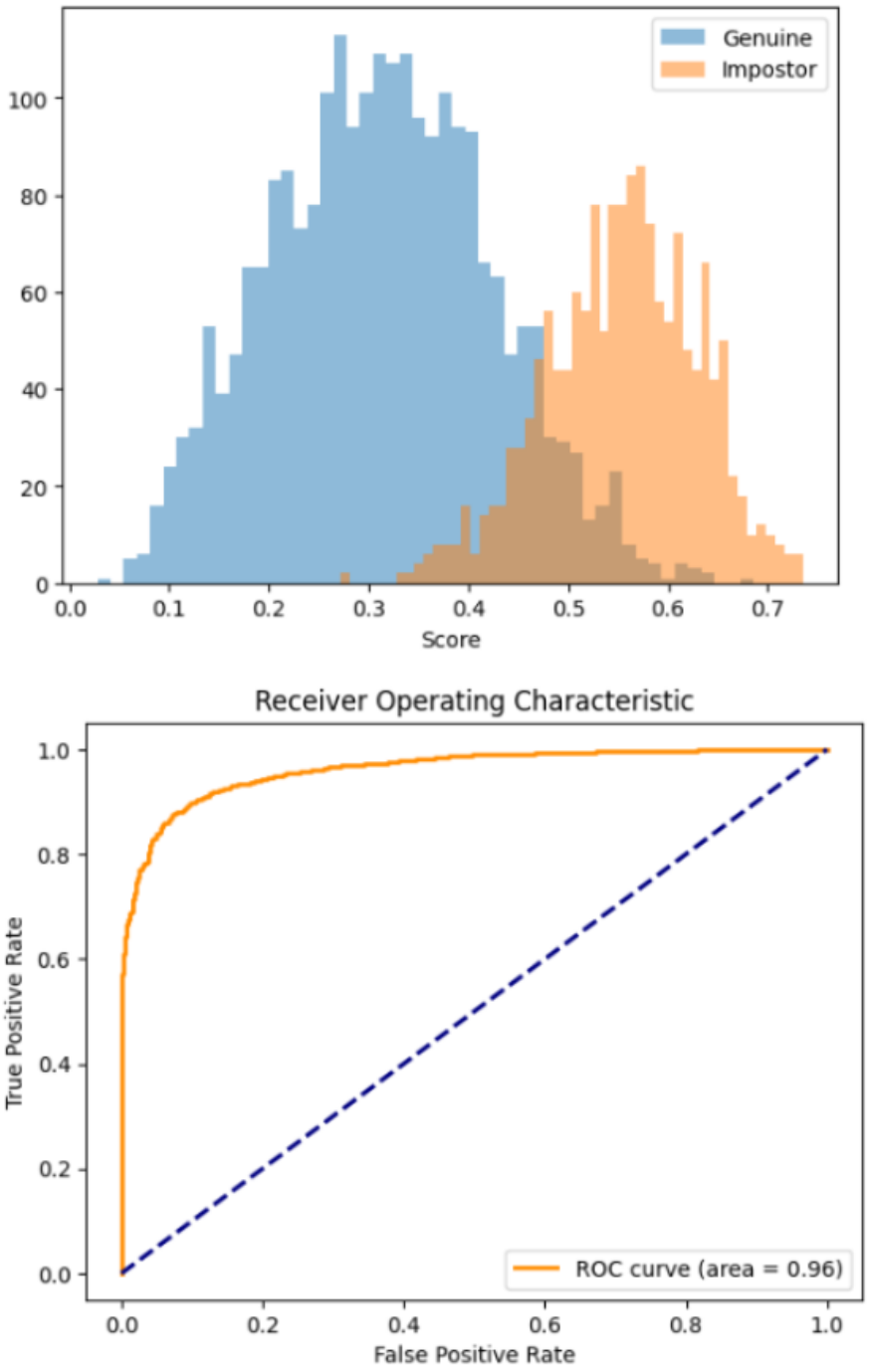

Fig. 4. Score Distribution and ROC for Subject-Independent Model

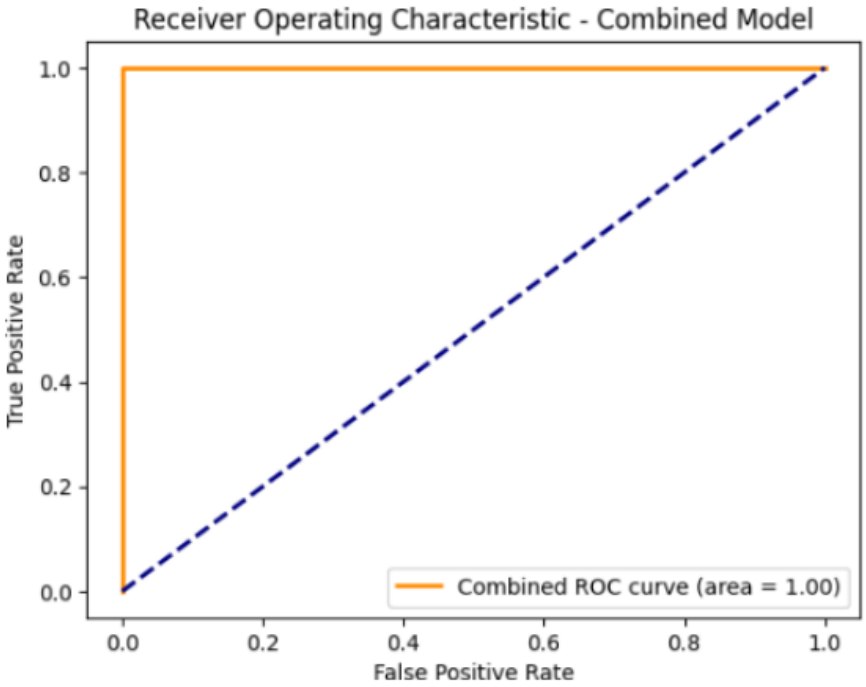

Fig. 5. ROC for composite model

Algorithm 8.7 The Adam algorithm

Require: Step size ϵ (Suggested default: 0.001)

Require: Exponential decay rates for moment estimates, ρ_1 and ρ_2 in $[0, 1)$. (Suggested defaults: 0.9 and 0.999 respectively)

Require: Small constant δ used for numerical stabilization (Suggested default: 10^{-8})

Require: Initial parameters $\boldsymbol{\theta}$

Initialize 1st and 2nd moment variables $\boldsymbol{s} = 0$, $\boldsymbol{r} = 0$

Initialize time step $t = 0$

while stopping criterion not met **do**

Sample a minibatch of m examples from the training set $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ with corresponding targets $\boldsymbol{y}^{(i)}$.

Compute gradient: $\boldsymbol{g} \leftarrow \frac{1}{m} \nabla_{\boldsymbol{\theta}} \sum_i L(f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}), \boldsymbol{y}^{(i)})$

$t \leftarrow t + 1$

Update biased first moment estimate: $\boldsymbol{s} \leftarrow \rho_1 \boldsymbol{s} + (1 - \rho_1)\boldsymbol{g}$

Update biased second moment estimate: $\boldsymbol{r} \leftarrow \rho_2 \boldsymbol{r} + (1 - \rho_2)\boldsymbol{g} \odot \boldsymbol{g}$

Correct bias in first moment: $\hat{\boldsymbol{s}} \leftarrow \frac{\boldsymbol{s}}{1 - \rho_1^t}$

Correct bias in second moment: $\hat{\boldsymbol{r}} \leftarrow \frac{\boldsymbol{r}}{1 - \rho_2^t}$

Compute update: $\Delta\boldsymbol{\theta} = -\epsilon \frac{\hat{\boldsymbol{s}}}{\sqrt{\hat{\boldsymbol{r}}} + \delta}$ (operations applied element-wise)

