symmetry

MDPI

Article

Anti-Software Attack Ear Identification System Using Deep Feature Learning and Blockchain Protection

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 recognition has made good progress as an emerging biometric technology. However, the recognition performance, generalization ability, and feature robustness of ear recognition systems based on hand-crafted features are relatively poor. With the development of deep learning, these problems have been partly overcome. However, the recognition performance of existing ear recognition systems still needs to be improved when facing unconstrained ear databases in realistic scenarios. Another critical problem is that most systems with ear feature template databases are vulnerable to software attacks that disclose users' privacy and even bring down the system. This paper proposes a software-attack-proof ear recognition system using deep feature learning and blockchain protection to address the problem that the recognition performance of existing systems is generally poor in the face of unconstrained ear databases in realistic scenarios. First, we propose an accommodative DropBlock (AccDrop) to generate drop masks with adaptive shapes. It has an advantage over DropBlock in coping with unconstrained ear databases. Second, we introduce a simple and parameterless attention module that uses 3D weights to refine the ear features output from the convolutional layer. To protect the security of the ear feature template database and the user's privacy, we use Merkle tree nodes to store the ear feature templates, ensuring the determinism of the root node in the smart contract. We achieve Rank-1 (R1) recognition accuracies of 83.87% and 96.52% on the AWE and EARVN1.0 ear databases, which outperform most advanced ear recognition systems.

Keywords: ear recognition; deep learning; accommodative DropBlock (AccDrop); attention module; smart contracts

Citation: Xu, X.; Liu, Y.; Liu, C.; Lu, L. Anti-Software Attack Ear Identification System Using Deep Feature Learning and Blockchain Protection. *Symmetry* **2024**, *16*, 85. https://doi.org/10.3390/sym16010085

Academic Editor: Douglas O'Shaughnessy

Received: 5 May 2023
Revised: 12 July 2023
Accepted: 13 July 2023
Published: 9 January 2024

1. Introduction

Ear recognition is a biometric technology that has emerged in recent years since ears, similar to fingerprints [1], irises [2], and faces [3], contain many specific and unique features [4] that can be used to identify a person [5]. The process of ear image acquisition is not dependent on the subject's cooperation and is non-invasive and non-contact. Capturing ear images from a distance covertly [6] is undoubtedly an attractive option favored for surveillance and security as well as other applications. Ear images extracted from side headshots or video clips can also be used in ear recognition systems. Studies have shown that age can impact biometric performance [7,8]. For example, fingerprints are subject to wear and tear with age, and wrinkles appear on a person's face, which can affect the system identification accuracy. By contrast, the most crucial characteristic of the human ear is that it does not age too significantly [9]. The shape of the ear is essentially set at birth and uniformly distributed in color, changing significantly only before age eight and after age 70, and this change is measurable. It is challenging for cybercriminals to replicate it [10]. In addition, ear biometric systems have been used in several fields, such as remote voting, authentication, attendance control, finance, and other transactions that require authorization. In automatic identification systems, the ear image can complement different

biometric patterns to provide identity cues if other biometric information is unreliable or corrupted. Early ear recognition systems used hand-crafted features for recognition. These systems had four main drawbacks: (1) most did not use a baseline ear database for system performance testing; (2) no standard performance evaluation metrics were used; (3) the ear database used for recognition was obtained in a constrained environment, resulting in poor system generalization and poor performance on unconstrained ear databases; and (4) recognition performance lags behind that of systems based on deep feature learning. Recent years have witnessed the development of deep learning [11–13], and the technique has been used in ear recognition. Ear recognition systems based on deep learning have excellent performance and are popular among researchers. Ear recognition systems have performed well in the last decade but face many challenges. Most ear recognition systems are vulnerable to software attacks [14], especially attacks against ear feature template databases. Such attacks can paralyze the entire ear recognition system and compromise the user's privacy. The stolen ear template data can be used illegally, adversely affecting individuals, businesses, and society. Blockchain is considered one of this era's most disruptive technologies [15,16]. The properties supported by blockchain technology, such as universal access, availability, accountability, and invariance [17], can be perfectly applied to protect a database of ear signature templates. Therefore, we propose an ear identification system using deep feature learning and blockchain protection against software attacks. The system protects users' privacy and prevents software attacks on the ear feature template database while providing superior recognition performance.

