

Article

LDA-CNN: Linear Discriminant Analysis Convolution Neural Network for Periocular Recognition in the Wild

Amani Alahmadi *, Muhammad Hussain  and Hatim Aboalsamh 

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

* Correspondence: amalahmadi@ksu.edu.sa

Abstract: Due to the COVID-19 pandemic, the necessity for a contactless biometric system able to recognize masked faces drew attention to the periocular region as a valuable biometric trait. However, periocular recognition remains challenging for deployments in the wild or in unconstrained environments where images are captured under non-ideal conditions with large variations in illumination, occlusion, pose, and resolution. These variations increase within-class variability and between-class similarity, which degrades the discriminative power of the features extracted from the periocular trait. Despite the remarkable success of convolutional neural network (CNN) training, CNN requires a huge volume of data, which is not available for periocular recognition. In addition, the focus is on reducing the loss between the actual class and the predicted class but not on learning the discriminative features. To address these problems, in this paper we used a pre-trained CNN model as a backbone and introduced an effective deep CNN periocular recognition model, called linear discriminant analysis CNN (LDA-CNN), where an LDA layer was incorporated after the last convolution layer of the backbone model. The LDA layer enforced the model to learn features so that the within-class variation was small, and the between-class separation was large. Finally, a new fully connected (FC) layer with softmax activation was added after the LDA layer, and it was fine-tuned in an end-to-end manner. Our proposed model was extensively evaluated using the following four benchmark unconstrained periocular datasets: UFPR, UBIRIS.v2, VISOB, and UBIPr. The experimental results indicated that LDA-CNN outperformed the state-of-the-art methods for periocular recognition in unconstrained environments. To interpret the performance, we visualized the discriminative power of the features extracted from different layers of the LDA-CNN model using the t-distributed Stochastic Neighboring Embedding (t-SNE) visualization technique. Moreover, we conducted cross-condition experiments (cross-light, cross-sensor, cross-eye, cross-pose, and cross-database) that proved the ability of the proposed model to generalize well to different unconstrained conditions.

Keywords: periocular biometric; mobile biometrics; deep learning; convolutional neural network; transfer learning; fine-tuning; linear discriminant analysis

MSC: 68T07



Citation: Alahmadi, A.; Hussain, M.; Aboalsamh, H. LDA-CNN: Linear Discriminant Analysis Convolution Neural Network for Periocular Recognition in the Wild. *Mathematics* **2022**, *10*, 4604. <https://doi.org/10.3390/math10234604>

Academic Editor: Dan Vilenchik

Received: 27 October 2022

Accepted: 30 November 2022

Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recognition of individuals using various biometric modalities in an unconstrained (i.e., in the wild) environment has emerged as an active research topic in the past decade [1–4]. Images in the wild environment are captured under uncontrolled conditions, as is common in surveillance-based applications. This includes variations in lighting, pose, expression, occlusion, and resolution. The human face has been proven to be the most popular and accurate biometric modality [5]. However, the performance of face recognition systems declines when the face is partially hidden or is in an unconstrained environment [6]. This limitation was addressed by the US police departments in 2013 after the investigation of the

Boston Marathon bombings [7]. Surveillance videos usually capture only a partial segment of the criminal's face. In some situations, the face is either covered by helmets, hair, glasses, or masks (COVID-19 pandemic response). Furthermore, some women cover their faces partially due to cultural and religious reasons. In these scenarios, the region around the eyes, periocular region, is the only visible trait that can be used as a biometric modality (see Figure 1). The periocular region contains the eye and its immediate vicinity, including eyelids, eyelashes, nearby skin area, and eyebrows (see Figure 2). It provides a trade-off between the whole face and the iris alone. Unlike other ocular biometrics (e.g., iris, retina, and sclera), the acquisition of the image of the periocular region does not require high user cooperation and close capture distance. It is also less impacted by aging and changing expressions than other facial areas [8]. In addition to serving as a stand-alone modality, periocular can be combined with other biometric traits, like face and/or iris, to improve recognition performance [9,10]. For all these reasons, periocular recognition emerged as an area of research interest in the biometrics community.



Figure 1. Examples of some cases in which using periocular biometrics is efficient.

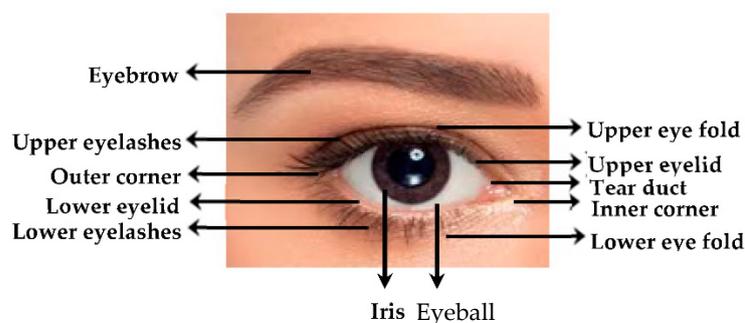


Figure 2. Periocular region and its elements.

The earlier studies on periocular biometrics adopted several handcrafted descriptors, such as local binary patterns (LBP) [11–14], histogram of oriented gradient (HOG) [15], scale-invariant feature transform (SIFT) [16], speeded up robust feature (SURF) [17] and phase intensive local pattern (PILP) [18]. However, these approaches were not robust to handling unconstrained variations, due to the inadequacy of handcrafted descriptors, which are custom designed to encode a specific representation [1]. A natural solution to overcome this disadvantage was to aggregate features from several descriptors; however, that led to the curse of the dimensionality problem [6].

In recent years, Convolutional Neural Networks (CNNs) have attracted many researchers in the field of visual recognition applications and have significantly outperformed traditional handcrafted methods or other learning-based approaches [19]. To train CNNs, a large-scale labeled database is needed [20]. This requirement is a major challenge for periocular recognition, as there is currently no public database with a huge number of labeled images [2,4]. To overcome this limitation, transfer learning-based methods were

proposed. These methods rely on the idea of pre-training a model on an independent task, having an existing large database, and then reusing that model on another related task [20]. In this direction, CNN-based periocular recognition methods were proposed and achieved a remarkable improvement over the hand-engineered techniques decade [2–4]. In spite of the significant CNN-based periocular recognition research, the development of robust and efficient periocular recognition techniques operating in the wild environment remains a challenging research problem [1]. This is largely due to the lack of information revealed by the periocular region, compared with the whole face, which makes it highly affected by variations in illumination, pose, resolution and occlusions which occur in wild environments [21,22]. Such variations increase the challenges of within-class variability and between-class similarity (as illustrated in Figure 3), which are considered important sources of recognition error [23,24].

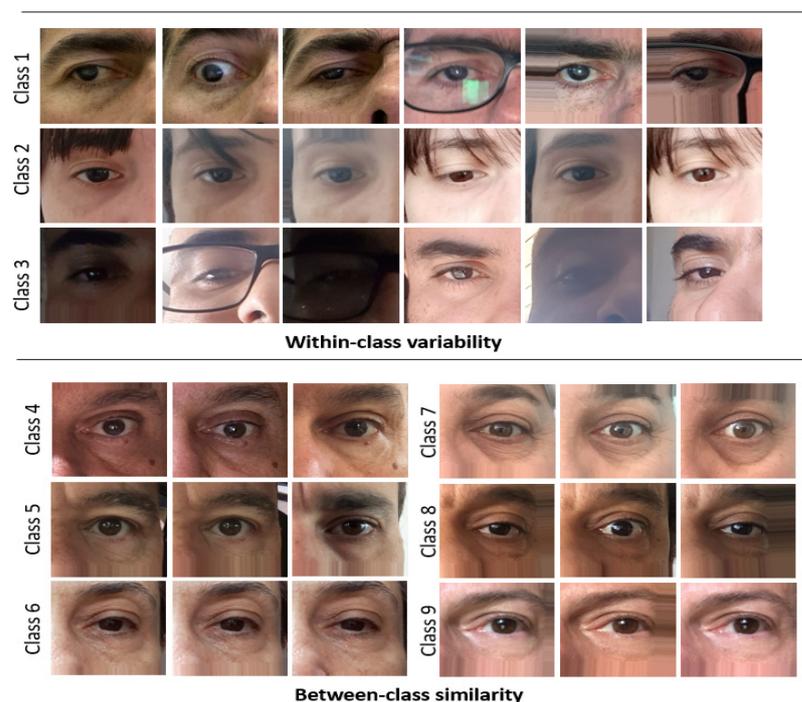


Figure 3. Within-class variability and between-class similarity challenges for periocular recognition in the wild. Here we show some examples for each challenge obtained from UFPR dataset.

To tackle this problem, in this work we proposed an end-to-end deep CNN periocular recognition model called linear discriminant analysis CNN (LDA-CNN), which introduces an LDA layer on top of the existing CNN architectures. The motivation for adding the LDA layer was to force the network to produce features with low variance within the same class and high variance between different classes. Specifically, the LDA layer was introduced after the last convolution layer of the fine-tuned backbone CNN model. The parameters of the LDA layer were trained by imposing the linear discrimination criterion [25,26] on the CNN features so that they had small within-class variation and large between-class separation. The greatest challenge that LDA encountered in many practical applications was the singularity, or small sample size, problem. When the dimension of the data is significantly greater than the number of training samples, the within-class scatter matrix is singular. As the dimension of the CNN features is very high and considerably larger than the number of training images, LDA is not applicable. To resolve the singularity issue, a principal component analysis (PCA) layer was introduced before the LDA layer to reduce the dimension of CNN features. Finally, a new fully connected (FC) layer was added after the LDA layer and fine-tuned in an end-to-end manner.

The proposed model was evaluated on four benchmark unconstrained periocular datasets: UFPR [27], UBIRIS.v2 [28], VISOB [29], and UBIPr [15]. It consistently outperformed the other state-of-the-art methods on all datasets. Even under challenging cross-condition (i.e., cross-light, cross-sensor, cross-eye, cross-pose, and cross-database) training and testing protocols, LDA-CNN exhibited a high level of robustness. To highlight the effect of the LDA layer, we conducted an ablation study to examine the performance of the proposed LDA-CNN model without the LDA and PCA layers. In this regard, we built and assessed two models: (1) FT-CNN, in which the LDA and PCA layers were both removed, and (2) PCA-CNN, in which just the LDA layer was removed. A new FC layer was added and adjusted end-to-end in both models.

To provide a visual explanation that elucidates the periocular areas our LDA-CNN model was concentrating on to make the prediction, we employed the gradient-weighted class activation mappings (Grad-CAM) [30]. Grad-CAM uses gradients to localize pixels in the activation map that strongly contribute to the model's prediction. We found that most of the strong features were concentrated on the eyes and the surrounding skin more than on eyebrows. Moreover, we visualized the features of different layers of the LDA-CNN model on 2D space using the t-distributed Stochastic Neighboring Embedding (t-SNE) visualization technique [31] and we found that the LDA layer's features were discriminative and had large between-class distance and small within-class distance.

