IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 July 2022, accepted 28 August 2022, date of publication 12 September 2022, date of current version 22 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3206024

RESEARCH ARTICLE

Image Quality Assessment for Effective Ear Recognition

SUSAN EL-NAGGAR[1] AND THIRIMACHOS BOURLAI[2]
[1]Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA
[2]School of Electrical and Computer Engineering, University of Georgia, Athens, GA 30602, USA

Corresponding author: Thirimachos Bourlai (thirimachos.bourlai@uga.edu)

ABSTRACT Due to the recent challenges in access control, surveillance and security, there is an increased need for efficient human authentication solutions. Ear recognition is an appealing choice where the data acquisition procedure is contactless, non-intrusive, and covert. This article proposes a deep learning-based solution for effective ear recognition. First, we explore multiple strategies to enhance learning using alternative ear datasets with a wide range of ear poses. Second, we investigate the performance of the proposed deep ear models in the presence of various image artifacts, which commonly occur in real-life recognition applications, to identify the robustness of the proposed ear recognition models in controlled and uncontrolled conditions (dataset dependent). Finally, we propose an efficient ear image quality assessment tool designed to guide the proposed ear recognition system. By performing a set of experiments on extended degraded ear datasets, we determine that the employment of the proposed ear image quality assessment tool improves ear identification performance from 58.72% to 97.25% for the USTB degraded dataset and from 45.80% to 75.11% for the degraded FERET dataset.

INDEX TERMS Biometrics, ear recognition, convolutional neural networks, image artifacts, quality assessment.

I. INTRODUCTION

The ongoing COVID-19 pandemic is changing the world and reshaping our society. The way people learn, work, and interact with each other has been affected in various ways. As a result, more sophisticated and flexible use of technology is designed, developed, and utilized to support online and distance interactions. Even for in-person activities in public spaces, advanced health-related safety measures are taken (such as wearing face masks and limiting contact with commonly touched surfaces). These measures help to avoid the spread of the infection and protect the health of individuals and communities. Therefore, the need for secure, efficient, and convenient human authentication applications is critical for various daily activities, including but not limited to banking, accessing medical records, border crossing, surveillance, and even personal access to mobile devices.

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

Biometrics is considered one of the leading technologies used to identify and authenticate individuals under various conditions. Biometric solutions offer a rapid and reliable way of human authentication by using unique personal physiological (face, voice, iris, and fingerprints) and/or behavioral (keystroke dynamics, signature, and gait) characteristics. Although face and fingerprints are among the most popular biometric modalities [1], there are several technological concerns (accuracy, efficiency, scalability, biometric attacks) as well as concerns with their usage, storage, and sharing, including privacy. For example, one of the technological-related concerns is using face masks, which has presented a serious challenge to face recognition systems [2]. Also, contact-based fingerprint scanners are not always preferable due to hygiene concerns. Thus, for some recognition scenarios, ear biometrics can provide a suitable alternative for human authentication.

Ear recognition has its advantages; it is passive, contactless, non-intrusive, and expressionless [3], [4], [5], [6]. In addition, it demonstrates high discriminative information

across individuals and has shown to be an efficient human authentication solution, even when used to distinguish identical twins [7]. An automatic ear recognition system is mainly *a pattern recognition system* that consists of three modules. First is the ear image pre-processing and detector module that provides the bounding box(s) of the ear(s) to localize them in images or videos. Second is the ear descriptor module that encodes the identity information from the detected ear. Last is the decision-making module that identifies or verifies the subject that the query ear belongs to.

Although conventional-based ear recognition systems are still being used and result in acceptable (dataset-dependent) recognition performance, deep learning methods have the potential to improve the current state-of-the-art further. Deep learning methods have already dramatically improved the efficiency of various computer vision systems and brought breakthrough solutions in processing images, videos, speech, and audio [9], [10], [11], [12], [13], [14], [15], [16]. Specifically, in biometrics, there has been progress in using deep learning-based models for different biometric applications such as face, fingerprint, periocular, and gait recognition [17], [18], [19], [20], [21]. Therefore, employing deep models for ear description and feature extraction became an attractive area of research.

This article is a follow-up to our previous work [22], where we investigated multiple convolutional neural networks for ear recognition and the optimum learning process setting. We propose deep ear recognition models and evaluate their performance. We first establish a baseline for the performance; then, we perform detailed experiments to quantitatively evaluate the performance of the different ear recognition models in the presence of image artifacts, which commonly occur in real-life recognition applications, to identify their shortcomings and draw conclusions for enhancement. The experimental results show that image artifacts significantly affect recognition performance and can cause an efficiency loss in the biometric system. Consequently, evaluating the quality of ear images before processing can benefit ear recognition systems.

There has been limited discussion on the quality assessment of ear images for recognition applications. A quality assessment algorithm evaluates an input sample to determine if it is suitable for automated matching [23]. This is also related to the recognition scenario (constrained vs. unconstrained), the biometric recognition system (COTS vs. academic) and cannot necessarily be aligned with the human perspective of ear image quality assessment. In this work, we develop a system for holistic ear image quality assessment. The system serves as a guide for ear recognition systems to enhance recognition accuracy.

Hence, we propose a set of efficient deep ear recognition models that offer high recognition accuracy under variable conditions when supported by an ear image quality assessment tool. The contributions of this work are summarized as follows:

- We provide a comparative evaluation of the performance of four deep CNN models: SqueezeNet, GoogLeNet, MobileNet, and DenseNet, for ear identification and verification tasks. For that purpose, we use an ear dataset with a wide range of pose angles.
- We quantitatively assess the impact of the quality of the ear image on the performance of deep ear models. In order to further explore the strengths and weaknesses of our proposed deep ear models, we evaluate the recognition performance in the presence of multiple ear image degradation factors, including blurriness, brightness, and contrast variations, to obtain if the performance of certain CNNs is more prone to degradation in response to specific grades of artifacts.
- We propose an automatic *ear image quality assessment* tool to act as a guide for improving ear recognition accuracy. Quality labels are obtained from scores yielded by an ear recognition matcher. Using predicted quality labels improves ear recognition performance and reduces error rates.

II. BACKGROUND

What follows is a discussion and literature review on the subject from the perspective of image quality assessment and ear recognition technologies.

A. INTRODUCTION TO BIOMETRIC QUALITY

A quality of a biometric sample is an indicator of how suitable it is for automated matching. The environmental image distortions such as noise, blur, and illumination variation are primary reasons for the deterioration in biometric identification accuracy. Therefore, there is a need for a quality assessment algorithm to produce a target quality that predicts the recognition performance of the biometric system when employing the sample regardless of human judgment. In some cases, comparing two biometric samples of low quality can produce high genuine similarity scores [24]; therefore, the biometric sample's quality needs to be evaluated without a reference or comparison with a second sample.

