



Cancelable multi-biometric recognition system based on deep learning

Essam Abdellatef¹ · Nabil A. Ismail² · Salah Eldin S. E. Abd Elrahman² · Khalid N. Ismail^{3,4} · Mohamed Rihan⁵ · Fathi E. Abd El-Samie⁵

Published online: 29 June 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

In this paper, we propose a cancelable multi-biometric face recognition method that uses multiple convolutional neural networks (CNNs) to extract deep features from different facial regions. We also propose a new CNN architecture that exploits batch normalization, depth concatenation and a residual learning framework. The proposed method adopts a region-based technique in which face, eyes, nose and mouth regions are detected from the original face images. Multiple CNNs are used to extract deep features from each region, and then, a fusion network combines these features. Moreover, to provide user's privacy and increase the system resistance against spoof attacks, a cancelable biometric technique using bio-convolving encryption is performed on the final facial descriptor. Our experiments on the FERET, LFW and PaSC datasets show excellent and competitive results compared to state-of-the-art methods in terms of recognition accuracy, specificity, precision, recall and f_{score} .

Keywords Face recognition · Deep learning · Fusion of features · Cancelable biometrics

1 Introduction

Biometrics such as face, fingerprint and iris are defined as the physiological characteristics of an individual. It has become a need now to get more efficient biometric systems that can be easily embedded in smart devices such as tablets and mobile phones. A human face is considered as one of the most effective biometric traits compared to other biometrics. Face recognition (FR) is characterized by its cheap cost, contactless nature and high acceptability during acquisition [1–3]. However, the performance of

FR systems is affected by changes in resolution, pose and illumination [4, 5]. To overcome this problem, most of the FR techniques extract features from face images using well-known algorithms such as scale-invariant feature transform (SIFT) [6], speeded up robust feature (SURF) extraction [7] or histogram of oriented gradients (HOG) [8]. In the last few years, researchers interested in the field of FR proved that the utilization of CNNs gives more robust, representative and detailed features that improve the overall system performance. Furthermore, to reach optimal results, a CNN has to be trained by a large number of face images without

✉ Essam Abdellatef
essam_abdellatef@yahoo.com

Nabil A. Ismail
Nabil.Ismail@el-eng.menofia.edu.eg

Salah Eldin S. E. Abd Elrahman
salaheldeem@el-eng.menofia.edu.eg

Khalid N. Ismail
Khalid.n.ismail@gmail.com

Mohamed Rihan
mohamed.elmelegy@el-eng.menofia.edu.eg

Fathi E. Abd El-Samie
fathi_sayed@yahoo.com

¹ Electronics and Communication Department, Delta Academy for Engineering, Mansoura, Egypt

² Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

³ Department of Computer Science, Durham University, Durham DH1 3LE, UK

⁴ Information Technology Department, Faculty of Computers and Information, Menoufia University, Menoufia, Egypt

⁵ Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

implementation of external pre-processing methods as the system accuracy could be affected by the easy failures of these methods [9–18].

Paying attention to enhancing the recognition accuracy should not make us overlook protection of biometric data from hackers. For this purpose, encryption techniques could be applied [19]. In practice, illumination variations have a great effect on biometric templates. So, different digests are produced when there is a minor change in the input. Another drawback of using encryption with biometrics is the need to apply data decryption that represents an attack point. Recently, cancelable biometrics has attracted a great attention. In cancelable biometrics, a one-way function transforms the biometric template instead of storing the original one. This way can provide revocability since another transformation is used to re-enroll a compromised biometric, and privacy since the original biometric cannot be recovered easily from a transformed one. Additionally, the recognition accuracy is not degraded as the statistical characteristics of features after transformation are approximately maintained [20–22].

Motivated by the previous observations, we propose a cancelable multi-biometric FR method in which a fusion network combines deep features which are extracted using multiple CNNs. Security and privacy of biometric data are provided by performing a bio-convolving encryption step on the final facial descriptor.

Our main contributions are as follows:

1. We propose a new, simple and efficient CNN architecture.
2. We present a FR method that achieves remarkable results compared to the state-of-the-art techniques in terms of recognition accuracy, specificity, precision, recall and f_{score} .
3. We provide protection of biometric data without degradation of the recognition accuracy.
4. We perform extensive experiments on different datasets.

This paper is organized as follows. Section 2 is about the related work. Section 3 describes the proposed method. Section 4 presents the experimental results and Sect. 5 gives the concluding remarks of the paper.

2 Related work

Recently, CNN-based FR methods have achieved remarkable results on different datasets [23, 24]. To study these methods, the following aspects should be taken into consideration: (a) number and architecture of CNNs; (b) dataset used for training; (c) type of loss function; and (d) the incorporated learning strategies. To enhance the system performance, some operations are applied on face images

to generate different types of inputs to train multiple CNNs. These operations could be: (a) cropping of certain face regions [25–29]; (b) utilization of synthesized or frontalized faces [30–32]; or (c) utilization of different numbers of training images from different datasets [26, 33]. Kurban et al. [34] proposed an approach in which score level fusion is applied between two different datasets to create a new one. In addition, energy imaging method is used for gesture feature extraction and principal component analysis (PCA) is used for dimensionality reduction. They achieved encouraging performance with high genuine match rate (GMR) and low false acceptance rate (FAR). Li et al. [35] designed a method that improves recognition accuracy through updating the classification model in real time by using self-detection, decision and learning (S-DDL).

Several researches are concerned with improving FR performance, but few are concerned with protecting biometric patterns and providing user's privacy. Generation of cancelable biometric templates [36–38] is an approach that is used to protect users' data. Traditional techniques reconstruct biometric templates by performing reverse engineering [39]. We can classify cancelable FR systems based on the number of biometrics into: (a) unimodal systems, in which cancelability is achieved using a single biometric trait, and (b) multi-biometric systems, in which multiple traits are used [22]. We can follow one of the two approaches to produce cancelable biometric templates: (a) cryptosystem approach [40–42], in which a key is obtained from the original data, or (b) transformation-based approach, in which discrimination between templates is enhanced using original templates with a specific key [43]. In practice, several techniques are used to provide cancelability for both unimodal and multi-biometric systems such as bio-convolving with random kernels, non-invertible geometric transforms, random projections or cancelable biometric filters [44]. In this work, bio-convolving with random kernels is adopted due to its simplicity and great success with CNN-based FR methods.

3 The proposed method

Figure 1 presents the proposed approach pipeline that consists of four main stages: (a) detections of different facial regions; (b) extraction of deep features using multiple CNNs; (c) utilization of a fusion network to form a discriminative facial descriptor; and (d) application of bio-convolving with random kernels to protect biometric data from different attacks.