Our main contributions are summarized as follows:

- We introduced an end-to-end LDA-CNN model that uses a pre-trained CNN model as a backbone and incorporates an LDA layer to ensure learning discriminative feature representation with small within-class scatter and large between-class separation.
- We validated the importance of the LDA layer by performing an ablation study in which the performance of the proposed LDA-CNN model was examined without the LDA and PCA layers. The results showed that the LDA layer is effective in enhancing the performance significantly.
- We used t-SNE visualization to interpret the discriminative power of the LDA-CNN features, and we discovered that the LDA layer features produce clear clusters with large between-class separation and small within-class distance.
- The Grad-Cam method was employed to highlight the most important periocular regions that our model focused on to make the prediction, and we discovered that the eyes and the surrounding skin are more significant than the eyebrows.
- The proposed model was extensively evaluated and compared with the state-of-the-art methods on four benchmark unconstrained periocular datasets: UFPR, UBIRIS.v2, VISOB, and UBIPr. The results indicated that LDA-CNN outperforms the state-of-the-art methods.
- We introduced a robustness analysis and proved the generalizability of the model to different wild environmental conditions. For this, we performed cross-condition experiments (i.e., trained the model using one condition and tested it with another one), which were: cross-light, cross-sensor, cross-database, cross-eye, and cross-pose.

The rest of this paper is organized as follows. Section 2 describes the proposed model. The experimental setup includes the evaluation protocol, and the dataset description is provided in Section 3. The design choices of the LDA-CNN model are discussed in Section 4, while Section 5 provides the experimental results and discussion. Finally, Section 6 presents the conclusion.

2. Proposed Model

A general framework of the proposed model (LDA-CNN) is shown in Figure 4. It consists of three main parts: the backbone CNN model, the PCA layer, and the LDA layer. First, the last FC layer of the pre-trained backbone model was removed, and a new FC layer was included to adjust the network for the new number of training classes on the periocular database. Once the transfer was finished and the model was adapted to the new periocular domain, two FC layers, i.e., PCA and LDA, were added after the last convolution layer. The

weights of the PCA and LDA layers were the principal eigenvectors, which were calculated from the activations of the last convolutional layer of the backbone model using PCA and LDA algorithms, respectively. Finally, again, a new FC layer was added after the LDA layer and fine-tuned by freezing the weights of the whole model, except that of the FC layer. A detailed description of each part of the proposed architecture is discussed in the following subsections.

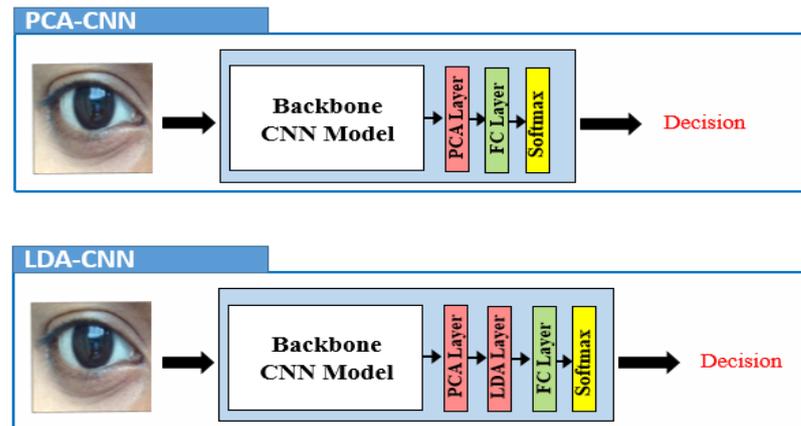


Figure 4. General framework of PCA-CNN and LDA-CNN (proposed) models.

2.1. Backbone CNN Model

Convolutional neural network (CNN) is a feedforward neural network with a deep structure and convolutional-based calculations. It is the most powerful deep learning algorithm used for visual recognition tasks [19]. Unlike conventional machine learning models that require a pipeline of multiple stages (i.e., preprocessing, feature extraction, feature selection, and classifier), CNN is an end-to-end learning model that replaces the pipeline with a single learning algorithm that accepts input from one end and produces output at the other end. CNN learns feature representations from data, based on a multilayer architecture consisting of convolution layers (CONV), pooling layers (POOL), and fully connected layers (FC), stacked alternately. In the literature, many CNN-based architectures have been proposed, such as: ResNet50 [19], GoogLeNet [32], VGG16 [33], DenseNet-201 [34], MobileNetV2 [35], EfficientNet-B0 [36], Xception [37] and InceptionResNet-v2 [38]. All these models are pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) database [39], which consists of more than a million images.

To overcome the lack of a large-scale periocular dataset, in this work, we exploited the transfer learning technique by using a CNN pre-trained model as a backbone model and then fine-tuned it using a periocular dataset to build an end-to-end CNN model. Similar to recent works on periocular recognition [40–43], we modified the backbone CNN model by discarding the last fully connected (FC) layer and adding a new FC layer with a number of neurons equal to the number of classes in the new periocular database. The remaining weights of the network were initialized from the pre-trained model, and then the entire network was retrained with the periocular database until convergence. We evaluated the following eight dominant CNN architectures: DenseNet-201, VGG16, ResNet50, GoogLeNet, MobileNetV2, EfficientNet-B0, Xception, and Inception-ResNet-v2. We found that DenseNet-201 achieved the best result (Section 4.1). DenseNet-201 (Dense Convolutional Network) is a 201-layers deep CNN architecture which connects each layer to every other layer in a feed-forward fashion. Each layer passes its output as an additional input to all the subsequent layers. These connections can mitigate the vanishing-gradient problem, increase feature propagation, and reduce the number of parameters. The internal description of DenseNet-201 is shown in Table 1. Figure 5 shows an instance of the LDA-CNN model using DenseNet-201 as a backbone model.

Table 1. The internal structure of DenseNet201 CNN model.

Layer	Input	Output	Description
Convolution	224×224	112×112	7×7 conv, stride 2
Pooling	112×112	56×56	3×3 max pool, stride 2
Denseblock (1)	56×56	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	56×56	$1 \times 1 \text{ conv}$
Denseblock (2)	28×28	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	28×28	$1 \times 1 \text{ conv}$
Denseblock (3)	14×14	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Transition Layer (3)	14×14	14×14	$1 \times 1 \text{ conv}$
Denseblock (4)	7×7	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$
Classification Layer	7×7	1×1	7×7 global average pool 1000D fully connected, softmax

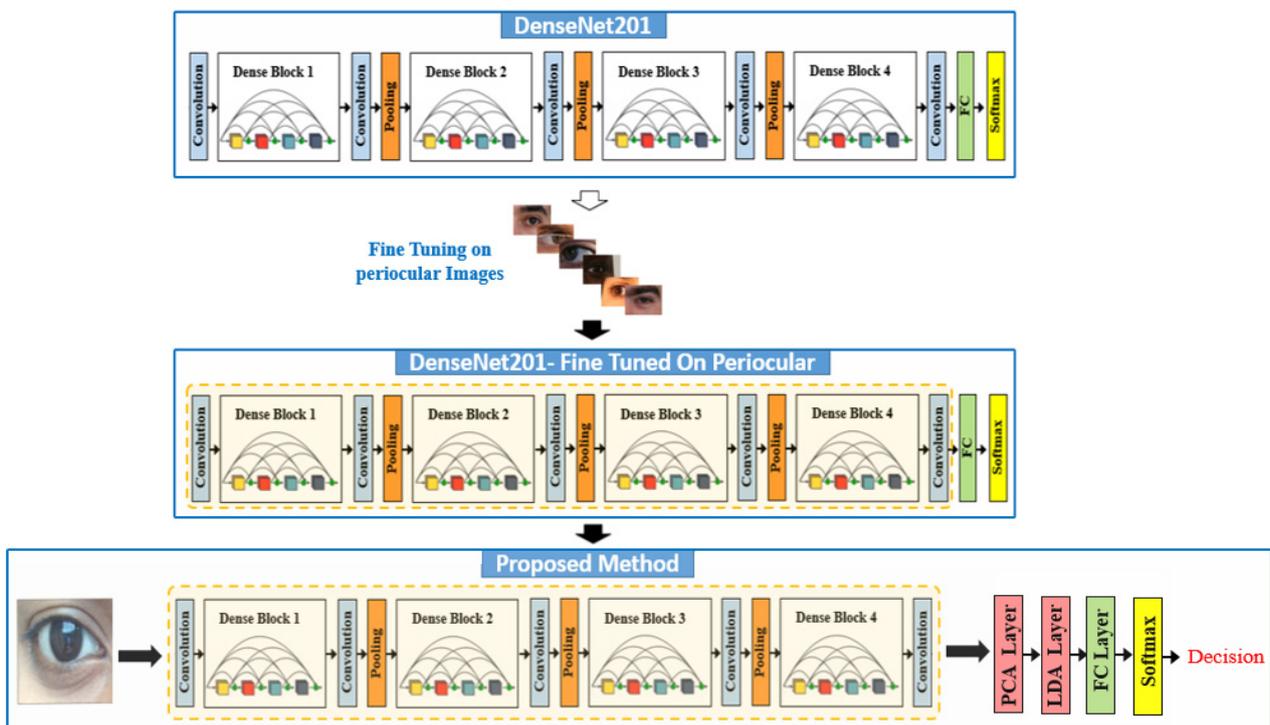


Figure 5. An instance of LDA-CNN model with DenseNet201 as backbone model.

2.2. PCA Layer

After fine-tuning the backbone model on the periocular dataset, the next step was to select the last convolution layer of the fine-tuned model (“conv5_block32” in Denseblock (4) of DenseNet-201) and feed its output to a new fully connected layer, known as the PCA layer. The reason for introducing this layer was to reduce the dimension of the features before passing them to the LDA layer to avoid the within-class singularity problem of the LDA algorithm. The weights and biases of PCA layer were calculated using the Principal Component Analysis (PCA) [44]. PCA is a dimensionality reduction technique that can be used to remove redundancy between correlated features in a dataset. It finds new

orthogonal features that are linear combinations of all input features. These new features are ranked, based on the amount of variance they can explain of the input data. For a X -dimensional dataset $\{x_i\}$ of size N , PCA [44] generates a Y -dimensional feature set $\{y_i\}$ of the same size, $X > Y$, by using the linear transformation:

$$y_i = W^T x_i \tag{1}$$

The new fully connected layer, PCA layer, uses the transformation matrix W to initialize the weights W_{PCA} and biases b_{PCA} of this layer. Algorithm 1 presents the detailed description of the steps of calculating W_{PCA} and b_{PCA} .

Algorithm 1: PCA Layer—Weights Computation

Input: Training images I_1, I_2, \dots, I_N , CNN model *DeepNet*, energy threshold t .

Output: The weights of PCA layer $W_{PCA} \in R^{K \times L}$, L is the size of vectorize (Conv), and the biases $b_{PCA} \in R^{k \times 1}$

Processing:

1. For $i = 1..N$
2. Conv = *DeepNet* (I_i)
3. $x_i = \text{vectorize}(\text{Conv})$
4. $x = x + x_i$
5. $\bar{x} = \frac{x}{N}$
6. For $i = 1..N$
7. $y_i = x_i - \bar{x}$
8. $A = [y_1, y_2, \dots, y_N]$
9. $C = A^T A$
10. Compute eigenvalues and eigenvectors of C
11. Select K principal eigenvectors corresponding to K largest eigenvalues v_1, v_2, \dots, v_K that satisfy the criteria

$$\frac{\sum_{i=1}^K v_i}{\sum_{i=1}^N v_i} > t$$

12. For $i = 1..K$
 13. $u_i = A v_i$
 14. Form PCA Matrix $M_{PCA} = [u_1, u_2, \dots, u_K]$
 15. PCA Layer Weights $W_{PCA} = M_{PCA}^T$
 16. PCA Layer Biases $b_{PCA} = -M_{PCA}^T \bar{x}$
-

2.3. LDA Layer

The output of the PCA layer is passed to another new fully connected layer, named the LDA layer. The weights and biases of LDA layer are computed using Linear discriminant analysis (LDA) [25]. LDA is used to increase the separation between different classes and reduce the variation within similar classes. This can be done by applying the Fisher’s linear discriminant criterion [26]. This criterion tries to maximize the ratio of the between-class scatter to the within-class scatter of the projected samples. Images are projected to $C - 1$ dimensional space (where C is the number of classes). Algorithm 2 presents the detailed description of the steps of calculating the wights W_{LDA} and biases b_{LDA} of LDA Layers.