The two modes of operation of a typical ear biometric system are enrollment and recognition (verification or identification). In the enrollment mode of operation, a user's ear biometric is captured to generate the template(s) to be stored in the system's database. In the recognition mode of operation, an input ear sample is processed to identify a subject or verify his/her identity. An ear image quality assessment can be helpful for one or more but not limited to the following scenarios:

- During enrollment, when the system determines that an enrollment (input) sample is of low quality; it can guide the user and recapture the sample.
- During verification, when genuine users are expected to provide an input ear biometric sample of high quality for recognition, ear image quality needs to be established as good before verification. The quality examination can be used to guide the recapture of the biometric sample or

to prevent spoofing by the presentation of a deliberately poor biometric sample from an imposter.
- In preprocessing of biometric samples, the evaluation of a biometric sample's quality can be used to initiate certain preprocessing algorithms.
- In surveillance or video-based ear recognition applications, the quality assessment is useful for the frame selection for the recognition operation.
- In the fusion of multiple images and/or biometric modalities, quality assessment can provide a guide for sample selection.

The target quality value can be a scaler prediction of the genuine score, a bin indicating that an image is poor/fair/good for matching, or a binary value of low-quality vs. high-quality images [23]. In this work, we develop a holistic ear image quality assessment without measuring individual factors. In most biometric applications, it is sufficient to detect low-quality biometric samples to reject them and initiate the proper action. The proposed system produces a binary value indicating whether an image is good or bad for matching.

B. RELATED WORK

In this section, we provide a brief review of relative work on ear recognition and the usage of image quality assessment in biometric applications.

1) EAR RECOGNITION

The potential of the human ear for personal identification was recognized by Alphonse Bertillon as early as 1890 [25]. In 1949, Alfred Iannarelli developed one of the first ear recognition systems. He used twelve measurements from the ear image to represent the ear [26]. Since then, multiple machine learning methods and conventional matchers have been used for ear recognition research studies. There have been multiple detailed reviews of ear recognition history, techniques, and their progress [32], [33], with a recent one [36].

After developing deep learning models and their improved performance for many machine vision applications, the ear recognition research shifted toward employing them for ear recognition systems. There have been multiple efficient systems based on CNNs for ear detection [37], [38], [39], [40] and ear segmentation [41], which is an important step that can be used towards deep learning-based ear recognition approaches. For recognition, the limited ear training data was the main obstacle in utilizing convolutional neural networks for ear recognition applications. Emeršič *et al.* [42] addressed this problem. They collected an uncontrolled ear dataset from the internet. The team presented the Unconstrained Ear Recognition Challenge (UERC), which was held twice in 2017 [44] and 2019 [45] to evaluate the state of the ear recognition technology for unconstrained ear images. Eyiokur *et al.* [46] presented a detailed ear recognition study using the UERC 2019 dataset. Whereas Dodge *et al.* [47] used a hybrid deep and shallow learning approach for ear recognition.

Zhang *et al.* [49] used three CNNs with different scales of ear images to obtain multi-scale ear representations for ear verification. They did their experiments on their new ear database named USTB-Helloear. Khaldi *et al.* [50] proposed a two-phase training method for the VGG16 architecture for ear classification. They also used Generative Adversarial Network to color the USTB II dataset images. The inceptionV3 deep learning model was used in [48] for recognition of the AMI ear database. They used the network as a feature extractor and principal component analysis to reduce the feature vector size. Alshazly *et al.* [51] achieved Rank-1 recognition accuracy of 93.45% for the EarVN1.0 dataset using the ResNeXt CNN, and they used the t-SNE algorithm to visualize the learned features. They also built ensembles of ResNet models with various depths for feature extraction, followed by SVM classifiers [52]. Finally, Meng *et al.* presented a study on distinctiveness and symmetry in Ear Biometrics [53]. In their experiments, they recognized the gender with a 90.9% success rate and confirmed the existence of symmetry between a subject's ears. Table 1 provides a comparative summary of ear recognition techniques in terms of Rank-1 (%) identification rate.

2) BIOMETRICS QUALITY

There has been plenty of studies that investigated the quality of face images for biometric recognition and the performance of face recognition algorithms concerning different covariates, on the contrary, there has been minimal work that analyzes the quality of ear images for recognition.

In the field of quality assessment for face recognition, Abaza *et al.* [54] examined the influence of face images' quality factors, such as contrast, brightness, sharpness, focus, and illumination, on recognition performance. They evaluated quality measures for each factor and proposed a face image quality index that combines multiple quality measures which reflects the changes of input quality factors in correlation with face recognition performance. In another work, Best-Rowden *et al.* [24] proposed a model for the automatic prediction of face image quality. They used two techniques for face image quality assessments: human ratings of face image quality and quality values computed from similarity scores from face matchers. For matcher-dependent face quality values, they used the normalized comparison of a sample's genuine score with its impostor distribution when compared to a gallery of samples. For both techniques, each face image was represented with a 320-dimensional feature vector extracted from face images using the ConvNet for face recognition. Using the face representations, they trained a support vector regression (SVR) model with a radial basis kernel function (RBF) to predict the normalized comparison scores from the face matcher or the human quality rating. In their experiments, they used the predicted face image quality to reject low-quality face samples, which reduced FRR at 1% FAR error rates by at least 13% for different face matchers.

TABLE 1. Comparative summary of 2D ear recognition performances in terms of identification rate at Rank-1 (in %).

Method	Year	Dataset	Rank-1
LBP, LPQ, HOG, BSIF [29]	2014	UND-J2	97.22, 98.73, 97.85 , 98.67
SqueezeNet [42]	2017	AWE + CVLE	62.00
GoogLeNet [46]	2017	Multi-PIE	99.32
VGG-16 + GoogLeNet [46]	2017	UERC	67.53
PHOG + LDA [30]	2017	IITD-I, IITD-II, UND-E	92.76, 95.77, 96.60
BSIF [34]	2017	USTB-I, IITD-I, IITD-II	98.97, 97.39, 97.63
SIFT, SURF, MLBP, LTP [31]	2018	FERET + WVU + UND + USTB	45.10, 52.60, 90.20, 85.37
Multiband PCA [27]	2018	USTB-II, IITD-II	51.95, 93.21
VGG-M + SVM [35]	2018	USTB-I, USTB-II, IITD-I, IITD-II	99.40, 99.60, 99.90, 99.80
ResNet18 [47]	2018	USTB-Helloear	97.40
VGG-Face [49]	2018	AWE + CVLE	80.03
MLBP [28]	2019	USTB-I, IITD-I, IITD-II	98.33, 98.40, 98.64
inceptionV3 [48]	2020	AMI	98.1
VGG16, ResNet, SqueezeNet [43]	2020	AWEx(Male)	43.7, 29.5, 52.6
ResNeXt [51]	2020	EarVN1.0	93.45
VGG16 [50]	2021	AMI, USTB-II, AWE	98.33, 100.00, 51.25
ResNet + SVM [52]	2021	AMI, AMIC, WPUT, AWE	99.64, 98.57, 81.89, 67.25
Our proposed approach	2022	**WVU, USTB-III**	**99.67, 99.35**

Ortega *et al.* [55], [56] proposed the FaceQnet for face image quality assessment for recognition purposes. The FaceQnet is based on the ResNet-50 architecture. The network was trained to output a quality measure between 0 and 1 related to face recognition accuracy. The authors labeled a subset from the VGGFace2 face database with quality scores for training. For each subject of the dataset, they used one face image with the highest compliance with ICAO (standards for machine-readable travel documents) as the perfect quality face image. The comparison scores between the other sample images of the subject with the high-quality face images were used as quality values for these face images. The FaceQnet was trained using the pairs of face images and their quality values. To evaluate their proposed system, they obtained the quality values for a test set of face images and performed verification. Their experiments showed a correlation between their quality measure and verification accuracy.

