symmetry

MDPI

Article

A Feature Fusion Human Ear Recognition Method Based on Channel Features and Dynamic Convolution

Xuebin Xu [1,2], Yibiao Liu [1,2,*], Chenguang Liu [1,2] and Longbin Lu [1,2]

[1] School of Computer Science and Technology, Xi'an University of Posts & Telecommunications, Xi'an 710121, China
[2] Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts & Telecommunications, Xi'an 710121, China
* Correspondence: liuyibiao@stu.xupt.edu.cn

Abstract: Ear images are easy to capture, and ear features are relatively stable and can be used for identification. The ear images are all asymmetric, and the asymmetry of the ear images collected in the unconstrained environment will be more pronounced, increasing the recognition difficulty. Most recognition methods based on hand-crafted features perform poorly in terms of recognition performance in the face of ear databases that vary significantly in terms of illumination, angle, occlusion, and background. This paper proposes a feature fusion human ear recognition method based on channel features and dynamic convolution (CFDCNet). Based on the DenseNet-121 model, the ear features are first extracted adaptively by dynamic convolution (DY_Conv), which makes the ear features of the same class of samples more aggregated and different types of samples more dispersed, enhancing the robustness of the ear feature representation. Then, by introducing an efficient channel attention mechanism (ECA), the weights of important ear features are increased and invalid features are suppressed. Finally, we use the Max pooling operation to reduce the number of parameters and computations, retain the main ear features, and improve the model's generalization ability. We performed simulations on the AMI and AWE human ear datasets, achieving 99.70% and 72.70% of Rank-1 (R1) recognition accuracy, respectively. The recognition performance of this method is significantly better than that of the DenseNet-121 model and most existing human ear recognition methods.

Keywords: ear recognition; dynamic convolution; ECA; max pooling; asymmetric

Citation: Xu, X.; Liu, Y.; Liu, C.; Lu, L. A Feature Fusion Human Ear Recognition Method Based on Channel Features and Dynamic Convolution. *Symmetry* **2023**, *15*, 1454. https://doi.org/10.3390/sym15071454

Academic Editor: Christos Volos

Received: 7 June 2023
Revised: 14 July 2023
Accepted: 15 July 2023
Published: 21 July 2023

1. Introduction

Human ear recognition is a biometric technology that emerged at the end of the last century. It has unique physiological characteristics and viewing angles [1]. This gives ear recognition technology natural advantages compared with other biometric technologies. Currently, relatively mature biometric technologies include face recognition, fingerprint recognition, iris recognition, etc. [2]. Among them, face recognition is influenced by various factors, such as changes in facial expression, whether or not to wear glasses, and whether or not to have a beard. In contrast, ear recognition is almost independent of these factors [3]. Acquiring human ear image information is much easier than receiving fingerprint information. This is because it can be collected secretly without a person's cooperation. Compared to iris recognition, the installation cost of ear image capture devices is relatively low. Moreover, acquiring iris information is more complex than ear image acquisition. Therefore, human ear recognition technology can be applied in many fast-paced identity identification scenarios. Although there is much research on human ear recognition at home and abroad, the technology could be more mature, and there is still a long way to go before it can be applied to real life. In-depth research on this technology can actively promote and improve contactless remote identification. The explosion of

COVID-19 worldwide in the past three years has affected many biometric systems. For example, facial recognition will be severely impacted by people wearing masks. At this time, ear recognition can benefit identity confirmation [4]. In addition, it performs well in financial and surveillance security [5].

Computer vision and machine learning techniques have been significantly developed in recent years. Among them, deep convolutional neural networks have been popular among most researchers and applied to almost all areas of computer vision, especially ear recognition tasks. Deep convolutional neural networks have the feature of fusing feature extraction and classification into an end-to-end model that can handle different practical problems by learning the representations of the input data. Most ear recognition methods based on hand-crafted features do not use standard performance evaluation metrics and baseline ear databases, and the variation in the collected subject ear images is slight. When these methods are confronted with an ear database with significant asymmetry in an unconstrained environment, the recognition performance is significantly worse than that of deep learning-based approaches. In deep feature extraction methods, the parameters of static convolution are artificially set and fixed, which can reduce the extraction effect of ear image features. However, dynamic convolution [6] can dynamically aggregate multiple parallel convolution kernels to adaptively adjust the convolution parameters to further refine ear features. The ECA [7] module can realize cross-channel information interaction, suppress invalid features, and improve the feature weights of the ear geometry region. The dynamic convolution and ECA modules can significantly enhance the feature representation ability of the model, which has shown excellent performance in the fields of CIFAR and ImageNet database classification [6–10], scene recognition [10], ancient Chinese character recognition [11], fine-grained image classification [12], and plant disease recognition [13]. Therefore, we propose a feature fusion human ear recognition method based on channel features and dynamic convolution (CFDCNet).

Our contributions can be summarized as follows: (1) a feature fusion human ear recognition method based on channel features and dynamic convolution [6] is proposed, which has good recognition performance in both constrained and unconstrained ear recognition scenarios; (2) in the case of significant differences in ear sample features between the same category and different categories. This paper introduces dynamic convolution to extract ear image features adaptively, enhancing the robustness of ear feature representation; (3) an ECA mechanism [7] is introduced to efficiently fuse the depth and spatial information of ear images and suppress invalid features such as background and noise; (4) we utilize the maximum pooling operation in the network to retain the primary feature information of the ear contour to the maximum extent and prevent the model from overfitting; and (5) we performed simulations on AMI [14] and AWE [15–17] human ear databases and achieved 99.70% and 72.70% Rank-1 (R1) recognition accuracy, respectively. The recognition performance of our method is significantly better than that of the DenseNet-121 [18] model and most existing human ear recognition methods.

The rest of this paper is organized as follows. Section 2 briefly reviews past work; Section 3 discusses our proposed method; Section 4 discusses the experimental results and analysis; and Section 5 presents a conclusion.

2. Related Work

Earlier researchers performed ear identification based on handcrafted features. In [19], the author used Haar wavelets for ear localization. Their method has good robustness against occlusions, and the recognition performance is significantly better than Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Orthogonal Locality Preserving Projection (OLPP). The drawback of this method is that it was evaluated on small datasets, and no fixed-value performance evaluation metrics were used to assess it. In [20], the authors proposed an ear recognition method combining homographic distance and Scale-invariant feature transform (SIFT) features. The method outperforms PCA in recognition and shows robustness to slight angle changes, background noise, and

occlusions. The drawback of their method is that it does not use a benchmark database or specify evaluation metrics to assess the model's performance. In [21], the authors first segmented the ear features using Fourier and morphological descriptors and then used log-Gabor, Gabor, and complex Gabor filters for local ear feature extraction. The method was evaluated on a private database containing 465 ear images. The results show that log-Gabor has the best feature extraction performance. The disadvantage is that no exact performance evaluation metrics or benchmark database were used. In [22], the authors proposed an ear recognition method using 2D orthogonal filters for ear feature extraction. It was evaluated on the IITD and UND ear databases. The method shows that the 2D orthogonal filter performs better than others. The disadvantage is that the database used to evaluate the method slightly varies. In [23], the authors first localize the ear information using the snake model and then use geometric features for ear identification. The model was evaluated on the IIT Delhi ear database. The drawback of their method is that it was validated only on a small database, and the ear images in this database were collected indoors with slight variation. In [24], the authors use a robust pattern recognition technique for human ear recognition. The method uses descriptors for ear feature extraction, and the extracted features are powerful for rotation and illumination. The authors tested it on AMI [14], IITD-II, and AWE [15–17] databases, and the recognition performance is significantly better than other descriptor methods. The disadvantage is that it needs better recognition performance on unconstrained datasets. In [25], the authors first extracted the local features of the ear using the local phase quantization operator, then removed the global features of the ear using the Gabor–Zernike operator, and finally put the optimal features of the ear together using a genetic algorithm. The recognition performance of this method evaluated on three constrained databases is ideal, but on unconstrained databases, the recognition performance is lower than that of the deep learning-based method.

Researchers have found some application-specific scenarios with high-security index requirements that require the combination of multiple biometrics, so they started to utilize multimodal approaches for ear recognition. In [26], the authors proposed a multi-modal biometric technique combining the ear and iris. They used a local feature descriptor, SIFT, for feature fusion. It was evaluated on the USTB-II ear database and the CASIA iris database. According to accuracy, the method is more accurate than ear biometric recognition alone. In [27], the authors propose a multimodal recognition system combining side faces and ears. They first augmented the images in the database, then obtained the local optima of the pictures using the Hessian matrix, and finally used Speeded Up Robust Features (SURF) to construct the scale space and localize the image feature points. The results of this method on three ear and side face databases show that multimodal recognition of ears and side faces performs better than ear recognition alone. In [28], the authors used ears and fingerprints for multi-pattern recognition. Local Binary Patterns (LBP) were used to extract the local texture features of the images. The system achieved an accuracy of 98.10%. The drawback is that they did not evaluate the system with a benchmark database.