Apply update: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \Delta\boldsymbol{\theta}$

end while

Fig. 6. The Adam Algorithm used for gradient-based optimization[2, page 306]

A.2 Tables

Table 1
AUC and d' for Model Variations

	Subj. Dependent	Subj. Dependent after PCA	Subj. Independent	Ensemble
AUC	0.97	0.96	0.96	1.00
d'	2.66	2.72	2.56	3.71

Table 2
Rank 1 vs Rank 5 Accuracy

	Rank 1 Accuracy	Rank 5 Accuracy
Subject Dependent	1.0	1.0
Subject Independent	1.0	1.0

A.3 Equations

$$L_{\text{bce}} = \sum_i y_i \log o_i + (1 - y_i) \log(1 - o_i) \tag{1}$$

Common Cross Entropy Gradient[5, slide 9]

$$\Omega_{\text{weights}} = \frac{1}{2} \sum_i^L \sum_j^n \sum_i^k \left(w_{ji}^{(l)} \right)^2 \tag{2}$$

L2 Regularization Term[6, slide 5]

$$\text{Accuracy} = \frac{1}{n} \sum_{i=1}^{n} 1(\hat{y}_i = y_i) \tag{3}$$

Accuracy[7, page 1]

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \tag{4}$$

Convolution[2, page 329]

$$p_k^{[l]} = \max_{i,j} a_{ijk}^{[l]} \tag{5}$$

The Max Pooling Layer[2, page 335-339]

$$p_k^{[l]} = \frac{1}{n_H \cdot n_W} \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} a_{ijk}^{[l]} \tag{6}$$

Average Pooling Layer[2, page 335-339]

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{7}$$

The SoftMax Layer[2, page 191]

$$\sigma = \sqrt{\epsilon + \frac{1}{m} \sum i(\mathbf{H} - \mu)_i^2} \tag{8}$$

Batch Normalization[2, page 315]

$$f(x) = max(0, x) \tag{9}$$

ReLu[8, slide 3]

$$cos(\theta) = \frac{\sum x_i * y_i}{\sqrt{\sum x_i^2} * \sqrt{\sum y_i^2}} \tag{10}$$

Cosine Distance[9]