3.1 Detection of facial regions

Face, nose, eyes and mouth regions are detected from the original face images [45]. Figure 2 shows examples of the detection and separation operations. Eyes, nose and mouth are very

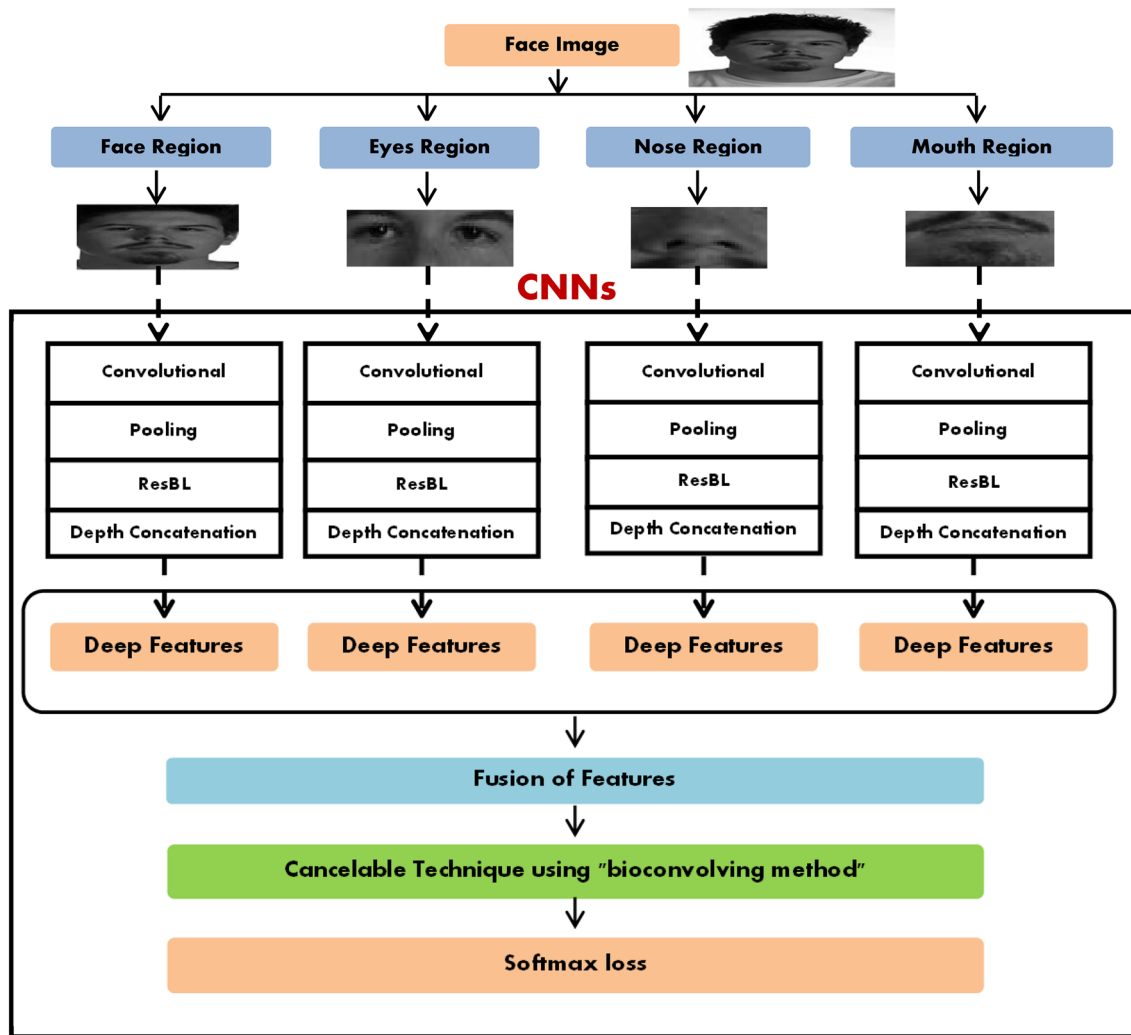


Fig. 1 The proposed FR method

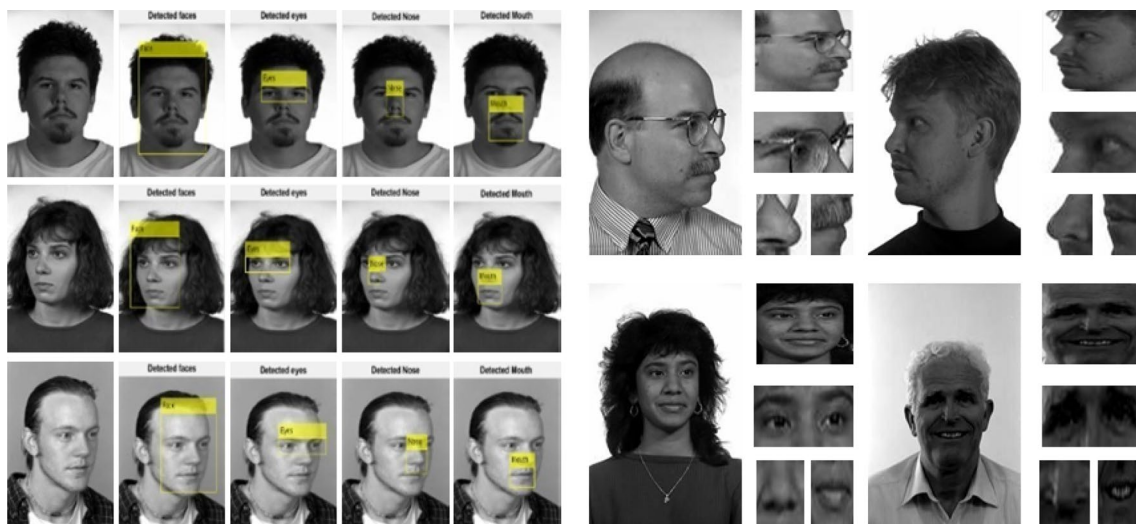


Fig. 2 Detection and separation of different facial regions

effective regions on which several changes are clearly noticed. These changes include laughing, closing eyes, opening mouth or wearing glasses. Figure 3 depicts the visual activations of the first five convolutional layers of the proposed CNN model for each facial region.

3.2 Deep feature extraction

Multiple CNNs of the same architecture are used to extract deep features from the detected facial regions. We propose a CNN model that includes “22” convolutional [46], “8” max pooling [46], “1” batch normalization [47], “2” residual learning framework “ResBL” [48], “3” depth concatenation [49], “1” fully connected [50], “1” feature normalization [51] and “1” softmax [52] layers as illustrated in Table 1. A multi-GPU platform could be implemented to limit the computation time consumption.

3.3 Fusion of deep features

We adopt a fusion network to combine the extracted features into a more representative, reliable, useful and detailed facial descriptor. This network consists of two layers: local and fusion layers. The local layer is composed of four parallel CNNs. If we consider that $\mathbf{F}^{(i)}(\cdot)$ represents the deep feature vector which is extracted from a CNN i , then the output of the fusion layer could be computed as illustrated in Eq. (1):

$$\text{Facial descriptor} = \sum_{i=1}^N \mathbf{W}^{(i)} \mathbf{F}^{(i)}(\cdot) + \mathbf{b}^{(i)} \quad (1)$$

where the corresponding weights and bias in the fusion layer are denoted by $\mathbf{W}^{(i)}$ and $\mathbf{b}^{(i)}$, respectively. The number of CNNs is represented by N ; here, $N = 4$.