For the quality of ear images for ear recognition, in an earlier study, Pflug *et al.* [57] investigated the impact of signal degradation on ear recognition performance. Their experiments examined the effect of noise and blur on descriptor-based ear recognition, including LBP, LPQ, and HOG. More recently, Emeršič *et al.* [43] performed a detailed study of the effect of subject-related covariates, including ethnicity, head rotation, gender and presence of occlusions, and accessories on the performance of ear recognition techniques.

III. METHODOLOGY

Our experiments examine multiple Convolutional Neural Network architectures to find the appropriate model for the ear recognition task. An overview of the models examined, the learning strategies implemented, and the ear image artifacts explored follows.

A. CONVOLUTIONAL NEURAL NETWORK MODELS LEARNING

A Convolutional Neural Network is a deep learning algorithm that processes input data, such as an image, to learn their spatial hierarchies and determine a set of distinguishing characteristics. The CNN consists of multiple layers (convolutional, pooling & fully connected) to filter images, extract their informative features, and classify them. Although Convolutional neural networks were introduced by LeCun *et al.* [61] in the 1980s, for the recognition of handwritten zip code digits, they became popular after the breakthrough they brought for image classification in 2014. Since then, Convolutional neural networks have emerged as a leading algorithm in computer vision. Advancements in computer hardware and larger datasets supported that advancement. In addition, there have been multiple studies to improve CNNs' architecture and enhance their performance for multiple machine learning applications, including biometrics. We examined multiple CNN models which represent the various developments in general CNN architectures and tuned them for ear recognition. Table 2 summarizes the main properties for each network.

For the model learning to overcome the limited size of the ear datasets available, we used multiple learning strategies, including data augmentation and transfer learning. For data augmentation, although we explored various data augmentation techniques, we concluded that the following ones resulted in improved accuracy of our models, namely rotation at random angles up to 40° in both directions (clockwise and counterclockwise) and translation horizontally or vertically with a random number of pixels in the range (−30° to +30°).

We also performed two phases of transfer learning:

1) ImageNet transfer learning: All CNN models used in this work are pre-trained on the ImageNet dataset [66].
2) Domain adaptation: The second phase of transfer learning, was domain adaptation where ear datasets were used to fine-tune the CNNs. We fine-tuned our CNN models using the training part of the AWE ear image data set [33]. The(AWE) dataset is an annotated ear dataset that was collected from web images of various quality and spatial resolution.

Fig. 1 presents an overview of the training for the deep learning models used for ear recognition.

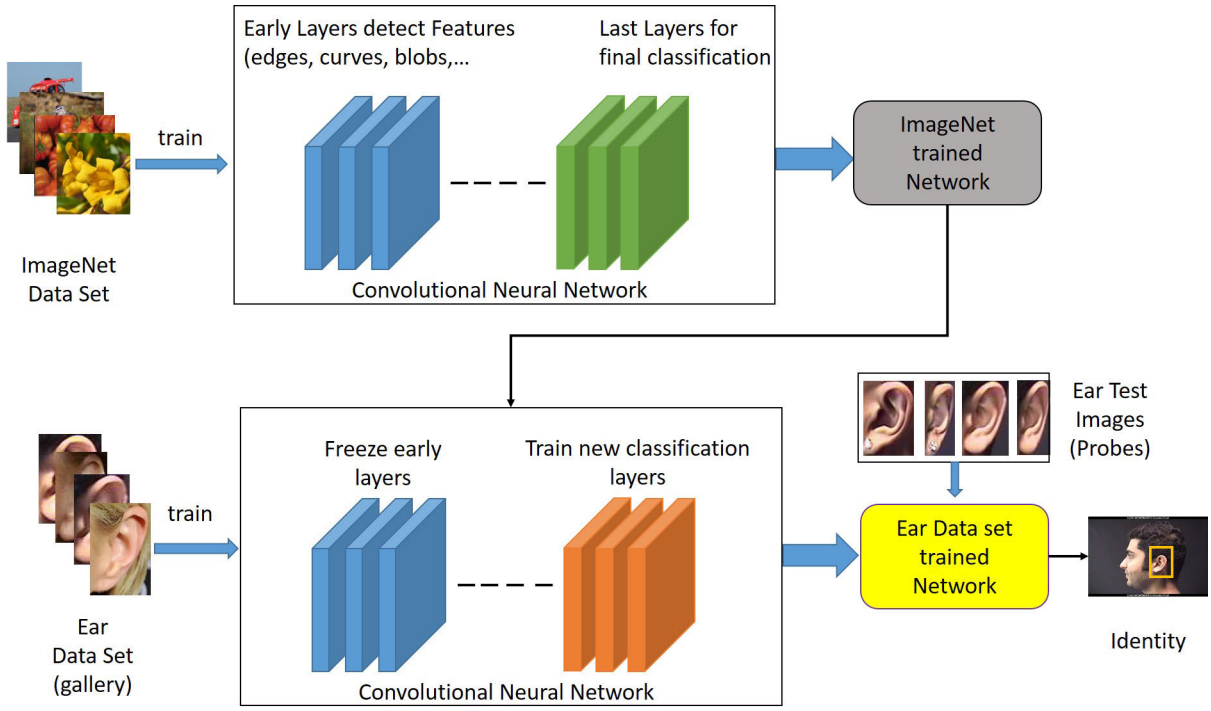


FIGURE 1. An overview of the deep learning-based ear recognition system. The proposed CNN is pre-trained on the ImageNet dataset and then fine-tuned using an ear dataset [22].

TABLE 2. Comparison of convolutional neural networks used. Note that parameters are in millions.

Network	Depth	Parameters	Size	Image Input Size	Convolutional layers
SqeezeNet [62]	18	1.24	4.6 MB	227×227×3	2 conv. and 8 fire
GoogLeNet [63]	22	7	27 MB	224×224×3	2 conv. and 9 inception
MobileNetV2 [64]	53	3.5	13 MB	224×224×3	1 conv. and 19 bottleneck res.
DenseNet [65]	201	20.0	77 MB	224×224×3	4 conv and 4 dense blocks

B. IMAGE DEGRADATIONS

The accuracy of Image-based biometric recognition systems is highly dependent on the quality of the input biometric images. Image degradation factors, such as out-of-focus, noise, and light alteration, commonly occur in real-life recognition applications and can affect the performance of biometric recognition systems. Therefore, assessing the conditions that can result in biometric image degradation manifested by the property of capture devices and conditions is helpful. In this work, we evaluate the impact of the variation of a set of image degradation factors on the performance of deep learning-based ear recognition systems. We systematically altered good quality ear probe images and, thus, generated a set of synthetic lower quality ear datasets. This was accomplished by adjusting the *contrast*, *brightness*, and *blurriness* of good quality ear images at different levels:

- Contrast: To adjust the contrast of ear probe images, we saturated ear images at low and high intensities in a 10% intensity degradation step.
- Brightness: The brightness of the probe ear images was artificially adjusted via a brightness (gamma γ) factor. This factor specifies the shape of the curve, describing the relationship between the values of the input and output images after the brightness level is manually adjusted. In case $\gamma < 1$, the mapping is weighted towards higher (brighter) output values, and if $\gamma > 1$, the mapping is weighted toward lower (darker) output values. We used γ values in the range [0.5, 1.4] with a uniform step size of 0.1 to generate nine probe sets for our brightness-related experiments.
- Blurriness: To generate the blurriness in probe ear images, we convolved them with a circular averaging filter and border replication. The value of diameter is

in the range [3, 19] pixels with a uniform step value of 2 pixels.