Our contributions can be summarized as follows: (1) We propose the accommodative DropBlock (AccDrop), which has advantages over DropBlock [18] in dealing with unconstrained ear databases. It can generate drop masks with adaptive shapes, which effectively improves the generally poor recognition performance of existing ear recognition systems when facing unconstrained ear databases in realistic scenarios. (2) We introduce a simple and parameter-free attention module [19] to infer the 3D attention weights of the ear feature maps output from the convolutional layer, enabling the ear recognition system to learn more critical feature regions. (3) We use Merkle tree [20] nodes to store the ear feature templates, ensuring the determinism of the root node in the smart contract and protecting the security of the ear feature template database and the user's privacy. (4) We conduct extensive experiments on two of the most exemplary unconstrained ear databases to further confirm the reliability of the proposed ear recognition system. The experimental results show that our system's recognition performance and generalization capability are more competitive than most state-of-the-art ear recognition systems.

The remainder of this paper is organized as follows: A brief review of related work is presented in Section 2. Section 3 describes our proposed method in detail. In Section 4, we evaluate the performance of the proposed method on two representative public unconstrained ear datasets, derive experimental results, and perform a correlation analysis. Finally, we conclude the whole paper in Section 5.

2. Related Work

Early researchers performed human ear recognition based on hand-made features, mainly geometric, holistic, local, and hybrid methods. Geometric methods: The computational process of geometric methods is relatively simple and mainly involves analyzing and extracting the geometric features of an ear image. Edge detection is the most common pre-processing operation of geometric methods. Edge information can provide geometric statistics for ear recognition and describe the ear's geometric features. Since the geometric method relies only on information related to the geometric features of the ear structure, it is robust to methods that do not produce geometric deformations, e.g., scaling, rotation, etc. However, geometric methods also have significant drawbacks in that the discriminative categorical feature information presented in the ear image is ignored. Many geometric methods have been proposed in the literature: Ref. [21] used information based on wrinkles and ear shape as well as outer ear point feature information for recognition, Ref. [22] pro-

posed a discrete geometric algorithm using open contour representation for ear biometrics, and Ref. [23] demonstrated that the geometric properties of ear subparts, such as earlobe shape, ear size, etc., have a significant impact on recognition. The reliance on edge detectors dramatically limits the development of geometric methods in the field of ear recognition. The edge detectors reduce feature robustness when subjected to noise and illumination changes. Holistic approach: The holistic approach treats the ear as a whole and globally encodes the entire ear structure. This method significantly degrades the feature extraction performance in the presence of significant differences in ear image angles and illumination, so it is usually necessary to apply normalization techniques as pre-processing operations to reduce the impact of these factors on performance before feature extraction. Ref. [24] proposed a (holistic) ear recognition method based on force-field transformation and kernel fisher discriminant analysis in zero space. This technique uses ear pixels as particles of force field source, extracts ear features by force field transformation, and is effective with good robustness for multi-angle ear recognition. The subspace projection technique is also widely used in ear recognition. With this technique, the ear image is represented as a linear combination of weights in the form of basis vectors in pixel space. Examples of such techniques are proposed in the literature, including principal component analysis (PCA) [25,26], enhanced local linear embedding (ELLE) [27], and non-negative matrix decomposition (NMF) [28]. Another overall technique for ear recognition is to operate in the frequency domain. Ref. [5] represents the ear image in the frequency domain using a generic Fourier descriptor, which shows good stability for ear image rotation. Local approach: this approach mainly uses the local features of an ear image for recognition and has achieved competitive results in the field of ear recognition. Ref. [29] extracts key point locations employing scale-invariant feature transform (SIFT) and then calculates descriptors for each key point detected. The advantage of this technique is that the extracted descriptors can be partially matched and therefore show some robustness to partial regional occlusions of the ear. The disadvantage is that there is a risk of ignoring discriminative global ear feature information. Ref. [30] extracts local grayscale phase information based on Gabor filters and then uses local features for recognition. This technique uses local descriptors to form the ear image representation and encode the global structure of the image, with the disadvantage that it is less robust to occlusions. Hybrid approach: a method to improve recognition performance by mixing multiple technical representations. Ref. [31] proposed a method that combines principal component analysis (PCA) and wavelets. Ref. [32] introduced a hybrid method based on the Haar transform and local binary patterns (LBP). Ref. [33] fused a Gabor filter and local binary patterns (LBP) for ear recognition. The hybrid method is competitive in ear recognition. Because it is a mixture of multiple technical representations, the method is much more computationally intensive than simple local or holistic methods.