3.4 Bio-convolving encryption

This method [44] adopts a convolution approach that leads to generating cancelable biometric templates. A transformed sequence $f[i]$, $i = 1, \dots, F$ is obtained using an original sequence $r[i]$, $i = 1, \dots, F$ through a convolution with a random kernel $h[i]$.

$$f[i] = r[i] * h[i] \quad (2)$$

4 Experimental results

To verify the performance of the proposed method, we present experiments on different datasets: FERET [53], LFW [54] and PaSC [55]. In addition, recognition accuracy, specificity, precision, recall and f_{score} are used for evaluation; see

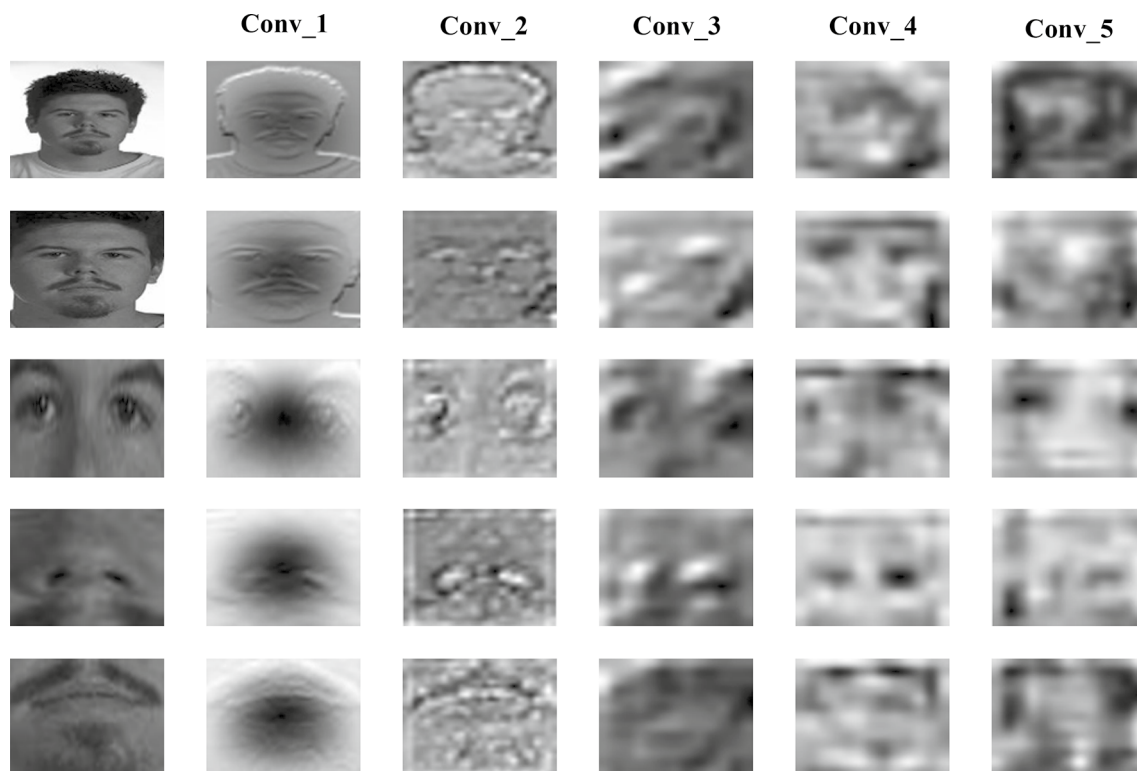


Fig. 3 Visual activations of the first five convolutional layers of the proposed CNN model

Table 1 Proposed CNN architecture

Layer name	No. of filters	Filter size	Stride size	Padding size
Conv1	64	7×7×3	2×2	3×3
ReLU	n/a	n/a	n/a	n/a
Max pooling	1	3×3	2×2	1×1
Batch normalization	Batch normalization			
Conv2	64	1×1×64	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv3	128	3×3×64	1×1	1×1
ReLU	n/a	n/a	n/a	n/a
Max pooling	1	3×3	2×2	1×1
ResBL	1	3×3×64	1×1	1×1
Conv4	192	3×3×64	1×1	1×1
ReLU	n/a	n/a	n/a	n/a
Max pooling	1	3×3	1×1	1×1
ResBL	1	3×3×64	1×1	1×1
Conv5	64	1×1×192	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv6	96	1×1×192	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv7	128	3×3×96	1×1	1×1
ReLU	n/a	n/a	n/a	n/a
Conv8	16	1×1×192	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv9	32	5×5×16	1×1	2×2
ReLU	n/a	n/a	n/a	n/a
Max pooling	1	3×3	1×1	1×1
Conv10	32	1×1×192	1×1	0
ReLU	n/a	n/a	n/a	n/a
Depth concatenation	Depth concatenation of 4 inputs			
Conv11	128	1×1×256	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv12	128	1×1×256	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv13	192	3×3×128	1×1	1×1
ReLU	n/a	n/a	n/a	n/a
Conv14	32	1×1×256	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv15	96	5×5×32	1×1	2×2
ReLU	n/a	n/a	n/a	n/a
Max pooling	1	3×3	1×1	1×1
Conv16	64	1×1×256	1×1	0
ReLU	n/a	n/a	n/a	n/a
Depth concatenation	Depth concatenation of 4 inputs			
Conv17	192	1×1×480	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv18	96	1×1×480	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv19	208	3×3×96	1×1	1×1
ReLU	n/a	n/a	n/a	n/a
Conv20	16	1×1×480	1×1	0
ReLU	n/a	n/a	n/a	n/a
Conv21	48	5×5×16	1×1	2×2

Table 1 (continued)

Layer name	No. of filters	Filter size	Stride size	Padding size
ReLU	n/a	n/a	n/a	n/a
Max pooling	1	3 × 3	1 × 1	1 × 1
Conv22	64	1 × 1 × 480	1 × 1	0
ReLU	n/a	n/a	n/a	n/a
Depth concatenation	Depth concatenation of 4 inputs			
Max pooling	1	3 × 3	2 × 2	0
Dropout	40% Dropout			
Fully connected layer	1000 fully connected layer			
Feature normalization	Feature normalization			
Softmax	n/a	n/a	n/a	n/a

Table 2 Platform specifications

System	Specifications
Type	64-bit Win 10
Processor	Intel Xeon 5670, 12 cores
Graphics card	NVIDIA GeForce GTX 1070
Installed memory (RAM)	48G memory

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F_{Score} = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{7}$$

where TP = true positive, FN = false negative, FP = false positive and TN = true negative.

Eqs. 3, 4, 5, 6 and 7. All experiments have been performed using a platform with the specifications shown in Table 2. We used the stochastic gradient descent method to train the CNN. Momentum is set to 0.9. We applied L_2 regularization with a weight decay of 5×10^4 . We began the CNN training with a learning rate equal to 0.1 and stopped the training after 5 epochs.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$Specificity = \frac{TN}{FP + TN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

4.1 Evaluation of unimodal and multi-biometric FR methods

As mentioned before, the proposed method adopts an FR technique in which four regions are detected from the original face images. Firstly, we studied the performance of the unimodal FR method to know the most effective part of a face image in the recognition process. The unimodal FR method uses a specific region: face, nose, eyes or mouth to train a single CNN. Table 3 illustrates the results of the unimodal FR method using different facial regions.