IV. EXPERIMENTAL SETUP AND RESULTS

In this section, first, we describe the data sets used in our experiments. Second, we explain the setup and the training procedure. Third, we present the performance of different deep models for ear recognition, including identification and verification. Afterward, we compare the ear recognition performance in the presence of image distortions.

A. DATASETS

- **WVU Ear Dataset:** For our experiments, we used the West Virginia University (WVU) Ear Dataset [58]. It was collected using a unique custom-made device. It consists of a moving arm holding a camera that captures video sequences. Each video begins at the left profile of a subject (0°) and terminates at the right profile (180°) in about 2 minutes. The WVU ear database consists of 460 video sequences for about 400 different subjects and multi-sequence for 60 subjects with an elapsed time period between them. We used the multiple sequences for our experiments. We used left ear images from one video sequence for each subject to generate the gallery ear dataset and images from the second video sequence as the probe ear dataset. For the gallery set, we extracted 20 ear images from the profile faces at different angles ranging from -10° passing by 0° (full profile) to about 60° (where the face is visible enough for face recognition). This process resulted in a training set of 1200 images for 60 subjects. For the probe set, we used five ear images per subject at about (−10°, 0°, 20°, 45° and 60°), which resulted in a testing set of 300 images.
- **USTB Ear Dataset:** The University of Science and Technology Beijing (USTB) collected multiple ear image datasets [59]. Dataset III contains ear images at multiple angles. Each subject rotates his/her head from 0° to 60° toward the right side, and from 0° to 45° toward the left side; two images were recorded at each angle. For our experiments, we used the ten left ear images for 77 subjects. The ear images were at angles 0°, 5°, 10°, 15°, and 20°. For each subject, eight images were used in the gallery set and two in the probe set.
- **FERET Dataset:** The FERET dataset [60] was part of the Face Recognition Technology Evaluation (FERET) program. For some individuals, images were collected at the right and left profiles (labeled pr and pl). From this dataset, we used left face profile(ear) images of 115 subjects to maintain two images per subject.

B. SETUP AND TRAINING FOR EAR RECOGNITION

We trained ear recognition CNN-based models using Stochastic Gradient Descent with Momentum (SGDM), learning rate 3×10^{-4}, and 20 maximum epochs. To speed up the network training and prevent it from overfitting to the new dataset, we froze the weights of the earlier layers in the network by setting the learning rates in those layers to zero. Specifically, we froze the weights for the first 5, 10, &17 layers of the network models. We trained the CNN models for ear identification that each subject represents a class. The final two layers were replaced with new layers to adapt to the new dataset. In SqueezeNet, the last convolutional layer was replaced with a new convolutional layer with the number of filters equal to the number of classes. For the other networks, the fully connected layers were replaced with new fully connected layers with outputs equal to the number of subjects in the dataset. When an unknown ear image (probe) is introduced to the network, the output is the subject (class) to which it is most likely the probe ear belongs, according to the probability from the SoftMax function. This was implemented for the WVU and the USTB datasets due to the availability of multiple samples per subject for training.

For the FERET dataset and verification experiments, we used the CNNs after domain adaptation to extract features for each ear image and generate image descriptors.

$$y = f(x), \tag{1}$$

where x is the input image, f(.) represents the CNN, and y is the image descriptor. The dimensionality of the image descriptor varies from model to model and depends on the design choices made during network construction.

C. EAR RECOGNITION PERFORMANCE

Our first experiment assesses the performance of the different CNN architectures for ear identification and verification tasks using the WVU and the USTB ear datasets. The Rank-1 identification scores for the four models examined are presented in Table 3.

TABLE 3. Ear identification performance for multiple models using Rank-1% scores.

Dataset	SqueezeNet	GoogLeNet	MobileNet	DenseNet
WVU	92.67	93.33	96.33	99.00
USTB	73.38	79.22	89.61	99.35

As shown in Table 3, the DenseNet model has the best identification performance for both the WVU & the USTB datasets, whereas SqueezeNet has the least Rank-1 scores.

For verification, each ear probe image descriptor is compared against each of the gallery images descriptors' using cosine similarity match scores. The match scores can be either genuine scores or imposter scores. Genuine scores are the scores when the gallery and probe ear images belong to the same subject, whereas imposter scores are when the gallery and probe ear images belong to different subjects. Match scores are compared against a numerical threshold. If the match score exceeds the threshold, it is classified as a match. Each input is either: True Positive (TP), True Negative (TN), False Positive (FP) or False Negative (FN). To analyze the verification performance, the False Accept Rate (FAR) and False Reject Rate (FRR) results are used, where:

- False Accept Rate (FAR): denotes the percentage of imposter ear images falsely recognized (FP) over the dataset's total number of ear images.
- False Reject Rate (FRR): denotes the percentage of genuine ear images (FN) falsely rejected over the total number of ear images in the dataset.

The Receiver Operating Characteristic (ROC) curves [8] relates the FAR to FRR at different thresholds to measure the verification performance of a biometric recognition system. Fig. 2 shows the ROC curves for the different ear recognition models. The MobileNet model has the best verification performance, followed by DenseNet. GoogLeNet and SqueezeNet has comparable verification performance.

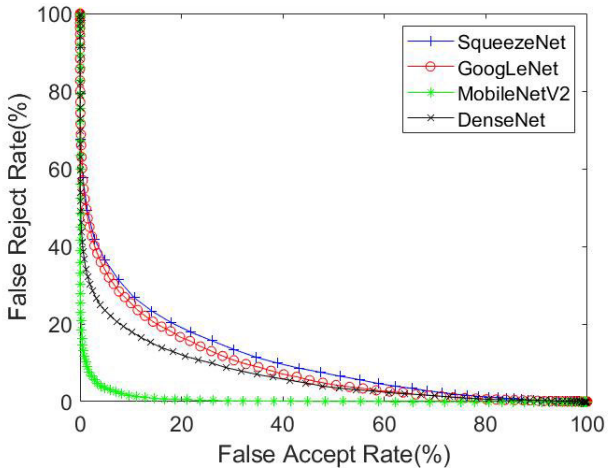


FIGURE 2. ROC comparison of four models SqeezeNet, GoogLeNet, MobileNet, and DenseNet for the WVU ear dataset.

D. EAR RECOGNITION WITH IMAGE DEGRADATIONS

For each of the datasets used for the recognition and quality experiments (WVU, USTB, and FERET), we kept the gallery part of the original ear images of the dataset. We applied image degradations to the ear images in the probe part of the datasets. That increased the probe part of the dataset, consisting of the combination of the original and degraded ear images. Table 4 shows the number of ear images in the original dataset and after adding the degraded images to the probe part of the dataset.

TABLE 4. Ear recognition datasets.

