Using Transfer Learning and BPDFHE to Improve Ocular Image Recognition Accuracy

Riddhiman Das
Department of Computer Science and Electrical Engineering
University of Missouri-Kansas City
Kansas City, USA
dasrid@mail.umkc.edu

Reza Derakhshani
Department of Computer Science and Electrical Engineering
University of Missouri-Kansas City
Kansas City, MO
reza@umkc.edu

***Abstract*—We used image enhancement algorithms along with transfer learning to fine-tune a deep convolutional neural network to perform ocular image recognition. To enhance the input images, we used a novel color image histogram equalization technique called Brightness Preserving Dynamic Fuzzy Histogram Equalization, which showed significant accuracy improvements: on the test data, using AlexNet, the ROC Area Under the Curve (AUC) increased to over 0.99, Equal Error Rate (EER) decreased 4-fold and dropped below 4%, and decidability (a measure of class separability) increased from 1.89 to 4.17**

I. Introduction

As smartphones proliferate through modern daily life, biometric identity verification for secure access to services has become commonplace. In addition to hardware biometric sensors built into some of the devices, the front facing cameras of these modern smartphones has been used to develop novel biometric modalities based on the ocular and periocular regions of the face. The biggest advantage of using the front facing camera on mobile devices for biometric authentication is that they enable secure biometric identity verification methodologies universally, without relying on specific hardware sensors for capturing the biometric sample template. The other main advantage is that they offer a seamless verification experience, because the front facing camera is always facing the ocular region of the face during use of the mobile device, and “selfies” for social media have made many end users comfortable with taking pictures of their face.

However, one of the major challenges with contemporary approaches to biometric verification with images captured with mobile device cameras is that the images can be degraded during acquisition due to noise, variation in lighting conditions, specular reflection and motion blur. Consequently, these images need to be pre-processed to mitigate these degradations before being used for matching and/or liveness detection.

This paper proposes a novel image contrast enhancement methodology based on using the Brightness Preserving Dynamic Fuzzy Histogram Equalization (BPDFHE) technique to pre-process the template images prior to matching [1]. BPDFHE introduces a novel modification of the brightness preserving dynamic histogram equalization technique, by using fuzzy statistics to represent and process digital images [2]. This improves the algorithm’s brightness preserving and contrast enhancement abilities while reducing its computational complexity at the same time. We show that this image enhancement technique can be combined with transfer learning to achieve significant improvement in ocular biometric verification accuracy.

II. Proposed Methodology

A. *Image Enhancement*

Convolutional deep neural networks for image recognition have exploded in popularity in the last few years due to their outstanding performance in a

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

variety of image recognition tasks. We conducted a survey of several image enhancement techniques for ocular image recognition based on deep neural nets. We started by evaluating several well-known convolutional neural networks out of the box for ocular matching. We chose template 1 as enrollment and template 2 as verification and matched them together. For this specific problem, we evaluated AlexNet [3] and VGG16 [4], although the concepts in this paper can be applied to any other deep convolutional neural network as well. AlexNet was chosen for its lower computational complexity compared to VGG16, which reduced training time on the hardware we had. We chose Cosine Similarity of the activations of the last fully connected layers of the neural network for the calculation of the match score. Moreover, the focus of this study was to assess the effects of image preprocessing methods on accuracy of deep neural net models for ocular biometric matching in visible light, for which a well-known model such as AlexNet seemed to be a good choice. The images from the input dataset [5] were split into enrollment and verification images for each subject and were resized per the input requirements of the input layer of the network. These enrollment and verification images were fed into the neural network and the activations of the last fully connected layer were deemed to be the features extracted by the neural network in its attempt to classify the image. The feature vectors for each enrollment image were compared with the feature vectors for each verification image using the cosine similarity metric, which was used as the match score. For each enrollment image of every subject, we also matched the verification features generated from every other subject's verification images with the enrollment features to generate the impostor match scores. With these verifications, we generated an ROC curve to measure AlexNet's out of the box performance for ocular image verification.