We observe from Table 3 that unimodal FR system based on face region achieves better results than those based on other regions as the face region contains plenty of features. The utilization of nose region comes next as the changes in nose are less than those of eyes and mouth

Table 3 Experimental results of unimodal FR using different face regions for different datasets

Dataset	Region	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F_{score} (%)
FERET	Face	97.14	97.23	94.93	96.23	95.57
	Eyes	94.87	94.93	92.75	93.67	93.2
	Nose	95.24	95.36	93.21	94.34	93.77
	Mouth	95.12	95.25	93.18	94.47	93.82
LFW	Face	97.94	98.09	96.15	97.33	96.73
	Eyes	95.19	95.26	93.28	94.47	93.87
	Nose	96.51	96.58	94.52	95.54	95.02
	Mouth	95.89	95.96	93.87	95.17	94.51
PaSC	Face	95.42	95.51	93.73	94.95	94.33
	Eyes	93.77	93.86	92.16	93.28	92.71
	Nose	94.11	94.18	92.41	93.54	92.97
	Mouth	93.93	94.07	92.79	93.89	93.33

Table 4 Comparison results of the unimodal FR system using state-of-the-art CNN models

Dataset	CNN	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F_{score} (%)
FERET	Arcface [16]	96.61	96.69	94.82	95.94	95.37
	CoCo loss [17]	97.02	97.13	95.34	96.45	95.89
	DeepVisage [52]	96.27	96.38	94.57	95.76	95.16
	29-layer [56]	92.62	92.76	90.97	92.14	91.55
	64-layer [57]	93.88	93.97	92.17	93.36	92.76
	Baidu [58]	95.73	95.82	93.94	95.05	94.49
	DeepID3 [59]	96.15	96.27	94.49	95.62	95.05
	TBE-CNN [60]	96.85	96.94	95.16	96.33	95.74
	FRP-VGG16 [61]	91.82	91.96	89.22	90.71	89.95
	BDL-PAP [62]	95.77	95.86	93.13	94.58	93.84
	KinGAP [63]	96.69	96.78	93.98	95.08	94.52
	Proposed CNN	97.14	97.23	94.93	96.23	95.57
LFW	Arcface [16]	97.48	97.55	95.77	96.96	96.36
	CoCo loss [17]	97.81	97.94	96.17	97.38	96.77
	DeepVisage [52]	97.33	97.45	95.62	96.84	96.22
	29-layer [56]	95.45	95.52	93.72	94.95	94.33
	64-layer [57]	96.12	96.23	94.48	95.62	95.04
	Baidu [58]	97.07	97.15	95.34	96.56	95.94
	DeepID3 [59]	97.26	97.38	95.63	96.83	96.22
	TBE-CNN [60]	97.64	97.74	95.93	97.14	96.53
	FRP-VGG16 [61]	94.34	94.47	91.87	93.04	92.45
	BDL-PAP [62]	97.23	97.32	95.61	96.61	96.1
	KinGAP [63]	97.39	97.49	95.85	96.88	96.36
	Proposed CNN	97.94	98.09	96.15	97.33	96.73
PaSC	Arcface [16]	94.85	94.93	93.07	94.18	93.62
	CoCo loss [17]	95.27	95.37	93.48	94.59	94.03
	DeepVisage [52]	94.47	94.56	92.63	93.75	93.18
	29-layer [56]	93.61	93.68	91.77	92.88	92.32
	64-layer [57]	93.78	93.85	91.96	93.02	92.48
	Baidu [58]	94.11	94.22	92.3	93.34	92.81
	DeepID3 [59]	94.35	94.46	92.53	93.65	93.08
	TBE-CNN [60]	95.04	95.17	93.28	94.36	93.81
	FRP-VGG16 [61]	93.79	93.91	91.86	93.06	92.45
	BDL-PAP [62]	94.23	94.36	92.62	93.56	93.08
	KinGAP [63]	94.83	94.94	93.12	94.25	93.68
	Proposed CNN	95.42	95.51	93.73	94.95	94.33

under different positions and expressions of persons in images. Table 4 gives the comparison results of the unimodal FR system based on face region using state-of-the-art CNNs. To guarantee a fair comparison, hyper-parameters are tuned for all methods.

Table 4 shows that the proposed CNN model is superior to other state-of-the-art CNN models. With the proposed CNN, the recognition accuracy reaches 97.14% on FERET dataset compared to 97.02% for CoCo loss, 97.94% on LFW

dataset compared to 97.81 for CoCo loss, and 95.42% on PaSC dataset compared to 95.27% for CoCo loss.

The performance of the proposed multi-biometric FR system has been studied. Table 5 provides the comparison results of the proposed multi-biometric FR system using the state-of-the-art CNN models.

From Table 5, compared to the state-of-the-art CNN models, the proposed CNN achieves remarkable results. With the proposed model, the recognition accuracy reaches 98.89% on FERET dataset compared to 98.73% for CoCo

Table 5 Comparison results of the proposed multi-biometric FR system using state-of-the-art CNN models

Dataset	CNN	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F_{score} ()
FERET	Arcface [16]	98.52	98.63	96.82	98.14	97.47
	CoCo loss [17]	98.73	98.88	97.16	98.38	97.76
	DeepVisage [52]	97.97	98.09	96.39	97.53	96.95
	29-layer [56]	94.81	94.88	93.18	94.38	93.77
	64-layer [57]	95.78	95.84	94.15	95.37	94.75
	Baidu [58]	97.61	97.69	95.91	97.12	96.51
	DeepID3 [59]	97.81	97.92	96.16	97.38	96.76
	TBE-CNN [60]	98.68	98.75	97.03	98.26	97.64
	FRP-VGG16 [61]	94.67	94.75	93.14	94.31	93.72
	BDL-PAP [62]	97.74	97.84	96.21	97.26	96.73
	KinGAP [63]	98.55	98.68	96.95	98.17	97.55
Proposed CNN	98.89	99.02	97.21	98.48	97.84	
LFW	Arcface [16]	98.66	98.74	97.14	98.25	97.69
	CoCo loss [17]	98.82	98.95	97.37	98.47	97.91
	DeepVisage [52]	98.26	98.33	96.74	97.84	97.28
	29-layer [56]	96.82	96.89	95.21	96.46	95.83
	64-layer [57]	97.24	97.35	95.79	96.83	96.3
	Baidu [58]	97.89	97.98	96.38	97.56	96.96
	DeepID3 [59]	98.19	98.28	96.63	97.87	97.24
	TBE-CNN [60]	98.71	98.8	97.22	98.38	97.79
	FRP-VGG16 [61]	96.76	96.91	95.16	96.38	95.76
	BDL-PAP [62]	97.92	98.05	97.43	97.67	97.54
	KinGAP [63]	98.63	98.78	97.12	98.24	97.67
Proposed CNN	98.93	99.08	97.51	98.77	98.13	
PaSC	Arcface [16]	97.17	97.27	95.66	96.88	96.26
	CoCo loss [17]	97.27	97.36	95.75	96.99	96.36
	DeepVisage [52]	96.68	96.75	95.04	96.25	95.64
	29-layer [56]	95.11	95.18	93.55	94.78	94.16
	64-layer [57]	95.67	95.74	94.04	95.21	94.62
	Baidu [58]	96.28	96.34	94.66	95.87	95.26
	DeepID3 [59]	96.51	96.57	94.95	96.18	95.56
	TBE-CNN [60]	97.24	97.33	95.64	96.91	96.27
	FRP-VGG16 [61]	95.29	95.36	93.71	94.83	94.26
	BDL-PAP [62]	96.43	96.58	94.75	95.93	95.33
	KinGAP [63]	97.18	97.29	95.41	96.72	96.06
Proposed CNN	97.38	97.48	95.89	97.12	96.5	

loss, 98.93% on LFW dataset compared to 98.82 for CoCo loss, and 97.38% on PaSC compared to 97.27% for CoCo loss. Figure 4 depicts the receiver operating characteristic (ROC) curves of the proposed CNN model and CoCo loss model for different datasets.