Dataset	Gallery	Probe	Probe with degradations
WVU	60 × 20 = 1200	60 × 5 =300	60 × 150 = 9000
USTB	77 × 8 = 616	77 × 2 = 154	77 × 60 = 4620
FERET	115	115	115 × 30 = 3450

In our experiments, we explore the impact of the degradation of the ear probe images on the Rank-1 identification accuracy of the different models examined in the previous section. First, we examine the effect of contrast alteration on ear recognition performance; the Rank-1 identification results are shown in Table 5. Fig. 3 shows sample images

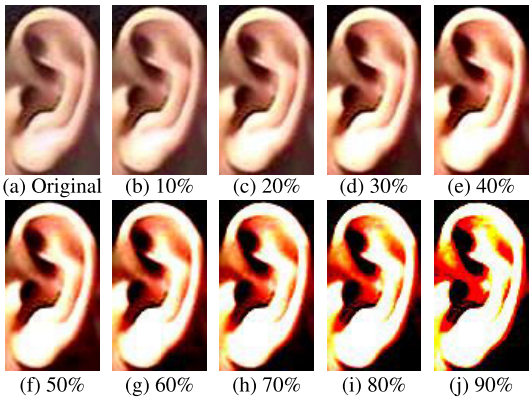


FIGURE 3. Contrast changes, where the percentage of the ear image intensity values are saturated.

TABLE 5. Ear recognition performance (Rank-1 %) using images where contrast was changed.

	SqueezeNet	GoogLeNet	MobileNet	DenseNet201
Normal	93.00	93.67	96.00	99.00
10% (0.05-0.95)	91.67	90.67	95.33	99.33
20% (0.1-0.9)	73.00	80.00	93.00	97.00
30% (0.15-0.85)	53.00	65.33	79.67	89.00
40% (0.2-0.8)	34.67	43.67	56.67	70.67
50% (0.25-0.75)	22.33	28.33	32.67	46.00
60% (0.3-0.7)	14.67	13.33	18.33	28.00
70% (0.35-0.65)	9.67	6.00	8.67	18.66
80% (0.4-0.6)	7.33	3.33	5.00	12.33
90% (0.45-0.55)	6.00	2.33	3.00	10.33

generated with different levels of contrast alteration. The results show that contrast increments affect the performance of all deep models. The DenseNet model performs the best, the performance with the 10% contrast increase gets better, and then accuracy decreases but remains acceptable with up to 30% of the contrast increase. However, the performance of all models falls fast after the 50% contrast increment, which can result from the clipping in pixel values, leading to image information loss.

TABLE 6. Ear recognition performance (Rank-1 %) using images where brightness was changed.

	SqueezeNet	GoogLeNet	MobileNet	DenseNet
$\gamma = 0.5$	17.67	33.33	71.67	62.67
$\gamma = 0.6$	29.67	58.67	88.67	83.67
$\gamma = 0.7$	52.33	78.33	94.33	97.33
$\gamma = 0.8$	74.67	88.67	95.67	99.33
$\gamma = 0.9$	89.00	93.00	96.67	99.33
Normal	93.00	93.67	96.00	99.00
$\gamma = 1.1$	90.33	91.33	95.00	98.67
$\gamma = 1.2$	77.00	88.67	92.67	99.00
$\gamma = 1.3$	59.00	82.67	88.67	98.00
$\gamma = 1.4$	45.00	77.67	83.00	93.67

Second, we examine the impact of brightness variation of the ear images. Fig. 4 shows sample images generated with different levels of brightness alteration. As the identification results in Table 6 show, the performance of the DenseNet is relatively robust, as well as the performance of

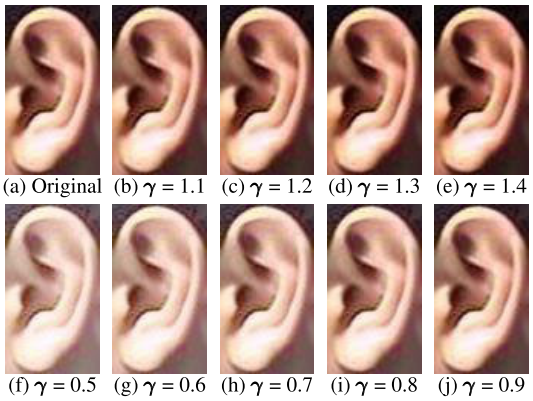


FIGURE 4. Brightness changes, where ear image intensity values are mapped to new values in the output image.

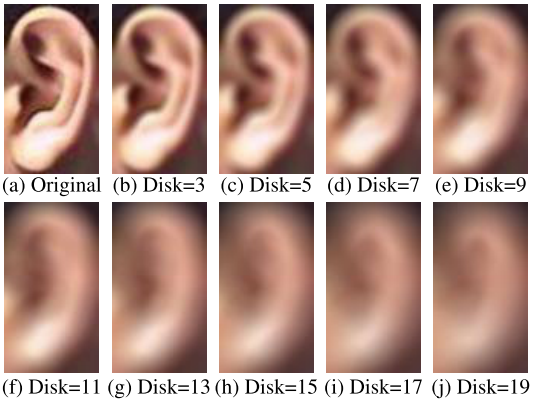


FIGURE 5. Blurring of the input ear images using a circular averaging filter with various diameters.

TABLE 7. Ear recognition performance (Rank-1 %) using images where blurriness intensity was changed.

	SqueezeNet	GoogLeNet	MobileNet	DenseNet201
Normal	95.67	94.67	98.00	99.00
Disk = 3	95.33	93.00	97.33	99.33
Disk = 5	89.33	84.67	92.67	99.33
Disk = 7	79.00	66.67	77.67	93.67
Disk = 9	54.67	45.67	51.00	86.00
Disk = 11	40.67	27.33	30.00	72.33
Disk = 13	26.33	21.33	18.67	58.00
Disk = 15	21.33	18.00	12.00	42.67
Disk = 17	16.33	13.67	8.00	31.67
Disk = 19	13.33	11.67	7.00	25.33

the MobileNet. On the other hand, the SqueezeNet suffers the most deterioration with the alteration of the brightness levels.

Third, the effect of image blurring was explored. Fig. 5 shows sample images generated with different levels of image blurring. As presented in Table 7, the performance is relatively robust with minor blurring, especially for the DenseNet model. With the increase in the radius of the blurriness, the performance of all four CNN models degrades. MobileNet & GoogLeNet performances decline rapidly with the blurring increase of the ear probe images.

V. EAR IMAGE QUALITY

A quality metric in biometrics is a function that takes a biometric sample as its input and returns an estimation of its quality level. Biometric quality measurement should be an indicator of recognition performance. A sample should be of good quality if it is suitable for automated matching. Automatic prediction of biometric quality (prior to matching and recognition) can be useful for several practical applications. A system with the ability to detect poor-quality images can subsequently process them accordingly, as explained in II-A. In this work, we develop an ear image quality assessment. The target quality value is a binary value indicating that an image is good or bad for matching. The process of building a model for automated biometric quality evaluation consists of two main steps:

1) The generation of ear quality ground truth labels for an ear dataset to train the model.
2) Train the model using pairs of images and quality labels to predict the quality of new unseen ear images automatically.