In recent years, ear recognition methods based on deep feature learning have achieved good human ear feature recognition results. In [29], the authors used a Convolutional Neural Network (CNN) consisting of convolutional, maximum pooling, and fully connected layers for ear feature extraction. The evaluation was performed on the USTB-III ear database. The disadvantage is that the method does not use standard evaluation parameters, and the database used for the assessment is constrained and small in number. In [30], the authors fine-tune the CNN frameworks of VGG face, VGG, ResNet, AlexNet, and GoogleNet to perform ear recognition. To enable the network to learn multi-scale information, the last pooling layer of each CNN model is replaced with a spatial pyramid pooling layer. A combination of softmax and center loss is used for training. The authors also created an unconstrained ear dataset called USTB HelloEar. The results show that the VGG face model has the best recognition performance. The drawback of this method is that performance evaluation metrics are not used to evaluate the model. In [31], the authors first used Refinet for ear detection and then hand-crafted feature-based and

ResNet models for ear recognition. The models were tested on the UERC database, and the recognition performance of the deep learning-based approach was significantly better than that of the hand-crafted feature-based approach. The disadvantage of the model is that the novelty could be better, and the ear detection and recognition are performed using existing models. In [32], the authors used integrated learning, feature extraction, and other learning strategies for ear recognition based on network models such as Inception, ResNext, and VGG. They evaluated the model's performance by resizing the image input network and achieved good recognition results on the EarVN1.0 unconstrained ear database. The drawback is that it was tested on only one dataset and not compared with other human ear recognition techniques. A CNN model that can be used for ear recognition was designed in [2]. It was evaluated on the AMI and IITD-II databases. The authors did not use standard performance evaluation metrics, and the database used was constrained, with slight variation in the ear images. In [33], ear recognition is performed with the NASNET model, and the performance is compared with MobileNet, VGG, and ResNet. The method was evaluated on the UERC-2017 unconstrained ear database and achieved the best recognition performance.

It is worth mentioning that most recognition methods based on hand-crafted features exhibit poor recognition performance in the face of human ear datasets with highly variable illumination, angle, occlusion, and background. Therefore, we propose a feature fusion human ear recognition method based on channel features and dynamic convolution (CFDCNet). Based on the DenseNet-121 [18] model, the robustness of ear feature representation is enhanced by replacing the original convolutional layer with dynamic convolution [6] for adaptive extraction of ear image features. Then the weights of the important ear features are increased by an efficient channel attention mechanism (ECA) [7]. Finally, we improved the model's generalization ability by using the maximum pooling operation to retain the ear's key features. We evaluated our model on two publicly available ear datasets exhibiting good recognition performance.

3. The Proposed Approach

3.1. Introduction to CFDCNet

Figure 1 is the overall structure diagram of CFDCNet, and the input is the ear image of the R, G, and B channels. The model mainly consists of 9 ECA blocks, 58 modified dense layers, and three modified transition layers. We can divide the model into the following parts: a shallow feature extraction block (SFE block), four deep feature extraction blocks (DFE block), a Max pooling layer, and a classification layer. Figure 2 shows the SFE block and the first DFE block.

First, the ear image is adjusted by a 7×7 convolutional layer in the SFE block and a 3×3 Max pooling layer to change the number of channels and extract the practical information. Then, the depth and spatial information of the ear image are efficiently fused by the ECA mechanism to suppress invalid features such as background and noise and achieve shallow feature extraction of the ear image. The previously extracted data features are used to extract the depth features of the ear image through four DFE blocks. The DFE blocks extract ear features adaptively through the modified dense and transition layers, increase the weight of essential features, make the ear features of the same category of samples more aggregated and different categories more dispersed, and enhance the robustness of the ear feature representation. Finally, the Max pooling layer retains the primary feature information of ear contour to prevent overfitting [34–38], and a linear classification layer is used to achieve ear image classification [18].

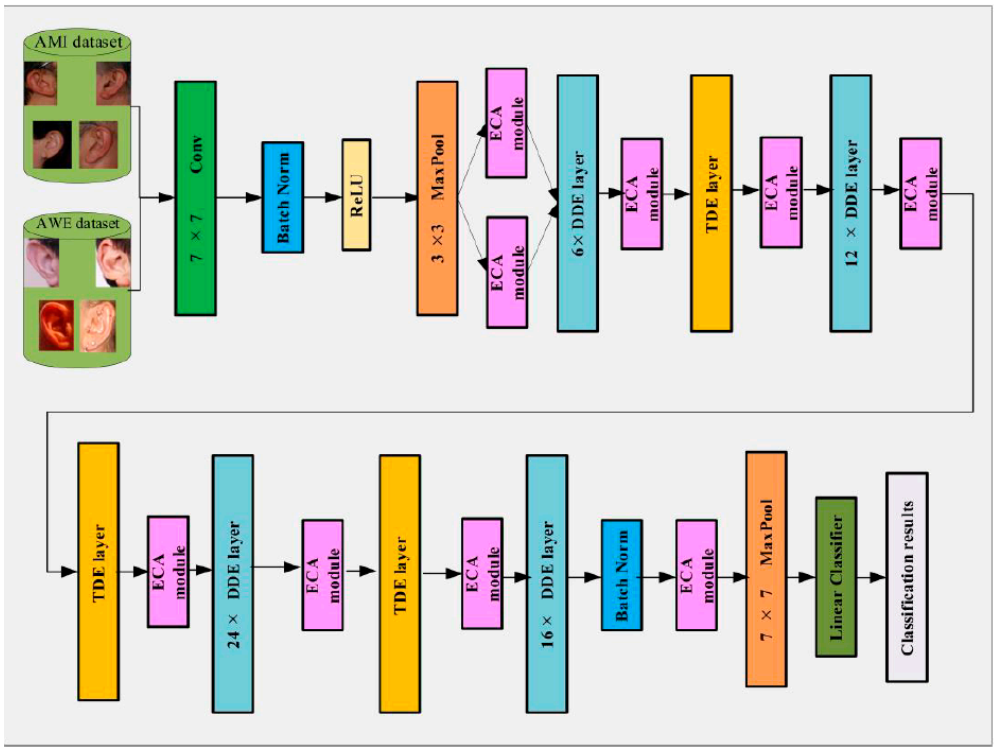


Figure 1. CFDCNet overall architecture diagram.

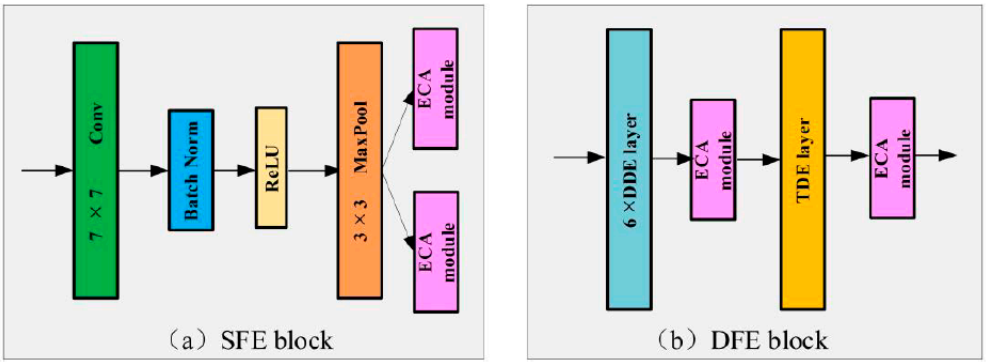
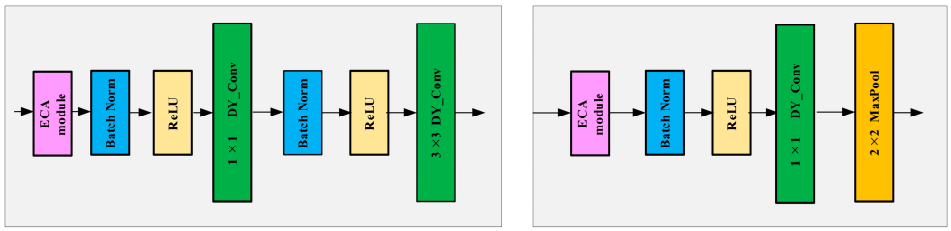


Figure 2. Schematic diagram of the SFE block and the first DFE block structure.

3.2. DDE Layer and TDE Layer

We improved the dense and transition layers. By introducing the ECA module, the main features of the ear image are preserved, and the unfavorable features, such as background and noise, are ignored. The robustness of ear feature representation is enhanced by dynamic convolution, which makes features of ear samples of the same category more aggregated and ear features of different types more dispersed. Figure 3a shows the dense layer fused with dynamic convolution and the ECA mechanism (DDE layer). We add an ECA module before each dense layer and replace the original 1×1 convolutional layer and 3×3 convolutional layers with 1×1 dynamic convolutional layer and 3×3 dynamic convolutional layers. Figure 3b shows the transition layer fused with dynamic convolution and the ECA mechanism (TDE layer). We added an ECA module before each transition layer and replaced the original 1×1 convolutional layer with a 1×1 dynamic convolutional layer. To reduce the parameters and computation of the model, prevent overfitting, and improve the model's generalization ability. We changed the 2×2 Avg pooling layer in the original network transition layer to the 2×2 Max pooling layer. The 7×7 Global Max pooling is used to replace the 7×7 Global Average pooling before the linear classifier.