As shown in Fig. 4, the proposed model gives better results than those of CoCo loss CNN on FERET and LFW datasets, while on PaSC dataset the performance of both CNN models is almost the same. Furthermore, Fig. 5 illustrates the ROC plots of both multi-biometric and unimodal systems on different datasets.

Figure 5 confirms the superiority of the proposed multi-biometric FR method to the unimodal one as the area under

the curve for the multi-biometric system is greater than that of the unimodal system.

4.2 Evaluation of bio-convolving method

The proposed method uses bio-convolving encryption to provide protection of biometric data without degradation in the system accuracy. To prove that, Table 6 illustrates the change in recognition accuracy for both unimodal and multi-biometric FR methods after applying bio-convolving and bloom filter [64] methods.

Table 6 shows that the recognition accuracy is slightly affected after applying bio-convolving. In addition, Fig. 6

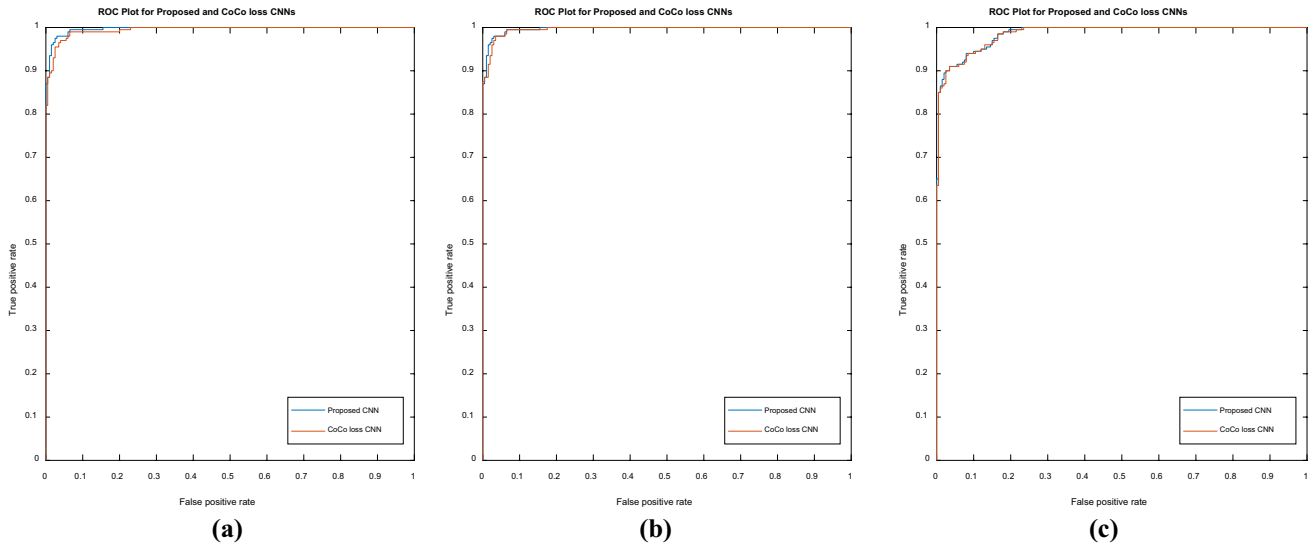


Fig. 4 ROC plots of the multi-biometric systems using the proposed and CoCo loss CNNs on **a** FERET dataset, **b** LFW dataset and **c** PaSC dataset

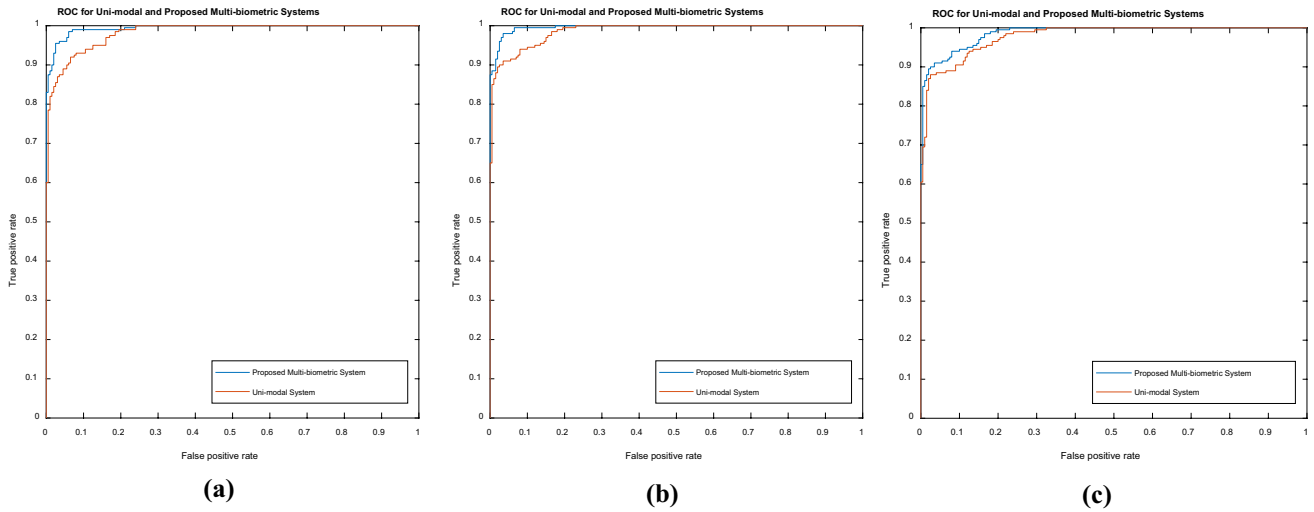


Fig. 5 ROC plots of both multi-biometric and unimodal systems on **a** FERET dataset, **b** LFW dataset and **c** PaSC dataset

Table 6 Cancelable biometric effect on recognition accuracy for uni-modal and multi-biometric systems

FR method	Dataset	Without cancelability (%)	Bioconvolving (%)	Bloom filter (%) [64]
Unimodal	FERET	97.92	97.14	93.18
	LFW	98.85	97.94	93.48
	PaSC	96.85	95.42	92.64
Multi-biometric	FERET	99.83	98.89	93.93
	LFW	99.89	98.93	94.02
	PaSC	98.31	97.38	93.4