Then we implemented fine-tuning using transfer learning, but without any image enhancement to obtain a baseline. Fine tuning offers several advantages over training a Convolutional Neural Network (CNN) from scratch, including faster training times and needing to only train the last few layers of the pre-trained CNN to learn the features of the new classification problem. We froze the first 10 layers of AlexNet and ran backpropagation for training on the last 13 layers, with a low learning rate for 13 epochs. The classification head was retained, and we used one-hot output encoding as the new fully connected layer. To achieve this one-hot output encoding, the output for the output line for the corresponding subject ID was set to 1, and the rest to 0, similar to the 1000 ImageNet classes classified by AlexNet. This style of classification may seem inefficient for a real-world biometric system, but given the focus of this study, which is the evaluation of the efficacy of the input preprocessing method, we deemed this methodology to serve the purpose. The input dataset was partitioned into a 70-30 split of training and validation images. The enrollment and feature vectors now were the activations of this new fully connected layer. The same matching methodology described above was performed again to determine improvement in the system after this transfer learning procedure. We generated the ROC curves and the distribution of the match scores to determine performance improvement.

We then considered several image enhancement algorithms to determine the appropriate image enhancement technique to apply to the dataset to improve the biometric verification rates. Based on visual inspection, the images in the input dataset did not seem to have any Gaussian blur or salt-and-pepper noise. All the images were RGB true-color of the ocular region, so we considered color image enhancement algorithms. Visual inspection combined with evaluation of the color histograms in RGB space showed skewed histograms revealing poor usage of the available visible range. Images with compact histograms or "unequal" histograms will often have poor visual contrast or a "washed-out appearance" [6]. If the image histogram space is filled out and equalized, the image will tend to have a higher contrast and a more distinctive appearance.

We started with a full-scale histogram stretch, also sometimes called a contrast stretch, across all three visible color planes, corresponding to the Red, Green and Blue color channels. However, the improvements we observed from a simple histogram stretch were insignificant.

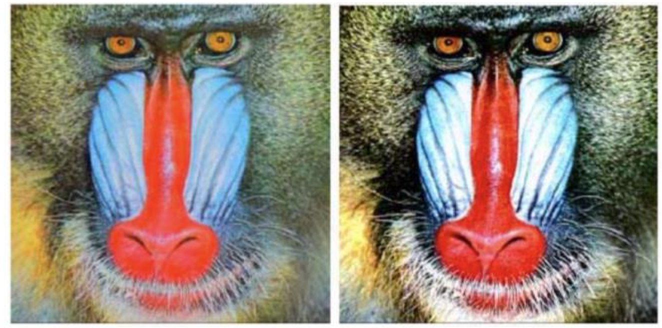
We then attempted a histogram equalization, which extends the idea of the histogram stretch: not only does the image color planes fill the available range, but also they should be uniformly distributed over that range. Although we should be careful in applying a powerful non-linear transformation to the

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

image that changes the shape of the image histogram rather than just stretching it, there are good mathematical reasons for regarding a flat histogram as a desirable goal because, theoretically, in the sense of maximum entropy, an image with a perfectly flat histogram contains the largest amount of information or complexity [6]. This technique actually changes the shape of the image histogram and can cause gaps to appear in the output histogram – which can become a problem if some of the histogram bins appear isolated, leading to false contouring in the image [6].

Building on the ideas of histogram stretching and equalization, we considered manipulation of the color spaces of the images, because the RGB color space is not always the most natural for human perception. We evaluated the Hue, Saturation and Value (HSV) color space, which is a perceptual color space and an alternative way of representing true color images in a manner that is more natural to the human perception and understanding of color than the RGB representation [7]. By examining the individual color channels in the HSV space, we see that objects in the image are more consistently contained in the hue field, despite varying lighting conditions on the scene.

