Attribute-Based Deep Periocular Recognition:
Leveraging Soft Biometrics to Improve Periocular Recognition

Veeru Talreja, Nasser M. Nasrabadi, and Matthew C. Valenti
West Virginia University
Morgantown, WV, USA
vtalreja@mix.wvu.edu,nasser.nasrabadi@mail.wvu.edu,valenti@ieee.org

Abstract

In recent years, periocular recognition has been developed as a valuable biometric identification approach, especially in wild environments (for example, masked faces due to COVID-19 pandemic) where facial recognition may not be applicable. This paper presents a new deep periocular recognition framework called attribute-based deep periocular recognition (ADPR), which predicts soft biometrics and incorporates the prediction into a periocular recognition algorithm to determine identity from periocular images with high accuracy. We propose an end-to-end framework, which uses several shared convolutional neural network (CNN) layers (a common network) whose output feeds two separate dedicated branches (modality dedicated layers); the first branch classifies periocular images while the second branch predicts soft biometrics. Next, the features from these two branches are fused together for a final periocular recognition. The proposed method is different from existing methods as it not only uses a shared CNN feature space to train these two tasks jointly, but it also fuses predicted soft biometric features with the periocular features in the training step to improve the overall periocular recognition performance. Our proposed model is extensively evaluated using four different publicly available datasets. Experimental results indicate that our soft biometric based periocular recognition approach outperforms other state-of-the-art methods for periocular recognition in wild environments.

1. Introduction

Traditionally, facial recognition systems (in cooperative settings) are presented with mostly non-occluded faces, which include all primary facial regions such as the eyes, nose, and mouth [33, 31, 32, 29, 27, 28, 30]. Inspired by the COVID-19 pandemic response, the widespread requirement that people should wear protective face masks in public places has driven a need to understand how co-operative facial recognition technology deals with occluded faces, such as when only the periocular region is visible. In this paper, we address this challenge by implementing a periocular recognition algorithm for unconstrained wild environments (i.e., masked faces).

In the recent past, periocular recognition in wild environments has garnered significant interest and has become a key area of research in biometric recognition [13, 2, 12, 20, 34, 10, 22, 24, 38, 17]. Despite the fact that there is no strict definition or standard from professional bodies like ISO/IEC or NIST, the periocular region usually refers to the region around the eye [38] as shown in Fig. 1 [35]. The periocular region is considered to be a highly discriminative biometric modality and a powerful alternative/compliment to face and/or iris recognition when accurate face/iris recognition cannot be guaranteed due to the unconstrained environment [20] or when the whole face or iris image is not clearly available [34, 36, 38], as illustrated in Fig. 2. Moreover, it has been shown in the literature that the periocular region is more robust to expression variation [24] and aging [6], when compared to other parts of the face. Nonetheless, periocular recognition in wild environments remains a challenging task, largely because the periocular region includes less information than the entire face and is highly susceptible to interference from occlusions (hair and glasses).

When a human recognizes a face, he or she not only analyzes the overall visual pattern but also analyzes semantic information, such as gender, ethnicity, age, etc., to judge whether the face belongs to a certain known person. Therefore, it is reasonable to hypothesize that semantic information is helpful for the task of automated visual identification. In this paper, we propose to use soft biometrics as semantic information for periocular recognition. Soft biometric information extracted from the periocular region of the face (e.g., gender, ethnicity, skin color, and so on) is ancillary information, which to some extent is easily distinguished at a distance but is not fully discriminative by itself during facial or periocular recognition tasks. However, soft biometrics can be explicitly incorporated into periocular recognition
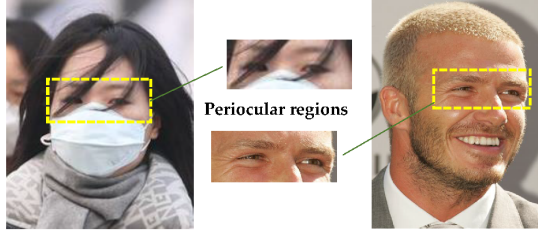
Figure 1: Periocular regions [35].

Figure 2: Masked Faces.

algorithms to improve the overall recognition performance when confronting highly variable conditions. We hypothesize that soft biometrics can provide valuable information for periocular recognition, where face images are usually captured in poor quality conditions due to variability in distance, illumination, and pose. We propose to complement 'hard' periocular facial signatures with soft biometrics to improve the overall periocular recognition performance.

We propose a new deep periocular recognition framework called Attribute-based Deep Periocular Recognition (ADPR), which simultaneously predicts soft biometrics and incorporates this ancillary information with a periocular recognition algorithm to determine identity from periocular images with higher accuracy. ADPR is an end-to-end framework, which uses several shared convolutional neural network (CNN) layers (a common network) whose output feeds into two separate branches (modality dedicated layers); the first branch classifies periocular images while the second branch predicts soft biometric attributes. Next, the features from these two branches are fused together for a final periocular recognition. Therefore, in contrast with other existing methods, which only use a shared CNN feature space to train these two tasks jointly, our proposed method fuses the predicted soft biometric features with periocular features in the training step to improve the overall periocular recognition performance. Thus, our proposed deep model predicts soft biometrics and simultaneously leverages the predicted soft biometric features as an auxiliary modality to improve periocular recognition performance. In summary, our major contributions are:

1. We design and implement an attribute-based deep periocular recognition framework, which is an end-to-end deep periocular recognition algorithm enhanced with the predicted soft biometrics (gender, ethnicity, etc.).