A. GROUND TRUTH LABELS

A quality model's role is to estimate new images' quality. The first step is the generation of quality ground truth labels for the images of the training dataset to serve as a reference for the model. A quality measure should be an indicator of the automated biometric matching performance for an input sample, which can be distinct from the human conception of quality.

The recognition operation is based mainly on comparing the gallery and the probe images. As mentioned in [56], image feature vectors contain quality information as well as identity information. Hence, when the gallery image is of good quality, the gallery/probe comparison outcome can be utilized to generate ground-truth quality labels for probe images. Since the comparison of two bad images can produce high genuine similarity scores, it is essential that the gallery images are of good quality to ensure that low genuine scores are not produced because of the low quality of gallery images [24].

For training, we generated the ground truth labels for the images of the WVU ear dataset. The gallery set consists of the original ear images of the dataset, and the probe set consists of the combination of the original and degraded images. The original dataset was collected with standard high-quality elements for biometric datasets. Also, the ear identification rate for the WVU is 99.00%, as shown in Table 3. Accordingly, we consider the original gallery images as good quality images. The last layer of a CNN is an activation function that gives us a discrete probability distribution over all the classes. We used these values to determine the ground truth quality labels for the training dataset. We used the output from the MobileNet model. The ear images correctly classified to their subject have a high probability value to the correct class. The probe dataset is binned into (good/bad) according to the probability of the input image being classified as its subject

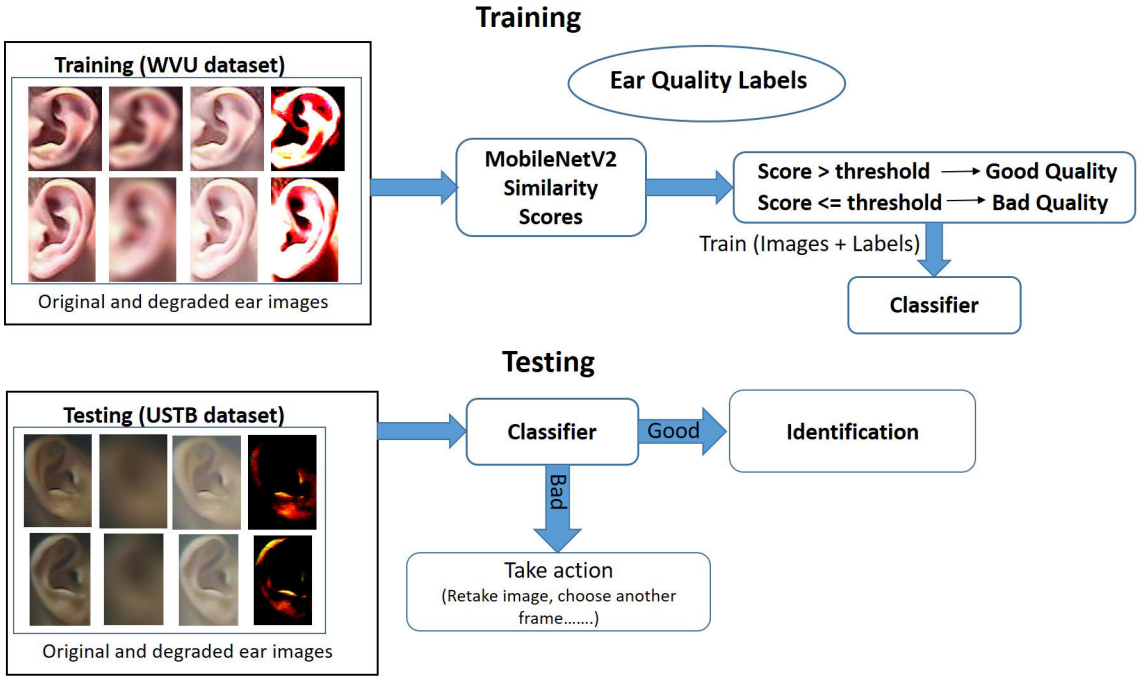


FIGURE 6. An overview of the tool for ear image quality evaluation. We used the scores from the MobileNetV2 model to generate the quality ground truth labels. The dataset is composed of a combination of the original and the degraded images of the WVU ear dataset. Using the training set containing ear images and their respective ground truth quality labels, we trained a classifier(AlexNet) to predict the quality label (good/bad) for an input ear image.

using multiple thresholds. A high threshold is most likely to decrease the recognition errors but will increase the number of samples classified as bad, which may be challenging or inconvenient for some applications. Conversely, decreasing the threshold reduces the number of samples classified as bad but may increase the recognition errors, which may comprise the security. Therefore, according to the application, the user should balance the system's errors and the amount of rejected samples. In our experiments, we examined multiple thresholds (0.8, 0.85, 0.9, and 0.95) to find a suitable threshold for rejecting low-quality samples. That gave us the ground truth labels for the probe set of the dataset that contains original and degraded ear images.

B. QUALITY ESTIMATION AND ASSESSMENT

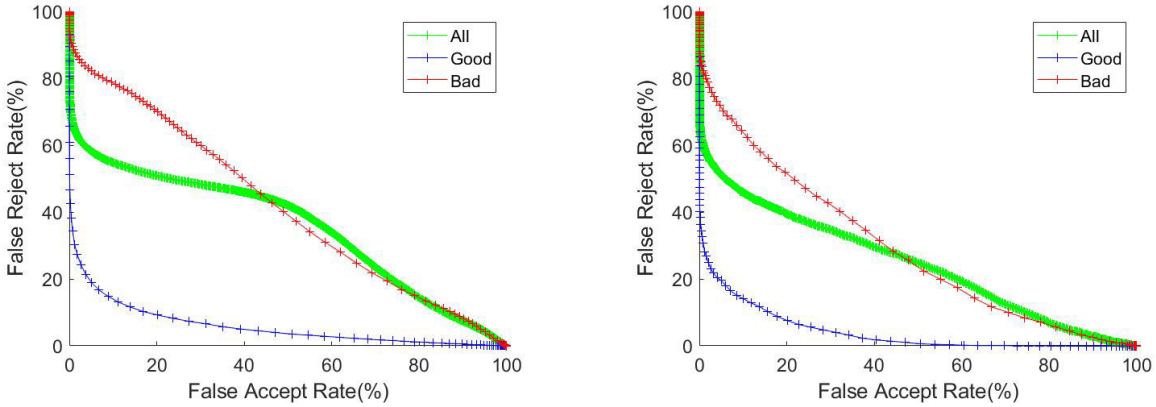
To develop a quality assessment tool for ear images, we used pairs of ear images with their ground truth quality class (good/bad)) generated in V-A. We tuned the pre-trained AlexNet classifier [67] to predict the quality labels for the images in the test dataset. We replaced the last three layers of the pre-trained network (the last fully connected layer, the SoftMax layer, and the final classification layer) to adapt the neural network for the ear image quality prediction task. Since we examined multiple thresholds for binning the matching scores for the training dataset, that gave us four sets of labeled ear images for classifier training. We trained the classifier using pairs of ear images with their quality class. We had four classifiers; each classifier was trained with one of the sets. A higher threshold is expected to keep ear images of high quality but may reject additional input ear images. Fig. 6 presents an overview of the tool for ear image quality evaluation.