Algorithm 2: LDA Layer—Weights Computation

Input: Training data of PCA layer $X \in \mathbb{R}^{K \times N}$, $x_i \in \mathbb{R}^{K \times 1}$ is the column of X , label $y_i \in \{1, 2, \dots, C\}$, $i = 1, \dots, N$, $N_c, c = 1, \dots, C$ is the number of samples of each class.

Output: The weights of LDA layer $W_{LDA} \in \mathbb{R}^{P \times K}$ and the biases $b_{LDA} \in \mathbb{R}^{P \times 1}$.

Processing:

1. Calculate mean vector of each class: $m_c = \frac{\sum_{i=1}^{N_c} x_i}{N_c} \in \mathbb{R}^{K \times 1}$
2. Calculate total mean vector: $m = \frac{\sum_{i=1}^N x_i}{N} \in \mathbb{R}^{K \times 1}$
3. Calculate within-class scatter: $S_w = \sum_{c=1}^C \sum_{j=1}^{N_c} (x_j - m_c)(x_j - m_c)^T \in \mathbb{R}^{K \times K}$
4. Calculate between-class scatter: $S_b = \sum_{c=1}^C \sum_{j=1}^{N_c} (m_c - m)(m_c - m)^T \in \mathbb{R}^{K \times K}$
5. Apply eigenvalue decomposition: $[V, D] = \text{eig}(S_w^{-1} S_b)$
6. Select P principal eigenvectors corresponding to P largest eigenvalues
7. Form LDA weights matrix $M_{LDA} = [V_1, V_2, \dots, V_P]$
8. LDA Layer Weights $W_{LDA} = M_{LDA}^T$
9. LDA Layer Biases $b_{LDA} = -M_{LDA}^T m$

2.4. Training of the Proposed Architecture

To deploy the model for classification, the LDA layer was followed by a new fully connected layer with softmax activation, and then the model was trained end-to-end for the final periocular recognition. The number of neurons in the FC layer was equal to the number of classes in the dataset. During this training, we only fine-tuned the last FC layer for classification while keeping the remaining network frozen.

3. Experimental Setup

We implemented the model using MATLAB R2021a on a PC with AMD Ryzon 9 5950X 16-Core CPU @ 3.40 GHz and 128 GB RAM. The identification performance of the system is reported in terms of accuracy.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of prediction}}$$

Further, the performance was also visualized using Cumulative Match Characteristic (CMC) curves.

3.1. Benchmark Datasets

The performance of the proposed periocular recognition system was evaluated using the most challenging benchmark periocular datasets: UBIPr, UFPR, UBIRIS.v2 and VISOB. These datasets are publicly available and commonly used in the literature where the performance of the proposed system can be fairly compared. In addition, these datasets contain several variabilities, like those existing in the wild environment, including illumination, pose, distance, expression, and occlusion. Table 2 reviews the number of subjects, number of images, image sizes, and the existing variations for these datasets. Sample images from each dataset are shown in Figure 6, and a brief description is provided in the following subsections.

3.1.1. University of Beira Interior Periocular (UBIPr)

This dataset consists of 10,252 periocular images (5126 left and 5126 right) from 344 subjects. The images are captured in the visible spectrum (VIS) with varying subject-camera distances (4 m–8 m), occlusions, levels of illumination, and poses [15]. Following the same protocol in [41], 80% of the dataset were used for training (80% for training and 20% for validation) and the remaining 20% were used for testing.

Table 2. Description of the collected benchmark databases.

Dataset	#Subjects	# Images	Image Size	Distance	Pose	Variations			
						Illumination	Expression	Occlusion	Makeup
VISOB	550	95,046	240 × 160	✓	×	✓	×	✓	✓
UBIPr	344	10,252	varied	✓	✓	✓	×	✓	×
UFPR	1122	33,660	256 × 256	✓	✓	✓	✓	✓	✓
UBIRIS.v2	261	11,102	400 × 300	✓	✓	✓	×	✓	×

**Figure 6.** Sample images from: (a) UBIPr (b) VISOB (c) UBIRIS.v2. (d) UFPR.

3.1.2. UFPR-Periocular Dataset

This dataset is designed to obtain images in unconstrained environments with realistic noise resulting from occlusion, blur, and fluctuations in lighting, distance, and angle. It consists of 16,830 images of both eyes from 1122 individuals. The periocular area is cropped and then divided into two patches to produce the left and right eye sides, yielding 33,660 periocular images from 2244 classes. The dataset's variance is mostly attributable to illumination, occlusion, blur, eyeglasses, off-angle, eye-gaze, cosmetics, and facial expression. Three sessions were used to capture the images. Using one session as a test set and the remaining two as a training/validation set generated three folds. More details about this dataset are provided in [27].

3.1.3. University of Beira Interior Iris (UBIRIS.v2)

This is acquired in VIS for the purpose of simulating the unconstrained environment of the real world. It is mostly used for evaluating at-a-distance iris recognition algorithms under visible illumination and difficult imaging conditions. There are various variations associated with the eye images, such as specular reflections, partial iris reflections, and poor focus of the iris, blur motion, and glare. It contains a total of 11,102 images of the left and right periocular regions from 261 subjects [28]. We made our dataset protocol consistent with [45] and we used all 11,102 images corresponding to the left and right eyes for the training and evaluation of the proposed model.

3.1.4. Visible Light Mobile Ocular Biometric Database (VISOB)

This is one of the competition datasets. It contains ocular images of 550 adult volunteers acquired using the front-facing cameras of three different smartphones, i.e., the Samsung Note 4, iPhone 5s, and Oppo N1. The images were captured under different lighting conditions, which were: office light, dim light, and daylight. The images varied in illumination, off-gaze angles, makeup, blur, and occlusion [29]. The dataset was divided into training and testing sets as listed in Table 3.

Table 3. Details of the training and testing sets of the VISOB dataset.

Phone	Day		Training Dim		Office		Day		Testing Dim		Office	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
Iphone	2622	2648	1865	1897	2522	2523	2536	2567	1763	1789	2261	2292
Samsung	1582	1648	2074	2220	2255	2418	1587	1678	2007	2175	2336	2456
Oppo	1963	1963	3749	3748	5284	5269	1985	1985	3742	3740	4962	4935
Total	6167	6259	7688	7865	10,061	10,210	6108	6230	7512	7704	9559	9683

3.2. Data Augmentation

Training deep models with millions of parameters needs a large-scale dataset. One solution for addressing this limitation of periocular datasets is to augment the training dataset by applying image transformation techniques. We carefully applied a small amount of augmentation to the data, which were the following: translation: (-30 to 30) and scaling factor: (0.9 to 1.1). Applying such a small transformation to the original images was shown to improve the accuracy.

4. Design Choices

The LDA-CNN model involves various parameters and design decisions which require adequate tuning in the following: the selection of the backbone model, the number of principal components (PCs) in PCA and LDA layers, fine tuning or transfer learning of the layers. For these decisions, in this section, a set of experiments were conducted using the UFPR dataset.

4.1. Selection of the Backbone Model

We employed different backbone CNN models that achieved promising results in the ImageNet dataset and were applied in recent works of ocular recognition [27,40,41]. These models were: DenseNet-201, VGG16, ResNet50, GoogLeNet, MobileNetV2, EfficientNet-B0, Xception, and Inception-ResNet-v2. The experiments were performed by extracting the features from the last convolution layer of the CNN model (without finetuning), and, then, PCA and LDA were applied. Sparse Augmented Collaborative Representation-based Classification (SA-CRC) [46] was used for recognition. It can be clearly noticed from Table 4 and Figure 7, that DenseNet-201 and EfficientNet-B0 achieved the best performance among the other CNN models. To decide, we implemented the LDA-CNN model on top of DenseNet-201 and EfficientNet-B0. The results of these models are provided in Table 5, which shows the superiority of using the DenseNet-201 model. The possible reason for such an achievement might be due to the fact that the DenseNet-201 model reused the features that made the parameter complexity of DenseNet-201 lower than that of the other models. Hence, we selected the model fine-tuned on DenseNet-201 for the remaining experiments.

Table 4. A comparison of the performance of the features extracted by applying PCA and LDA on the last convolution layers of different CNN backbone models without finetuning. The results were obtained using SA-CRC on UFPR dataset.

Backbone CNN Model	Accuracy (%)
Xception	70.69
MobileNetV2	77.19
DenseNet201	93.92
EfficientNetB0	94.15
Googlenet	66.738
ResNet50	84.42
InceptionResNet	74.77
VGG16	90.87

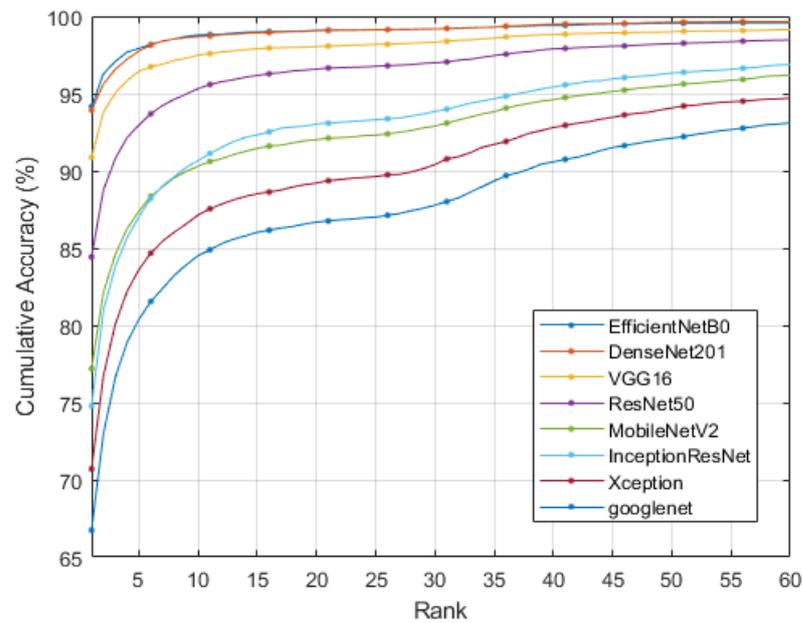


Figure 7. The CMC curves for the feature extracted from the last convolution layer of different backbone model using UFPR dataset.

Table 5. Performance comparison of using DenseNet-201 and Efficientnet-B0 as backbone model for LDA-CNN model. The results were obtained using UFPR dataset.