The specific impact of the pooling method on the simulation has been given in Table 3 in Section 4.



(a) DDE layer (b) TDE layer

Figure 3. Structure diagram of the DDE layer and the TDE layer.

3.3. DenseNet-121

The connection mechanism of the residual network [39] is a short-circuit connection between each layer and one or two layers in front of it through element-by-element addition, i.e., Equation (1):

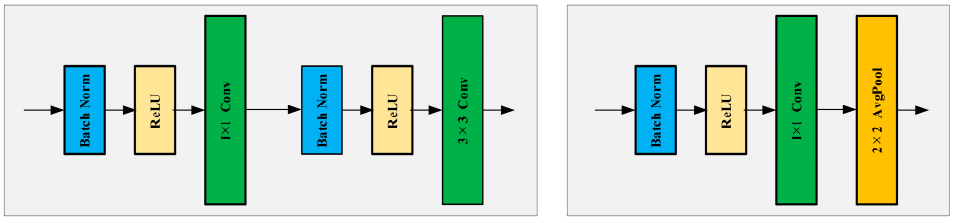
$$P_z = Q_z(P_{z-1}) + P_{z-1} \tag{1}$$

In contrast, DenseNet [18] is layer-to-layer interconnected, where each layer accepts as input the feature information of all previous layers and connects the feature maps of each layer in the channel dimension, i.e., Equation (2):

$$P_z = Q([P_0, P_1, \cdots, P_{z-1}]) \tag{2}$$

where $Q_z(\cdot)$ is the nonlinear transformation function, which is a combined operation containing the batch normalization (BN), the rectified linear unit (ReLU), and the convolution layer, and z is the index of the layer. P_z represents the output of the z layer and $[P_0, P_1, \cdots, P_{z-1}]$ represents the feature map spliced from the 0 layers to the $(z-1)$ layer.

The DenseNet-121 network structure is mainly composed of multiple densely connected blocks (Dense blocks) and Transition layer composition, and each Dense block is composed of multiple dense layers. The Dense layer and Transition layer before improvement are shown in Figure 4.



(a) Dense layer (b) Transition layer

Figure 4. Structure diagram of the Dense layer and Transition layer before improvement.

3.4. Dynamic Convolution

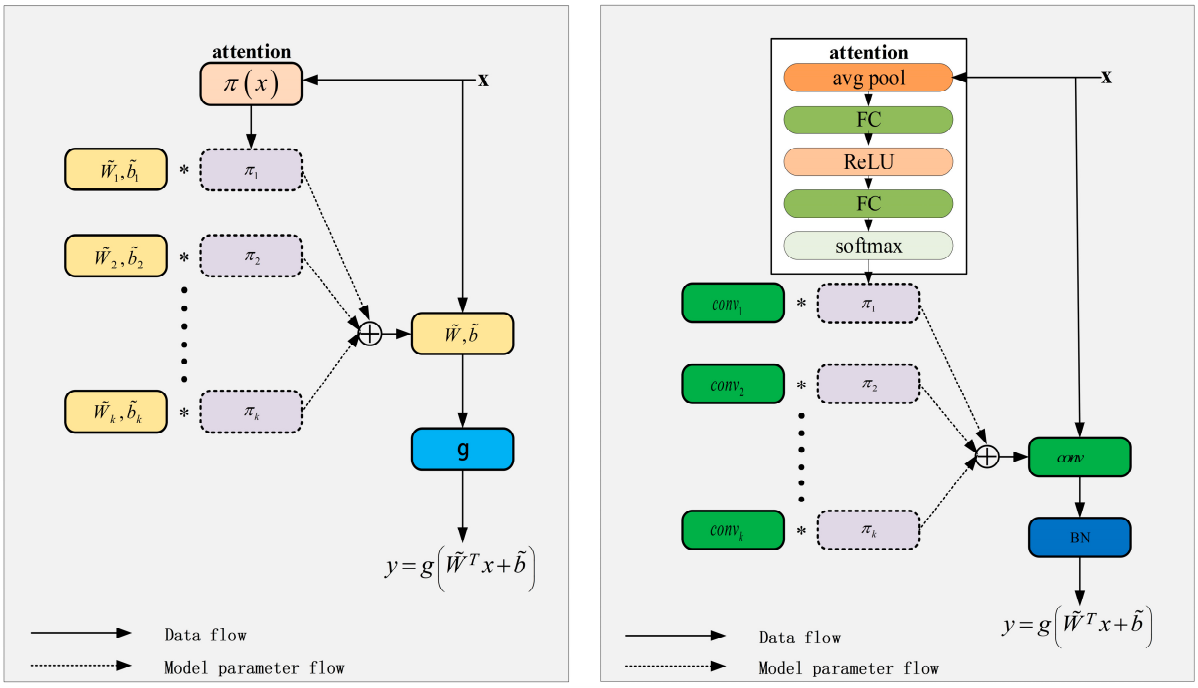
Due to the variety of image types in actual situations, there will be significant differences between samples of different categories and even between samples of the same type. This phenomenon is prevalent in the unconstrained ear dataset. Static convolution usually uses a single convolution kernel to perform the same operation on all input images, making it difficult to perform accurate feature extraction on images. Equations (3) and (4) list the conventional convolutional and dynamic convolutional [6], respectively. In (3) W and b are the weight matrix and bias vector and g is the activation function. In (4), π_k is the attention weight of the k^{th} linear function $\hat{W}_k^T x + \hat{b}_k$. As the input x changes, the attention weight $\pi_k(x)$ also changes, resulting in an optimal aggregation of linear models $\left\{\hat{W}_k^T x + \hat{b}_k\right\}$ [6].

However, the aggregation model $\hat{W}^T(x)x + \hat{b}(x)$ is a nonlinear function, so the dynamic perceptron has a more robust feature representation ability than the static perceptron [6].

$$\begin{aligned} y &= g\left(W^T x + b\right), \\ y &= g\left(\hat{W}^T x + \hat{b}\right), \end{aligned} \tag{3}$$

$$\begin{aligned} \hat{W} = \sum_{k=1}^{K} \pi_k(x)\hat{W}_k,\ \hat{b} = \sum_{k=1}^{K} \pi_k(x)\hat{b}_k, \\ s.t. 0 \le \pi_k(x) \le 1,\ \sum_{k=1}^{K} \pi_k(x) = 1 \end{aligned} \tag{4}$$

With the deepening of the network depth, the image resolution and feature details after the deep network will suffer considerably. When training many images, more advanced general rules can be obtained, and the feature extraction mode of each image sample can be obtained through multi-layer combination adjustment. The basic idea of dynamic convolution is to adaptively adjust the convolution parameters of the input image according to attention and dynamically aggregate multiple parallel convolution kernels. The convolution kernels after aggregation are small in size and computationally efficient, and they are aggregated in a non-linear manner through attention. They are more capable of feature representation. Figure 5 shows a dynamic perceptron and a dynamic convolution layer (DY_Conv).



(a)Dynamic perceptron **(b)Dynamic convolution layer(DY_Conv)**

Figure 5. Structure diagram of a dynamic perceptron and a dynamic convolution layer (DY_Conv). The asterisk represents multiplication.

3.5. Efficient Channel Attention Mechanism

Efficient Channel Attention (ECA) [7], used in this paper, is a lightweight module. It can suppress the ear image's background, noise, and other invalid features. In addition, this mechanism can efficiently fuse the depth and spatial information of the ear image and focus on extracting the main features of the ear, thus improving the recognition accuracy of the ear image. Figure 6 shows the structure diagram of Efficient Channel Attention (ECA). It first performs a global average pooling of the input feature maps with a value representing each channel's feature layer. Then a one-dimensional convolution of size three is used to

generate weights for each channel to obtain the interdependencies between each channel and normalize them using a Sigmoid activation function. Finally, the weights generated for each channel are multiplied by the input feature map to enhance the extraction of essential features in the ear. In Figure 6, *C* is the number of channels, H is the height of the input data, *W* is the width of the input data, GAP is the global average pool, *K* is the one-dimensional convolutional cross-channel interaction size (*K* = 3 in this paper), and σ is the Sigmoid activation function.