We finally chose a Lumosity based color space, and the Brightness Preserving Dynamic Fuzzy Histogram Equalization (BPDFHE) image enhancement technique [1]. BPDFHE introduces modifications to other well-known histogram equalization techniques by introducing fuzzy statistics of digital images (fuzzy histogram). The imprecision in gray levels is handled well by fuzzy statistics, and the fuzzy histogram, when computed with appropriate fuzzy membership function, does not have random fluctuations or missing intensity levels and is essentially smooth. This helps in obtaining its meaningful partitioning required for brightness preserving equalization. Contrast enhancement of color images is typically done by converting the image to a color space that has image luminosity as one of its components, such as the L*a*b* color space. Contrast adjustment is performed on the luminosity layer 'L*' only, and then the image is converted back to the RGB color space. Manipulating luminosity affects the intensity of the pixels, while preserving the original colors.



Original BPDFHE

B. *Pre-Trained Models*

Several pre-trained deep learning models were tested on VISOB dataset on the left and right ocular regions for each subject [5]. Receiver Operating Characteristic (ROC) curves were utilized as the measure of accuracy of the model on the input dataset. Our objective was to maximize the area under the curve (AUC), with an AUC of 0.5 referring to pure chance decision and an AUC of 1 denoting a perfect classification.

The following pre-trained models were tested on the input dataset:

AlexNet [3]: AlexNet is a convolutional deep neural network named after, and designed by, Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton that has been trained on the approximately 1.2 million images from the ImageNet database. AlexNet is a simple yet effective deep neural network with 23 layers that can classify images into 1000 object categories.

VGG16 [4]: VGG16 (also called Oxford Net) is a convolutional neural network architecture named after its developer, Visual Geometry Group. VGG network is characterized by its simplicity, using only 3×3 convolutional layers stacked on top of each other. Reducing the volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a SoftMax classifier.

We tested both networks on a subset of the data and chose AlexNet because of the reduced training time on the hardware we had available for training. VGG16 seemed to have higher accuracy, but at the cost of significantly higher training times.

III. Experimental Validation

A. Dataset

We utilized the VISOB dataset in this study [5]. Visible light mobile Ocular Biometric (VISOB) Dataset ICIP2016 Challenge Version is a publicly available database consisting of eye images from 550 healthy adult volunteers acquired using three different smartphones (iPhone 5s, Samsung Note 4 and Oppo N1). The iPhone was set to capture bursts of still images at 720p resolution, while the Samsung and Oppo devices were capturing bursts of still images at 1080p resolution using pixel binning. Volunteers' data were collected during two visits (Visit 1 and Visit 2), 2 to 4 weeks apart. At each visit, volunteers were asked to take selfie-like captures using front facing cameras of the aforementioned three mobile devices in two different sessions (Session 1 and Session 2) that were about 10 to 15 minutes apart. The volunteers used the mobile phones naturally, holding the devices 8 to 12 inches from their faces. During each session, a number of images were captured under three lighting conditions: regular office light, dim light (office lights off but dim ambient lighting still present), and natural daylight settings (next to large sunlit windows). The collected database was preprocessed to crop and retain only the eye regions of size 240 x160 pixels using a Viola-Jones based eye detector. Figure 1 below shows sample eye images from VISOB Dataset exhibiting variations such as light and dark irises, reflection, make-up and imaging artifacts.

Figure 1 - Sample Eye Images from VISOB dataset

For the purposes of our experiments in this study, we used a random subset of 100 subjects out of the 550 total subjects in this dataset.

B. Results

- Out of the box, AlexNet had relatively poor performance. Figure 2 shows the match score distributions for genuine vs impostor matches, and Figure 3 shows the corresponding ROC curve.
- The ROC curve in Figure 3 shows a 0.9 AUC. The corresponding equal error rate (EER) of over 16%, which is not quite satisfactory.
- These figures show that the classifier is not adequately able to distinguish between genuine vs. impostor matches out of the box.
- To interpret the performance, higher AUC in the ROC curve is better, and for the EER lower is better. The EER is an operating point of the classifier where sensitivity and specificity are equal.