Model	Train	Accuracy (%) Test
DenseNet-201	99.24	97.68
Efficient-B0	-	90.37

4.2. Fine-Tuning/Transfer Learning

The first step of the proposed model was finetuning the backbone model to adapt it to the periocular dataset. This was done by removing the last FC layer and adding a new FC layer with a number of neurons equal to the number of classes in the periocular dataset. In transfer learning, the previous n layers (i.e., before the FC layer) of the target network were randomly initialized by the previous n layers of the pretrained network. Then back-propagation training was applied, either by freezing the previous n layers (fixing their weight values) or fine-tuning them (adjusting their weight values). Experiments were conducted using the DenseNet-201 model and the UFPR dataset to examine the above approaches by selecting different values of the learning rate factor (0, 1, or 4; where 0 meant freeze the layer) of the last FC layer and the layers before it. It can be noticed from Table 6 that the model performed better if the entire model was fine-tuned with the learning rate factor of the last FC layer (i.e., equals 4) higher than that of the layers (i.e., equals 1) before it. This was mostly due to the random initialization of the weights of the new final FC layer, whereas the weights of the preceding layers had already been trained on ImageNet and simply required fine-tuning for better adaptation to the periocular domain. The learning rate (LR) was used for 100 epochs, with LR = 0.0001. We stopped the training after 5 epochs without improvement in validation error. The remaining parameters were inherited from the CNN backbone model and were kept.

Table 6. Performance comparison of freezing (learning rate factor equals 0) and unfreezing (learning rate factor equals 1 and 4) the layers on transfer learning of DenseNet-201 model using UFPR dataset.

Learning Rate Factor		Accuracy (%)	
All layers before last FC layer	Last FC layer	Validation	Test
0	1	62.37	31.16
	4	80.62	37.86
1	1	97.99	81.21
	4	98.71	84.50

4.3. PCA and LDA Weights

The proposed approach included PCA and LDA layers. In both layers, the critical point was the selection of the number of principal components (PCs) (i.e., FC layer weights) corresponding to the largest eigenvalues. For the PCA layer, we compared the proportion of the variance of different periocular datasets (i.e., UFPR, UBIPr, UBIRIS.v2, and VISOB). As can be seen from Figure 8a, we found that the selection of the largest 900, 1100, 1300, or 1500 PCs was a good trade-off between the variance proportion and the computational complexity for the four datasets. Figure 8b demonstrates that the largest 1300 PCs provided the maximum accuracy. Hence, the number of PCs for the PCA layer was fixed at 1300 for onward experiments.

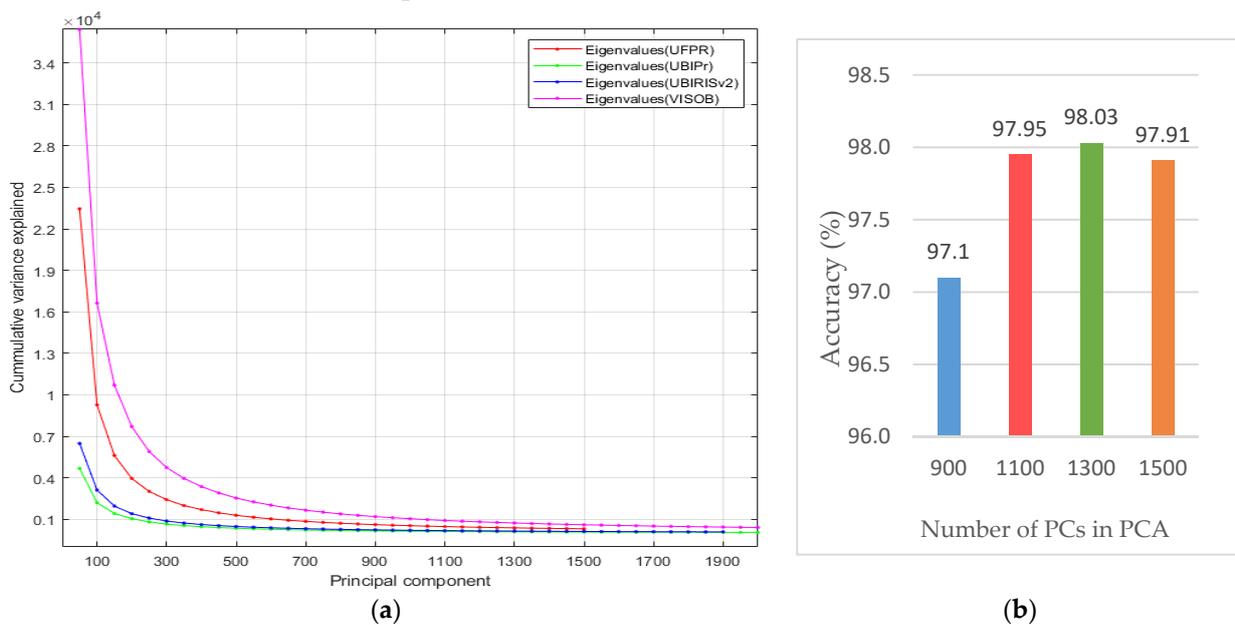


Figure 8. The selection of the number of principal components (PCs): (a) The proportion of the variance of the principal components for UFPR, UBIPr, UBIRIS.v2 and VISOB datasets; (b) The performance of LDA-CNN model with different number of PCs using fold 1 UFPR dataset.

As the maximum number of PCs in LDA was $C - 1$ (where C is the number of classes), we selected C to be the smallest number of classes in all datasets, which was 261 in UBIRIS.v2. Hence, the number of PCs selected for the LDA layer was 260.

5. Experimental Results and Discussion

In this section, thorough experiments were performed to evaluate the LDA-CNN model from different aspects, and comparisons made with state-of-the-art methods on UBIPr, UFPR, UBIRIS.v2 and VISOB datasets. Extensive experiments were also conducted to demonstrate the robustness of the proposed model. Further, an ablation study was applied to investigate the effect of PCA and LDA layers on the performance of the LDA-

CNN model. Finally, distinct layers of the LDA-CNN model were visualized, using the t-SNE algorithm, to show the discriminative power of LDA-CNN features.

5.1. Comparison with State-of-the-Art Methods

The performance of the proposed approach was compared with state-of-the-art methods in the literature on four unconstrained periocular datasets: UBIPr, UFPR, UBIRIS.v2 and VISOB. Figure 9a provides the identification result on UBIRIS.v2 (99.82%) compared to [40,45,47–50], while Figure 9b compares the identification result of UBIPr (99.17%) to those of [21,51,52]. The comparison results on VISOB and UFPR datasets are listed in Tables 7 and 8, respectively. From these results, it can be observed that our approach consistently outperformed the state-of-the-art methods on all datasets.

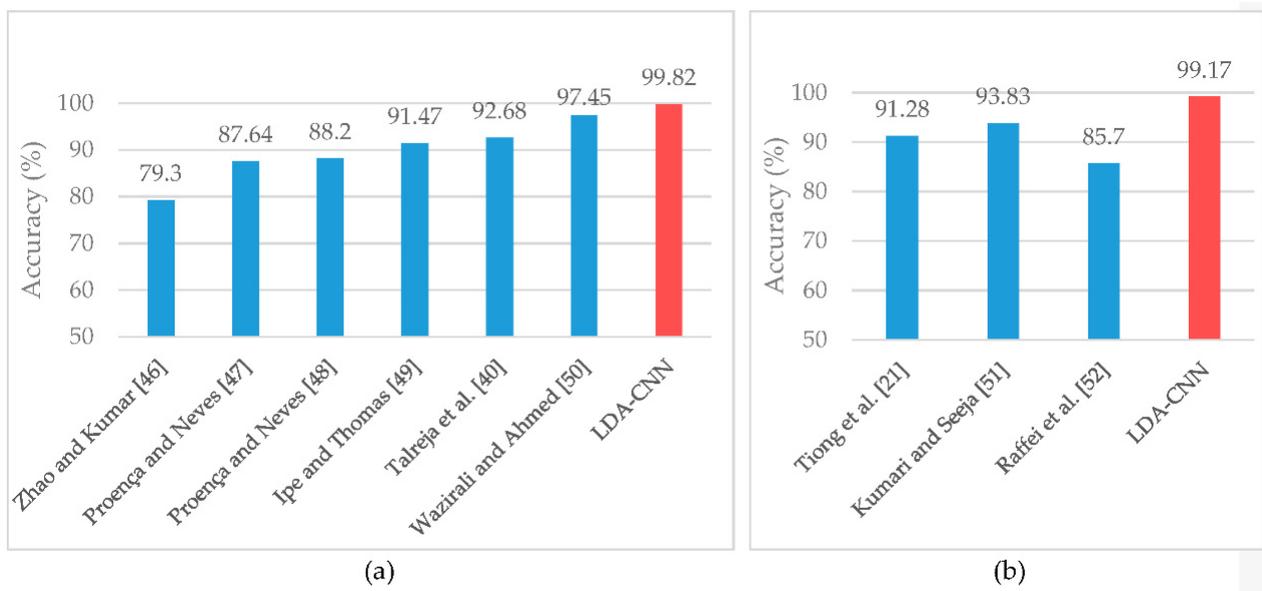


Figure 9. Comparison of the performance of LDA-CNN model with state-of-the-art methods using: (a) UBIRIS.v2; (b) UBIPr datasets.

Table 7. Comparison of the performance of LDA-CNN model with existing method using VISOB dataset.

Side	Phone	Light Condition	Ahuja et al. [53]	Ahuja et al. [54]	ConvSRC [55]	LDA-CNN
Left	Samsung	Day	64.18	92.44	98.1843	99.75
		Dim	63.97	93.12	98.5234	99.45
		Office	48.76	90.45	96.9789	98.89
	iPhone	Day	71.05	95.98	99.7541	99.57
		Dim	59.97	96.00	98.9298	99.55
		Office	55.37	93.54	99.0625	99.34
	Oppo	Day	64.48	94.21	98.9583	99.50
		Dim	78.51	96.31	97.4012	98.69
		Office	61.25	90.79	96.3258	97.72
Right	Samsung	Day	68.22	92.97	98.4217	99.88
		Dim	67.40	93.61	98.1733	99.40
		Office	55.05	91.53	97.0143	98.82
	iPhone	Day	70.13	94.82	99.4309	99.26
		Dim	56.34	96.14	98.7719	99.39
		Office	55.32	93.89	98.7556	99.21
	Oppo	Day	65.54	94.81	98.8542	99.45
		Dim	79.49	96.15	97.6977	98.88
		Office	57.34	90.23	96.2194	96.56

Table 8. Comparison of the performance of LDA-CNN model with existing method using UFR dataset.

Method	Accuracy (%)
Zanlorensi et al. [27]	84.32 ± 0.71
LDA-CNN	97.68 ± 0.31

5.2. Robustness Analysis

As the focus of this work was to design a periocular model for the unconstrained environment, in this section, several experiments were performed to evaluate the robustness of the proposed method and its ability to generalize to deferent non-ideal conditions.

5.2.1. Cross-Eye

The human body is essentially a bilaterally symmetrical structure, which makes it appear more appealing and function more effectively [56]. Left and right sides of the body, like the eyes, ears, and limbs, always appear identical when viewed from the middle vertical plane. The interocular symmetry of a variety of biometric measures between both eyes was assessed in [56]. The results of this study indicated that a normal person's eyes are highly symmetrical. Notably, the experiments on UBIPr and UBIRIS.v2 assumed bilateral symmetry between the left and right eyes, as the protocol for these datasets treated the left and right eyes as a single class and included them both in training and testing the model. This section looked at how well the model generalized when trained and tested on various periocular sides, as well as how it performed when trained and tested on the same periocular sides. To accomplish this, we tested the following cases on the UBIPr dataset using the same protocol as that in [41]:

- Case 1: Right-side periocular images were used for training, validation, and testing.
- Case 2: Left-side periocular images were used for training, validation, and testing.
- Case 3: Left-side images were used for training and validation, while right-side images were used for testing.
- Case 4: Left-side images were used for training and validation, while the mirror of the right-side images was used for testing.
- Case 5: Right-side images were used for training and validation, while left-side images were used for testing.
- Case 6: Right-side images were used for training and validation, while the mirror of the left-side images was used for testing.