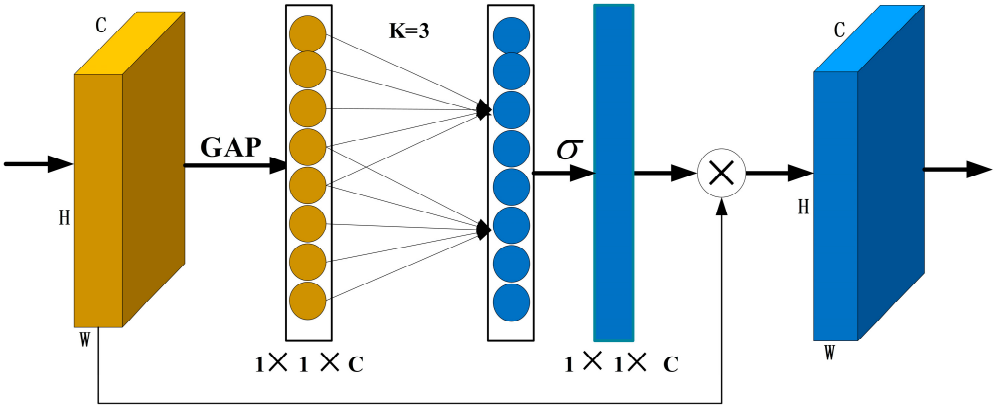


Figure 6. Efficient Channel Attention (ECA) module structure diagram.

ECA generates weights for each channel through one-dimensional convolutional cross-channel interactions of size *K*, i.e.,

$$\omega = \delta(CID_K(y)) \tag{5}$$

ω is the channel weight, δ is the sigmoid Activation function, *CID* is the one-dimensional convolution, and $y \in \mathbb{R}^C$ is the aggregation feature. As the number of input feature map channels *C* increases, the *K* value rises, i.e.,

$$C = 2^{(\gamma * K - b)} \tag{6}$$

In this paper, the value of *K* is determined adaptively by a function related to the number of channel dimensions, i.e.,

$$K = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{7}$$

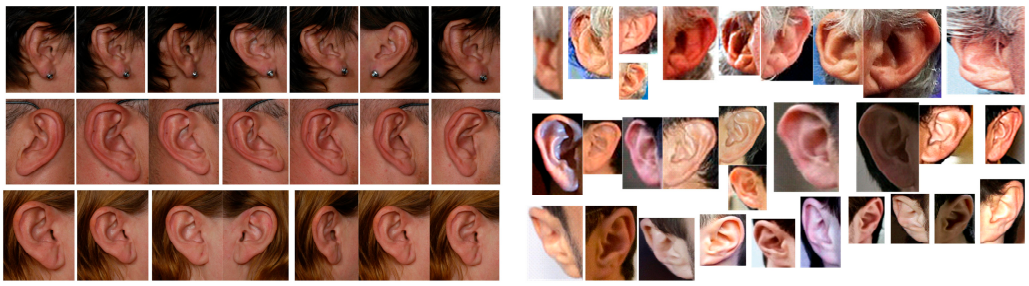
In the equation, $|t|_{odd}$ is the odd number closest to t, γ and b are set to 2 and 1, respectively.

4. Simulation Results

4.1. Dataset Introduction

To evaluate the model's performance, we use two human ear datasets in our simulations. The first dataset is the AMI human ear dataset [14]. A total of 100 subjects' ear images were collected; each person has six right ear images and one left ear image, and the total number of images is 700. The five images of the right ear slightly change the shooting angle. The sixth image of the right ear shows the subject looking forward but with a different focus. The last is an image of the left ear, with the subject looking forward. The dataset was taken indoors with a Nikon D100 camera under constant lighting conditions. All images have a resolution of 492 × 702 pixels. It is a constrained dataset. In Figure 7a, we randomly selected the ear images of three subjects. The second dataset is the AWE Human Ear Database [15–17], which is one of the most challenging datasets for ear recognition. Most of its images are collected from the Internet, including 100 subjects; each subject has

ten images, and the total number of images is 1000. This dataset's image resolution, head angle, and posture vary greatly, and the lighting conditions are also different. The left and right ears are distinguished, the image contrast is poor, and even earrings, accessories, and hair in individual images will cause severe occlusion. In short, significant changes in the same category and between different categories in this dataset dramatically increase the difficulty of identification. This dataset is a typical unconstrained dataset, and in Figure 7b, we randomly select the ear images of three subjects for display.

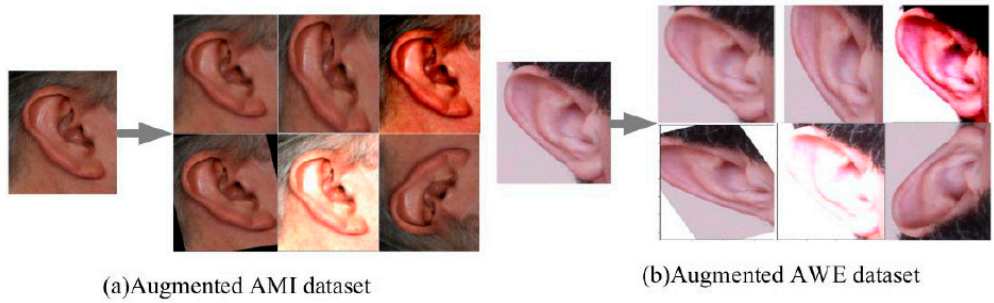


(a)AMI dataset (b)AWE dataset

Figure 7. Ear images from three subjects in AMI and AWE.

4.2. Data Augmentation

To suppress the overfitting of the model during the training process, we performed random cropping, random angle rotation, random brightness change, random scaling, random contrast change, and other operations on the original images of the AMI dataset and AWE dataset to achieve data expansion. Figure 8 shows a series of different images after data augmentation. This ensures that the images received by our model during training are all different, which dramatically improves the robustness of feature extraction and generalization.



(a)Augmented AMI dataset (b)Augmented AWE dataset

Figure 8. The result of image data augmentation.

4.3. Parameter Settings

The simulations in this paper are run on NVIDIA Tesla V100 SXM2 16G, and the PyTorch open-source framework is used to verify the recognition effect of the CFDCNet network model on ear images. Regarding the optimizer, we choose the stochastic gradient descent method (SGD), and the parameters are set to support momentum parameters, the learning decay rate, and the Nesterov momentum. Regarding the learning rate, we set the cosine scheduler and defined the learning rate decay. The specific changes in the learning rate are shown in Figure 9. Regarding the number of training iterations, the number of iterations we train in the AMI dataset is 100. In the AWE dataset, the number of training iterations is 500. The batch size is set to 16. Considering that CFDCNet belongs to a deep network and uses the ReLU Activation function to facilitate model convergence, we adopt the weight initialization method introduced in [40]. To objectively and effectively reflect the model's performance, we partition the dataset using hold-out cross-validation [41] and use the average of 10 tests to evaluate the model's performance.

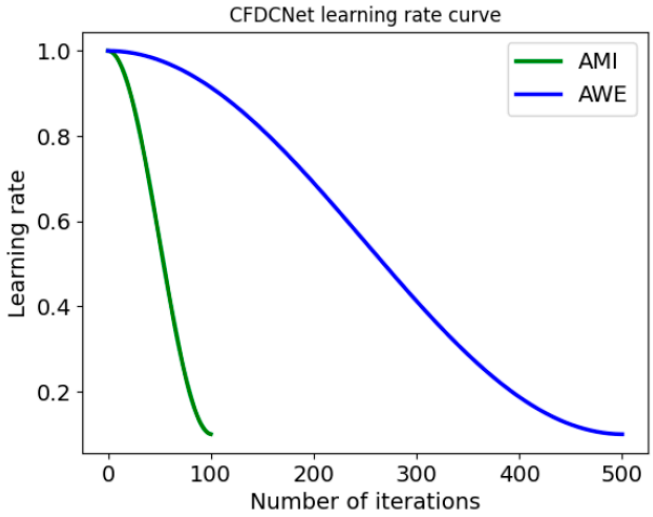


Figure 9. The learning rate curve of CFDCNet under different datasets.

4.4. Evaluation Metrics

We used quantitative performance evaluation metrics (R1 [42–46], R5 [42–46], and AUC [42–46]) to evaluate the performance of each ear recognition model and plotted the cumulative matching characteristics (CMC [42–46]) curve for each ear recognition model.

The Rank-1 (R1) recognition rate is the percentage of correct identities found to be the best matching probe ear images in the ear database.

The Rank-5 (R5) recognition rate is the percentage of probe ear images where the correct identity is found as the top five matches in the ear database.

The AUC is the area under the cumulative matching feature (CMC) curve.

Cumulative matching feature (CMC) curve: the probability that the model returns a correct identity within the first z ($z \leq N$) ranks, where N is the number of subjects in the entire ear database.

4.5. The Impact of Data Augmentation

To demonstrate the effectiveness of data augmentation in the ear recognition task, the original and data-augmented ear databases were input to the CFDCNet model for training and testing, respectively. The experimental results are shown in Table 1. The original images were directly input into the CFDCNet model for training, and the R1 recognition accuracies were 91.25% and 55.10%, respectively. After the data augmentation, the R1 recognition accuracy of CFDCNet increases to 99.70% and 72.70%, respectively. Similarly, the R5 recognition accuracy also improved significantly, indicating that data augmentation can enhance the generalization ability and robustness of the model and effectively suppress overfitting.