To evaluate the efficiency of our ear quality system, we used it to predict the quality of the ear images from the USTB and the FERET ear datasets, which were not used during the training phase. We performed the image degradations to both datasets to expand them that the test dataset consists of a combination of the original standard quality ear images and the degraded quality ear images.

The evaluation methodology for the performance of the ear quality system proposed included two main sets of results:

1) The Receiver Operating Characteristic (ROC) curves.
2) The recognition accuracy for the dataset with and without using the quality labels before performing the recognition.

We used our ear quality evaluation model to assess each image of the test dataset and generate a quality label for it. According to the quality labels of the test samples, we divided the test dataset into three groups: The good quality ear images test set, the bad quality ear images test set, and the total ear images test set, which is the combination of the two other groups. Fig. 7 shows the ROCs curves of the DenseNet for the three test groups. The verification performance improves when using the group of good quality ear images and the correlation between the quality estimation and the verification performance is apparent.



(a) USTB (b) FERET

FIGURE 7. ROC curves obtained with the DenseNet for the quality subsets of data (Bad, Good, and All) of the USTB & the FERET ear datasets.

For the second set of results, we used the recognition accuracy for the dataset with and without using the quality labels before performing the recognition. Table 8 shows the recognition accuracy for the USTB & FERET ear datasets for the whole probe set (original + degraded), and after rejecting the bad probe images according to the quality labels generated using each of the four classifiers with the fraction of probes removed. As mentioned in Table 4, the degraded ear images form 96.67% of the probe test set. The recognition accuracy is close for the multiple classifiers but the accuracy overall increased by 38.53 % and 29.31 % for the USTB and the FERET datasets, respectively. This shows that the quality assessment tool is beneficial, and the quality labels are correlated with the recognition performance.

TABLE 8. Recognition accuracy for the USTB & FERET ear datasets for the whole probe set (original + degraded), and after rejecting the bad probe images using target quality labels generated using each of the four classifiers with the fraction of probes removed.

	USTB		FERET	
	Accuracy	Rejected %	Accuracy	Rejected %
All	58.72%		45.80%	
Classifier1	94.75%	48.85%	73.47%	52.57%
Classifier2	94.70%	48.52%	75.60%	52.00%
Classifier3	96.01%	52.27%	74.44%	52.21%
Classifier4	97.25%	59.09%	75.11%	54.57%

VI. CONCLUSION

In this article, first, we presented the baseline ear recognition (identification/verification) performance using four deep CNN models: SqueezeNet, GoogLeNet, MobileNetV2, and DenseNet. To overcome the limited training data, we used data augmentation, transfer learning, and domain adaptation for the learning process. The DenseNet yielded the highest identification rate of 99.00% and 99.35% for the WVU and the USTB datasets, respectively.

Second, we evaluated the performance of deep ear recognition models in case of image artifacts such as blurriness or degradation in contrast and brightness. The performance of the studied models was affected by image quality to different degrees. The DenseNet model was the most robust, followed by the MobileNetV2 model, then GoogLeNet and SqueezeNet models. The limited brightness, contrast, and blur alteration resulted in slight degradation, but the performance declined with significant artifacts.

Finally, we provided a tool for automatically detecting low-quality ear images. Detection of low-quality ear images has advantages. It prevents spoofing, recommends re-capture, or initiates sample preprocessing. The proposed approach uses a CNN classifier model to automatically predict ear quality before matching. We performed several experiments on extended degraded ear datasets. The results show that the proposed tool can predict low-quality ear images and improve ear recognition performance. It increased the recognition performance by 38.53 % and 29.31 % for the USTB and the FERET degraded datasets, respectively.

As for future work, we plan to study other quality factors that affect ear recognition performance (such as image compression and spatial resolution) on synthetic and real-world data. Further, we plan to examine different quality measures and use them to alternate the ear recognition models, i.e., changing the ear recognition model based on the detected artifact.

REFERENCES

[1] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognit. Lett.*, vol. 79, pp. 80–105, Aug. 2016.

[2] M. Ngan, P. Grother, and K. Hanaoka, "Ongoing face recognition Vendor test (FRVT) part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms," U.S. Dept. Commerce, Nat. Inst.of Standards Technol., Tech. Rep., 2020.

[3] D. Hurley, M. Nixon, and J. Carter, "Force field feature extraction for ear biometrics," *Comput. Vis. Image Understand.*, vol. 98, no. 3, pp. 491–512, Jun. 2005.

[4] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1297–1308, Aug. 2007.