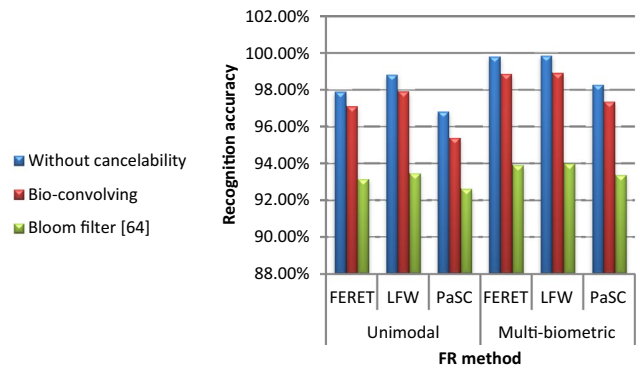


Fig. 6 Cancelable biometric influence on recognition accuracy for unimodal and multi-biometric FR

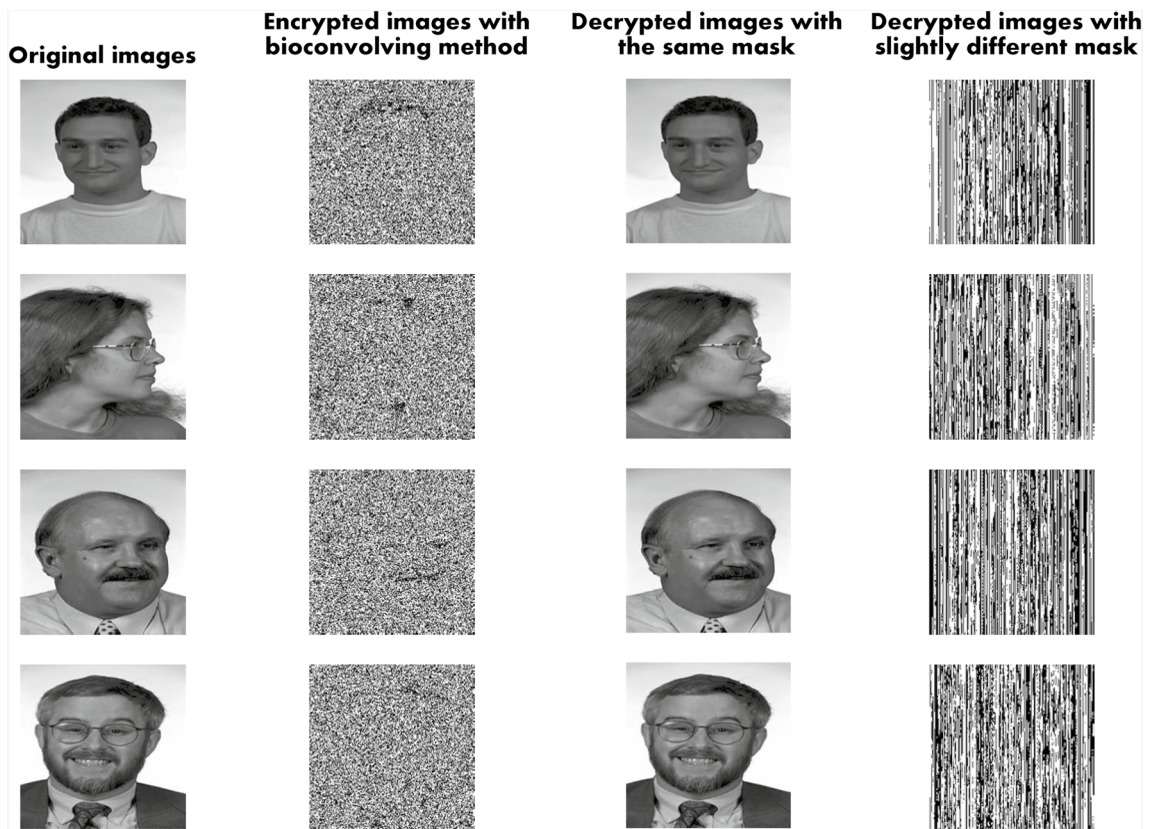


Fig. 7 Encryption and decryption operations on a number of facial images using different de-convolution masks

shows a graphical comparison between different cancelable techniques and their influence on recognition accuracy for both unimodal and multi-biometric FR systems.

The ability of bioconvolving method to provide security and privacy of user’s data can be verified through performing encryption and decryption operations on a number of

facial images using different de-convolution masks as shown in Fig. 7. Figure 8 presents the probability density functions (PDFs) of the mean square error (MSE) and normalized absolute error (NAE) for both correct and slightly different de-convolution masks.

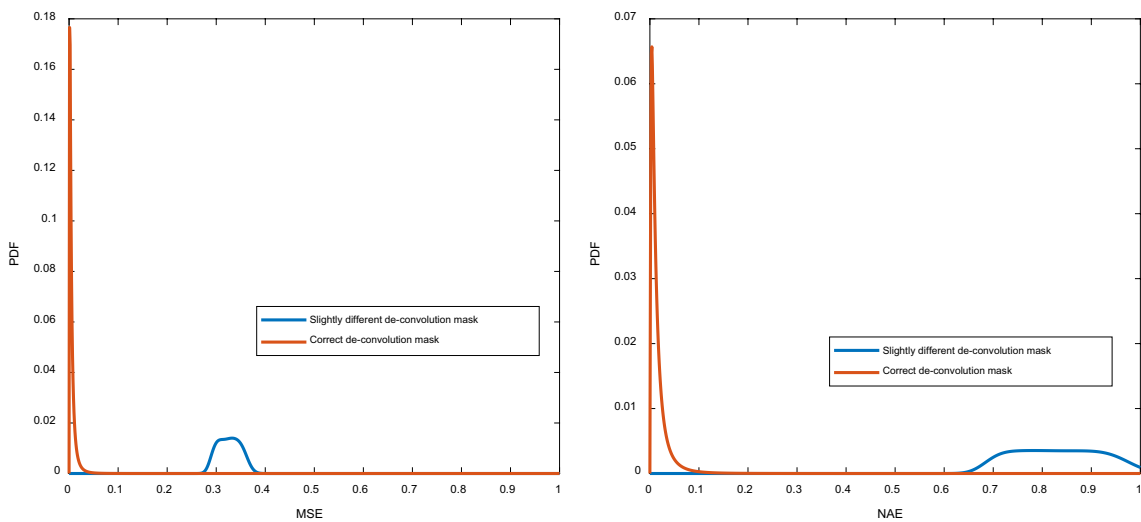


Fig. 8 PDF of MSE and NAE

As illustrated in Figs. 7 and 8, it is clear that we can distinguish easily between correct and incorrect de-convolution results. Overall, the experimental results prove that the proposed multi-biometric FR method is superior to the unimodal one due to exploiting multiple CNNs to obtain a variety of features from different facial regions and applying fusion to get more appropriate and sufficient facial descriptors. On the other hand, the utilization of four CNNs adds complexity to the proposed FR method. Furthermore, a single CNN model takes about 4 h in the training process. So, the proposed method suffers from a time consumption issue.