To test all potential cases, we added two additional cases (Cases 5 and 6) that were not included in [41]. Table 9 lists the divisions of the training, validation, and testing sets for each case, while Table 10 provides the identification accuracy and comparisons with state-of-the-art methods for these cases. Table 10 demonstrates that when we conducted experiments for the right side only (Case 1) and the left side only (Case 2), the identification accuracy increased to 100% and 99.92%, respectively, compared to 99.82% when the model was trained and tested on both sides. In addition, based on the results of the remaining cases (Cases 3–6), we concluded that the proposed model trained on one periocular side generalized well compared to other state-of-the-art methods and was capable of predicting the images of the other side with an identification accuracy of 65.02% and 66.76% in Cases 3 and 5, respectively. When the tested images were reflected (see Figure 10), there was an improvement of between 5.03% and 9.07% in Case 5 and Case 3, respectively. From these results, we concluded that the left and right eyes shared a degree of symmetry but were not identical.

Table 9. Division of the UBIPr dataset into training, validation, and testing sets for the different cases used in the cross-eye experiment.

		Training	Validation	Testing
Case 1	Side	Right	Right	Right
	# Images	3280	820	1026
Case 2	Side	Left	Left	Left
	# Images	3280	820	1026
Case 3	Side	Left	Left	Right
	# Images	4096	1030	5126
Case 4	Side	Left	Left	Right mirror
	# Images	4096	1030	5126
Case 5	Side	Right	Right	Left
	# Images	4096	1030	5126
Case 6	Side	Right	Right	Left mirror
	# Images	4096	1030	5126

Table 10. Comparison of the results of the LDA-CNN model with other existing methods on cross-eye experiments on UBIPr using the protocol in Table 9.

Method	Case 1		Case 2		Identification Accuracy (%)				Case 5		Case 6	
	Validation	Testing	Validation	Testing	Case 3	Case 4	Case 4	Case 4	Validation	Testing	Validation	Testing
Punam an Seeja [41]	93.67	89	92	88	-	40.2	-	61.6	-	-	-	-
Smereka et al. [57]	-	78.59	-	84.14	-	-	-	-	-	-	-	-
Dozier et al. [58]	-	88.3	-	87.1	-	-	-	-	-	-	-	-
Woodard et al. [59]	-	88.60	-	88.40	-	-	-	-	-	-	-	-
LDA-CNN	100	100	100	99.90	99.80	65.02	99.80	74.76	99.90	66.76	99.90	71.79

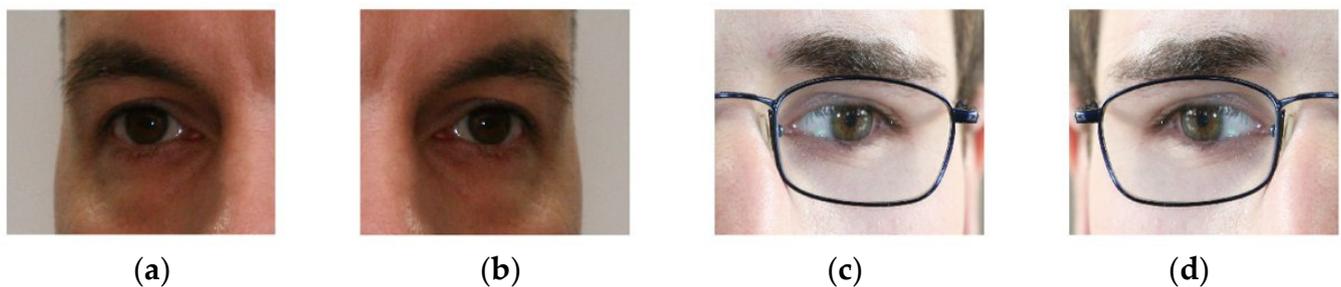


Figure 10. An example of left and right periocular images along with their mirror images from UBIPr dataset. (a) Left image; (b) Left mirror image; (c) Right image; (d) Right mirror image.

5.2.2. Cross-Pose

Pose variation is one of the most prevalent degradation factors in biometrics recognition that occurs frequently in the wild. This section examines the ability of the LDA-CNN model trained on frontal periocular images (0 degree pose) to recognize images with two different pose angles, 30 and -30 degrees. Figure 11 depicts an example of a periocular image at three different pose degrees (0, 30 and -30) from the UBIPr dataset. Experiments were conducted on the UBIPr dataset using the same protocol as [41] in which the model was trained and validated on 0 degree pose and tested with 30 degree (Testing: Case1) and -30 degree (Testing: Case2) poses. According to [41], Table 11 depicts the division of the training, validation, and testing sets being used. The results listed in Table 12 demonstrated the robustness of our method under cross pose conditions, with an improvement in accuracy over state-of-the-art methods ranging from 2.89% to 5.09% for -30- and 30-degree poses, respectively.

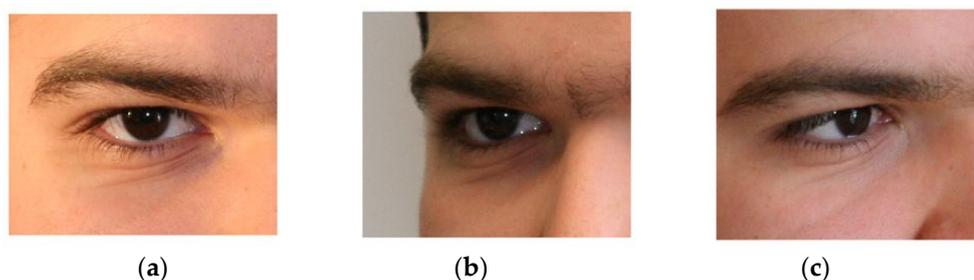


Figure 11. Example of a periocular image from UBIPr dataset with different pose variations: (a) 0-degree pose; (b) 30-degree pose (c) -30 -degree pose.

Table 11. Division of the UBIPr dataset into training, validation, and testing sets for the different cases used in the cross-pose experiment.

		Training	Validation	Testing
Case 1	Pose	0	0	30
	# Images	2735	683	3418
Case 2	Pose	0	0	-30
	# Images	2735	683	3416

Table 12. Comparison of the results of the LDA-CNN model with other existing methods on cross-pose experiments on UBIPr using the protocol in Table 11.

Method	Validation	Identification Accuracy (%)	
		Case 1: Testing	Case 2: Testing
Punam and Seeja [41]	100	94	96
Park et al. [6]	—	50	—
Sharma et al. [60]	—	71	71
Cao et al. [61]	—	94	—
LDA-CNN	99.56	99.09	98.89

5.2.3. Cross-Sensor

To investigate the consistency and generalizability of the LDA-CNN model in greater depth, we conducted cross-sensor experiments in which the gallery and probe sets were captured using different devices. These experiments were performed on the VISOB dataset, which contains images captured by iPhone, Samsung, and Oppo smartphones. As the number of classes in each device varies, we were unable to construct an end-to-end model. Instead, we trained and validated the LDA-CNN model on a single device, and then used that model as a feature extractor for the other devices. Particularly, the LDA layer's features were extracted and then passed to the SA-CRC for recognition. It is evident from Figure 12 that there was no significant difference between the results obtained using the same and different devices for training and testing, which demonstrated the robustness of LDA-CNN.

5.2.4. Cross-Light

In this section, we investigate the behavior of the proposed model in recognizing periocular images when trained and tested under identical and varying light conditions. We employed the VISOB dataset for these experiments, which contained three light conditions: day, dim, and office. As in the preceding section, the LDA-CNN model trained on a single condition was used as a feature extractor (i.e., LDA Layer) and SA-CRC was employed for recognition. Figure 13 depicts the cross-light experiment results. Overall, same light and cross light conditions yielded excellent results. However, we observed that the best results were obtained when the model was trained on the office light set, with 97.84%, 98.32%, and 96.98% for day, dim, and office light sets, respectively. This might be due to the fact that office lighting was considered to be the most challenging of all lighting conditions.

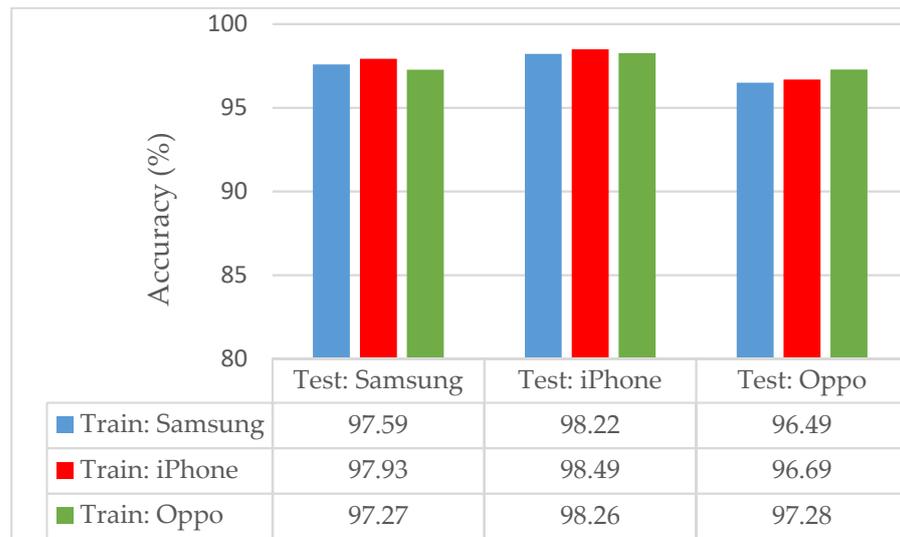


Figure 12. The performance of LDA-CNN model on cross-sensor experiments using VISOB dataset.

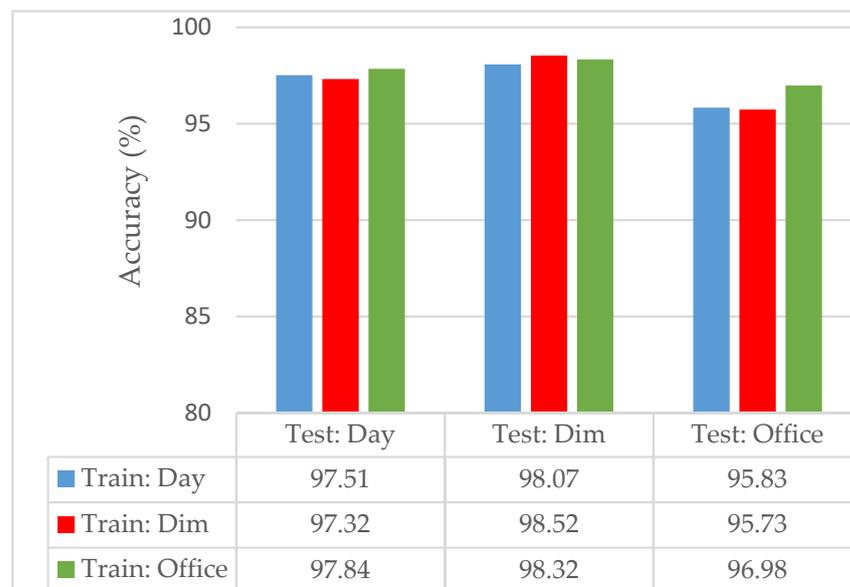


Figure 13. The performance of LDA-CNN model on cross-light experiments using VISOB dataset.