2. We train the proposed model to effectively predict soft biometrics while simultaneously being trained with the loss from the periocular recognition task. The model shares learned parameters to train both tasks and also fuses soft biometric information with the periocular features to improve the overall periocular recognition performance.

3. We present an ADPR-based Siamese architecture using contrastive loss, which makes it possible for the ADPR to be used for cross-dataset periocular recognition, which implies that the ADPR does not require training samples from target datasets.

2. Related Work

Periocular recognition in unconstrained conditions has been the subject of ongoing research. Surveys of various periocular recognition algorithms can be found in [1, 11, 18]. A seminal paper investigating the feasibility of using the periocular region for human recognition under various conditions is given by Park *et al*. [13]. Further research in this area is found in [2], which illustrates the usefulness of periocular recognition when iris recognition fails. There has also been research on cross-spectrum periocular matching [22], which are based on neural network techniques.

Further research into periocular recognition has focused on using hand-crafted features [10, 24, 25]. For example, [10] extracts periocular features using DSIFT and exploits K-means clustering for dictionary learning and representation. However, DSIFT feature extraction and K-mean clustering are computationally expensive and time-consuming. In [24], a Periocular Probabilistic Deformation Model (PPDM) was proposed, which utilizes a probabilistic inference model to evaluate the matching scores from correlation filters on periocular image pairs. The same research group later improved PPDM by selecting discriminative patch regions for more accountable matching [25]. However, both of the patch-based matching schemes [24, 25] are less resistant to scale variance among samples that often exists in challenging forensic and security scenarios [38].

Recent developments in periocular recognition techniques are more focused on deep learning-based methods [38, 17]. Proenca and Neves [17] proposed Deep-PRWIS, where a deep CNN model is trained in such a way that the recognition is exclusively based on information surrounding the eye, and the iris and sclera regions features are degraded during learning. Zhao and Kumar [38] proposed a deep-learning based model called Semantics-Assisted CNN (SCNN), which incorporates explicit semantic information

(gender and side) of the training samples to extract more comprehensive periocular features, helping to improve the CNN's performance. While the proposed ADPR framework is motivated by SCNN, and similarly leverages soft biometrics for periocular recognition, it additionally fuses soft biometric features with periocular features in a joint learning framework to predict soft biometrics and also improve periocular recognition accuracy.

3. Proposed Method

The proposed end-to-end deep learning framework is shown in Fig. 3. It consists of four important blocks: The backbone network, the only periocular recognition (PR) block (green dashed section), the soft biometric classifier (orange dashed section), and the joint periocular recognition (JPR) block (red dashed section). Specifically, the architecture uses several CNN layers for the backbone network whose output feeds into two separate branches (modality specific layers); the first branch is for only periocular recognition while the second branch is for soft biometric classification. The features from the soft biometric classifier and the PR block are fused in the JPR block for a final joint periocular recognition.

The backbone network is formed by only the convolutional layers of VGG-16 [23] pretrained on ImageNet [4]. ResNet[5] can also be used in place of VGG-16 for the backbone network. For the PR block and soft biometric classifier, a new set of fully connected layers (green and orange dashed sections) are integrated with the backbone network. During optimization, the periocular layers (green) are optimized for only periocular recognition followed by optimization of the soft biometric layers (orange) for only soft biometric prediction.

For the JPR block, the features from the periocular layers are fused with the features from the soft biometric layers (shown in bold black arrows) in the fusion layer followed by additional fully connected layers and a Softmax layer for periocular recognition again. However, this time the periocular recognition uses features from both the soft biometric layers and the periocular layers. This is why it is called the *joint* periocular recognition block. Finally, the whole network (except the Softmax layer in PR block) is trained end-to-end for joint periocular recognition and soft biometric classification, and the loss from both the joint periocular recognition and soft biometric classification is back propagated through the network. The novelty is that, because of this feature fusion and loss propagation, the soft biometrics features enhance the discriminative power of the periocular recognition network and therefore improve the overall periocular recognition performance. The implementation details for each block of the proposed architecture are defined in the following subsections.
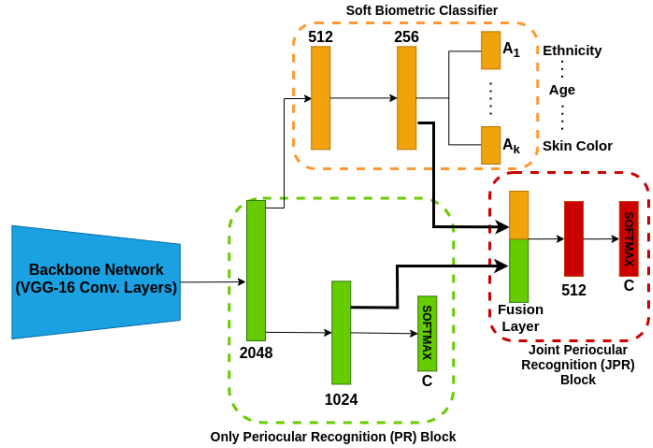

Figure 3: Attribute-based Deep Periocular Recognition (ADPR). The rectangular solid colored boxes with numbers depict fully connected layers.