Table 1. The impact of data augmentation on the recognition performance of the proposed CFDC-Net model.

Database	Augmentation	R1	R5
AMI	×	91.25%	96.47%
	✓	99.70%	100.00%
AWE	×	55.10%	72.21%
	✓	72.70%	89.90%

4.6. Ablation Studies

To ensure the rigor of the simulations, we conducted an ablation study before discussing the impact of the pooling approach on the simulations. The simulations were done

on the AMI and AWE datasets, respectively, and the quantitative performance metrics (R1, R5, AUC, and FLOPs) are presented in Table 2. We also plotted CMC curves to show the differences in recognition performance for different simulation cases, as shown in Figure 10. The results show that the original DenseNet-121 model exhibits the worst recognition performance. When we introduce dynamic convolution or an efficient channel attention mechanism alone, the recognition performance of the model is slightly improved. However, when we consider both dynamic convolution and efficient channel attention mechanisms, optimal recognition performance is obtained, highlighting the effectiveness of our method. The efficient channel attention mechanism will not increase computational complexity, while dynamic convolution will increase computational complexity by 13 M. CFDCNet has increased the computational complexity by only 13 M compared to DenseNet-121, but it has significantly improved recognition performance.

Table 2. We compared the quantitative performance metrics (R1, R5, AUC, and FLOPs) under different ablation studies and highlighted the best values of the performance metrics in bold.

DY_Conv	ECA	MFLOPs	AMI			AWE		
			R1	R5	AUC	R1	R5	AUC
		142	97.00%	99.11%	98.94%	62.00%	76.38%	95.29%
✓		155	98.00%	99.62%	98.95%	70.00%	86.98%	96.95%
	✓	142	98.50%	99.83%	98.96%	71.45%	88.51%	96.99%
✓	✓	155	**99.40%**	**99.90%**	**98.98%**	**72.38%**	**89.70%**	**97.05%**

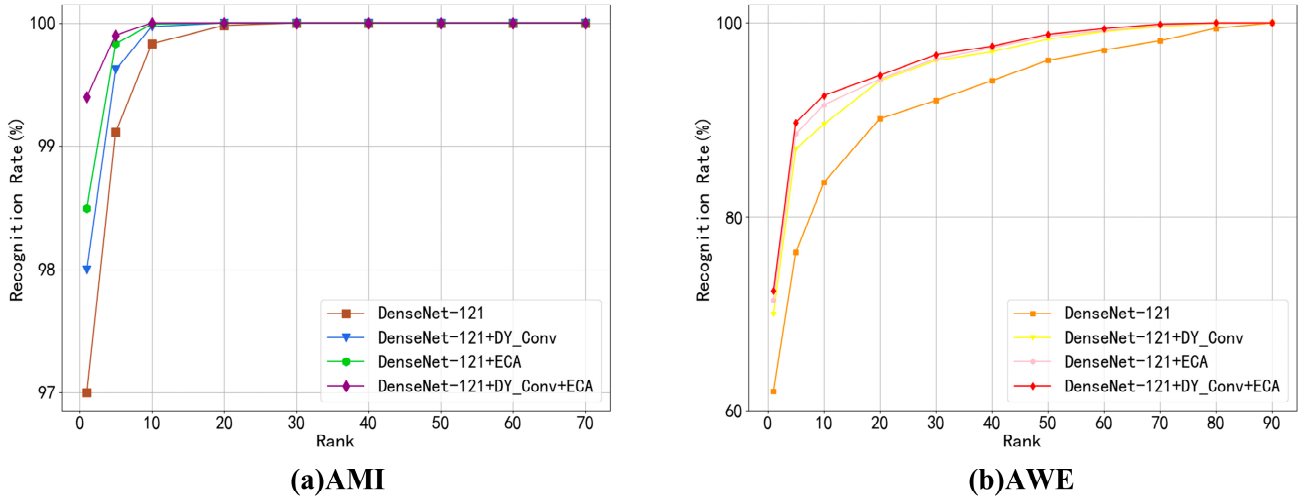


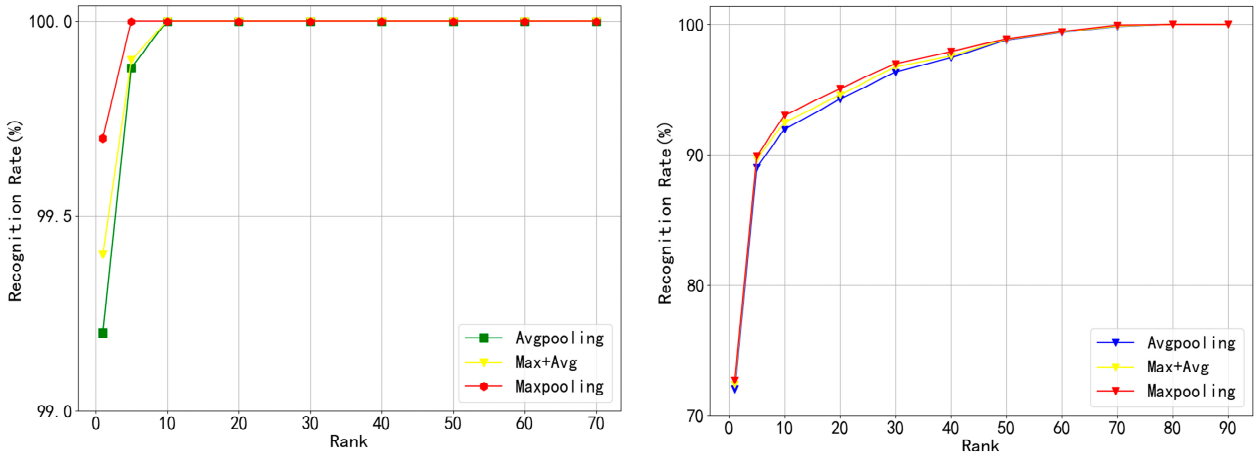
Figure 10. Comparison of CMC curves of different models in ablation studies.

4.7. The Effect of Pooling Methods

Since different pooling methods have a particular impact on feature extraction, affecting the accuracy of human ear recognition, after the ablation studies, we used three pooling methods: Avg pooling, Max + Avg, and Max pooling to conduct simulations. The final simulation results show that only using the average pooling Rank-1 (R1) recognition accuracy is the lowest. On the contrary, only using the maximum pooling Rank-1 (R1) recognition accuracy is the highest. The specific quantitative performance metrics (R1, R5, and AUC) are given in Table 3. We also plot the CMC curves for different pooling methods, as shown in Figure 11.

Table 3. We compare the quantitative performance metrics (R1, R5, and AUC) for different pooling methods, and the best values of the performance metrics are marked in bold.

Pooling Methods	AMI			AWE		
	R1	R5	AUC	R1	R5	AUC
Avg pooling	99.20%	99.88%	98.97%	72.00%	89.03%	97.02%
Max + Avg	99.40%	99.90%	98.98%	72.38%	89.70%	97.05%
Max pooling	**99.70%**	**100.00%**	**99.01%**	**72.70%**	**89.90%**	**97.08%**



(a)AMI **(b)AWE**

Figure 11. The CMC curves compare the recognition performance under different pooling methods.

4.8. The Impact of the Dataset Division Ratio

We set up four different dataset division ratios to verify the influence of this factor on simulations. The specific divisions are as follows: Training set:Validation set:Test set = 2:4:4, Training set:Validation set:Test set = 4:3:3, Training set:Validation set:Test set = 6:2:2, and Training set:Validation set:Test set = 8:1:1. The specific quantitative performance metrics (R1, R5, and AUC) are given in Table 4. We also plotted the CMC curves for different dataset segmentation ratios, as shown in Figure 12. From the analysis results, it can be concluded that the recognition accuracy of the original model is more sensitive to the division ratio of the dataset. When the training sample data input to the network is relatively small, the recognition accuracy will fluctuate greatly, and the simulation results are not very satisfactory. On the contrary, our proposed network (CFDCNet) performs very well on both the constrained AMI dataset and the unconstrained AWE dataset, and the recognition accuracy of CFDCNet fluctuates relatively little under different division ratios of the dataset. The recognition accuracy of CFDCNet in the case of a small number of samples input (Training set:Validation set:Test set = 2:4:4) is the same as that of DenseNet-121 in the case of a large number of samples input (Training set:Validation set:Test set = 8:1:1). The recognition accuracy is the same or even higher; that is to say, CFDCNet can accurately extract the characteristics of ear images through a small number of ear samples and improve the accuracy of human ear recognition.