5.2.5. Cross-Database

In the previous sections we examined the robustness of the LDA-CNN model under various non-ideal conditions using a single dataset. In this section, we increase the challenging level by evaluating the model performance using different periocular datasets for training and testing. The experiments were conducted using VISOB, UBIPr and UFPR datasets, with features extracted from the LDA layer and recognition accuracy calculated with SA-CRC. The results of the same-dataset and cross-dataset experiments were compared in Table 13. Clearly, the proposed method exhibited a high degree of robustness in all datasets. In particular, the LDA-CNN model trained on the UFPR dataset showed the highest robustness level compared to the others. This might be due to the fact that UFPR contains a greater variety of unrestricted conditions than other datasets.

Table 13. The performance of LDA-CNN model on cross-database experiments using VISOB, UFPR and UBIPr datasets.

Training Dataset	Testing Dataset	Accuracy (%)
VISOB	VISOB	98.53
	UFPR	86.41
	UBIPr	98.73
UFPR	VISOB	94.9
	UFPR	98.06
	UBIPr	99.27
UBIPr	VISOB	90.38
	UFPR	81.43
	UBIPr	99.85

5.3. Ablation Study: The Impact of PCA and LDA Layers

To understand the contribution of the LDA layer to the overall model, we performed an ablation study that investigated the performance of the proposed LDA-CNN model without the LDA and PCA layers. In this direction, we implemented and evaluated two models: (1) FT-CNN, in which both the LDA and PCA layers were removed, and (2) PCA-CNN, in which only the LDA layer was removed. In both models, a new FC layer was added and fine-tuned in an end-to-end manner. Figure 14 shows the performance comparison of FT-CNN, PCA-CNN and LDA-CNN models on the UFPR dataset. It can be noticed from Figure 14 that the addition of the LDA layer significantly improved the identification performance on all CNN backbone models. Moreover, in Table 14, we compared the average recognition time needed by these three models to predict an image and discovered that LDA-CNN and PCA-CNN required nearly the same amount of time, which was slightly better than that required by FT-CNN.

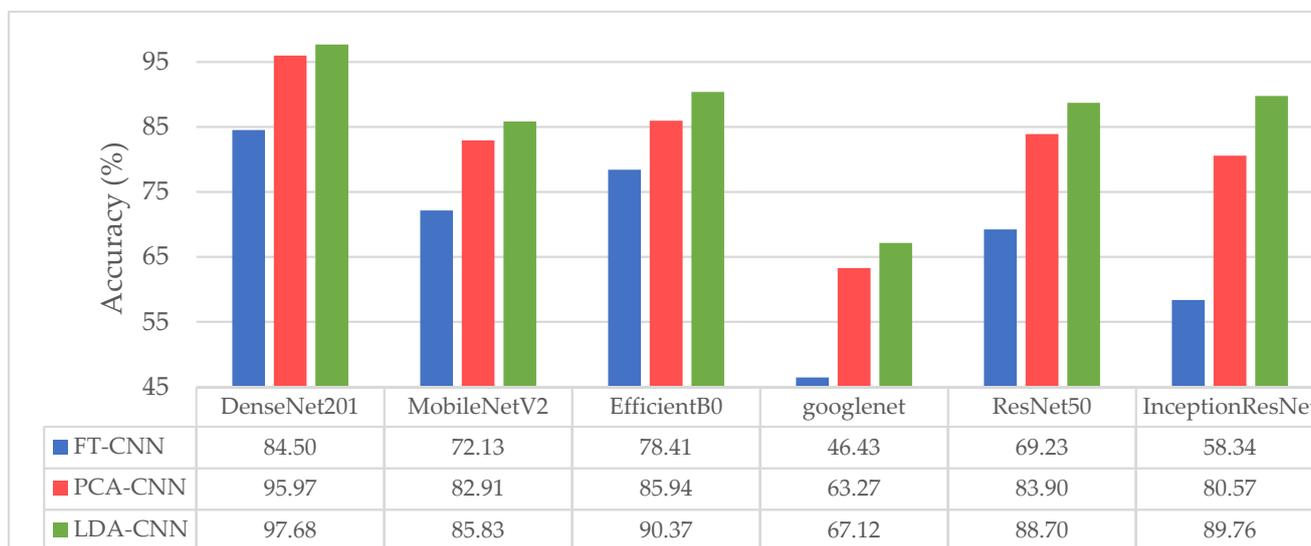


Figure 14. The performance comparison of FT-CNN, PCA-CNN and LDA-CNN models using different backbone models on the UFPR dataset.

Table 14. Comparison of the average recognition time needed by FT-CNN, PCA-CNN and LDA-CNN models using UFPR dataset.

Recognition Time (Seconds)		
FT-CNN	PCA-CNN	LDA-CNN
0.0104	0.0091	0.0093

For deep analysis, in Figure 15, we illustrate how the three models behave under within-class variability and between-class similarity challenges. Observing the variations in the images of the same class (Figure 15a) and their similarities with images of other classes (Figure 15e), we can see that the proposed method LDA-CNN successfully overcame these challenges and predicted the correct class, whereas FT-CNN and PCA-CNN failed to make the correct predictions.

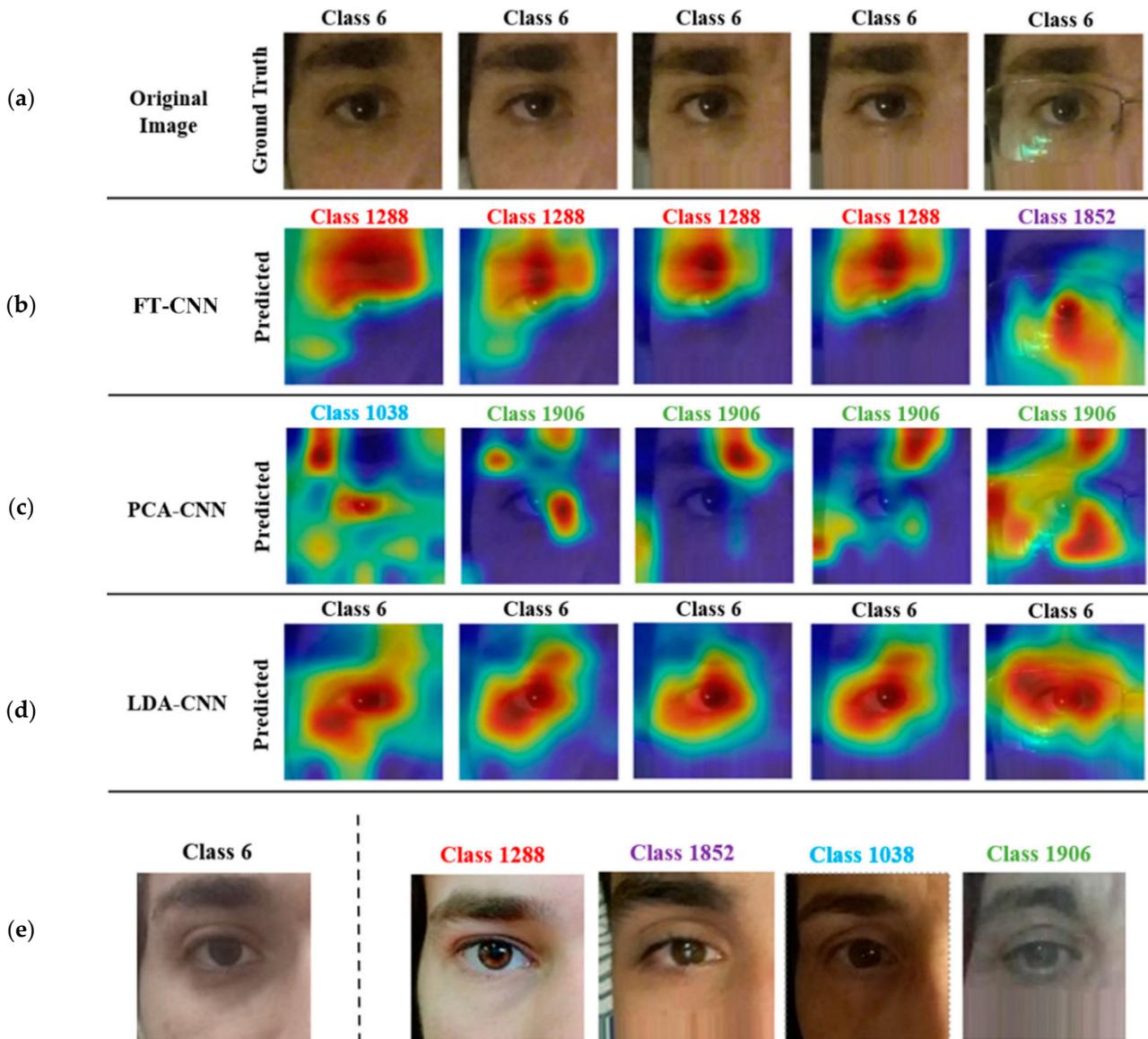


Figure 15. A comparison of the performance and the Grad-CAM explanations for the deep features of (b) FT-CNN, (c) PCA-CNN and (d) LDA-CNN (proposed) on predicting the correct class of (a) samples of Class 6 from UFPR dataset under within-class variability and (e) between-class similarity challenges.

Figure 15 also shows the Grad-CAM heatmaps that visualize the importance of pixels in the input image to the recognition; the closer a pixel is to the color red, the greater its significance. The heatmaps of LDA-CNN (Figure 15d) revealed that the information within the eye and surrounding skin was more significant than that of the eyebrow and other regions. This was supported by noticing the false predictions of FT-CNN (Figure 15b) and PCA-CNN (Figure 15c), where the heatmaps did not concentrate on the eye. The same conclusion could be drawn from the LDA-CNN heatmaps presented in Figure 16, where

the images were selected from the UFPR dataset under various variations that occur in wild environment for the periocular region: different poses, gazes, with/without makeup, and partial occlusions (hair and glasses).