5 Conclusions

In this paper, we presented a cancelable multi-biometric FR method that achieves a promising performance across different datasets. Thanks to depth concatenation and the residual learning framework, we proposed an efficient CNN architecture. In the proposed method, multiple CNNs extract deep features from different facial regions, a fusion network combines these features into a more representative, reliable and detailed output, and finally, a bio-convolving method maintains privacy and security of biometric templates without affecting the recognition accuracy. The experiments on the FERET, LFW and PaSC datasets demonstrated that (a) the proposed method outperforms the state-of-the-art techniques in the presence of cancelability and (b) the multi-biometric system achieves excellent results compared to the unimodal one.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Choi, J.Y.: Spatial pyramid face feature representation and weighted dissimilarity matching for improved face recognition. *Vis. Comput.* **34**, 1535–1549 (2018)
- Yongjie, Ch., Lindu, Z., Touqeer, A.: Multiple feature sub-spaces analysis for single sample per person face recognition. The visual figure 7: encryption and decryption operations on a number of facial images using different de-convolution masks. *Computer* **35**(2), 239–256 (2019)
- Li, K., Yi, J., Muhammad, W., Ruize, H., Jiongwei, Ch.: Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. *Vis. Comput.* (2019). <https://doi.org/10.1007/s00371-019-01627-4>
- Cheng, Y., Jiao, L., Cao, X., Li, Z.: Illumination-insensitive features for face recognition. *Vis. Comput.* **33**, 1483–1493 (2017)
- Yueming, W., Gang, P., Jianzhuang, L.: A deformation model to reduce the effect of expressions in 3D face recognition. *Vis. Comput.* **27**(5), 333–345 (2011)
- Yujie, L., Deng, Y., Xiaoming, Ch., Zongmin, L., Jianping, F.: TOP-SIFT: the selected SIFT descriptor based on dictionary learning. *Vis. Comput.* **35**, 667–677 (2019)
- Leila, K., Mehrez, A., Ali, D.: Image classification by combining local and global features. *Vis. Comput.* **35**(5), 679–693 (2019)
- Baozhen, L., Hang, W., Weihua, S., Wenchang, Z., Jinggong, S.: Rotation-invariant object detection using Sector-ring HOG and boosted random ferns. *Vis. Comput.* **34**(5), 707–719 (2018)
- Fengping, A., Zhiwen, L.: Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM. *Vis. Comput.* (2019). <https://doi.org/10.1007/s00371-019-01635-4>
- Jianfeng, Z., Xia, M., Jian, Z.: Learning deep facial expression features from image and optical flow sequences using 3D CNN. *Vis. Comput.* **34**(10), 1461–1475 (2018)
- He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016)
- Gaurav, T., Kuldeep, S., Dinesh, V.: Convolutional neural networks for crowd behaviour analysis: a survey. *Vis. Comput.* **35**(5), 753–776 (2019)
- Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: *CVPR*, pp. 815–823 (2015)
- Heekyung, Y., Kyungha, M.: Classification of basic artistic media based on a deep convolutional approach. *Vis. Comput.* (2019). <https://doi.org/10.1007/s00371-019-01627-4>
- Wu, X., He, R., Sun, Z.,; Tan, T.: A light CNN for deep face representation with noisy labels. In: *Proceedings of the IEEE Transactions on Information Forensics and Security*, pp. 2884–2896 (2018)
- Deng, J., Guo, J., Zafeiriou, S.: Arcface: additive angular margin loss for deep face recognition. *arXiv preprint arXiv:1801.07698* (2018)
- Liu, Y., Li, H., Wang, X.: Rethinking feature discrimination and polymerization for large-scale recognition. *arXiv preprint arXiv:1710.00870* (2017)
- Zheng, Y., Pal, D., Savvides, M.: Ring loss: convex feature normalization for face recognition. In: *CVPR* (2018)
- Lobache, O.: Direct visualization of cryptographic keys for enhanced security. *Vis. Comput.* **34**, 1749–1759 (2018)
- Ratha, N., Chikkerur, S., Connell, J., Bolle, R.: Generating cancelable fingerprint templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 561–572 (2007)
- Teoh, A., Goh, A., Ngo, D.: Random multispace quantization as an analytic mechanism for bihashing of biometric and random identity inputs. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 1892–1901 (2006)
- Polash, P., Gavrilova, M., Klimenko, S.: Situation awareness of cancelable biometric system. *Vis. Comput.* **30**, 1059–1067 (2014)
- Klare, B., Klein, B., Jain, A.: Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In: *Proceedings of the IEEE CVPR*, pp. 1931–1939 (2015)
- Je-Kang, P., Dong, K.: Unified convolutional neural network for direct facial keypoints detection. *Vis. Comput.* (2018). <https://doi.org/10.1007/s00371-018-1561-3>