4.9. Compared with Other Methods

To evaluate the performance of the CFDCNet model on human ear recognition, we summarize past work and compare the recognition accuracy of CFDCNet on the AMI and AWE datasets with past methods. The performance evaluation metrics (R1, R5, and AUC) of the various methods are given in Tables 5 and 6 in percentage form. Among them, the CFDCNet model has the best performance for human ear recognition on the AMI and AWE datasets, with Rank-1 accuracy of 99.70% and 72.70%, respectively. By analyzing the prediction results of the models on the test set, it can be concluded that incorrect

predictions often occur when the ears are heavily obscured by hair and accessories. Correct predictions are made if there are apparent hair and accessories as auxiliary information. For images with small head rotation angles, the effect of occlusion on the recognition rate is insignificant. On the contrary, for images with large head rotation angles, occlusion has a significant impact on the recognition rate. The presence of earplugs and eyeglass frames has a slight negative impact on the recognition rate, as do background features such as facial skin texture and hair color that strongly contrast with ear features.

Table 4. Comparison of quantitative performance metrics (R1, R5, and AUC) of DenseNet-121 and CFDCNet with different dataset division ratios.

Dataset Division Ratio	Method	AMI			AWE		
		R1	R5	AUC	R1	R5	AUC
2:4:4	DenseNet-121	60.00%	82.43%	96.51%	20.65%	49.95%	84.89%
	CFDCNet	97.00%	99.41%	98.95%	62.12%	76.41%	95.32%
4:3:3	DenseNet-121	81.25%	90.02%	98.01%	43.97%	63.21%	93.11%
	CFDCNet	99.58%	99.94%	98.99%	70.02%	87.01%	96.95%
6:2:2	DenseNet-121	96.40%	98.39%	98.87%	58.95%	70.42%	94.67%
	CFDCNet	99.62%	99.98%	99.00%	72.61%	89.73%	97.07%
8:1:1	DenseNet-121	97.00%	99.11%	98.94%	62.00%	72.38%	95.29%
	CFDCNet	99.70%	100.00%	99.01%	72.70%	89.90%	97.08%

Table 5. The quantitative performance metrics (R1, R5, and AUC) of our method (CFDCNet) on the AMI ear database are compared with previous work. The best values of the performance metrics are marked in bold.

Previous Work	AMI		
	R1	R5	AUC
Raghavendra et al. [47]	86.36%	-	-
Alshazly et al. [48]	70.20%	-	-
Chowdhury et al. [49]	67.26%	-	-
Hassaballah et al. [50]	73.71%	-	-
Alshazly et al. [42]	94.50%	99.40%	98.90%
Alshazly et al. [43]	97.50%	99.64%	98.41%
Omara et al. [51]	97.84%	-	-
Zhang et al. [52]	93.96%	-	-
Omara et al. [53]	96.82%	-	-
Khaldi et al. [44]	96.00%	99.00%	94.47%
Hassaballah et al. [24]	72.29%	-	-
Ahila et al. [2]	96.99%	-	-
Khaldi et al. [54]	98.33%	-	-
Alshazly et al. [45]	99.64%	100%	98.99%
Aiadi et al. [55]	97.67%	-	-
Sharkas [56]	99.45%	-	-
Ebanesar et al. [57]	98.99%	-	-
Kohlakala et al. [58]	99.20%	-	-
Our method (CFDCNet)	**99.70%**	**100%**	**99.01%**

Table 6. The quantitative performance metrics (R1, R5, and AUC) of our method (CFDCNet) on the AWE ear database are compared with previous work. The best values of the performance metrics are marked in bold.

Previous Work	AWE		
	R1	R5	AUC
Hassaballah et al. [50]	49.60%	-	-
Emersic et al. [16]	49.60%	-	-
Dodge et al. [59]	56.35%	74.80%	-
Dodge et al. [59]	68.50%	83.00%	-
Zhang et al. [30]	50.00%	70.00%	-
Emersic et al. [46]	62.00%	80.35%	95.51%
Khaldi et al. [44]	50.53%	76.35%	80.97%
Hassaballah et al. [24]	54.10%	-	-
Khaldi et al. [60]	48.48%	-	-
Khaldi et al. [54]	51.25%	-	-
Alshazly et al. [45]	67.25%	84.00%	96.03%
Regouid et al. [61]	43.00%	-	-
Kacar et al. [62]	47.80%	72.10%	95.80%
Sajadi et al. [25]	53.50%	-	-
Omara et al. [63]	72.22%	-	-
Our method (CFDCNet)	**72.70%**	**89.90%**	**97.08%**