3.1. Backbone Network

The backbone network uses the 14 convolutional layers from the VGG-16 architecture pre-trained on the ImageNet dataset as a starting point. The fully connected layers of the VGG-16 are discarded and not part of the backbone network. There are multiple reasons for using the VGG-16 pre-trained on the ImageNet dataset for the backbone network. In the proposed architecture, the backbone network is only used as feature-extractor for periocular recognition and soft biometric classification, and it can be seen from the previous literature [9, 39, 8, 21, 26, 14] that the features provided by VGG-16 pre-trained on ImageNet and fine-tuned on different biometric image data are very discriminative and therefore can be used as a starting point. Moreover, starting with a well-known architecture makes the work highly reproducible.

3.2. Only Periocular Recognition Block

For the PR block, a new set of fully connected layers (green blocks in the green dashed section) for only periocular recognition are integrated with the backbone network. Specifically, we add two fully connected layers with the first layer having 2,048 nodes and the second layer having 1,024 nodes. The final layer in the PR block is the Softmax layer, where the size of the layer depends on the number of classes C in the given dataset.

3.3. Soft Biometric Classifier

In addition to improving periocular recognition, another important objective of our proposed method is effective prediction of soft biometrics using only the periocular region. However, separating these two objectives by training multiple CNNs individually is not optimal since different objectives may share common features and have hidden re-

lationship, which can be leveraged to jointly optimize the objectives. This notion of joint optimization has been used in [40], where they train a CNN for face recognition, and utilize the features for attribute prediction.

Therefore, for this task, we use the feature set output by the first fully connected layer (size 2,048) of the PR block to also predict the soft biometrics for a given periocular image. For the soft biometric classifier we add two fully connected layers, a first layer of size 512 and a second layer of size 256, as shown in Fig. 3. The final layer in the soft biometric classifier is the set of k binary classifiers, where k denotes the number of soft biometrics predicted.

3.4. Joint Periocular Recognition

After optimizing the soft biometric classifier and PR block individually, the features from the soft biometric layers are fused with the features from the periocular layers at the fusion layer of the JPR block. The fusion layer is followed by additional fully connected layers and a Softmax layer for the final periocular recognition. Thus, the final periocular recognition utilizes the fused features from the periocular layers and the soft biometric layers. The whole system is then trained end-to-end and the loss from both the joint periocular recognition block and soft biometric classification is back propagated through the network.

In our implementation, we use the output features from the second fully connected layer (size 256) of the soft biometric classifier and vertically concatenate it with the features of the second fully connected layer (size 1,024) of the PR block. This vertical concatenation is the fusion layer. The output of the fusion layer is of size 1,280. In addition to this fusion layer, we add another fully connected layer of size 512 followed by a Softmax layer, whose output is equal to the number of classes C in the given dataset.

4. Training of the Proposed Architecture

The proposed architecture is trained in three steps:

1. In the first step, only the PR block is optimized for the periocular classification of the input images. Consider that the input image to ADPR is a periocular image denoted by x^i, where the class label for the image is given by $y^i \in C$ for $i = 1, \cdots, N$ where N is the number of training images in a mini-batch. In this step, we only optimize the PR block's layers for classification while keeping the backbone network frozen. The classification formulation has been incorporated into the PR block by adding the Softmax layer as shown in Fig. 3. Let E_1 denote the objective function required for classification:

$$E_1(\mathbf{w}_{PR}) = \frac{1}{N} \sum_{i=1}^{N} L_i(f(x^i, \mathbf{w}_{PR}), y^i) + \lambda ||\mathbf{w}_{PR}||^2, \quad (1)$$

where the first term $L_i(.)$ is the classification loss for training instance i and N is the number of training images in a

mini-batch. $f(x^i, \mathbf{w}_{PR})$ is the predicted softmax output of the PR network and is a function of the input training image x^i and the weights of the PR network $\mathbf{w}_{PR}$. The last term is the regularization function where λ governs the relative importance of the regularization.

The choice of the loss function $L_i(.)$ depends on the application. We use a classification loss function that uses softmax outputs to minimize the cross-entropy error function. Let the predicted softmax output $f(x^i, \mathbf{w}_{PR})$ be denoted by $\hat{y}^i$. The classification loss for the ith training instance is given as:

$$L_i(\hat{y}^i, y^i) = - \sum_{m=1}^{C} y_m^i \ln \hat{y}_m^i, \quad (2)$$

where y_m^i and $\hat{y}_m^i$ are the ground truth and the prediction result for the mth softmax output of the ith training instance, respectively and C is the number of softmax outputs.

2. In the second step, only the soft biometric classifier block is optimized for the prediction of the soft biometrics. For the soft biometric prediction task, a periocular image is given as input to predict a set of soft biometrics. The periocular image is again denoted by x^i, the number of different soft biometrics being predicted is denoted by k, and A_t^i denotes the ground truth soft biometric label for training sample i and attribute t for $t = 1, \cdots, k$. In this case, using the feature set from the second fully connected layer (size 256), the soft biometric prediction loss function is denoted by E_2, and is given as:

$$E_2(\mathbf{w}_t) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{k} L_t^i(f^t(z_1(x^i) \times \mathbf{w}_t), A_t^i), \quad (3)$$

where $f^t(.)$ is a binary classifier for the soft biometric t operating on the feature set from the second fully connected layer (size 256) of the soft biometric classifier. The classifier is learned by using a binary cross-entropy loss function $L_t^i(.)$. $\mathbf{w}_t$ represents the weight parameters for the classifier, and these parameters are learned separately for each soft biometric attribute.

3. In the final step, the JPR block and the soft biometric classifier are optimized together. The loss function used to train the network in this step is a combination of E_1 and E_2, which implies it is a combination of classification loss and soft biometric prediction loss functions. Let E_3 denote this combination given as:

$$E_3(\mathbf{w}_{JPR}, \mathbf{w}_t) = E_1(\mathbf{w}_{JPR}) + E_2(\mathbf{w}_t), \quad (4)$$

where $\mathbf{w}_{JPR}$ signifies that the classification loss formulation E_1 is a function of the JPR block layers which includes the fused feature vector of the soft biometric classifier features and the periocular recognition features. Finally, the whole network (except the Softmax layer in the PR block) is trained end-to-end using the E_3 loss function.

For each training step, we have used the stochastic gradient descent (SGD) algorithm, with a batch size of 64 samples. The learning rate was 10^{-3}, with a momentum of 0.9 and a weight decay of 5×10^{-4}.

5. Open vs. Closed-World Setting

When CNNs are used for recognition, it is important to understand and decide if the implemented system is expected to work in an open-world or closed-world mode; i.e., if the system has access to all the training time samples from all the classes that will be seen at the testing time or not. In a scenario where the CNNs are trained in a classification protocol (i.e., the identity or the category of the input data is known), the closed-world mode can be enabled and the output of the nodes in the final Softmax layer can be used as probabilities for each class label. However, in the open-world mode, one-to-one matching for probably unseen subjects is the key problem and needs to be evaluated. Therefore, in the open-world setting, the model needs to be trained to be able to generalize to unseen subjects that are not included in the training set. In this paper, we have evaluated our proposed architecture for both open-world and closed-world settings.

5.1. Siamese Network for Open-World Setting

For the open-world setting, we use a Siamese network [3] to train the ADPR, generalize to unseen subjects, and predict the similarity between a pair of feature vectors. The Siamese network is primarily designed for verification scenarios. The ADPR-based Siamese network architecture is shown in Fig. 4. The Siamese network requires genuine and impostor pairs of images. In the Siamese architecture, we use identical ADPR networks with shared weights for both the input images, and these two networks are coupled together using a contrastive loss function (L_{cont}). We use the output of the fully connected layer (size 512) of the JPR block in ADPR as the feature representation of the input data and for coupling the two networks using L_{cont}. While the Softmax layer in the JPR block represents the class probabilities during the training process, the second-to-last layer should contain the most relevant and aggregated information that can contribute to distinguishing the classes or identities [38]. Therefore, it is reasonable to use the output of the fully connected layer (size 512) of the JPR block as the feature representation and generalize the model to unseen subjects.

The loss function L_{cont} is minimized so as to drive the genuine pairs (i.e., both the periocular images belonging to the same subject) towards each other in the feature domain, and at the same time, push the impostor pairs (i.e., the periocular images belonging to different subjects) away from each other. Let x^i denote the one input periocular image, and x^j denote the other input periocular image as shown in
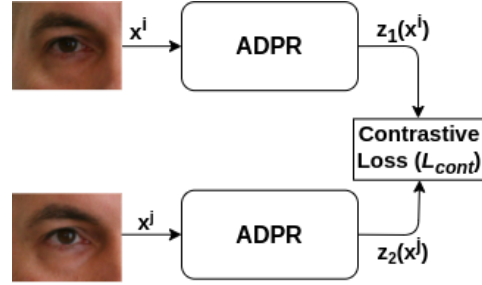

Figure 4: ADPR-based Siamese

Fig. 4. Define $c(i, j)$ to be a binary label, which is equal to 0 if x^i and x^j belong to a genuine pair, and equal to 1 if x^i and x^j belong to an impostor pair. Let $z_1(.)$ and $z_2(.)$ denote the ADPR network to transform x^i and x^j, respectively, into their feature representation of size 512. If $c(i, j) = 0$ (i.e., genuine pair), the contrastive loss function is:

$$L_{cont}(z_1(x^i), z_2(x^j), c(i, j)) = \frac{1}{2} \left\| z_1(x^i) - z_2(x^j) \right\|_2^2.$$
(5)

Similarly if $c(i, j) = 1$, then contrastive loss function is:

$$L_{cont}(z_1(x^i), z_2(x^j), c(i, j)) =$$
$$\frac{1}{2}\max\left(0, m - \left\| z_1(x^i) - z_2(x^j) \right\|_2^2 \right),$$
(6)

where m is the contrastive margin used to "tighten" the constraint. Therefore, the total loss function for training the Siamese architecture is denoted by L_{cpl} and is given as:

$$L_{cpl} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} L_{cont}(z_1(x^i), z_2(x^j), c(i, j)),$$
(7)

where N is the number of training samples. The main motivation for using the coupling loss is that it has the capacity to find the discriminative embedding subspace because it uses the class labels implicitly, which may not be the case with some other metrics such as the Euclidean distance.

Our approach for periocular recognition in the open-world setting using the trained Siamese ADPR network does not require that the training and testing samples belong to the same dataset, which is a key advantage over other state-of-the-art (SOTA) approaches [34, 17, 24]. In our experiments for the open-world setting, the ADPR is trained with one database and tested on a separate dataset. The testing and training sets have mutually exclusive subjects and highly different image qualities, conditions, and/or sensors.

6. Performance Evaluation

In this section, we discuss the datasets and the performance evaluation of our proposed architecture when com-

pared to other state-of-the-art methods. For fair comparison, we have maintained consistency with [17] and [38] with respect to the dataset and the testing protocol.

6.1. Datasets

We have used the following publicly available databases for the experiments:

1) **UBIPr** [12]: We have used the UBIPr database only for training our proposed ADPR architecture in the open-world setting. Originally, this database contains 5,126 images for each of the left and right perioculars from 344 subjects. However, following the protocol in [38], we have removed those subjects that are also in UBIRISV2 [16]. Finally, for each of the left and right perioculars, we end up with 3,359 images from 224 subjects.

2) **UBIRISV2** [16]: The UBIRISV2 database contains periocular images, and is mainly used for assessment of at-a-distance iris recognition algorithms under visible illumination and challenging imaging environments [38]. We have used this dataset to evaluate the performance of both open-world and closed-world settings. For the open-world setting, using the protocol defined in [38], we have used a subset of about 1,000 images from this dataset for the performance evaluation. This subset contains both left and right periocular images corresponding to 161 subjects, that are captured from a distance of 3 to 8 meters. For the closed-world setting, we have made our dataset protocol consistent with [17] and used all 11,100 images corresponding to 522 different eyes for the training and evaluation of the proposed architecture. The UBIPr and UBIRISV2 datasets are annotated with only gender as a soft biometric.

3) **Face Recognition Grand Challenge (FRGC)** [15]: The FRGC dataset consists of full facial images. We have used Multi-task Cascaded Convolutional Networks (MTCNN) [37] to crop out left and right periocular images. We have used the FRGC dataset for both open-world and closed-world settings. For the open-world setting, consistent with [38], we have used only right periocular images from the Fall 2002 subset with subject IDs from 202-269 to 202-317. This corresponds to about 540 right eye images from 163 subjects. For the closed-world setting, we have used the Spring 2004 subset which consists of about 25,000 images corresponding to 690 classes. The FRGC dataset is annotated with only gender and ethnicity as soft biometrics.

For open-world setting experiments, it is important to clarify the difference in the training techniques for four methods (Our proposed approach and the three comparable approaches from [38, 24, 34]). For our approach and [38], the model is trained on the UBIPr database and tested on the UBIRISV2 and FRGC databases, which is identical to the train/test configuration in [38] and therefore provides a fair comparison. The other two methods [24, 34] require within-database training and testing and this offers better

results for these two methods. Therefore, the training and testing are performed on the same dataset for them. Additionally, for fair comparison, the train/test configuration for these two methods is similar to the configuration in [38].

6.2. Open-World Performance

We first examine the advantage of the fusion of soft biometric features and periocular recognition features in the JPR block. We have compared the performance of a periocular recognition at the output of the PR block (ADPR (PR)) with the performance of a periocular recognition at the output of the JPR block (ADPR (JPR)). We have used the proposed Siamese architecture with contrastive loss, which is designed for open-world verification. The model is trained on the UBIPr dataset and tested on the UBIRISV2, and FRGC datasets. When training with contrastive loss, the margins are discretely tuned in the range [0.5, 5], and the results providing best performance are used for comparison. The results from the verification experiments using receiver operating characteristic (ROC) curves are illustrated in Fig. 5. It can be observed that the proposed ADPR (JPR) consistently outperforms the output of the PR block (ADPR (PR)). This observation suggests that, due to the feature fusion and back propagation of the combined loss function, the soft biometrics are directly influencing the discriminative power of the network and therefore improving the overall periocular recognition performance.

We also compared the performance of our approach with other SOTA approaches [38, 24, 34] on the periocular recognition problem. We have made the test protocols consistent with [38]. The verification results (ROC) for these comparisons are also shown in Fig. 5. It can be observed from Fig. 5 that the proposed approach using ADPR consistently outperforms the three SOTA approaches; SCNN [38], PPDM [24], Texton [34]. For instance, in case of the UBIRISV2 dataset, the EER for our proposed ADPR (JPR) is improved by 3% over the second best method SCNN [38].

6.3. Closed-World Performance

As discussed earlier, the proposed Siamese ADPR network is primarly designed for open-world verification. However, some recent techniques [19, 17] also focus on the closed-world setting, where there is an overlap of subjects during training/testing phase. Under the closed-world setting, we have maintained the ADPR architecture from Fig. 3. The size of the Softmax layer C is consistent during training and test phases in this closed-world setting. The output for each neuron of the Softmax layer is considered the probability that the input sample belongs to a specific subject, and therefore is used as the verification score. Fig. 6 provides ROCs for the verification results under the closed-world setting on UBIRISV2 and FRGC, and includes a comparison to the performance when using only
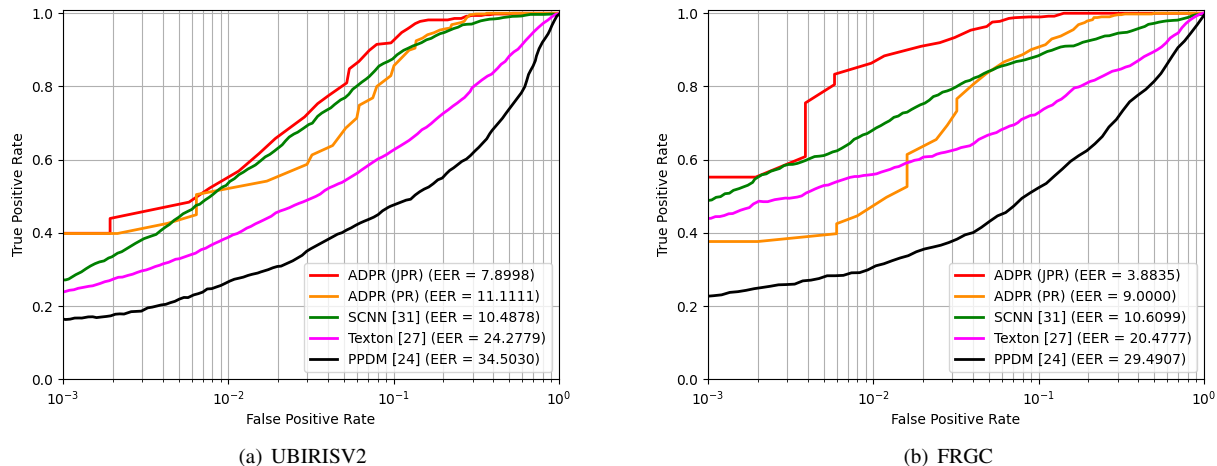
(a) UBIRISV2

(b) FRGC

Figure 5: ROC curve comparison for open-world setting against the output of PR block and other state-of-the-art methods for two different datasets.

the PR block (ADPR (PR)), and three other SOTA methods.

From Fig. 6, we observe that our approach consistently outperforms the recent SOTA method DEEP-PRWIS [17]. Under the closed-world settings, our results have scored significantly low EER (1.73%). This is due to the fact that, with the feature fusion and back propagation of the combined loss function, the soft biometric features enhance the discriminative power of the network and therefore improve the overall periocular recognition performance.

We have also evaluated the Area Under the Curve (AUC), Rank-1 accuracy, and Equal Error Rate (EER) of our proposed system and compared it against the earlier SOTA methods [38, 34, 17]. Table 1 summarizes the performance observed in our experiments, for the three algorithms and two data sets considered. The proposed method outperforms the earlier methods, with the true identity being reported at the first position (Rank-1) over 92.68% of the time. In all performance measurements, the differences with respect to the second-best method [17] is evident.

6.4. Soft Biometric Prediction

In this section, we evaluate the effect of fusing the soft biometric features with the periocular recognition features on periocular recognition performance and on soft-biometric prediction performance. In this experiment, we evaluated the soft biometric prediction performance using classification accuracy before the fusion of the features (ADPR (SB)) and also after the feature fusion with end-to-end model training (ADPR (JPR)). This experiment has been performed on the UBIRISV2 and the FRGC datasets under the closed-world setting. Additionally, for the FRGC dataset we have compared the performance with another

Table 1: Performance comparison between proposed method and other sate-of-the-art methods for closed-world setting. Values given are percentages.

Method	AUC	Rank-1	EER
UBIRISV2			
ADPR (JPR)	**99.9 ± 0.02**	**92.68 ± 1.04**	**1.73 ± 0.04**
ADPR (PR)	98.2 ± 0.07	83.95 ± 0.87	4.85 ± 0.12
DEEP-PRWIS [17]	99.8 ± 0.04	87.64 ± 1.68	1.89 ± 0.04
SCNN [38]	98.6 ± 0.05	79.3 ± 1.6	5.4 ± 0.04
Texton [34]	84.3 ± 0.09	64.6 ± 2.3	17.57 ± 0.32
FRGC			
ADPR (JPR)	**99.9 ± 0.01**	**93.65 ± 0.88**	**0.58 ± 0.02**
ADPR (PR)	98.7 ± 0.04	88.25 ± 0.87	2.19 ± 0.09
DEEP-PRWIS [17]	99.9 ± 0.04	92.05 ± 0.92	0.71 ± 0.02
SCNN [38]	99.1 ± 0.04	89.2 ± 0.79	2.25 ± 0.04
Texton [34]	94.1 ± 0.05	73 ± 2.1	7.9 ± 0.45

earlier method based on LBP features [7].

Table 2 provides the comparisons of gender accuracy and ethnicity accuracy for the different models discussed above. It can be observed from Table 2 that after feature fusion, there is an improvement of at least 1.5% in soft biometric prediction accuracy. This improvement can be attributed to the feature fusion and end-to-end model training of the ADPR, where the periocular recognition features enhance the soft biometric discriminative space and therefore improve the soft biometric prediction accuracy.

6.5. Effect of the Number of Soft Biometrics on Periocular Recognition

We have also evaluated the effect of the number of soft biometrics being used on the periocular recognition performance. We have performed this experiment on the FRGC
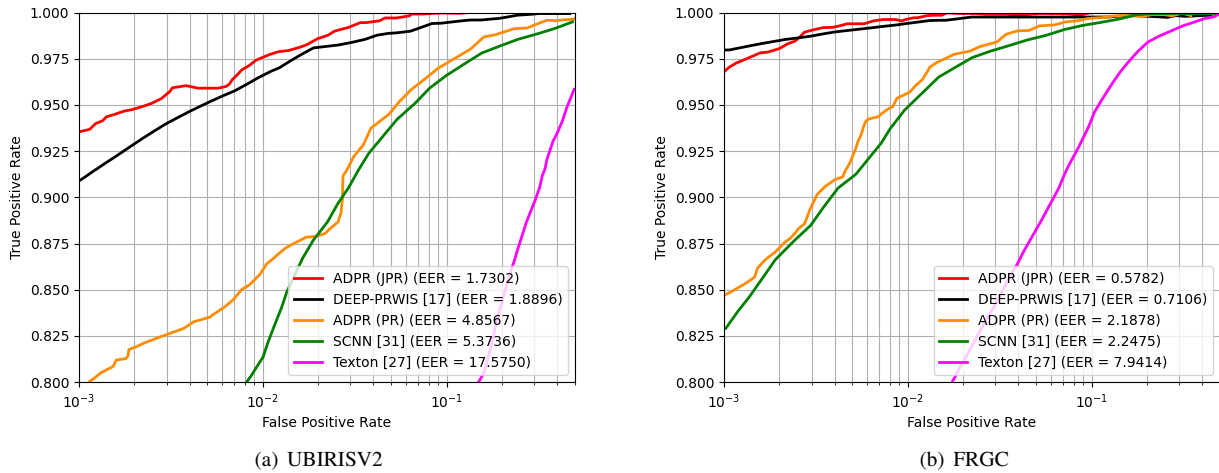
(a) UBIRISV2

(b) FRGC

Figure 6: ROC curve comparison for closed-world setting against the ADPR (PR) and other state-of-the-art methods.

Table 2: Soft biometric prediction performance comparison.

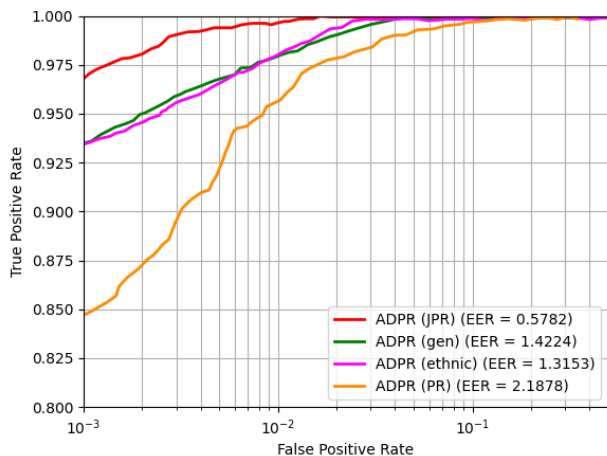| Method	Gender Acc.	Ethnicity Acc.
UBIRISV2		
ADPR (JPR)	**97.6**	-
ADPR (SB)	95.76	-
FRGC		
ADPR (JPR)	**97.5**	**98.7**
ADPR (SB)	96.1	97.1
LBP [7]	92.5	94.5

Figure 7: Performance comparison for different number of soft biometrics.

dataset under a closed-world setting. For this experiment,

we have trained the ADPR model using only one soft biometric (gender or ethnicity) and tested it for verification. We have compared the performance of no soft biometric fusion (ADPR (PR)), only the gender being used as the soft biometric (ADPR (gender)), only the ethnicity being used as the soft biometric (ADPR (ethnicity)), and both gender and ethnicity being used (ADPR (JPR)). Fig. 7 provides the ROC curve comparison for different models. It can be observed from Fig. 7 that fusion of soft biometrics with the periocular features helps to improve the overall performance and also that using more soft biometric attributes improves the performance. This is because with more soft biometrics, it becomes easier for the network to the learn the discriminative space, which improves the overall performance.

7. Conclusion

We have presented a periocular recognition framework based on a convolutional neural network (CNN) architecture and the fusion of soft biometric features with periocular features. The utility of this framework is that, due to the fusion of soft biometrics and periocular features, along with end-to-end model training, the soft biometric features enhance the discriminative power of the network and therefore improve the overall periocular recognition performance. We observed an improvement in EER of at least 3% in the open-world setting verification performance and an improvement in Rank-1 accuracy of at least 2% in the closed-world setting, when compared to the state-of-the-art methods. We have also evaluated the soft biometric prediction performance and observed an improvement of at least of 1.5% in accuracy due to the fusion of periocular features with the soft biometric features.

References

[1] Fernando Alonso-Fernandez and Josef Bigun. A survey on periocular biometrics research. *Pattern Recognit. Lett.*, 82:92–105, 2016.

[2] Samarth Bharadwaj, Himanshu S. Bhatt, Mayank Vatsa, and Richa Singh. Periocular biometrics: When iris recognition fails. In *Proc. IEEE Int'l Conf. on Biometric Theory, Appl. and Sys.*, 2010.

[3] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 539–546 vol. 1, 2005.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[6] Felix Juefei-Xu, Khoa Luu, Marios Savvides, Tien D. Bui, and Ching Y. Suen. Investigating age invariant face recognition based on periocular biometrics. In *Proc. Int'l Joint Conf. on Biometrics*, 2011.

[7] Jamie R. Lyle, Philip E. Miller, Shrinivas J. Pundlik, and Damon L. Woodard. Soft biometric classification using periocular region features. In *Proc. IEEE Int'l Conf. on Biometric Theory, Appl. and Sys.*, 2010.

[8] Shervin Minaee, Amirali Abdolrashidiy, and Yao Wang. An experimental study of deep convolutional features for iris recognition. In *Proc. IEEE Signal Processing in Medicine and Biology Symposium*, 2016.

[9] Kien Duy Nguyen, Clinton Fookes, Arun Ross, and Sridha Sridharan. Iris recognition with off-the-shelf CNN features: A deep learning perspective. *IEEE Access*, 6:18848–18855, 2017.

[10] Lei Nie, Ajay Kumar, and Song Zhan. Periocular recognition using unsupervised convolutional rbm feature learning. In *Proc. IEEE Int'l Conf. on Pattern Recognition*, pages 399–404, 2014.

[11] Ishan Nigam, Mayank Vatsa, and Richa Singh. Ocular biometrics: A survey of modalities and fusion approaches. *Information Fusion*, 26:1–35, 2015.

[12] Chandrashekhar N. Padole and Hugo Proenca. Periocular recognition: Analysis of performance degradation factors. In *Proc. IAPR Int'l Conf. on Biometrics*, pages 439–445, 2012.

[13] Unsang Park, Arun Ross, and Anil K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *Proc. IEEE Int'l Conf. on Biometric Theory, Appl. and Sys.*, 2009.

[14] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. *Proc. of the British Machine Vision Conf.*, 2015.

[15] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the face recognition grand challenge. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 947–954, 2005.

[16] Hugo Proenca, Silvio Filipe, Ricardo Santos, Joao Oliveira, and Luis A. Alexandre. The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(8):1529–1535, 2010.

[17] Hugo Proença and João C. Neves. Deep-PRWIS: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Trans. on Information Forensics and Security*, 13(4):888–896, 2018.

[18] Ajita Rattani and Reza Derakhshani. Ocular biometrics in the visible spectrum: A survey. *Image and Vision Computing*, 59:1–16, 2017.

[19] Ajita Rattani, Reza Derakhshani, Sashi K. Saripalle, and Vikas Gottemukkula. ICIP 2016 competition on mobile ocular biometric recognition. In *Proc. IEEE Int'l Conf. on Image Processing*, pages 320–324, 2016.

[20] Gil Santos and Hugo Proença. Periocular biometrics: An emerging technology for unconstrained scenarios. In *Proc. IEEE Symposium on Computational Intelligence in Biometrics and Identity Management*, pages 14–21, 2013.

[21] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 815–823, 2015.

[22] Anjali Sharma, Shalini Verma, Mayank Vatsa, and Richa Singh. On cross spectral periocular recognition. In *Proc. IEEE Int'l Conf. on Image Processing*, pages 5007–5011, 2014.

[23] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proc. Int'l Conf. on Learning Representations*, 2015.

[24] Jonathon M. Smereka, Vishnu Naresh Boddeti, and B. V. K. Vijaya Kumar. Probabilistic deformation models for challenging periocular image verification. *IEEE Trans. on Information Forensics and Security*, 10(9):1875–1890, 2015.

[25] Jonathon M. Smereka, B. V. K. Vijaya Kumar, and Andres Rodriguez. Selecting discriminative regions for periocular verification. In *Proc. IEEE Int'l Conf. on Identity, Security and Behavior Analysis*, 2016.

[26] Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang. DeepID3: Face recognition with very deep neural networks. *CoRR*, abs/1502.00873, 2015.

[27] Fariborz Taherkhani, Veeru Talreja, Jeremy Dawson, Matthew C. Valenti, and Nasser M. Nasrabadi. Pf-cpgan: Profile to frontal coupled gan for face recognition in the wild. In *Proc. Int'l Joint Conf. on Biometrics*, Oct. 2020.

[28] Fariborz Taherkhani, Veeru Talreja, Matthew C. Valenti, and Nasser M. Nasrabadi. Error-corrected margin-based deep cross-modal hashing for facial image retrieval. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(3):279–293, Apr. 2020.

[29] Veeru Talreja, Terry Ferrett, Matthew C Valenti, and Arun Ross. Biometrics-as-a-service: A framework to promote innovative biometric recognition in the cloud. In *Proc. IEEE International Conference on Consumer Electronics*, Jan. 2018.

[30] Veeru Talreja, Sobhan Soleymani, Matthew C. Valenti, and Nasser M. Nasrabadi. Learning to authenticate with deep

multibiometric hashing and neural network decoding. In *Proc. IEEE Int'l Conf. on Commun.*, May 2019.

[31] Veeru Talreja, Fariborz Taherkhani, Matthew C. Valenti, and Nasser M. Nasrabadi. Attribute-guided coupled GAN for cross-resolution face recognition. In *Proc. IEEE Int'l Conf. on Biometric Theory, Appl. and Sys.*, Sept. 2019.

[32] Veeru Talreja, Matthew C Valenti, and Nasser M Nasrabadi. Zero-shot deep hashing and neural network based error correction for face template protection. In *Proc. IEEE Int'l Conf. on Biometric Theory, Appl. and Sys.*, Sept. 2019.

[33] Veeru Talreja, Matthew C. Valenti, and Nasser M. Nasrabadi. Deep hashing for secure multimodal biometrics. *IEEE Trans. on Information Forensics and Security*, 16:1306–1321, 2021.

[34] Chun-Wei Tan and Ajay Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *IEEE Trans. on Image Processing*, 22(10):3751–3765, 2013.

[35] Leslie Ching Ow Tiong, Yunli Lee, and Andrew Beng Jin Teoh. Periocular recognition in the wild: Implementation of RGB-OCLBCP dual-stream CNN. *Applied Sciences*, 9(13), 2019.

[36] Damon L. Woodard, Shrinivas Pundlik, Philip Miller, Raghavender Jillela, and Arun Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *Proc. IEEE Int'l Conf. on Pattern Recognition*, pages 201–204, 2010.

[37] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multi-task cascaded convolutional networks. *CoRR*, abs/1604.02878, 2016.

[38] Zijing Zhao and Ajay Kumar. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Trans. on Information Forensics and Security*, 12(5):1017–1030, 2017.

[39] Zijing Zhao and Ajay Kumar. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 3809–3818, 2017.

[40] Yang Zhong, Josephine Sullivan, and Haibo Li. Face attribute prediction using off-the-shelf CNN features. In *Proc. IAPR Int'l Conf. on Biometrics*, 2016.