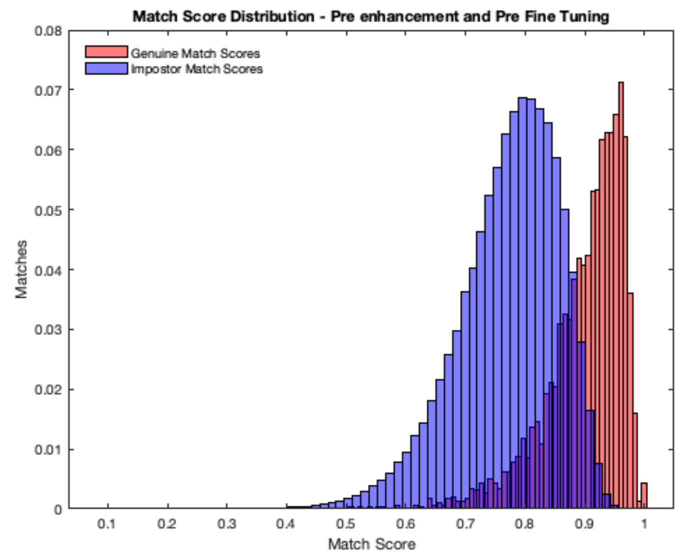


Figure 2 - Match Score Distribution (out of the box) on test data.

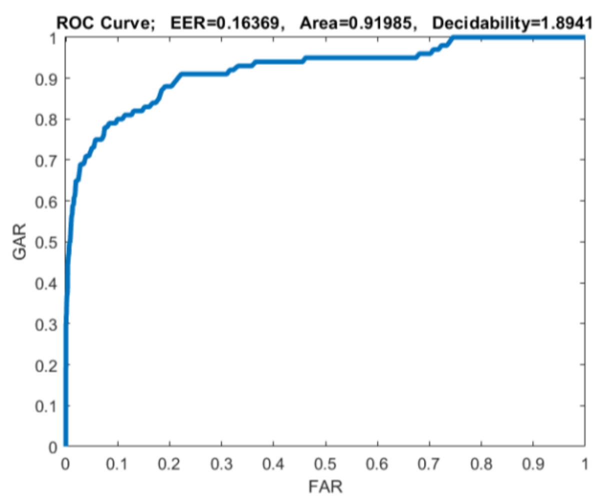


Figure 3 - ROC Curve for Out of the box AlexNet performance

- We show our efforts to resolve these using a combination of neural network fine-tuning and image enhancement techniques.
- The enhancement followed by neural network fine tuning using transfer learning shows significant improvement in AlexNet's ability to distinguish genuine vs impostor verifications. The AUC increased to over 0.99, EER decreased 4-fold and dropped below 4%, and decidability (a measure of class separability) increased from 1.89 to 4.17.

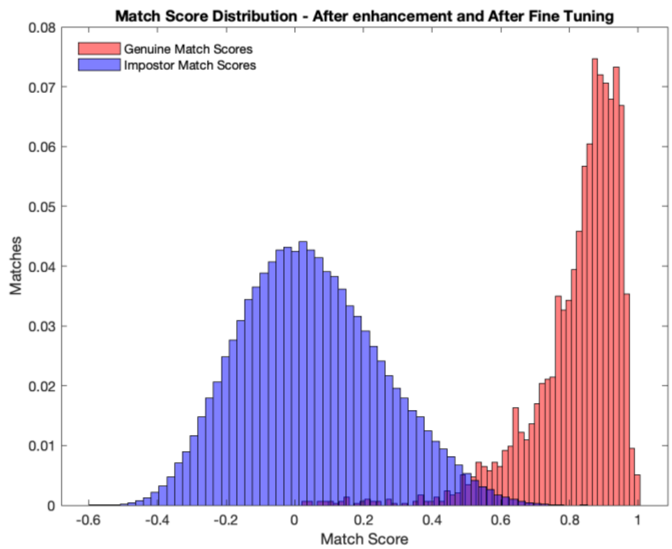


Figure 4 Match Score Distribution after fine-tuning with image enhancement on test data.