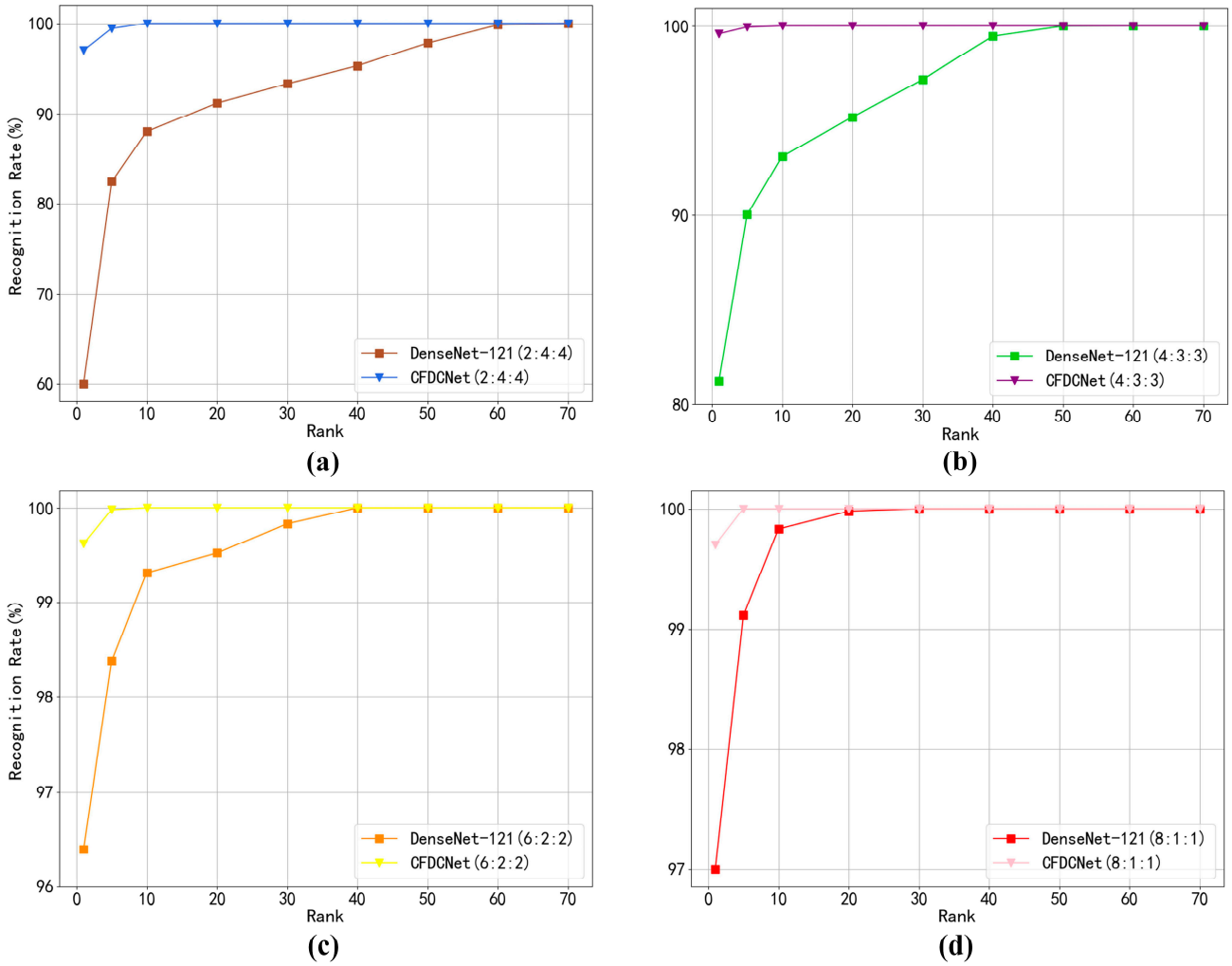


Figure 12. *Cont.*

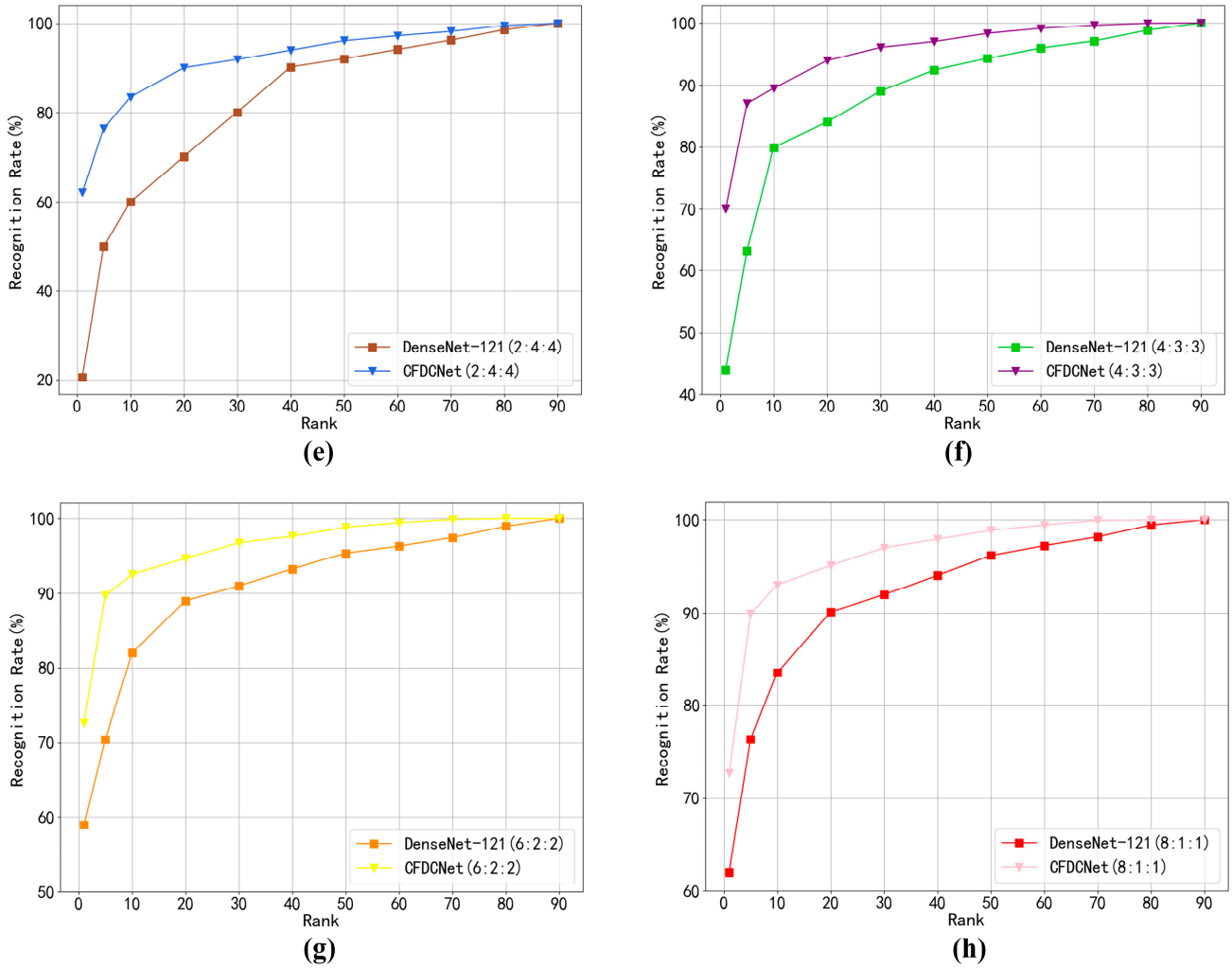


Figure 12. The CMC curves compare the recognition performance of DenseNet-121 and CFDCNet with different dataset division ratios. Where (**a**–**d**) are the CMC curves for the AMI dataset with different division ratios, and (**e**–**h**) are the CMC curves for the AWE dataset with different division ratios.

5. Conclusions

In this paper, we propose a feature fusion human ear recognition method based on channel features and dynamic convolution. Favorable ear features are extracted adaptively by dynamic convolution and an efficient channel attention mechanism and then combined with maximum pooling operations to retain the main features of the ear contour. The method achieves good recognition results on the AMI and AWE human ear databases, with Rank-1 (R1) recognition accuracies of 99.70% and 72.70%, respectively. Compared with the DenseNet-121 model and other methods, this method can extract feature information from human ear images more accurately and has better recognition performance. We will continue to optimize our approach and improve its recognition performance on the challenging unconstrained ear dataset.

Author Contributions: All of the authors made significant contributions to this work. Conceptualization and data analysis, X.X., Y.L. and L.L.; experimental design and manuscript—preparation, advice, and revision, Y.L., C.L. and L.L.; experimental, Y.L. and C.L.; manuscript—writing, X.X. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61673316); by the Scientific Research Project of the Education Department of Shaanxi Province (21JK0921); by

the Key Research and Development Program of Shaanxi Province, under Grant 2017GY-071; by the Technical Innovation Guidance Special Project of Shaanxi Province, under Grant 2017XT-005; and by the research program of Xian Yang City, under Grant 2017K01-25-3.

Data Availability Statement: AMI can be found at https://ctim.ulpgc.es/research_works/ami_ear_database/, accessed on 12 November 2021. AWE can be found at http://awe.fri.uni-lj.si/, accessed on 12 November 2021.

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

References

1. Jain, A.; Bolle, R.; Pankanti, S. *Introduction to Biometrics*; Springer: Berlin/Heidelberg, Germany, 1996.
2. Ahila Priyadharshini, R.; Arivazhagan, S.; Arun, M. A deep learning approach for person identification using ear biometrics. *Appl. Intell.* **2021**, *51*, 2161–2172. [CrossRef] [PubMed]
3. Olanrewaju, L.; Oyebiyi, O.; Misra, S.; Maskeliunas, R.; Damasevicius, R. Secure ear biometrics using circular kernel principal component analysis, Chebyshev transform hashing and Bose–Chaudhuri–Hocquenghem error-correcting codes. *Signal Image Video Process.* **2020**, *14*, 847–855. [CrossRef]
4. Bokade, G.U.; Kanphade, R.D. Secure multimodal biometric authentication using face, palmprint and ear: A feature level fusion approach. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; pp. 1–5.
5. Petaitiemthong, N.; Chuenpet, P.; Auephanwiriyakul, S.; Theera-Umpon, N. Person identification from ear images using convolutional neural networks. In Proceedings of the 2019 9th IEEE international conference on control system, computing and engineering (ICCSCE), Penang, Malaysia, 29 November–1 December 2019; pp. 148–151.
6. Chen, Y.; Dai, X.; Liu, M.; Chen, D.; Yuan, L.; Liu, Z. Dynamic convolution: Attention over convolution kernels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11030–11039.
7. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
8. Zhang, Y.; Zhang, J.; Wang, Q.; Zhong, Z. Dynet: Dynamic convolution for accelerating convolutional neural networks. *arXiv* **2020**, arXiv:2004.10694.
9. Tian, Y.; Shen, Y.; Wang, X.; Wang, J.; Wang, K.; Ding, W.; Wang, Z.; Wang, F.-Y. Learning Lightweight Dynamic Kernels With Attention Inside via Local–Global Context Fusion. *IEEE Trans. Neural Netw. Learn. Syst.* 2022; *online ahead of print*. [CrossRef]
10. Liu, K.; Moon, S. Dynamic Parallel Pyramid Networks for Scene Recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–11. [CrossRef]
11. Sun, J.; Li, P.; Wu, X. Handwritten Ancient Chinese Character Recognition Algorithm Based on Improved Inception-ResNet and Attention Mechanism. In Proceedings of the 2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI), Xiamen, China, 10–12 June 2022; pp. 31–35.
12. Shang, Y.; Huo, H. A study on fine-grained image classification algorithm based on ECA-NET and multi-granularity. *Int. J. Front. Eng. Technol.* **2023**, *5*, 31–38.
13. Liu, S.; Bai, H.; Li, F.; Wang, D.; Zheng, y.; Jiang, Q.; Sun, F. An apple leaf disease identification model for safeguarding apple food safety. *Food Sci. Technol.* **2023**, *43*, e104322. [CrossRef]
14. González Sánchez, E. Análisis biométrico de la Orejas. Ph.D. Thesis, Universidad de las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain, 2008.
15. Emeršič, Ž.; Meden, B.; Peer, P.; Štruc, V. Evaluation and analysis of ear recognition models: Performance, complexity and resource requirements. *Neural Comput. Appl.* **2020**, *32*, 15785–15800. [CrossRef]
16. Emeršič, Ž.; Štruc, V.; Peer, P. Ear recognition: More than a survey. *Neurocomputing* **2017**, *255*, 26–39. [CrossRef]
17. Emeršič, Ž.; Gabriel, L.L.; Štruc, V.; Peer, P. Convolutional encoder–decoder networks for pixel-wise ear detection and segmentation. *IET Biom.* **2018**, *7*, 175–184. [CrossRef]
18. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
19. De Marsico, M.; Michele, N.; Riccio, D. HERO: Human ear recognition against occlusions. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 178–183.
20. Bustard, J.D.; Nixon, M.S. Toward unconstrained ear recognition from two-dimensional images. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2010**, *40*, 486–494. [CrossRef]
21. Kumar, A.; Wu, C. Automated human identification using ear imaging. *Pattern Recognit.* **2012**, *45*, 956–968. [CrossRef]
22. Chan, T.-S.; Kumar, A. Reliable ear identification using 2-D quadrature filters. *Pattern Recognit. Lett.* **2012**, *33*, 1870–1881. [CrossRef]
23. Anwar, A.S.; Ghany, K.K.A.; Elmahdy, H. Human ear recognition using geometrical features extraction. *Procedia Comput. Sci.* **2015**, *65*, 529–537. [CrossRef]