25. Li, L., Jun, Z., Fei, J., Li, S.: An incremental face recognition system based on deep learning. In: Proceedings of the Fifteenth IAPR International Conference on Machine Vision Applications (MVA) Nagoya University, Nagoya, Japan, pp. 238–241 (2017)
26. Xujie, L., Hui, H., Hanli, Z., Yandan, W., Mingxiao, H.: Learning a convolutional neural network for propagation-based stereo image segmentation. *Vis. Comput.* (2018). <https://doi.org/10.1007/s00371-018-1582-y>
27. Xiaoxiao, L., Qingyang, X., Ning, W.: A survey on deep neural network-based image captioning. *Vis. Comput.* **35**(3), 445–470 (2019)
28. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE CVPR, pp. 2892–2900 (2015)
29. Sun, Y., Wang, X., Tang, X.: Sparsifying neural network connections for face recognition. In: Proceedings of the IEEE CVPR (2016)
30. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE CVPR, pp. 1701–1708 (2014)
31. Hao, G., Baozhong, C.: How do deep convolutional features affect tracking performance: an experimental study. *Vis. Comput.* **34**(12), 1701–1711 (2018)
32. Masi, I., Rawls, S., Natarajan, P.: Pose-aware face recognition in the wild. In: Proceedings of the IEEE CVPR, pp. 4838–4846 (2016)
33. Taigman, Y., Yang, M., Wolf, L.: Web-scale training for face identification. In: Proceedings of the IEEE CVPR, pp. 2746–2754 (2015)
34. Kurban, O., Yildirim, T., Bilgiç, A.: A multi-biometric recognition system based on deep features of face and gesture energy image. In: Proceedings of the IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA), pp. 361–364 (2017)
35. Li, L., Jun, Z., Fei, J., Li, S.: An incremental face recognition system based on deep learning. In: Proceedings of the Fifteenth IAPR International Conference on Machine Vision Applications (MVA), pp. 238–241 (2017)
36. Feng, Y.C., Yuen, P.C., Jain, A.K.: A hybrid approach for generating secure and discriminating face template. *IEEE Trans. Inf. Forensics Secur.* **5**(1), 103–117 (2010)
37. Yang, W., Wang, S., Zheng, G., Chaudhry, J., Valli, C.: ECB4CI: an enhanced cancelable biometric system for securing critical infrastructures. *J. Supercomput.* **74**, 4893–4909 (2018)
38. Paul, P.P., Gavrilova, M.: A novel cross folding algorithm for multimodal cancelable biometrics. *Int. J. Softw. Sci. Comput. Intell.* **4**(3), 21–38 (2012)
39. Alder, A.: Sample images can be independently restored from face recognition templates. *Electr. Comput. Eng.* **2**, 1163–1166 (2003)
40. Bedad, F., Adjoudj, R.: Multi-biometric template protection: an overview. In: Proceedings of International Conference in Artificial Intelligence in Renewable Energetic Systems, pp. 70–79 (2018)
41. Natgunanathan, I., Mehmood, A., Xiang, Y., Hua, G., Li, G., Bangay, S.: An overview of protection of privacy in multibiometrics. *Multimed. Tools Appl.* **77**, 6753–6773 (2018)
42. Droandi, G., Barni, M., Lazzeretti, R., Pignata, T.: SEMBA:SEcure multi-biometric authentication. [arXiv:1803.10758](https://arxiv.org/abs/1803.10758) (2018)
43. Teoh, A., Ngo, D., Goh, A.: Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognit.* **37**(11), 2245–2255 (2004)
44. Patel, V.M., Ratha, N.K., Chellappa, R.: Cancelable biometrics: a review. *IEEE Signal Process. Mag.* **32**(5), 54–65 (2015)
45. Marco, C., Oscar, D., Cayetano, G., Mario, H.: ENCARA2: real-time detection of multiple faces at different resolutions in video streams. *J. Vis. Commun. Image Represent.* **18**(2), 130–140 (2007)
46. Yuan, Z., Liu, Y., Yue, J., Li, J., Yang, H.: CORAL: coarse-grained reconfigurable architecture for convolutional neural networks. In: Proceedings of the IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), pp. 1–6 (2017)
47. Ioffe, S., Szegedy, Ch.: Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167v3 [cs.LG]*, pp. 1–11 (2015)
48. Zhang, X., Ren, Sh., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
49. Liu, W., Jia, Y., Sermanet, P., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
50. Xiaoliang, Z., Zijian, C.: Dual-modality spatiotemporal feature learning for spontaneous facial expression recognition in e-learning using hybrid deep neural network. *Vis. Comput.* (2019). <https://doi.org/10.1007/s00371-019-01660-3>
51. Dandan, L., Huangang, L., Zhenbo, Y., Yipu, Z.: Deep convolutional BiLSTM fusion network for facial expression recognition. *Vis. Comput.* (2019). <https://doi.org/10.1007/s00371-019-01636-3>
52. Hasnat, A., Bohné, J., Milgram, J., Gentric, S., Chen, L.: Deep-Visage: making face recognition simple yet with powerful generalization skills. In: Proceedings of the CVPR, pp. 1–12 (2017). <https://www.nist.gov/itl/iad/image-group/color-feret-database>
53. Huang, G., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (2007)
54. Beveridge, J., Phillips, J., Bolme, D., Draper, B., Givens, G., Lui, Y., Teli, M., Zhang, H., Scruggs, W., Bowyer, K.: The challenge of face recognition from digital point-and-shoot cameras. In: Proceedings of IEEE International Conference on Biometrics: Theory, Applications Systems, pp. 1–8 (2013)
55. Wu, X., He, R., Sun, Z., Tan, T.: A Light CNN for deep face representation with noisy labels. In: CVPR, pp. 1–13 (2017)
56. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: SphereFace: deep hypersphere embedding for face recognition. In: CVPR (2017)
57. Liu, J., Deng, Y., Huang, C.: Targeting ultimate accuracy: face recognition via deep embedding. [arXiv:1506.07310](https://arxiv.org/abs/1506.07310) (2015)
58. Sun, Y., Liang, D., Wang, X., Tang, X.: DeepID3: face recognition with very deep neural networks. [arXiv:1502.00873](https://arxiv.org/abs/1502.00873) (2015)
59. Ding, C., Tao, D.: Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 1002–1014 (2017)
60. Shu, X., Liyan, Z., Jinhui, T., Guo-Sen, X., Shuicheng, Y.: Computational face reader. In: MMM, Part I LNCS 9516, pp. 114–126 (2016)
61. Shu, X., Jinhui, T., Zechao, L., Hanjiang, L., Liyan, Z., Shuicheng, Y.: Personalized age progression with bi-level aging dictionary learning. *IEEE Trans. Pattern Anal. Mach. Intell.* (2017)
62. Shu, X., Jinhui, T., Hanjiang, L., Zhiheng, N., Shuicheng, Y.: Kinship-guided age progression. *Pattern Recognit.* **59**, 156–167 (2016). <https://doi.org/10.1016/j.patcog.2015.12.015>
63. Lin, Y., Xun, L.: A Cancelable multi-biometric template generation algorithm based on bloom filter. In: International Conference on Algorithms and Architectures for Parallel Processing, pp. 547–559 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Essam Abdellatef received the Master degree in Electronics and Communications Engineering from the Faculty of Engineering, Mansoura University, Egypt by 2015. Currently, he is pursuing his PhD Degree in Menoufia University, Egypt. He is now an assistant lecturer at the electronics and communications engineering department, Delta Higher Institute for Engineering and Technology, Egypt. His main research area includes Deep Learning applications, 3D object reconstruction, and image processing.



Khalid N. Ismail is an Assistant professor at the Faculty of Computers and Information, Menoufia University, Egypt. He is also a research associate at the Department of Computer Science, Durham University, UK. He has been working as an academic research fellow at Nottingham, Bath, and Leeds Universities. He obtained his PhD from University of Leeds, UK, in 2013. His research interests include advanced computer vision and image processing, 3D image reconstruction, tomography,

and deep machine learning applied to complex real-world problems.



Nabil A. Ismail was the Dean of the Faculty of Computers and Information, Menoufia University (August 2006–August 2008). He is now a Professor of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University. Prof. Ismail obtained his PhD degree in Computer Eng. from Durham University, England, in 1983. His main research area includes Deep Learning applica-

tions and computer vision, Tomography, Computer security, Computer architecture, Elliptic curve cryptography, Processor design, Light-power smart devices, Security applications, EIT algorithms, and Multi-core/many-core parallel programming.



Mohamed Rihan received his B.Sc. in electronics and communication engineering, with honors from Menoufia University, Egypt. He received his M.Sc. and PhD in electronics and communication engineering from Egypt-Japan University of Science and Technology in 2012 and 2015, respectively. Currently, he is a postdoctoral research fellow with College of Information Engineering, Shen-

zhen University, China. He is also serving as an assistant professor at the Faculty of Electronic Engineering (FEE), Menoufia University, Egypt. His research interests include massive MIMO and mmWave communications, interference alignment, resource allocation, and cognitive heterogeneous networks.



Salah Eldin S. E. Abd Elrahman received the B.Sc. degree in Industrial Electronics Engineering on May 1988, from the Faculty of Electronic Engineering, Menouf, Menoufia University, Egypt. He received the M. Sc. degree in Computer Science and Engineering on July 9, 1994, from the Faculty of Electronic Engineering, Menouf, Menoufia University, Egypt. The thesis subject was “A Design of an Arabic Programming Language and its Compiler.” He obtained his Ph.D. degree in Computations

on July 11, 2003, from the Dept. of Computation, University of Manchester Institution of Science and Technology (UMIST), UK. The thesis subject was “Object-Oriented Technology for System-Level Design.” His main research area is developing Computer Systems that includes Embedded Systems, Real-Time Systems, Concurrent Systems, and Software Systems such as compilers.



Fathi E. Abd El-Samie received the B.Sc. (Honors), M.Sc., and Ph.D. from the Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt, in 1998, 2001, and 2005, respectively. He has received the most cited paper award from Digital Signal Processing journal for the year 2008. His current research areas of interest include image enhancement, image restoration, image interpolation, super resolution reconstruction of images,

data hiding, multimedia communications, medical image processing, optical signal processing, and digital communications.