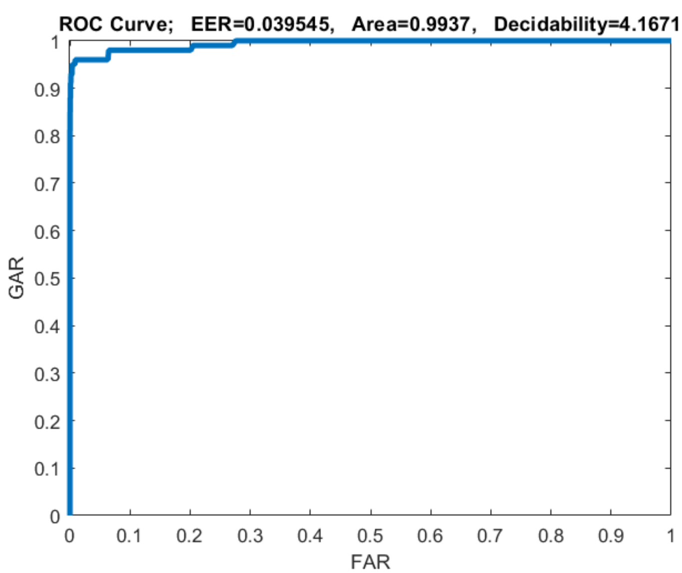


Figure 5 ROC Curve after fine tuning with enhancement on test data.

- Table 1 below summarizes the improvement achieved by using the image enhancement techniques described in this paper.

	AUC	**EER**	**Decidability**
Before Enhancement	0.91985	0.16369	1.8941
After Enhancement & fine-tuning	0.9937	0.039545	4.1671

Table 1 Summary of performance results before & after image enhancement with fine tuning (transfer learning) on AlexNet

IV. Conclusions and Future Work

We show that image contrast enhancement pre-processing techniques combined with transfer learning can be utilized to significantly increase the accuracy of pre-trained neural networks in classifying ocular images. More specifically, we showed that BPDFHE is an appropriate image contrast enhancement technique for ocular images obtained through RGB selfie captures images. The focus of this work was not building a real-world biometric system, but rather showing what image preprocessing combined with model fine-tuning can do to improve off the shelf deep neural network models' accuracy for image-based biometrics.

For future work, we would like to evaluate these image enhancement techniques in a system that supports enrollment and verification of new individuals, without re-training against all prior individuals. We would also like to explore additional motion blur reduction techniques along with fine-tuning with newer convolutional neural network models to further improve the accuracy of the proposed methods.

References

[1] D. Sheet, H. Garud, A. Suveer, J. Chatterjee and M. Mahadevappa, "Brightness Preserving Dynamic Fuzzy Histogram Equalization", *IEEE Trans., Consumer Electronics*, vol. 56, no. 4, pp. 2475 - 2480, Nov. 2010.

[2] H. Ibrahim, N.S.P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement", *IEEE Trans. Consumer Electron.*, vol. 53, pp. 1752-1758, 2007.

[3] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Advances in neural*

information processing systems, pp. 1097-1105, 2012.

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

[5] A. Rattani, R. Derakhshani, S. K Saripalle, and V. Gottemukkula, "ICIP 2016 Competition on Mobile Ocular Biometric Recognition", *IEEE Conference on Image Processing*, 320-324, 2016

[6] A. Bovik, The Essential Guide to Image Processing, 2nd Edition, Elsevier, 2009.

[7] C. Solomon and T. Breckon, Fundamentals of Digital Image Processing, Wiley-Blackwell, 2011.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE