


## Article

# Periocular Recognition in the Wild: Implementation of RGB-OCLBCP Dual-Stream CNN

Leslie Ching Ow Tiong <sup>1,†</sup>, Yunli Lee <sup>2,†</sup> and Andrew Beng Jin Teoh <sup>3,\*,†</sup> 

<sup>1</sup> Computational Science Research Center, Korea Institute of Science and Technology (KIST), Building L0243 14 gil, 5 Hwarangro, Seongbukgu, Seoul 02792, Korea

<sup>2</sup> School of Science and Technology, Sunway University, 5 Jalan Universiti, Bandar Sunway, Petaling Jaya 47500, Selangor, Malaysia

<sup>3</sup> School of Electrical and Electronic Engineering, Yonsei University, 50 Yonsei-ro, Sinchon-dong, Seodaemun-gu, Seoul 03722, Korea

\* Correspondence: bjteoh@yonsei.ac.kr

† These authors contributed equally to this work.

Received: 2 June 2019; Accepted: 1 July 2019; Published: 3 July 2019



**Featured Application:** The proposed periocular biometric network can apply to any application that requires identity management, such as homeland security, border controls, access control, criminal investigation, etc.

**Abstract:** Periocular recognition remains challenging for deployments in the unconstrained environments. Therefore, this paper proposes an RGB-OCLBCP dual-stream convolutional neural network, which accepts an RGB ocular image and a colour-based texture descriptor, namely Orthogonal Combination-Local Binary Coded Pattern (OCLBCP) for periocular recognition in the wild. The proposed network aggregates the RGB image and the OCLBCP descriptor by using two distinct late-fusion layers. We demonstrate that the proposed network benefits from the RGB image and the OCLBCP descriptor can gain better recognition performance. A new database, namely an Ethnic-ocular database of periocular in the wild, is introduced and shared for benchmarking. In addition, three publicly accessible databases, namely AR, CASIA-iris distance and UBIPr, have been used to evaluate the proposed network. When compared against several competing networks on these databases, the proposed network achieved better performances in both recognition and verification tasks.

**Keywords:** periocular recognition in the wild; convolutional neural network; colour-based local binary coded pattern

## 1. Introduction

Biometric systems have been widely deployed since the late 1990s worldwide for identity management, banking, homeland security, etc. [1]. Among different biometric systems, face recognition enjoys flexibility, availability, and user-friendly [2]. However, biometrics experts and the police departments of the United States have agreed that the face recognition technology remains challenging after the “Boston Marathon bombings” in 2013 [3]. For instance, the appearances of subjects such as cosmetic products, plastic surgery or wearing masks may cause the failure of identifying the suspects. To hinder the complexity of the facial region, periocular recognition is gaining attention these days attributed to its promising recognition performance [4].

What does periocular refer to? According to the definition in [5], periocular defines the region around the eyes, which includes the eyelids, eyelashes, and eyebrows (see Figure 1). The periocular

region demonstrates more tolerance of variability in expression and occlusion, such as crime scene where perpetrators intentionally mask part of their faces. This creates more capability of matching partial faces [6,7]. In addition, due to the rapid growth of camera use in social networks, surveillance, and smartphones, this arguably increases the interest of periocular recognition [8,9]. For all these reasons, periocular recognition has become an area of intense study in the biometrics and computer vision communities.



**Figure 1.** Samples of periocular regions. We demonstrate sample images of the periocular region that including eyebrows. The images are collected from The Korea Times [10] and Kitchen Decor [11].

In this paper, we address the challenges of periocular recognition in the unconstrained or “in-the-wild” environments that remain not well-addressed by the current works [12,13]. This challenge is associated with the issue of dissimilarities in periocular images due to the placement of sensors, pose alignments, illumination levels, occlusions, etc. Thus, we study this problem by means of a fusion approach with dual-stream Convolutional Neural Network (CNN), which accepts RGB ocular image and a novel colour-based texture descriptor, known as Orthogonal Combination-Local Binary Coded Pattern (OCLBCP). We have also developed and shared a new database, namely Ethnic-Ocular database, by collecting the periocular region images in the wild to validate the proposed network.

### 1.1. Related Works

The early study on periocular biometrics presented in [5] shows promising results in human recognition. The authors adopted several handcrafted descriptors such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Scale-Invariant Feature Transform (SIFT) as periocular feature representation, followed by the score fusion for classification. Fernandez et al. [14] and Cao et al. [15] also introduced a similar approach, which convolves periocular features extracted from HOG or LBP feature matrix with Gabor filters and followed by score fusion. There are several research articles focused on combinations of texture descriptors with fusion algorithm for periocular representation and recognition [16–20]. All these approaches are mainly focused on amalgamation of various handcrafted texture descriptors and followed by learning machines for decent performances in periocular recognition. However, these approaches are less robust to “in the wild” variations such as resolutions, levels of illumination, poses, and occlusions due to inadequacy and inflexibility of handcrafted texture descriptors in representing periocular features. Therefore, the periocular recognition in the wild remains a challenge.

In recent years, CNNs have gained escalating attention in image classification [21,22]. CNNs can be used to extract image texture features from different layers while handcrafted texture descriptor are only limited to low-level features, which is equivalent to the first convolutional (*conv*) layer features of CNNs. Apart from *conv* layers at different level, the features can be extracted from max pooling (*maxpool*) and fully-connected (*fc*) layers of CNNs. Several researchers have employed CNNs for periocular recognition. For instance, Gangwar et al. [23] proposed two CNNs (for left and right oculars), namely DeepIrisNet, which extracts comprehensive information to boost recognition

performance. Other studies, e.g., by Proença et al. [24] and Zhao et al. [25], have demonstrated enhanced CNN frameworks for periocular recognition where the prior knowledge is exploited to discard unnecessary information. Proença et al. [24] suggested removal/separation of the iris and sclera from the periocular regions, while Zhao et al. [25] identified the critical regions (only included eyebrow and eye region) that can extract more discriminative information to improve periocular recognition. However, these networks were found to underperform when there are misalignments of periocular images, images missing the eyebrows and images missing ocular.

The relevant works that deal with non-ideal ocular are those by Zhang et al. [26] and Soleymani et al. [27]. Zhang et al. [26] fused iris and periocular modalities through a weighted concatenation. The network achieved significant results when compared to other CNNs. Similarly, Soleymani et al. [27] invented a new multimodal CNN, namely multi-fusion CNN, where the iris, face and fingerprint features are fused at *fc* layer. A fusion layer is designed to fuse different levels of *fc* layers as multi-feature representations with the sole RGB image. However, these works leveraged several biometrics where all of them may not always be available such as occluded face with mouth covered or iris from a distance. Furthermore, the use of multiple biometrics modality may jeopardise the usability of the system such as fingerprint and iris need cooperation from the users.

In the previous work of CNN that consumes face texture descriptor, Levi et al. [28] demonstrated the use of colour-based LBP descriptor as input to CNN rather than raw RGB face image for emotional recognition. The authors showed that colour-based texture descriptor is useful to train their network in the wild environment. This work motivates us to investigate and analyse the impact of colour-based texture descriptor within CNN for periocular recognition in the wild.

## 1.2. Motivation and Contributions

In the early days of periocular recognition, the problems were mostly concerned about what was the best way to handle periocular in the presence of illuminations, pose alignments, and occlusions [5,6]. Many periocular databases were built using carefully controlled images for each of these issues. UBIPr [12], CASIA-iris database [29], and MICHE database [30] are the most comprehensive efforts in this direction and created in a well-controlled environment.

Presently, the challenges of periocular recognition concern about images that having large variations due to in the wild environments, such as ageing, appearances, cameras location, level of illuminations, occlusions, pose alignments, and others [18,31]. In addition, many existing databases [12,13,29,30] and research communities [18,23,27] still yet to prepare for periocular recognition in the wild challenge. Especially, the appearances of periocular with cosmetic products, and plastic surgery can affect the recognition performance negatively.

This paper offers a solution for periocular recognition in the wild by investigating the fusion of RGB periocular images and a novel texture descriptor, i.e., OCLBCP, by means of a dual-stream CNN. OCLBCP exploits the colour information in the periocular texture to better represent the periocular features for recognition in the wild. The two networks share the parameters and a late fusion takes place at the last *conv* layer before *fc* layer.

For validation of the proposed network, a new database is introduced, namely Ethnic-ocular, by collecting the periocular region images in the wild setup. The databased includes five ethnic groups: *African*, *Asian*, *Latin American*, *Middle Eastern*, and *White*. The database is created in such a way that each ethnic group has a unique shape of periocular and skin texture of periocular regions [32]. Therefore, the database avoids unbalanced selection, as there are differences in the configuration of oculars among different ethnicities.

Hence, the contributions of this paper are as follows:

- To study complementarity between CNN and input features, we investigate and analyse the combination of RGB image and a novel texture descriptor, namely OCLBCP for periocular recognition in the wild.

- Two distinct late-fusion layers are introduced in the proposed CNN. The role of the late-fusion layers is to aggregate the RGB image and OCLBCP descriptor. Hence, the proposed two-stream CNN is beneficial from these new features of the late-fusion layers to deliver better accuracy performance.
- A new periocular in the wild database, namely Ethnic-ocular, is created and shared in [33]. The images were collected across highly uncontrolled subject–camera distances, appearances, resolutions, locations, levels of illumination, and so on. The database includes training and testing schemes for performance analysis and evaluation.

The paper is organised as follows: Section 2 describes the structure of the proposed colour-based Orthogonal Combination-Local Binary Coded Pattern (OCLBCP) texture descriptor. The proposed network with fusion algorithm is presented in Section 3 and the detailed database information is presented in Section 4. Section 5 discusses the experimental results and analysis. A conclusion is summarised in Section 6.

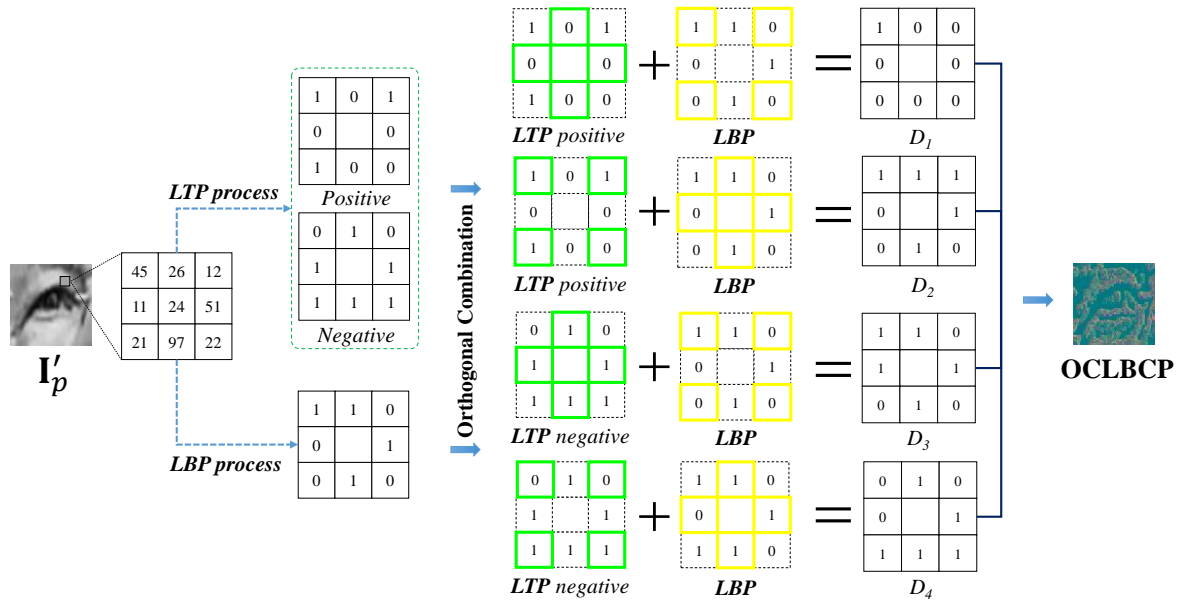
## 2. Colour-Based Orthogonal Combination—Local Binary Coded Pattern

This section introduces a new colour-based texture descriptor known as Orthogonal Combination—Local Binary Coded Pattern (OCLBCP). OCLBCP is devised based on the notion of an orthogonal combination of Local Binary Pattern (LBP) [34] and Local Ternary Pattern (LTP) [35]. The OCLBCP descriptor yields a more vibrant texture representation since it is less sensitive to the image noise and levels of illuminations.

Let  $I_p \in \mathbb{R}^{x \times y}$  be the periocular grayscale image, where  $x$  and  $y$  are the width and height of  $I_p$ , respectively. The apparent changes in the images are related to illuminations and poses, thus we deploy the pre-processing method used in [36] to reduce the noise from  $I_p$ . First, we transform the  $I_p$  into Fourier domain as  $Z$ . Furthermore, we apply the Butterworth filter ( $B$ ) to  $Z$  by reducing the illumination noise and enhancing the reflectance [37]. After that, we apply an inverse Fourier transform to obtain the filtered image  $I'_p$ .

To construct the OCLBCP descriptor,  $I'_p$  has to be proposed first according to the LBP [34] and LTP [35] transformation. LBP summarises the local structure in an image by comparing each pixel with its neighbourhood [34]. This descriptor works by thresholding a neighbourhood matrix using the grey level of the central pixel in the binary code. LTP is an extension of the LBP with three-valued codes [35]. The descriptor works by comparing each pixel with its neighbouring pixels. Then, they are combined after thresholding into a ternary pattern. The ternary pattern is split into two binary patterns and called positive and negative matrices.

In this paper, the LBP consists of the  $3 \times 3$  neighbourhood matrix, and the LTP consists of the positive and negative matrices. To do so,  $I'_p$  is partitioned into sub-matrix with size  $3 \times 3$  and the neighbourhood values of sub-matrix is binarised according to the centre value of the sub-matrix, which serves as a reference value for thresholding. After that, the descriptor combines the sub-matrix of LBP and LTP into four orthogonal groups:  $D_1$ ,  $D_2$ ,  $D_3$ , and  $D_4$  (see Figure 2). The orthogonal groups serve to achieve illumination invariance and uncover better texture information by removing outlying disturbances. Specifically, to obtain  $D_1$ , the bits from the yellow boxes in the LBP and the bits from green boxes in LTP positive in Figure 2 are combined. The same processes are repeated for  $D_2$ ,  $D_3$ , and  $D_4$ .



**Figure 2.** Illustration of Orthogonal Combination-Local Binary Coded Pattern (OCLBCP).

Suppose  $\theta$  is the OCLBCP descriptor, we first convert the binary codes  $D_k$  into a decimal number  $D_{ck}$ ,  $k = 1, 2, 3$ , and  $4$ , and then choose the largest value from all the orthogonal groups. Specifically, the  $\theta$  is formed by combining the groups as follows:

$$\theta(i, j) = \max [D_{c1}(i, j), D_{c2}(i, j), D_{c3}(i, j), D_{c4}(i, j)], \quad (1)$$

where  $i$  and  $j$  are the indices of  $\theta$ .

To map  $\theta(i, j)$  into a colour-based texture descriptor, we create a distance pattern matrix  $\Delta$  to represent the similarity of the image intensity patterns across all possible pixel values based on [28]:

$$\Delta := \begin{bmatrix} \delta_{1,1} & \delta_{1,2} & \delta_{1,3} & \cdots & \delta_{1,c} \\ \delta_{2,1} & \delta_{2,2} & \delta_{2,3} & \cdots & \delta_{2,c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{r,1} & \delta_{r,2} & \delta_{r,3} & \cdots & \delta_{r,c} \end{bmatrix}, \quad (2)$$

where  $r$  and  $c$  are defined as the indices of  $\delta$ .  $\delta_{r,c}$  is calculated by Earth Mover's Distance. After that, the Multi-Dimensional Scaling (MDS) algorithm is adopted to seek the mapping of  $\Delta$  to the low-dimensional metric space (colour pattern matrix  $\mathcal{M}$ ) [38]:

$$\mathcal{M} = [\text{MDS}(\theta) + \|\min(\text{MDS}(\theta))\|] * \left[ \frac{255}{\|\max(\text{MDS}(\theta))\|} \right], \quad (3)$$

$$\text{MDS}(\cdot) = \sqrt{\frac{\sum_r \sum_c f(\delta_{r,c})}{q}}, \quad (4)$$

where  $q$  is scale factor and  $f(\delta_{r,c})$  is a monotonic transformation function of  $\delta_{r,c}$ . In this paper, we set  $q$  to three due to RGB channels in the colour image. Note that  $\mathcal{M}$  is a three-colour channels matrix that outputs from  $\text{MDS}(\cdot)$ , which contains R, G, and B pixel values. Finally, we map  $\theta(i, j)$  with  $\mathcal{M}$  to generate colour-based texture descriptor OCLBCP. The mapping process uses the given pixel values of  $\theta(i, j)$  to match the pixel values from the R channel of  $\mathcal{M}$ . After that,  $\theta(i, j)$  is converted with the RGB values from  $\mathcal{M}$ . Algorithm 1 summarises the process of generating OCLBCP.

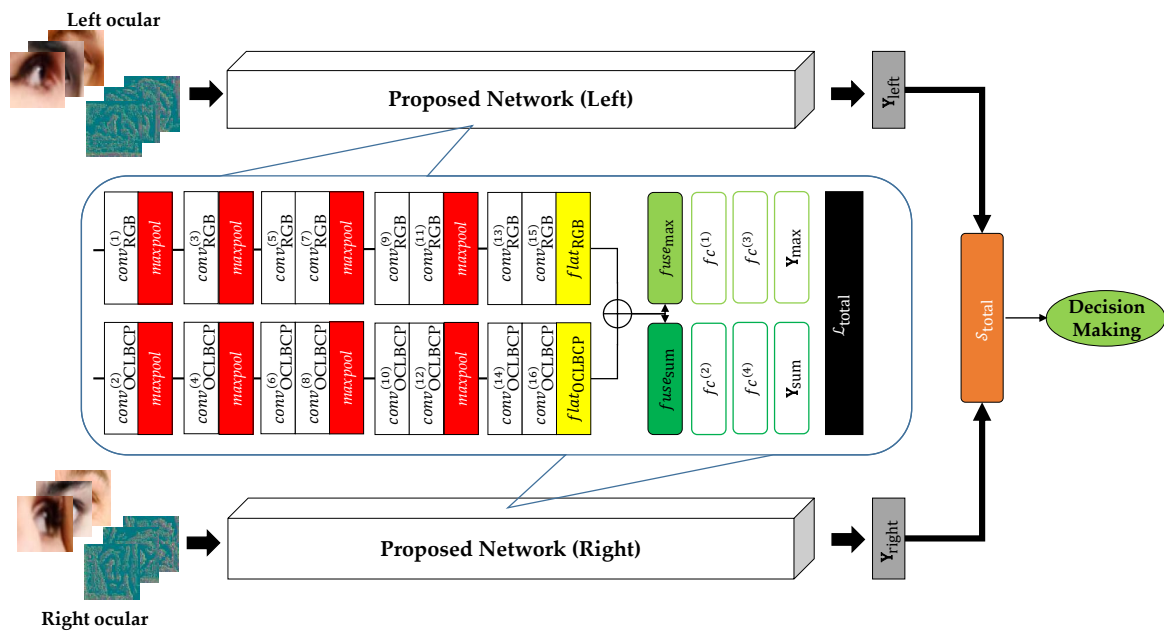
**Algorithm 1** Creating colour-based texture description OCLBCP.**Input:**  $I_p \in \mathbb{R}^{x \times y}$ **Output:** OCLBCP

- 1: Perform preprocessing to  $I_p$  and obtain the filtered image  $I'_p$
- 2: Construct LBP and LTP process on  $I_p$
- 3: Perform Equation (1) with the LBP, LTP positive and LTP negative matrices to obtain  $\theta$
- 4: Construct distance pattern matrix  $\Delta$  using Equation (2)
- 5: Generate the colour-based pattern matrix  $\mathcal{M}$  with  $\delta$  by using Equations (3) and (4)
- 6: Map  $\theta$  with  $\mathcal{M}$  to generate OCLBCP

**3. RGB-OCLBCP of Dual-Stream CNN**

We propose a dual-stream CNN that conceives the periocular RGB image and OCLBCP descriptor as the first and second stream to the network. Note that the dual-stream CNN was originally proposed by Feichtenhofer et al. [39] for action detection and recognition. The two input streams refer to temporal and structural streams. In our work, the network accepts and processes periocular colour image and texture descriptor, and then feature fusion layers are devised to extract better feature representation for ocular recognition.

As shown in Figure 3, the architecture of the proposed network consists of 16 convolutional (*conv*) layers and 8 max-pooling (*maxpool*) layers. The *conv* layers are designed to learn the correspondence between the RGB image and OCLBCP descriptor and to discriminate between themselves with the shared weights. Table 1 tabulates the architecture of the proposed network.



**Figure 3.** The architecture of the proposed network.



**Table 1.** Configurations of each layer for the proposed network.

Network Layers	Configurations
$conv_{RGB}^{(1)}, conv_{OCLBCP}^{(2)}$	$f^1: 64@80 \times 80; k^2: 2 \times 2; maxpool: 2 \times 2$
$conv_{RGB}^{(3)}, conv_{OCLBCP}^{(4)}$	$f: 128@40 \times 40; k: 2 \times 2; maxpool: 2 \times 2$
$conv_{RGB}^{(5)}, conv_{OCLBCP}^{(6)}$	$f: 256@20 \times 20; k: 2 \times 2$
$conv_{RGB}^{(7)}, conv_{OCLBCP}^{(8)}$	$f: 256@20 \times 20; k: 2 \times 2; maxpool: 2 \times 2$
$conv_{RGB}^{(9)}, conv_{OCLBCP}^{(10)}$	$f: 512@10 \times 10; k: 2 \times 2$
$conv_{RGB}^{(11)}, conv_{OCLBCP}^{(12)}$	$f: 512@10 \times 10; k: 2 \times 2; maxpool: 2 \times 2$
$conv_{RGB}^{(13)}, conv_{OCLBCP}^{(14)}$	$f: 512@10 \times 10; k: 2 \times 2$
$conv_{RGB}^{(15)}, conv_{OCLBCP}^{(16)}$	$f: 512@10 \times 10; k: 2 \times 2$
$flat_{RGB}, flat_{OCLBCP}$	$1 \times 1 \times 12,800$
$fuse_{max}, fuse_{sum}$	$1 \times 1 \times 4096$
$fc^{(1)}, fc^{(2)}$	$1 \times 1 \times 4096$
$fc^{(3)}, fc^{(4)}$	$1 \times 1 \times 4096$
$Y_{max}, Y_{sum}$	$1 \times 1 \times C$

<sup>1</sup>  $f$  refers to the size of the feature map in  $conv$  layers. <sup>2</sup>  $k$  is defined as the filter size.

### 3.1. Fusion Layers

Two fusion layers, namely  $fuse_{max}$  and  $fuse_{sum}$ , are designed to aggregate the information from the RGB image and OCLBCP descriptor, as shown in Figure 3. The  $fuse_{max}$  layer takes the largest activation from the  $flat_{RGB}$  and  $flat_{OCLBCP}$  layers with  $m$  nodes, where both of them are flattened to  $conv_{RGB}^{(15)}$  and  $conv_{OCLBCP}^{(16)}$ , respectively. The  $fuse_{max}$  can be represented as:

$$fuse_{max}(i) = \max[flat_{RGB}(i), flat_{OCLBCP}(i)], \quad i = 1, \dots, m. \quad (5)$$

On the other hand,  $fuse_{sum}$  takes a sum of activations of  $flat_{RGB}$  and  $flat_{OCLBCP}$ . The layer is defined as follows:

$$fuse_{sum}(i) = flat_{RGB}(i) + flat_{OCLBCP}(i), \quad i = 1, \dots, m. \quad (6)$$

### 3.2. Total Loss for Training

For training, we define a total loss function,  $\mathcal{L}_{total}$ , which is composed of a summation of softmax cross entropy  $\mathcal{L}$  of logit vector and their respective encoded label:

$$\mathcal{L}_{total} = \mathcal{L}(\mathbf{V}_{max}) + \mathcal{L}(\mathbf{V}_{sum}), \quad (7)$$

$$\mathcal{L}(\mathbf{V}) = - \sum_n^N \sum_c^C L_{nc} \log[\text{softmax}(\mathbf{V})_{nc}], \quad (8)$$

$$\text{softmax}(\mathbf{V})_{nc} = \frac{\exp \mathbf{V}_{nc}}{\sum_c^C \exp \mathbf{V}_{nc}}, \quad (9)$$

where  $\mathbf{V} \in \{\mathbf{V}_{max}, \mathbf{V}_{sum}\}$ .  $\mathbf{V}_{max}$  and  $\mathbf{V}_{sum}$  are defined as the features of  $fuse_{max}$  and  $fuse_{sum}$  layers in the training samples  $\mathbf{V}$ , respectively.  $L$ ,  $N$ , and  $C$  denote class labels, the number of training samples in  $\mathbf{V}$ , and the number of classes, respectively. Note that a periocular region contains left and right oculars; we therefore train each side with separate networks (Figure 3).

### 3.3. Score Fusion Layer for Recognition

To recognise an unknown identity, a score fusion layer  $\mathcal{S}_{total}$  is devised to merge the distance scores from the softmax vectors for decision-making. Let  $\mathbf{Y}_{max} = \text{softmax}(\mathbf{V}_{max}) \in \mathbb{R}^C$  and  $\mathbf{Y}_{sum} = \text{softmax}(\mathbf{V}_{sum}) \in \mathbb{R}^C$  be the softmax vectors of  $fc^{(3)}$  and  $fc^{(4)}$ , respectively. Since we train the

proposed network for left and right ocular, we thus differentiate the softmax vector  $\mathbf{Y}$  to  $\mathbf{Y}_{\text{left}}$  and  $\mathbf{Y}_{\text{right}}$ . Note each individual  $\mathbf{Y}$  to  $\mathbf{Y}_{\text{left}}$  and  $\mathbf{Y}_{\text{right}}$  is still the sum of its corresponding  $\mathbf{Y} = \mathbf{Y}_{\text{max}} + \mathbf{Y}_{\text{sum}}$ .

We evaluated the proposed system in two common biometric working modes: recognition and verification. For the former, the testing data are divided into a gallery set and a probe set. Each subject in the gallery set is composed of his/her left and right softmax vectors as  $\mathbf{Y}_j^G = \{\mathbf{Y}_{j,\text{left}}^G, \mathbf{Y}_{j,\text{right}}^G\}$ , where  $j = 1, \dots, C$ ; the probe set is defined as  $\mathbf{Y}^P = \{\mathbf{Y}_{\text{left}}^P, \mathbf{Y}_{\text{right}}^P\}$ . The score fusion layer is computed with the sum rule as follows:

$$\mathcal{S}_{\text{total}}(\mathbf{Y}^P, \mathbf{Y}_j^G) = s(\mathbf{Y}_{\text{left}}^P, \mathbf{Y}_{j,\text{left}}^G) + s(\mathbf{Y}_{\text{right}}^P, \mathbf{Y}_{j,\text{right}}^G), \quad (10)$$

where  $s(\mathbf{Y}_*^P, \mathbf{Y}_{j,*}^G) = 1 - \cos(\mathbf{Y}_*^P, \mathbf{Y}_{j,*}^G)$  is defined as cosine similarity distance and  $*$   $\in$  {left, right}. To identify  $\mathbf{Y}^P$ ,  $\phi$  is decided as follows:

$$\phi = \max_j [\mathcal{S}_{\text{total}}(\mathbf{Y}^P, \mathbf{Y}_j^G)] \quad (11)$$

Verification protocol refers to verifying a person's identity that is claimed as a genuine or an impostor. Let  $\mathbf{Y}^R = \{\mathbf{Y}_{\text{left}}^R, \mathbf{Y}_{\text{right}}^R\}$  as the reference set (template) and  $\mathbf{Y}^A = \{\mathbf{Y}_{\text{left}}^A, \mathbf{Y}_{\text{right}}^A\}$  as the query set, to decide the  $\mathbf{Y}^A$  is a genuine or an impostor,  $\zeta$  is decided by using Equation (12) as follows:

$$\zeta(\mathbf{Y}^A, \mathbf{Y}^R) = \begin{cases} 1, & \mathcal{S}_{\text{total}}(\mathbf{Y}^A, \mathbf{Y}^R) \leq \tau \\ 0, & \mathcal{S}_{\text{total}}(\mathbf{Y}^A, \mathbf{Y}^R) > \tau \end{cases} \quad (12)$$

where  $\tau$  is training dataset dependence threshold value.

#### 4. Database

A large-scale collection of periocular in the wild images from different ethnic groups was created, namely Ethnic-ocular database. This database is built for periocular recognition, which contains left and right oculars that were extracted from 85,394 images downloaded from the web. All images were collected in the wild, with uncontrolled subject-camera distances, poses, appearances with and without make-up, and levels of illumination.

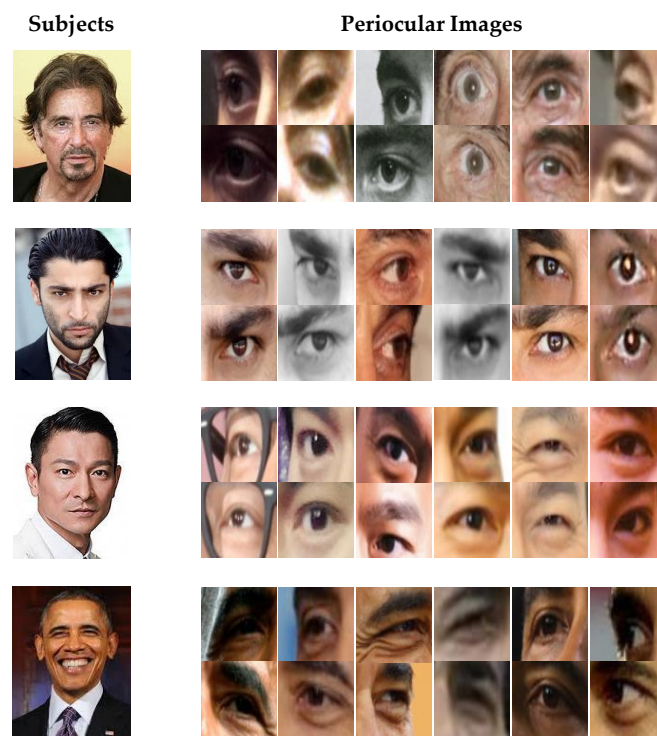
We propose this new database to support balanced selection in the configuration of oculars among different ethnicities, and also to stimulate research for periocular recognition in the wild that all periocular images are taken in common and everyday settings. Figure 4 demonstrates several samples of images.

##### 4.1. Collection Setup

To create our database, we selected subject names randomly from BBC News [40], CNN News [41], Naver News [42], and FaceScrub database [43]. The subjects were randomly selected based on different ethnicities. They mostly are celebrities, politicians, athletes, etc.

From the search result, the top 300 images for each subject were downloaded using Python scripts. After that, the images were manually verified to ensure that the subjects correctly labelled the images. We firstly extracted facial regions in these images by using the face detector from Matlab [44] for periocular region extraction. Then, the coordinates of facial feature points were fixed based on the face detector bounding box for image alignment. Then, the images of subjects were labelled manually. After that, we implemented the technique from [45], which allowed us to crop images into left and right oculars. The database contains 85,394 images (including left and right oculars images) of 1034 subjects. Note that the views of these images are between  $-45^\circ$  and  $45^\circ$ .





**Figure 4.** Samples of periocular images in the wild.

#### 4.2. Training Protocol

For the training protocol, 623 subjects were randomly selected. Note that no subjects for training overlapped with the subjects for benchmarking. To develop or train our own models, we designed the protocol by dividing the images for each subject with the ratio of training, testing, and validations as 70:15:15.

#### 4.3. Benchmark Protocol

We selected the remaining 411 subjects as benchmarks. In the benchmarking scheme, we created recognition and verification tasks. For recognition task, images about a specific set of individuals to be recognised (gallery set) were gathered and a new image (the probe set) was presented; the task was to decide which of the gallery identities was represented by the probe set. In the experiments, we divided the images per subject with the ratio of the gallery set to probe set as 50:50. This division process was repeated three times.

For verification task, the task was to analyse two sets of periocular images and decide whether they represent the same person or two different people. In the experiments, we randomly selected 1200 pairs as “same” labels and 1200 pairs as “not same”. This selection process was repeated three times.

### 5. Experiments

We conducted several experiments to evaluate the performance comparisons of recognition and verification between our network and other benchmark networks. All configurations of the networks are described in Section 5.1 and the experimental results are presented in Section 5.2.

#### 5.1. Experimental Setup

##### 5.1.1. Configuration of Proposed Network

The proposed network was implemented using the open source deep learning toolkit TensorFlow [46]. About the configurations, we applied an annealed learning rate and it was started

from  $1.0 \times 10^{-3}$ . The rate was subsequently reduced by  $10^{-1}$  for every 10 epochs. The minimum learning rate was defined as  $1.0 \times 10^{-5}$ . We applied an Adam optimiser in this network, where the weight decay and momentum were set to  $1.0 \times 10^{-4}$  and 0.9, respectively.

In our experiments, the batch size was set to 64 and the training was carried out across 200 epochs. The training was done by using our database and following the protocols mentioned in Section 4.2 and it was performed by an NVidia Titan Xp GPU.

### 5.1.2. Configuration of Benchmark Networks

We selected several deep networks to evaluate the performance of periocular recognition: AlexNet [21], DeepIrisNet-A [23], DeepIrisNet-B [23], FaceNet [47], LCNN29 [48], Multi-fusion CNN [27], and VGG16 [49]. Inspired by the work of Gangwar et al. [23], Soleymani et al. [27], Schroff et al. [47], Wu et al. [48], and Hernandez et al. [50], these networks have been proven to be successful in very large recognition tasks. In the experiments, we utilised the pre-trained models that were provided by the authors to fine-tune and improve the networks themselves by training the left and right oculars, respectively. In the cases of DeepIrisNet-A, DeepIrisNet-B, and Multi-fusion CNN, the networks are not publicly available. Therefore, we did our best effort to implement these networks from scratch by following Gangwar et al. [23] and Soleymani et al. [27], respectively.

## 5.2. Experimental Results

We present the experimental results on the tasks of periocular recognition and verifications by conducting the databases on periocular recognition in the wild and controlled environments. For the recognition, we evaluated the performance by using Cumulative Matching Characteristic (CMC) curve with 95% confidence interval (CI). For the verification, we evaluated the performance using Receiver Operating Characteristic (ROC) curve with Equal Error Rate (EER) and Area under the ROC curve (AUC).

### 5.2.1. Performance Analysis on Proposed Network

This section analyses the robustness and performance of our network and other networks using Ethnic-ocular database, which reports the experimental results in Table 2.

**Table 2.** Performance analysis on the proposed network and other networks with Rank-1 and Rank-5 recognition accuracies. The highest accuracy is highlighted in bold.

Networks	Accuracy (%)		<i>t.w.</i> <sup>2</sup>	<i>flops</i> <sup>3</sup>
	Rank-1	Rank-5		
CNN <sup>1</sup> with RGB image	80.79 ± 1.43	90.42 ± 1.29	131.1 M	2.22 GFLOPS
CNN with OCLBCP	66.65 ± 2.22	89.73 ± 1.91	131.1 M	2.22 GFLOPS
Dual-stream CNN (using unshared weights)	82.09 ± 1.59	92.11 ± 1.32	250.8 M	1.90 GFLOPS
<b>Proposed network</b>	<b>85.03 ± 1.88</b>	<b>94.23 ± 1.26</b>	<b>126.1 M</b>	<b>0.90 GFLOPS</b>

<sup>1</sup> CNN is defined as single-stream CNN; <sup>2</sup> *t.w.*, total weight number; <sup>3</sup> *flops*, floating points operation.

Table 2 shows the proposed network achieved the highest Rank-1 and Rank-5 recognition accuracies with  $85.03 \pm 1.88\%$  and  $94.23 \pm 1.26\%$ , respectively. As compared to CNN, this network using the RGB image only achieved the Rank-1 and Rank-5 accuracies of  $80.79 \pm 1.43\%$  and  $90.42 \pm 1.29\%$ , respectively. In addition, CNN using the OCLBCP can only achieved  $66.65 \pm 2.22\%$  and  $89.73 \pm 1.91\%$  for Rank-1 and Rank-5 accuracies, respectively. These results indicate that our network provides more complementary information than CNN. This leads to the proposed late-fusion layers that significantly correlate the RGB image and OCLBCP for achieving better recognition performance.

Furthermore, we also evaluated the dual-stream CNN without using shared weights. However, this network only achieved  $82.09 \pm 1.59\%$  and  $92.11 \pm 1.32\%$  at Rank-1 and Rank-5 accuracies (see Table 2), respectively. The experimental results prove that the proposed network performed

well with at least 2.9% improvement as compared to dual-stream CNN without using shared weight. As can be observed, the shared *conv* layers and the fusion layers were utilised in the network to aggregate the RGB image and OCLBCP. Thus, the proposed network successfully transformed new knowledge representations to perform better recognition in the wild.

In Table 2, we also notice the space complexity (total weight number) and time complexity (flops) of the proposed network are significantly smaller than its single network and dual-stream unshared weights networks counterparts while still outperforming them.

### 5.2.2. Performance Evaluation on Recognition and Verification Tasks

We used Ethnic-ocular, as well as three public databases, the AR [51], CASIA-iris distance [29], and UBIPr [12], to evaluate the performances of the proposed network and other benchmark networks. All the experimental results are outlined in the following sections.

#### Evaluation on AR Database

The AR database is designed under a constrained environment, which consists of 117 subjects with varying neutrals, expressions, illuminations, and occlusion conditions, who were captured across two sessions. We opted for this database as it provides a good baseline to evaluate the robustness and performance in constrained environments, such as different levels of illuminations and expressions in an indoor environment. Extraction for the periocular regions was done by using the method in [45].

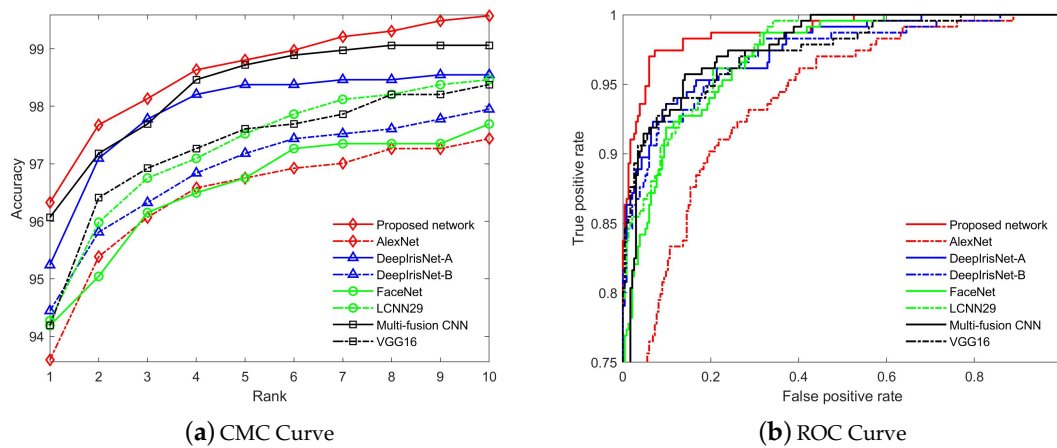
The experimental protocol for recognition was as follows: ten images for each subject were used as gallery sets from Session 1 and another ten per subject as probe sets from Session 2. On the other hand, the verification protocol was designed by randomly selecting 250 reference-query pairs as “same” and another 250 pairs as “not same”.

Table 3 presents the performance comparisons on recognition. As can be seen in the table, our network achieved the highest Rank-1 and Rank-5 recognition accuracies with 96.32% and 98.80%, respectively. Likewise, DeepIrisNet-A had the best performance on Rank-1 and Rank-5 among the other benchmark approaches, which only achieved accuracies of 95.24% and 98.38%, respectively. Figure 5a illustrates that the proposed network outperformed other approaches with respect to all the benchmarks from Rank-1 to Rank-10 recognition.

For the verification task, we report the experimental results in Table 4. The proposed network also achieved the best EER and AUC with 5.13% and 0.9880, respectively. DeepIrisNet-A, Multi-fusion CNN, and VGG16 achieved the second-best performances among the other benchmark approaches with 7.69% for EER. Figure 5b illustrates the ROC curve and shows that the proposed network (red solid line with diamond) outperformed the benchmark approaches.

**Table 3.** Performance evaluation of the recognition task on the AR database, CASIA-iris distance database, UBIPr database, and Ethnic-ocular database. The highest accuracy is written in bold.

Networks	AR		CASIA-iris		UBIPr		Ethnic-Ocular	
	Rank-1	Rank-5	Rank-1	Rank-5	Rank-1	Rank-5	Rank-1	Rank-5
AlexNet	93.59	96.75	95.00 ± 1.8	96.98 ± 2.5	84.88 ± 2.5	96.01 ± 1.8	64.72 ± 3.3	82.98 ± 2.5
DeepIrisNet-A	95.24	98.38	95.95 ± 2.1	98.15 ± 0.6	90.30 ± 1.2	97.41 ± 1.1	79.54 ± 3.1	90.43 ± 2.4
DeepIrisNet-B	94.44	97.18	95.79 ± 2.6	97.75 ± 0.6	90.20 ± 1.7	97.43 ± 0.5	81.13 ± 3.1	92.37 ± 1.2
FaceNet	94.19	97.75	96.09 ± 2.1	98.10 ± 0.4	90.24 ± 1.4	97.36 ± 0.4	78.71 ± 3.7	92.19 ± 1.6
LCNN29	94.27	97.52	96.01 ± 2.0	97.85 ± 0.9	90.28 ± 1.7	97.18 ± 0.7	79.35 ± 2.6	92.17 ± 1.8
Multi-fusion CNN	96.07	98.71	95.81 ± 1.9	97.67 ± 1.0	90.75 ± 1.0	97.44 ± 0.3	81.79 ± 3.5	93.03 ± 1.3
VGG16	94.20	97.61	95.88 ± 0.1	97.99 ± 0.5	90.24 ± 1.4	97.09 ± 1.1	76.43 ± 2.2	91.29 ± 1.5
<b>Proposed Network</b>	<b>96.32</b>	<b>98.80</b>	<b>96.62 ± 1.3</b>	<b>98.45 ± 0.4</b>	<b>91.28 ± 1.2</b>	<b>98.59 ± 0.4</b>	<b>85.03 ± 1.9</b>	<b>94.23 ± 1.3</b>



**Figure 5.** Performances of recognition and verification tasks on AR database. The figures are best viewed in colour.

**Table 4.** Performance evaluation of the verification task on the AR database, CASIA-iris distance database, UBIPr database, and Ethnic-ocular database. The highest accuracy is written in bold.

Networks	AR		CASIA-Iris		UBIPr		Ethnic-Ocular	
	EER (%)	AUC	EER (%)	AUC	EER (%)	AUC	EER (%)	AUC
AlexNet	14.53	0.9363	8.06 $\pm$ 5.3	0.9533	7.11 $\pm$ 2.9	0.9805	16.47 $\pm$ 1.6	0.9139
DeepIrisNet-A	7.69	0.9751	7.51 $\pm$ 1.1	0.9674	5.07 $\pm$ 2.2	0.9877	8.79 $\pm$ 1.7	0.9689
DeepIrisNet-B	8.12	0.9741	5.87 $\pm$ 1.5	0.9756	4.29 $\pm$ 0.9	0.9890	8.77 $\pm$ 1.1	0.9693
FaceNet	9.40	0.9692	6.10 $\pm$ 2.2	0.9738	5.46 $\pm$ 1.5	0.9870	11.67 $\pm$ 1.2	0.9489
LCNN29	9.39	0.9737	6.34 $\pm$ 1.6	0.9719	6.34 $\pm$ 2.1	0.9849	10.95 $\pm$ 1.6	0.9536
Multi-fusion CNN	7.69	0.9756	8.69 $\pm$ 1.1	0.9594	4.09 $\pm$ 2.1	0.9913	8.63 $\pm$ 1.3	0.9681
VGG16	7.69	0.9747	7.42 $\pm$ 1.7	0.9681	4.38 $\pm$ 1.3	0.9892	9.43 $\pm$ 2.5	0.9553
<b>Proposed Network</b>	<b>5.13</b>	<b>0.9882</b>	<b>4.35 <math>\pm</math> 0.5</b>	<b>0.9860</b>	<b>3.41 <math>\pm</math> 1.8</b>	<b>0.9938</b>	<b>6.63 <math>\pm</math> 1.5</b>	<b>0.9818</b>

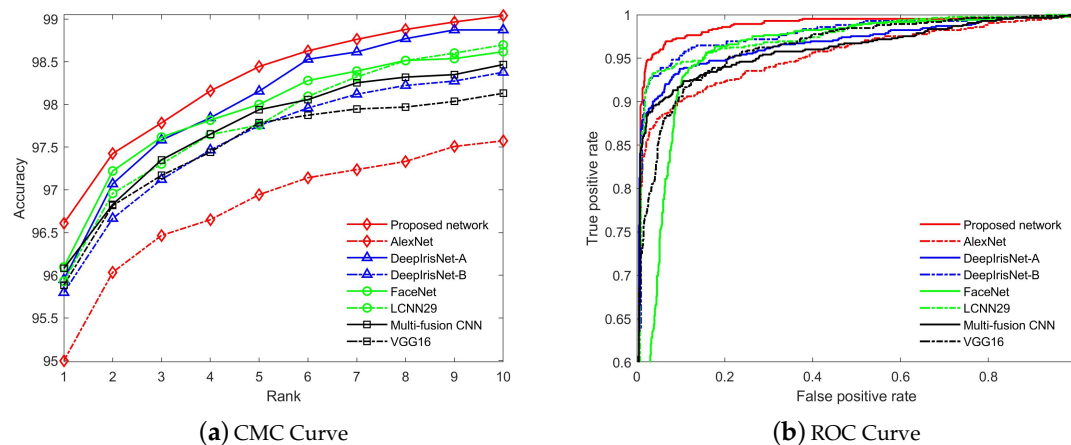
#### Evaluation on CASIA-Iris Distance Database

To evaluate whether our approach performs well on another standard database, we also tested its performance in a more subjective experiment with CASIA-iris distance database. This database consists of 142 subjects under a long-range subject–camera distance and indoor environment. The images were captured by a high-resolution camera so both dual-eye iris and periocular are included in the image region of interest. The further details of the database can be found in [29].

The experimental protocol for recognition was designed with the ratio of the gallery set to probe set as 50:50 and the division process was repeated three times. The experimental protocol for the verification was designed by randomly selecting 250 reference–query pairs as “same” and another 250 pairs as “not same”. This selection process was repeated three times.

According to Table 3, the proposed network achieved the highest average accuracies for Rank-1 and Rank-5 recognitions with  $96.62 \pm 1.3\%$  and  $98.45 \pm 0.4\%$ , respectively. Besides, FaceNet achieved the second-best performance with  $96.09 \pm 2.1\%$  and  $98.10 \pm 0.4\%$  for Rank-1 and Rank-5 recognition accuracies, respectively. We also present in Figure 6a the Rank-1 to Rank-10 recognition results. As can be seen, our network achieved the best results among the benchmark networks.

For the verification, the proposed network achieved the lowest EER accuracy as  $4.35 \pm 0.5\%$  and AUC as 0.9860. Interestingly, DeepIrisNet-B attained second lowest performance with  $5.87 \pm 1.5\%$  for EER and 0.9756 as AUC. Figure 6b illustrates the ROC curve, which demonstrates that our network obtained the best performance of AUC and the lowest EER. Both recognition and verification results indicate that the proposed network is capable of learning the features of the RGB image and OCLBCP decently for improving the performance of recognition and verification tasks.



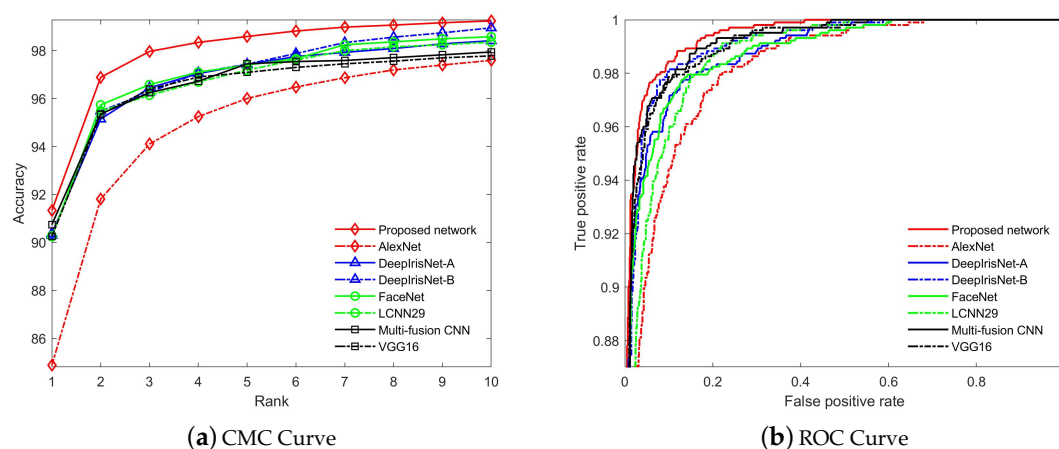
**Figure 6.** Performances of recognition and verification tasks on CASIA-iris distance database. The figures are the best to view in colour.

### Evaluation on UBIPr Database

We also conducted another more challenge experiment with the UBIPr database to verify the robustness of the proposed network. This database consists of 342 subjects with varying subject–camera distances, levels of illumination, and poses [12]. This experiment evaluated the performance of all the networks with varying poses and subject–camera distances. Six images from each subject were randomly divided as a gallery set; the remaining images were used as a probe set. The division process was repeated three times. For the verification, we randomly selected 600 reference–query pairs as “same” and another 600 pairs as “not same”. This selection process was also repeated three times.

Table 3 presents that our network achieved the highest average Rank-1 and Rank-5 recognition accuracies with  $91.28 \pm 1.2\%$  and  $98.59 \pm 0.4\%$ , respectively. The second best was achieved by multi-fusion CNN with  $90.75 \pm 1.0\%$  and  $97.44 \pm 0.3\%$  as Rank-1 and Rank-5 accuracies, respectively. Besides, Figure 7a also illustrates the CMC curve and shows that our network achieved the best performance of recognition for all ranks.

For the verification, Table 4 reveals that our network achieved the lowest EER with  $3.41 \pm 1.8\%$  and AUC was 0.9938. This is concrete evidence to demonstrate that the proposed network can verify the unconstrained periocular robustly. Figure 7b shows that our network outperformed most of the benchmark networks and achieved the highest recall rate against all other approaches.



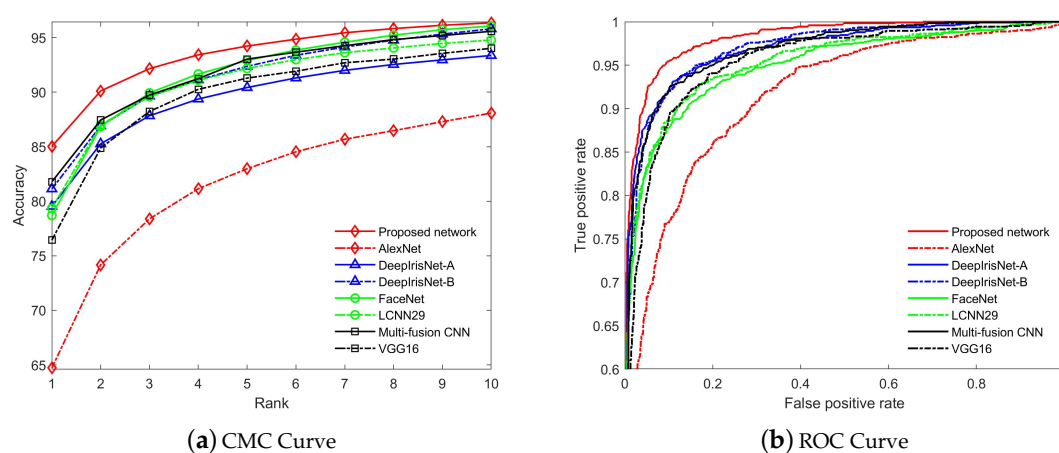
**Figure 7.** Performances of recognition and verification tasks on UBIPr distance database. The figures are the best to view in colour.



## Evaluation on Ethnic-Ocular Database

We present the experimental results in Table 3 by following the recognition protocol mentioned in Section 4.3. To evaluate the performance of the proposed approach, we compared our results with seven benchmark approaches (see Table 3). For the results of recognition, our network achieved  $84.79 \pm 1.9\%$  and  $94.23 \pm 1.3\%$  as Rank-1 and Rank-5 accuracies, respectively. Figure 8a illustrates the CMC curve of the proposed network, showing that the proposed method outperformed other benchmark methods from Rank-1 to Rank-10 recognition accuracies. The results indicate that the late-fusion layers are capable of correlating the RGB image and OCLBCP descriptor.

Table 4 also shows that the proposed network achieved the lowest EER accuracy with  $6.63 \pm 1.5\%$  for verification. Figure 8b illustrates the ROC curve, showing that our network outperformed all benchmark networks. The results prove that our approach can learn new features from the late-fusion layers in order to transfer knowledge between the networks to perform better performance of recognition.



**Figure 8.** Performances of recognition and verification tasks on Ethnic-ocular database. The figures are the best to view in colour.

### 5.2.3. Discussion

Through the experimental analysis and results, we observed that having access to the RGB image and OCLBCP descriptor can exploit the discriminatory features as inputs for a better periocular recognition. In addition, the proposed network utilises the colour-based texture information, which contributes to a more robust feature representation for the challenges in recognition and verification in the wild. This is because handcrafted texture descriptor can offer latent and complement information for complex data learning.

By evaluating across constrained environments, our results score higher accuracies consistently. Periocular recognition and verification in the wild bring more challenges as compared to the constrained environment. The experimental results prove that our network is able to perform better recognition due to its ability to learn new features from the proposed late-fusion layers. The effectiveness of fusion layers in the network supports our assumption firmly that multi-feature learning can work much better than just using RGB image in periocular recognition.

## 6. Conclusions

This paper proposed a dual-stream CNN, which accepts RGB ocular image and OCLBCP for periocular recognition in the wild. By aggregating the RGB image and OCLBCP features into two distinct late-fusion layers, these features offer robust and better recognition performance. We collected and shared a new Ethnic-ocular database, which consists of a large collection of periocular images in the wild based on different ethnic groups. Through extensive experiments by comparing against several



competing networks on new Ethnic-ocular database and publicly available databases, the proposed network achieved better performance in both recognition and verification tasks.

In the near future, we plan to investigate different kinds of fusion stages and fusion layers in CNNs, which could improve the performance of multi-feature learning. Periocular recognition is futile for subjects with “wearing sunglasses”. As a remedy, we shall incorporate the Generative Adversarial Model, which is useful to recover the periocular area in the face image.

**Author Contributions:** The authors have equivalent contributions.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NO. NRF-2019R1A2C1003306).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jain, A.K.; Nandakumar, K.; Ross, A. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recog. Lett.* **2016**, *79*, 80–105. [CrossRef]
2. Klare, B.F.; Klein, B.; Taborsky, E.; Blanton, A.; Cheney, J.; Allen, K.; Grother, P.; Mah, A.; Jain, A.K. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus benchmark A. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1931–1939.
3. Klontz, J.C.; Jain, A.K. A case study of automated face recognition: The Boston Marathon bombings suspects. *Computer* **2013**, *46*, 91–94. [CrossRef]
4. Barroso, E.; Santos, G.; Cardoso, L.; Padole, C.; Proença, H. Periocular recognition: How much facial expressions affect performance? *Pattern Anal. Appl.* **2016**, *19*, 517–530. [CrossRef]
5. Park, U.; Jillela, R.R.; Ross, A.; Jain, A.K. Periocular biometrics in the visible spectrum: A feasibility study. In Proceedings of the International Conferences on Biometrics: Theory, Applications and Systems (BTAS), Washington, DC, USA, 28–30 September 2009; pp. 1–6.
6. Bharadwaj, S.; Bhatt, H.S.; Vatsa, M.; Singh, R. Periocular biometrics: When iris recognition fails. In Proceedings of the International Conferences on Biometrics: Theory, Applications and Systems (BTAS), Washington, DC, USA, 27–29 September 2010; pp. 1–6.
7. Park, U.; Jillela, R.R.; Ross, A.; Jain, A.K. Periocular biometrics in the visible spectrum. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 96–106. [CrossRef]
8. Raja, K.B.; Raghavendra, R.; Stokkenes, M.; Busch, C. Smartphone authentication system using periocular biometrics. In Proceedings of the International Conferences on Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 10–12 September 2014; pp. 1–8.
9. Mokhayeri, F.; Granger, E.; Bilodeau, G. Synthetic face generation under various operational conditions in video surveillance. In Proceedings of the International Conferences on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 4052–4056.
10. The Korea Times. Available online: [https://www.koreatimes.co.kr/www/nation/2019/01/371\\_262460.html](https://www.koreatimes.co.kr/www/nation/2019/01/371_262460.html) (accessed on 12 February 2019).
11. Kitchen Decor. Available online: <https://kitchendecor.club/files/now-beckham-hairstyle-david.html> (accessed on 12 February 2019).
12. Padole, C.N.; Proença, H. Periocular recognition: Analysis of performance degradation factors. In Proceedings of the International Conferences on Biometrics (ICB), New Delhi, India, 29 March–1 April 2012; pp. 439–445.
13. Raja, K.B.; Raghavendra, R.; Stokkenes, M.; Busch, C. Collaborative representation of deep sparse filtered features for robust verification of smartphone periocular images. In Proceedings of the International Conferences on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 330–334.
14. Alonso-Fernandez, F.; Bigun, J. Periocular recognition using retinotopic sampling and Gabor decomposition. In Proceedings of the European International Conferences on Vision (ECCV), Firenze, Italy, 7–13 October 2012; pp. 309–318.
15. Cao, Z.; Schmid, N.A. Fusion of operators for heterogeneous periocular recognition at varying ranges. *Pattern Recognit. Lett.* **2016**, *82*, 170–180. [CrossRef]

16. Mahalingam, G.; Ricanek, K. LBP-based periocular recognition on challenging face datasets. *EURASIP J. Image Video Process.* **2013**, *36*, 1–13. [[CrossRef](#)]
17. Tan, C.-W.; Kumar, A. Towards online iris and periocular recognition under relaxed imaging constraints. *IEEE Trans. Image Process.* **2013**, *22*, 3751–3765.
18. Nigam, I.; Vatsa, M.; Singh, R. Ocular biometrics: A survey of modalities and fusion approaches. *Inf. Fusion* **2015**, *26*, 1–35. [[CrossRef](#)]
19. Raghavendra, R.; Busch, C. Learning deeply coupled autoencoders for smartphone based robust periocular verification. In Proceedings of the International Conferences on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 325–329.
20. Cho, S.R.; Nam, G.P.; Shin, K.Y.; Nguyen D.T.; Pham, T.D.; Lee, E.C.; Park, K.R. Periocular-based biometrics robust to eye rotation based on polar coordinates. *Multimed. Tools Appl.* **2017**, *76*, 11177–11197. [[CrossRef](#)]
21. Krizhevsky, A.; Sutskever, I.; Geoffrey, H. Imagenet classification with deep convolutional neural networks. In Proceedings of the International Conferences on Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
22. Anwer, R.M.; Khan, F.S.; van de Weijer, J.; Molinier, M.; Laaksonen, J. Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *138*, 74–85. [[CrossRef](#)]
23. Gangwar, A.; Joshi, A. DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In Proceedings of the International Conferences on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 2301–2305.
24. Proença, H.; Neves, J.C. Deep-PRWIS: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 888–896. [[CrossRef](#)]
25. Zhao, Z.; Kumar, A. Improving periocular recognition by explicit attention to critical regions in deep neural network. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2937–2952. [[CrossRef](#)]
26. Zhang, Q.; Li, H.; Sun, Z.; Tan, T. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2897–2912. [[CrossRef](#)]
27. Soleymani, S.; Dabouei, A.; Kazemi, H.; Dawson, J.; Nasrabadi, N.M. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification. In Proceedings of the International Conferences on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 3469–3476.
28. Levi, G.; Hassner, T. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In Proceedings of the International Conferences on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015; pp. 503–510.
29. CASIA-Iris Distance Database. Available online: <http://www.cbsr.ia.ac.cn/china/Iris%20Databases%20CH.asp> (accessed on 12 December 2018).
30. Marsico, M.D.; Nappi, M.; Riccio, D.; Wechsler, H. Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recognit. Lett.* **2015**, *57*, 17–23. [[CrossRef](#)]
31. Alonso-Fernandez, F.; Raja, K.B.; Raghavendra, R.; Busch, C.; Bigun, J.; Vera-Rodriguez, R.; Fierrez, J. Cross-sensor periocular biometrics: A comparative benchmark including smartphone authentication. *arXiv* **2019**, arXiv:1902.08123.
32. Rhee, S.C.; Woo, K.S.; Kwon, B. Biometric study of eyelid shape and dimensions of different races with references to beauty. *Aesthetic Plast. Surg.* **2012**, *36*, 1236–1245. [[CrossRef](#)] [[PubMed](#)]
33. Ethnic-Ocular Database. Available online: <https://www.dropbox.com/sh/vgg709to25o01or/AAB4-20q0nXYmgDPTYdBejg0a?dl=0> (accessed on 29 January 2019).
34. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
35. Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **2010**, *19*, 1635–1650.
36. Tiong, L.C.O. Multimodal Biometrics Recognition Using Multi-Layer Fusion Convolutional Neural Network with RGB and Texture Descriptor. Ph.D. Thesis, KAIST, Daejeon, Korea, 15 February 2019.
37. Delac, K.; Grgic, M.; Kos, T. Sub-image homomorphic filtering technique for improving facial identification under difficult illumination conditions. In Proceedings of the International Conferences on Systems, Signals and Image Processing, Budapest, Hungary, 21–23 September 2006; pp. 95–98.

38. Martinez W.L.; Martinez, A.R.; Solka, J. Chapter 3 Dimensionality reduction—Nonlinear methods. In *Exploratory Data Analysis with MATLAB*; Martinez W.L., Martinez, A.R., Solka, J., Eds.; CRC Press LLC: Boca Raton, FL, USA, 2005; pp. 61–68.
39. Feichtenhofer, C.; Pinz, A.; Zisserman, A. Convolutional two-stream network fusion for video action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1933–1941.
40. BBC News. Available online: <http://www.bbc.com/news> (accessed on 10 October 2018).
41. CNN News. Available online: <https://edition.cnn.com/> (accessed on 11 October 2018).
42. Naver News. Available online: <http://news.naver.com/> (accessed on 11 October 2018).
43. Ng, H.W.; Winkler, S. A data-driven approach to cleaning large face datasets. In Proceedings of the International Conferences on Image Processing (ICIP), CNIT La Défense, Paris, France, 27–30 October 2014; pp. 343–347.
44. Matlab Object Detector. Available online: <https://uk.mathworks.com/help/vision/ref/vision.cascadeobjectdetector-system-object.html> (accessed on 10 October 2018).
45. Štruc, V.; Pavešić, N. The complete Gabor-fisher classifier for robust face recognition. *EURASIP J. Adv. Signal Process.* **2010**, 1–26.
46. TensorFlow. Available online: <https://tensorflow.org> (accessed on 21 November 2018).
47. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823.
48. Wu, X.; He, R.; Sun, Z.; Tan, T. A light CNN for deep face representation with noisy labels. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2884–2896. [[CrossRef](#)]
49. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep face recognition. In Proceedings of the British Machine Vision Conference, Swansea, UK, 7–10 September 2015; pp. 1–12.
50. Hernandez-Diaz, K.; Alonso-Fernandez, F.; Bigun, J. Periocular recognition using CNN features off-the-shelf. *arXiv* **2018**, arXiv:1809.06157.
51. Martínez A.; Benavente, R. *The AR Face Database*; CVC Technical Report #24; Robot Vision Lab; Purdue University: Barcelona, Spain, 1998.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).