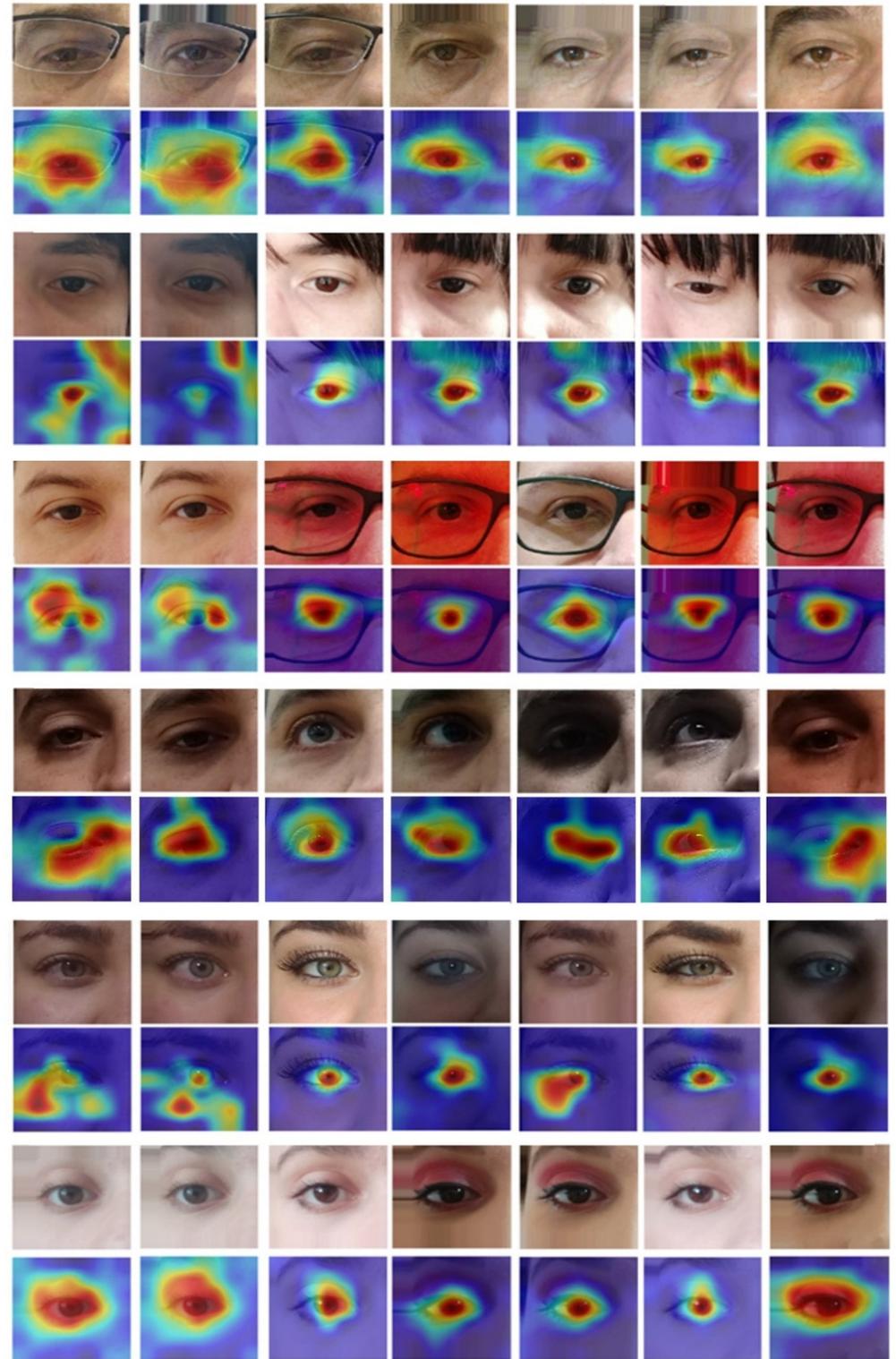


Figure 16. The Grad-CAM heatmaps of the LDA-CNN model for selected images were selected from the UFPR dataset under various variations that occur in the wild environment: different poses, gazes, with/without makeup, and partial occlusions (hair and glasses).

5.4. Visualization of LDA-CNN Using t-SNE

After demonstrating the superiority of the proposed model over FT-CNN and PCA-CNN models, it was evident that the LDA layer was responsible for this advancement. Our primary justification for these results was the role of the LDA layer in projecting the samples to a separable space with small within-class variation and large between-class separation. To find a meaningful explanation for this justification, we visualized different features extracted from different layers of the LDA-CNN model and compared them to those extracted from the LDA layer using the t-SNE algorithm [31], which maps the data of a high-dimensional space to a low-dimensional space, while maintaining the local characteristics of the dataset. The separability of data in a low-dimensional space can be used to determine whether the data can be separated in a high-dimensional space. We displayed t-SNE visualizations on five randomly selected classes of the VISOB dataset since it had a higher sample size per class than the other datasets, which made it simpler to qualitatively observe how the individuals clustered.

The mapping of Pooling 1, Denseblock (1), Denseblock (2), Denseblock (3), Denseblock (4), PCA, and LDA in 2D space by the t-SNE algorithm are depicted in Figure 17a–g, respectively. As can be seen in Figure 17, no clusters could be seen in the layers before Denseblock (3) because their features were too general to be used to distinguish classes. However, the PCA and Denseblock (4) features started to cluster, whereas the LDA features showed a highly distinct cluster with large between-class separation and tiny within-class distance.

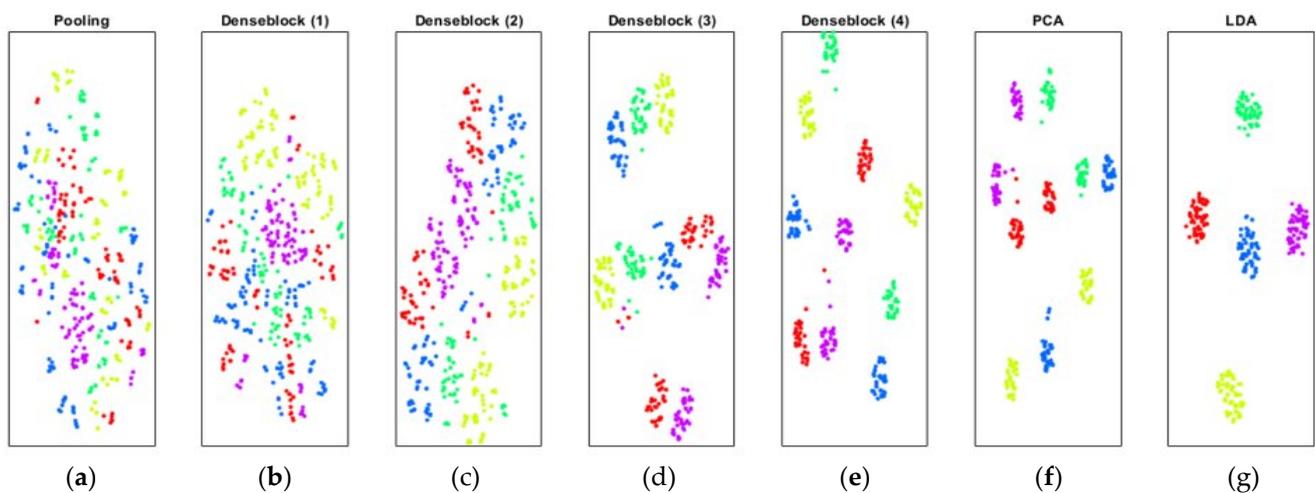


Figure 17. Visualization using t-SNE plots of the features extracted from various layers of the LDA-CNN model: (a) Pooling 1; (b) Denseblock 1; (c) Denseblock 2; (d) Denseblock 3; (e) Denseblock 4; (f) PCA and (g) LDA for five classes of the VISOB dataset.

6. Conclusions

This paper proposed an end-to-end deep CNN periocular recognition model, called linear discriminant analysis CNN (LDA-CNN), that is robust under the variations in the wild environment. The LDA-CNN model introduces an LDA layer on top of the existing CNN model to force the network to produce discriminative features with low variance within the same class and high variance between different classes. Extensive experiments on four unconstrained periocular datasets demonstrated that the proposed model achieved significantly better results than several state-of-the-art methods, even in the difficult cross-condition (i.e., cross-light, cross-sensor, cross-eye, cross-pose, and cross-database) experiments. The effectiveness of the LDA layer was separately validated by examining the performance of the model with and without the LDA layer and was also visually proved by visualizing the discriminative power of the LDA layer's features using the t-SNE algorithm. The LDA layer features showed a highly distinct cluster with a large between-class separation and a small within-class distance. From the Grad-CAM heatmap

analysis of the proposed model, we concluded that the information within the eyes and the surrounding skin are critical to periocular recognition, more so than that of the eyebrows. Therefore, binding the proposed model with a visual attention mechanism that focuses on the important periocular region is expected to further improve the performance and could be developed as a future extension of this work. Extending our work to incorporate more visualization approaches to improve our understanding of the decisions underlying the model's predictions is an additional interesting future research topic. Lastly, it would be interesting to investigate the impact of sparse PCA [62,63] and sparse LDA [64] in place of PCA and LDA for learning the PCA and LDA layers in the proposed model, as these have been shown to give better performance in some applications [62,64–66].

Author Contributions: Conceptualization, A.A., M.H. and H.A.; methodology, A.A. and M.H.; software, A.A.; validation, A.A. and M.H.; formal analysis, A.A. and M.H.; investigation, A.A. and M.H.; resources, A.A. and M.H.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, A.A. and M.H.; visualization, A.A. and M.H.; supervision, M.H. and H.A.; project administration, M.H. and H.A.; funding acquisition, A.A., M.H. and H.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported under Researchers Supporting Project number (RSP-2021/109) King Saud University, Riyadh, Saudi Arabia. The first author was also supported by IBM PhD Fellowship Awards Program.

Data Availability Statement: In this article, the publicly available UFPR (<https://web.inf.ufpr.br/vri/databases/ufpr-periocular/>, accessed on 26 October 2022), UBIPr (<http://iris.di.ubi.pt/ubipr.html>, accessed on 26 October 2022), UBIRIS.v2 (<http://iris.di.ubi.pt/ubiris2.html>, accessed on 26 October 2022), and VISOB [29] datasets were used.

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

References

1. Zanlorensi, L.A.; Laroca, R.; Luz, E.; Britto, A.S.; Oliveira, L.S.; Menotti, D. Ocular recognition databases and competitions: A survey. *Artif. Intell. Rev.* **2022**, *55*, 129–180. [\[CrossRef\]](#)
2. Kumari, P.; Seeja, K.R. Periocular biometrics: A survey. *J. King Saud Univ. Comput. Inf. Sci.* **2019**, *34*, 1086–1097. [\[CrossRef\]](#)
3. Badejo, J.; Akinrinmade, A.; Adetiba, E. Survey of Periocular Recognition Techniques. *J. Eng. Sci. Technol. Rev.* **2019**, *12*, 214–226. [\[CrossRef\]](#)
4. Rattani, A.; Derakhshani, R. Ocular biometrics in the visible spectrum: A survey. *Image Vis. Comput.* **2017**, *59*, 1–16. [\[CrossRef\]](#)
5. Masi, I.; Wu, Y.; Hassner, T.; Natarajan, P. Deep Face Recognition: A Survey. In Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); IEEE: Parana, Brazil, 2018; pp. 471–478.
6. 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. [\[CrossRef\]](#)
7. Klontz, J.C.; Jain, A.K. A Case Study of Automated Face Recognition: The Boston Marathon Bombings Suspects. *Computer* **2013**, *46*, 91–94. [\[CrossRef\]](#)
8. Smereka, J.M.; Boddeti, V.N.; Vijaya Kumar, B.V.K. Probabilistic Deformation Models for Challenging Periocular Image Verification. *IEEE Trans. Inform. Forensic Secur.* **2015**, *10*, 1875–1890. [\[CrossRef\]](#)
9. Nigam, I.; Vatsa, M.; Singh, R. Ocular biometrics: A survey of modalities and fusion approaches. *Inf. Fusion* **2015**, *26*, 1–35. [\[CrossRef\]](#)
10. Zhang, Q.; Li, H.; Sun, Z.; Tan, T. Deep Feature Fusion for Iris and Periocular Biometrics on Mobile Devices. *IEEE Trans. Inform. Forensic Secur.* **2018**, *13*, 2897–2912. [\[CrossRef\]](#)
11. Miller, P.E.; Rawls, A.W.; Pundlik, S.J.; Woodard, D.L. Personal Identification Using Periocular Skin Texture. In Proceedings of the Proceedings of the 2010 ACM Symposium on Applied Computing; ACM: New York, NY, USA, 2010; pp. 1496–1500.
12. Genetic-Based Type II Feature Extraction for Periocular Biometric Recognition: Less is More—IEEE Conference Publication. Available online: <https://ieeexplore.ieee.org/abstract/document/5597604> (accessed on 30 October 2019).
13. Santos, G.; Hoyle, E. A fusion approach to unconstrained iris recognition. *Pattern Recognit. Lett.* **2012**, *33*, 984–990. [\[CrossRef\]](#)
14. Ambika, D.R.; Radhika, K.R.; Seshachalam, D. Fusion of Shape and Texture for Unconstrained Periocular Authentication. *World Acad. Sci. Eng. Technol.* **2017**, *11*, 7.
15. Padole, C.N.; Proenca, H. Periocular recognition: Analysis of performance degradation factors. In Proceedings of the 2012 5th IAPR International Conference on Biometrics (ICB); IEEE: New Delhi, India, 2012; pp. 439–445.