24. Hassaballah, M.; Alshazly, H.A.; Ali, A.A. Robust local oriented patterns for ear recognition. *Multimed. Tools Appl.* **2020**, *79*, 31183–31204. [CrossRef]
25. Sajadi, S.; Fathi, A. Genetic algorithm based local and global spectral features extraction for ear recognition. *Expert Syst. Appl.* **2020**, *159*, 113639. [CrossRef]
26. Ghoualmi, L.; Chikhi, S.; Draa, A. A SIFT-based feature level fusion of iris and ear biometrics. In Proceedings of the Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction: Third IAPR TC3 Workshop, MPRSS 2014, Stockholm, Sweden, 24 August 2014; pp. 102–112.
27. Rathore, R.; Prakash, S.; Gupta, P. Efficient human recognition system using ear and profile face. In Proceedings of the 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 29 September–2 October 2013; pp. 1–6.
28. Kumar, A.M.; Chandralekha, A.; Himaja, Y.; Sai, S.M. Local binary pattern based multimodal biometric recognition using ear and FKP with feature level fusion. In Proceedings of the 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 11–13 April 2019; pp. 1–5.
29. Tian, L.; Mu, Z. Ear recognition based on deep convolutional network. In Proceedings of the 2016 9th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI), Datong, China, 15–17 October 2016; pp. 437–441.
30. Zhang, Y.; Mu, Z.; Yuan, L.; Yu, C. Ear verification under uncontrolled conditions with convolutional neural networks. *IET Biom.* **2018**, *7*, 185–198. [CrossRef]
31. Emeršič, Ž.; Križaj, J.; Štruc, V.; Peer, P. Deep ear recognition pipeline. *Recent Adv. Comput. Vis. Theor. Appl.* **2019**, *804*, 333–362.
32. Alshazly, H.; Linse, C.; Barth, E.; Martinetz, T. Deep convolutional neural networks for unconstrained ear recognition. *IEEE Access* **2020**, *8*, 170295–170310. [CrossRef]
33. Radhika, K.; Devika, K.; Aswathi, T.; Sreevidya, P.; Sowmya, V.; Soman, K. Performance analysis of NASNet on unconstrained ear recognition. *Nat. Inspired Comput. Data Sci.* **2020**, *871*, 57–82.
34. Koniusz, P.; Yan, F.; Mikolajczyk, K. Comparison of mid-level feature coding approaches and pooling strategies in visual concept detection. *Comput. Vis. Image Underst.* **2013**, *117*, 479–492. [CrossRef]
35. Zhao, Z.; Ma, H.; Chen, X. Protected pooling method of sparse coding in visual classification. In Proceedings of the International Conference on Computer Vision and Graphics, Warsaw, Poland, 15–17 September 2014; pp. 680–687.
36. Boureau, Y.-L.; Ponce, J.; LeCun, Y. A theoretical analysis of feature pooling in visual recognition. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israe, 21–24 June 2010; pp. 111–118.
37. Koniusz, P.; Gosselin, P.-H.; Mikolajczyk, K. *Higher-Order Occurrence Pooling on Mid-and Low-Level Features: Visual Concept Detection*; HAL Open Science: Paris, France, 2013; p. 20.
38. Avila, S.; Thome, N.; Cord, M.; Valle, E.; AraúJo, A.D.A. Pooling in image representation: The visual codeword point of view. *Comput. Vis. Image Underst.* **2013**, *117*, 453–465. [CrossRef]
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
41. Efron, B.; Gong, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* **1983**, *37*, 36–48.
42. Alshazly, H.; Linse, C.; Barth, E.; Martinetz, T. Handcrafted versus CNN features for ear recognition. *Symmetry* **2019**, *11*, 1493. [CrossRef]
43. Alshazly, H.; Linse, C.; Barth, E.; Martinetz, T. Ensembles of deep learning models and transfer learning for ear recognition. *Sensors* **2019**, *19*, 4139. [CrossRef]
44. Khaldi, Y.; Benzaoui, A. A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions. *Evol. Syst.* **2021**, *12*, 923–934. [CrossRef]
45. Alshazly, H.; Linse, C.; Barth, E.; Idris, S.A.; Martinetz, T. Towards explainable ear recognition systems using deep residual networks. *IEEE Access* **2021**, *9*, 122254–122273. [CrossRef]
46. Emeršič, Ž.; Štepec, D.; Štruc, V.; Peer, P. Training convolutional neural networks with limited training data for ear recognition in the wild. *arXiv* **2017**, arXiv:1711.09952.
47. Raghavendra, R.; Raja, K.B.; Busch, C. Ear recognition after ear lobe surgery: A preliminary study. In Proceedings of the 2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), Sendai, Japan, 29 February–2 March 2016; pp. 1–6.
48. Alshazly, H.A.; Hassaballah, M.; Ahmed, M.; Ali, A.A. Ear biometric recognition using gradient-based feature descriptors. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018, Cairo, Egypt, 1–3 September 2018; pp. 435–445.
49. Chowdhury, D.P.; Bakshi, S.; Guo, G.; Sa, P.K. On applicability of tunable filter bank based feature for ear biometrics: A study from constrained to unconstrained. *J. Med. Syst.* **2018**, *42*, 11. [CrossRef]
50. Hassaballah, M.; Alshazly, H.A.; Ali, A.A. Ear recognition using local binary patterns: A comparative experimental study. *Expert Syst. Appl.* **2019**, *118*, 182–200. [CrossRef]

51. Omara, I.; Hagag, A.; Ma, G.; Abd El-Samie, F.E.; Song, E. A novel approach for ear recognition: Learning Mahalanobis distance features from deep CNNs. *Mach. Vis. Appl.* **2021**, *32*, 38. [CrossRef]
52. Zhang, J.; Yu, W.; Yang, X.; Deng, F. Few-shot learning for ear recognition. In Proceedings of the 2019 International Conference on Image, Video and Signal Processing, Shanghai, China, 25–28 February 2019; pp. 50–54.
53. Omara, I.; Hagag, A.; Chaib, S.; Ma, G.; Abd El-Samie, F.E.; Song, E. A hybrid model combining learning distance metric and DAG support vector machine for multimodal biometric recognition. *IEEE Access* **2020**, *9*, 4784–4796. [CrossRef]
54. Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear recognition based on deep unsupervised active learning. *IEEE Sens. J.* **2021**, *21*, 20704–20713. [CrossRef]
55. Aiadi, O.; Khaldi, B.; Saadeddine, C. MDFNet: An unsupervised lightweight network for ear print recognition. *J. Ambient. Intell. Humaniz. Comput.* 2022; *online ahead of print*. [CrossRef] [PubMed]
56. Sharkas, M. Ear recognition with ensemble classifiers; A deep learning approach. *Multimed. Tools Appl.* **2022**, *81*, 43919–43945. [CrossRef]
57. Ebanesar, T.; Bibin, A.; Jalaja, J. Human Ear Recognition Using Convolutional Neural Network. *J. Posit. Sch. Psychol.* **2022**, *6*, 8182–8190.
58. Kohlakala, A.; Coetzer, J. Ear-based biometric authentication through the detection of prominent contours. *SAIEE Afr. Res. J.* **2021**, *112*, 89–98. [CrossRef]
59. Dodge, S.; Mounsef, J.; Karam, L. Unconstrained ear recognition using deep neural networks. *IET Biom.* **2018**, *7*, 207–214. [CrossRef]
60. Khaldi, Y.; Benzaoui, A. Region of interest synthesis using image-to-image translation for ear recognition. In Proceedings of the 2020 International Conference on Advanced Aspects of Software Engineering (ICAASE), Constantine, Algeria, 28–30 November 2020; pp. 1–6.
61. Regouid, M.; Touahria, M.; Benouis, M.; Mostefai, L.; Lamiche, I. Comparative study of 1D-local descriptors for ear biometric system. *Multimed. Tools Appl.* **2022**, *81*, 29477–29503. [CrossRef]
62. Kacar, U.; Kirci, M. ScoreNet: Deep cascade score level fusion for unconstrained ear recognition. *IET Biom.* **2019**, *8*, 109–120. [CrossRef]
63. Omara, I.; Zhang, H.; Wang, F.; Hagag, A.; Li, X.; Zuo, W. Metric learning with dynamically generated pairwise constraints for ear recognition. *Information* **2018**, *9*, 215. [CrossRef]