[5] A. Ross and A. Abaza, "Human ear recognition," *Computer*, vol. 44, no. 11, pp. 79–81, 2011.
[6] S. El-Naggar, A. Abaza, H. Ammar, and T. Bourlai, "On a taxonomy of ear features," in *Proc. IEEE Symp. Technol. Homeland Secur. (HST)*, Waltham, MA USA, May 2016, pp. 1–6.
[7] H. Nejati, L. Zhang, T. Sim, E. Martinez-Marroquin, and G. Dong, "Wonder ears: Identification of identical twins from ear images," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, 2012, pp. 1201–1204.
[8] A. Jain, A. Ross, and K. Nandakumar, "Introduction to biometrics," in *Biometrics*. Boston, MA, USA: Springer, 2011.
[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.
[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[11] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2012.
[12] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1725–1732.
[13] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8599–8603.
[14] D. Yu and L. Deng, *Automatic Speech Recognition*, vol. 1. Berlin, Germany: Springer, 2016.
[15] K. Noda, Y. Yamaguchi, K. Nakadai, H. G. Okuno, and T. Ogata, "Audio-visual speech recognition using deep learning," *Appl. Intell.*, vol. 42, no. 4, pp. 722–737, Oct. 2015.
[16] S. Zhang, S. Zhang, T. Huang, and W. Gao, "Multimodal deep convolutional neural network for audio-visual emotion recognition," in *Proc. ACM Int. Conf. Multimedia Retr.*, 2016, pp. 281–284.
[17] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.
[18] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis.*, vol. 1, no. 3, 2015, p. 6.
[19] L. Jiang, T. Zhao, B. Tong, Y. Chaochao, and M. Wu, "A direct fingerprint minutiae extraction approach based on convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2016, pp. 571–578.
[20] Z. Zhao and A. Kumar, "Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 5, pp. 1017–1030, May 2017.
[21] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Mar. 2017.
[22] S. El-Naggar and T. Bourlai, "Evaluation of deep learning models for ear recognition against image distortions," in *Proc. Eur. Intell. Secur. Inform. Conf. (EISIC)*, 2019, pp. 85–93.
[23] P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 531–543, Apr. 2007.
[24] L. Best-Rowden and A. K. Jain, "Learning face image quality from human assessments," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 12, pp. 3064–3077, Dec. 2018.
[25] A. Bertillon and R. McClaughry, *Signaletic Instructions Including the Theory and Practice of Anthropometrical Identification*. Itasca, IL, USA: Werner Company 1896.
[26] A. V. Iannarelli, *Ear Identification*. Hyderabad, Telangana, Paramont, 1964.
[27] M. Zarachoff, A. Sheikh-Akbari, and D. Monekosso, "2D multi-band PCA and its application for ear recognition," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2018, pp. 1–5.
[28] Z. Youbi, L. Boubchir, and A. Boukrouche, "Human ear recognition based on local multi-scale LBP features with city-block distance," *Multimedia Tools Appl.*," vol. 78, no. 11, pp. 14425–14441, Jun. 2019.
[29] A. Pflug, P. Paul, and C. Busch, "A comparative study on texture and surface descriptors for ear biometrics," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, 2014, pp. 1–6.
[30] P. Sarangi, B. Mishra, and S. Dehuri, "Ear recognition using pyramid histogram of orientation gradients," in *Proc. 4th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, 2017, pp. 590–595.
[31] S. El-Naggar, A. Abaza, and T. Bourlai, "A study on human recognition using auricle and side view face images," *Surveillance in Action*, P. Karampelas and T. Bourlai, Eds. Springer, 2018, pp. 77–104.
[32] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Comput. Surv.*, vol. 45, no. 2, p. 22, Feb. 2013.
[33] Ž. Emeršić, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, Sep. 2017.
[34] A. Benzaoui, I. Adjabi, and A. Boukrouche, "Experiments and improvements of ear recognition based on local texture descriptors," *Opt. Eng.*, vol. 56, no. 4, 2017, Art. no. 043109.
[35] I. Omara, X. Wu, H. Zhang, Y. Du, and W. Zuo, "Learning pairwise SVM on hierarchical deep features for ear recognition," *IET Biometrics*, vol. 7, no. 6, pp. 557–566, Nov. 2018.
[36] Z. Wang, J. Yang, and Y. Zhu, "Review of ear biometrics," *Arch. Comput. Methods Eng.*, vol. 45, no. 5, pp. 1–32, 2019.
[37] Y. Zhang and Z. Mu, "Ear detection under uncontrolled conditions with multiple scale faster region-based convolutional neural networks," *Symmetry*, vol. 9, no. 4, p. 53, 2017.
[38] S. El-Naggar, A. Abaza, and T. Bourlai, "Ear detection in the wild using faster R-CNN deep learning," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 1124–1130.
[39] I. I. Ganapathi, S. Prakash, I. R. Dave, and S. Bakshi, "Unconstrained ear detection using ensemble-based convolutional neural network model," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 1, p. e5197, 2020.
[40] A. Kamboj, R. Rani, A. Nigam, R. R. Jha, and R. Ranjan, "CED-Net: Context-aware ear detection network for unconstrained images," *Pattern Anal. Appl.*, vol. 24, no. 2, pp. 779–800, 2021.
[41] V. Z. Emeršič, D. Sušanj, B. Meden, P. Peer, and V. Štruc, "ContexedNet: Context–aware ear detection in unconstrained settings," *IEEE Access*, vol. 9, pp. 145175–145190, 2021.
[42] V. Z. Emeršič, D. Štepec, V. Štruc, and P. Peer, "Training convolutional neural networks with limited training data for ear recognition in the wild," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2017, pp. 987–994.
[43] V. Z. Emeršič, B. Meden, P. Peer, and V. Štruc, "Evaluation and analysis of ear recognition models: Performance, complexity and resource requirements," *Neural Comput. Appl.*, vol. 32, no. 20, pp. 15785–15800, 2020.
[44] V. Z. Emeršič, D. Štepec, V. Štruc, P. Peer, A. George, A. Ahmad, E. Omar, T. E. Boult, R. Safdari, Y. Zhou, S. Zafeiriou, D. Yaman, F. I. Eyiokur, and H. K. Ekenel, "The unconstrained ear recognition challenge," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 715–724.
[45] Ž. Emeršič, "The unconstrained ear recognition challenge 2019," in *Proc. Int. Conf. Biometrics*, Crete, Greece, 2019, pp. 4–7.
[46] F. I. Eyiokur, D. Yaman, and H. K. Ekenel, "Domain adaptation for ear recognition using deep convolutional neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 199–206, 2017.
[47] S. Dodge, J. Mounsef, and L. Karam, "Unconstrained ear recognition using deep neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 207–214, 2018.
[48] H. Mehraj and A. H. Mir, "Human recognition using ear based deep learning features," in *Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI)*, 2020, pp. 357–360.
[49] Y. Zhang, Z. Mu, L. Yuan, and C. Yu, "Ear verification under uncontrolled conditions with convolutional neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 185–198, 2018.
[50] Y. Khaldi, A. Benzaoui, A. Ouahabi, S. Jacques, and A. Taleb-Ahmed, "Ear recognition based on deep unsupervised active learning," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20704–20713, Sep. 2021.
[51] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Deep convolutional neural networks for unconstrained ear recognition," *IEEE Access*, vol. 8, pp. 170295–170310, 2020.
[52] H. Alshazly, C. Linse, E. Barth, S. Idris, and T. Martinetz, "Towards explainable ear recognition systems using deep residual networks," *IEEE Access*, vol. 9, pp. 122254–122273, 2021.
[53] D. Meng, M. Nixon, and S. Mahmoodi, "On distinctiveness and symmetry in ear biometrics," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 2, pp. 155–165, Feb. 2021.
[54] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biometrics*, vol. 3, no. 4, pp. 314–324, Dec. 2014.

[55] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "FaceQnet: Quality assessment for face recognition based on deep learning," in *Proc. Int. Conf. Biometrics (ICB)*, 2019, pp. 1–8.
[56] J. Hernandez-Ortega, J. Galbally, J. Fierrez, and L. Beslay, "Biometric quality: Review and application to face recognition with FaceQnet," 2020, *arXiv:2006.03298*.
[57] A. Pflug, J. Wagner, C. Rathgeb, and C. Busch, "Impact of severe signal degradation on ear recognition performance," in *Proc. 37th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, 2014, pp. 1342–1347.
[58] G. Fahmy, A. El-Sherbeeny, S. Mandala, M. Abdel-Mottaleb, and H. Ammar, "The effect of lighting direction/condition on the performance of face recognition algorithms," in *Proc. SPIE*, Orlando, FL, USA, 2006, pp. 188–200.
[59] *University of Science and Technology Beijing USTB Database*. [Online]. Available: http://www1.ustb.edu.cn/resb/en/index.htm
[60] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.
[61] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
[62] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.
[63] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2015, pp. 1–9.
[64] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks title," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[65] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.
[66] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[67] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, Dec. 2012, pp. 1097–1105.

SUSAN EL-NAGGAR received the B.Sc. and M.Sc. degrees in systems and biomedical engineering from Cairo University, Cairo, Egypt, in 2000 and 2005, respectively. She is currently pursuing the Ph.D. degree with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown. Her current research interests include image processing, deep learning, and ear biometrics.

THIRIMACHOS BOURLAI is an Associate Professor with the School of Electrical and Computer Engineering, University of Georgia. He is the Founder and the Director of the Multi-Spectral Imagery Laboratory. He has published four books, three patents, and more than 120 publications. He is a member of the Board of Directors at the Document Security Alliance, the former VP on Education of the IEEE Biometrics Council, the Executive Secretary of the FI Subcommittee at NIST, and a member of the Academic Research and Innovation Expert Group of the Biometrics Institute. He is a Springer Nature Series Editor of the *Advanced Sciences and Technologies for Security Applications*.