16. Ross, A.; Jillela, R.; Smereka, J.M.; Boddeti, V.N.; Kumar, B.V.K.V.; Barnard, R.; Hu, X.; Pauca, P.; Plemmons, R. Matching highly non-ideal ocular images: An information fusion approach. In Proceedings of the 2012 5th IAPR International Conference on Biometrics (ICB); IEEE: New Delhi, India, 2012; pp. 446–453.
17. Karahan, Ş.; Karaöz, A.; Özdemir, Ö.F.; Gü, A.G.; Uludag, U. On identification from periocular region utilizing SIFT and SURF. In Proceedings of the 2014 22nd European Signal Processing Conference (EUSIPCO); IEEE: Piscataway, NJ, USA, 2014; pp. 1392–1396.
18. Bakshi, S.; Sa, P.K.; Majhi, B. A novel phase-intensive local pattern for periocular recognition under visible spectrum. *Biocybern. Biomed. Eng.* **2015**, *35*, 30–44. [[CrossRef](#)]
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); IEEE: Las Vegas, NV, USA, 2016; pp. 770–778.
20. Cao, X.; Wipf, D.; Wen, F.; Duan, G.; Sun, J. A Practical Transfer Learning Algorithm for Face Verification. In Proceedings of the 2013 IEEE International Conference on Computer Vision; IEEE: Sydney, Australia, 2013; pp. 3208–3215.
21. Tiong, L.C.O.; Lee, Y.; Teoh, A.B.J. Periocular Recognition in the Wild: Implementation of RGB-OCLBCP Dual-Stream CNN. *Appl. Sci.* **2019**, *9*, 2709. [[CrossRef](#)]
22. Jung, Y.G.; Low, C.Y.; Park, J.; Teoh, A.B.J. Periocular Recognition in the Wild With Generalized Label Smoothing Regularization. *IEEE Signal Process. Lett.* **2020**, *27*, 1455–1459. [[CrossRef](#)]
23. Bhattacharyya, S.; Rahul, K. Face recognition by linear discriminant analysis. *Int. J. Commun. Netw. Secur.* **2013**, *2*, 31–35. [[CrossRef](#)]
24. Mahmud, F.; Khatun, M.T.; Zuhori, S.T.; Afroge, S.; Aktar, M.; Pal, B. Face recognition using Principle Component Analysis and Linear Discriminant Analysis. In Proceedings of the 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, 21–23 May 2015; pp. 1–4.
25. Izenman, A.J. Linear Discriminant Analysis. In *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*; Izenman, A.J., Ed.; Springer Texts in Statistics; Springer: New York, NY, USA, 2008; pp. 237–280. ISBN 978-0-387-78189-1.
26. Xanthopoulos, P.; Pardalos, P.M.; Trafalis, T.B. Linear Discriminant Analysis. In *Robust Data Mining*; Xanthopoulos, P., Pardalos, P.M., Trafalis, T.B., Eds.; SpringerBriefs in Optimization; Springer: New York, NY, USA, 2013; pp. 27–33. ISBN 978-1-4419-9878-1.
27. Zanolensi, L.A.; Laroca, R.; Lucio, D.R.; Santos, L.R.; Britto, A.S., Jr.; Menotti, D. UFPR-Periocular: A Periocular Dataset Collected by Mobile Devices in Unconstrained Scenarios. *arXiv* **2020**, arXiv:2011.12427.
28. Proenca, H.; Filipe, S.; Santos, R.; Oliveira, J.; Alexandre, L.A. The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-the-Move and At-a-Distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1529–1535. [[CrossRef](#)]
29. Rattani, A.; Derakhshani, R.; Saripalle, S.K.; Gottemukkula, V. ICIIP 2016 competition on mobile ocular biometric recognition. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 320–324.
30. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); IEEE: Venice, Italy, 2017; pp. 618–626.
31. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
32. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
33. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
34. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–27 July 2017; pp. 2261–2269.
35. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
36. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
37. Chollet, F. Xception: Deep Learning With Depthwise Separable Convolutions. *arXiv* **2017**, arXiv:1610.02357.
38. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence; AAAI Press: San Francisco, CA, USA, 2017; pp. 4278–4284.
39. ImageNet Classification with Deep Convolutional Neural Networks | Communications of the ACM. Available online: <https://dl.acm.org/doi/abs/10.1145/3065386> (accessed on 23 October 2022).
40. Talreja, V.; Nasrabadi, N.M.; Valenti, M.C. Attribute-Based Deep Periocular Recognition: Leveraging Soft Biometrics to Improve Periocular Recognition. *arXiv* **2022**, arXiv:2111.01325.
41. Kumari, P.; Seeja, K.R. Periocular Biometrics for non-ideal images: With off-the-shelf Deep CNN & Transfer Learning approach. *Procedia Comput. Sci.* **2020**, *167*, 344–352. [[CrossRef](#)]

42. Periocular Recognition Using CNN Based Feature Extraction and Classification | IEEE Conference Publication | IEEE Xplore. Available online: <https://ieeexplore.ieee.org/abstract/document/9509734> (accessed on 23 October 2022).
43. Ipe, V.M.; Thomas, T. Periocular Recognition Under Unconstrained Image Capture Distances. In Proceedings of the Advances in Signal Processing and Intelligent Recognition Systems; Thampi, S.M., Krishnan, S., Hegde, R.M., Ciunzo, D., Hanne, T., Kannan, R.J., Eds.; Springer: Singapore, 2021; pp. 175–186.
44. Abdi, H.; Williams, L.J. Principal component analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459. [[CrossRef](#)]
45. Ipe, V.M.; Thomas, T. Periocular Recognition Under Unconstrained Conditions Using CNN-Based Super-Resolution. In *Applied Soft Computing and Communication Networks: Proceedings of ACN 2019*; Thampi, S.M., Sherly, E., Dasgupta, S., Lloret Mauri, J., H. Abawajy, J., Khorov, E., Mathew, J., Eds.; Lecture Notes in Networks and Systems; Springer: Singapore, 2020; pp. 235–246. ISBN 9789811538520.
46. Akhtar, N.; Shafait, F.; Mian, A. Efficient classification with sparsity augmented collaborative representation. *Pattern Recognit.* **2017**, *65*, 136–145. [[CrossRef](#)]
47. Zhao, Z.; Kumar, A. Accurate Periocular Recognition Under Less Constrained Environment Using Semantics-Assisted Convolutional Neural Network. *IEEE Trans. Inform. Forensic Secur.* **2017**, *12*, 1017–1030. [[CrossRef](#)]
48. 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](#)]
49. Proença, H.; Neves, J.C. A Reminiscence of “Mastermind”: Iris/Periocular Biometrics by “In-Set” CNN Iterative Analysis. *IEEE Trans. Inform. Forensic Secur.* **2019**, *14*, 1702–1712. [[CrossRef](#)]
50. Wazirali, R.; Ahmed, R. Hybrid feature extractions and CNN for enhanced periocular identification during Covid-19. *Comput. Syst. Sci. Eng.* **2022**, *41*, 305–306. [[CrossRef](#)]
51. Kumari, P.; Seeja, K.R. A novel periocular biometrics solution for authentication during Covid-19 pandemic situation. *J. Ambient. Intell. Hum. Comput.* **2021**, *12*, 10321–10337. [[CrossRef](#)]
52. Raffei, A.F.M.; Sutikno, T.; Asmuni, H.; Hassan, R.; Othman, R.M.; Kasim, S.; Riyadi, M.A. Fusion Iris and Periocular Recognitions in Non-Cooperative Environment. *Indones. J. Electr. Eng. Inform.* **2019**, *7*, 543–554. [[CrossRef](#)]
53. Ahuja, K.; Bose, A.; Nagar, S.; Dey, K.; Barbhuiya, F. ISURE: User authentication in mobile devices using ocular biometrics in visible spectrum. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 335–339.
54. Ahuja, K.; Islam, R.; Barbhuiya, F.A.; Dey, K. Convolutional neural networks for ocular smartphone-based biometrics. *Pattern Recognit. Lett.* **2017**, *91*, 17–26. [[CrossRef](#)]
55. Alahmadi, A.; Hussain, M.; Aboalsamh, H.; Azmi, A. ConvSRC: SmartPhone-based periocular recognition using deep convolutional neural network and sparsity augmented collaborative representation. *IFS* **2020**, *38*, 3041–3057. [[CrossRef](#)]
56. Li, Y.; Bao, F.J. Interocular symmetry analysis of bilateral eyes. *J. Med. Eng. Technol.* **2014**, *38*, 179–187. [[CrossRef](#)] [[PubMed](#)]
57. Smereka, J.M.; Kumar, B.V.K.V.; Rodriguez, A. Selecting discriminative regions for periocular verification. In Proceedings of the 2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), Sendai, Japan, 29 February–2 March 2016; pp. 1–8.
58. Dozier, G.; Purrington, K.; Popplewell, K.; Shelton, J.; Abegaz, T.; Bryant, K.; Adams, J.; Woodard, D.L.; Miller, P. GEFES: Genetic & evolutionary feature selection for periocular biometric recognition. In Proceedings of the 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM); IEEE: Paris, France, 2011; pp. 152–156.
59. Woodard, D.L.; Pundlik, S.J.; Lyle, J.R.; Miller, P.E. Periocular region appearance cues for biometric identification. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops; IEEE: San Francisco, CA, USA, 2010; pp. 162–169.
60. Sharma, A.; Verma, S.; Vatsa, M.; Singh, R. On cross spectral periocular recognition. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP); IEEE: Paris, France, 2014; pp. 5007–5011.
61. Cao, Z.; Schmid, N.A. Fusion of operators for heterogeneous periocular recognition at varying ranges. *Pattern Recognit. Lett.* **2016**, *82*, 170–180. [[CrossRef](#)]
62. Deshpande, Y.; Montanari, A. Sparse PCA via covariance thresholding. *J. Mach. Learn. Res.* **2016**, *17*, 4913–4953.
63. Holtzman, G.; Soffer, A.; Vilenchik, D. A Greedy Anytime Algorithm for Sparse PCA. In Proceedings of the Thirty Third Conference on Learning Theory, Graz, Austria, 9–12 July 2020; pp. 1939–1956.
64. Shao, J.; Wang, Y.; Deng, X.; Wang, S. Sparse linear discriminant analysis by thresholding for high dimensional data. *Ann. Statist.* **2011**, *39*, 1241–1265. [[CrossRef](#)]
65. Zou, H.; Xue, L. A Selective Overview of Sparse Principal Component Analysis. *Proc. IEEE* **2018**, *106*, 1311–1320. [[CrossRef](#)]
66. Zou, H.; Hastie, T.; Tibshirani, R. Sparse Principal Component Analysis. *J. Comput. Graph. Stat.* **2006**, *15*, 265–286. [[CrossRef](#)]