Some ear recognition methods based on hand-crafted features show near-perfect recognition performance on constrained datasets. However, their performance on unconstrained datasets is poor compared to in-depth feature-learning-based methods. Deep learning has developed rapidly in recent years, and many human ear recognition methods based on deep feature learning have been proposed and achieved excellent recognition performance. In [34], a deep convolutional neural network model was designed to perform ear recognition. The system's robustness was verified by evaluating it on a constrained ear dataset. The disadvantages are that the authors do not use standard performance evaluation parameters, and the database used has a slight variation in the internal ear images, which does not reflect the generalization ability of the proposed system. A two-path convolutional neural network model was proposed in [35]. The network focuses on discriminative image regions by pooling the information related to patches, ensuring good recognition performance. The disadvantage is that only one ear database is used to evaluate the system's performance. Ref. [36] used the NASNet network model for ear recognition. An optimized network is provided by reducing the number of operations and the number of learnable parameters to achieve excellent recognition performance.

Ear recognition systems have evolved rapidly in the last decade but face many challenges [37,38]. Almost all ear recognition systems are vulnerable to software attacks [14] or physical attacks [39]. In recent years, researchers have made some progress in ear recognition systems against physical attacks and proposed some anti-fraud methods. Ref. [40] collected a database of ear presentation attack detection, including three types of fake ear attacks—display attacks, print attacks, and video attacks—and used image quality assessment techniques to extract ear features of interest. Ref. [41] proposed a three-level fusion-based ear anti-spoofing system based on image quality assessment techniques. The system performance was evaluated using printed photo attack images, and the results showed that the system could distinguish well between real and fake ear images. Ref. [42] proposed a light-field ear artifact database containing images from multiple types of attack devices, such as cell phones, laptops, and tablets. The problem of detecting ear presentation attacks was solved, and promising results were reported. Software attacks, especially attacks on ear feature template databases, can disrupt the working process of ear recognition systems and harm users' privacy. However, there is still a gap in the existing literature and a relative lack of techniques to prevent software attacks on ear recognition systems. Properties such as invariance and accountability [17] supported by blockchain technology meet the need for ear feature template database protection.

This paper proposes an ear identification system against software attacks using deep feature learning and blockchain protection. We evaluate the system performance on two representative unconstrained ear databases, AWE [43–45] and EARVN1.0 [46], respectively. Our proposed AccDrop can generate drop masks with adaptive shapes for ear images, which effectively improves the recognition performance of the ear recognition system when facing unconstrained ear databases in realistic scenarios. To address the problem of large intra- and inter-class variation in unconstrained ear databases, we introduce an attention mechanism at the back end of the convolutional layer of the feature extraction network to refine the ear features output from the convolutional layer using 3D weights, without increasing the parameters and the computational effort of the network. To ensure the determinism of the root node in the smart contract, we use Merkle tree nodes to store the ear feature templates to further protect personal privacy and ensure the stability of the ear feature template database.

3. The Proposed Approach

Our proposed ear identification system against software attacks using deep feature learning and blockchain protection is shown in Figure 1. As shown in the figure, the system consists of four phases: input, feature extraction, matching, and decision-making. Among them, the matching phase is vulnerable to software attacks, and tampering with the template in the database and intercepting the channel between the database and the matcher are common types of software attacks. Therefore, we introduce a public blockchain when exchanging the ear feature template database and use Merkle tree blockchain storage technology to enhance the protection of the ear feature templates, which significantly improves the security of the ear recognition system. Merkle tree is a binary data structure using recursive construction, in which each node contains a cryptographic hash of its child node content. Since the root node counts the information of all the child nodes except the root node, any tampering with the content of the child nodes will cause the root node value to change. Arranging the data in the form of a Merkle tree can ensure the integrity of the root node data in the smart contract and achieve the purpose of securely storing the content of the ear feature template. In the feature extraction phase of the human ear recognition system, we use a deep-feature-learning-based approach for ear feature extraction. The details of the network architecture for this phase are shown in Figure 2. It consists of five convolutional layers, five attention modules, three pooling layers, three fully connected layers, and an AccDrop module. Each convolutional layer is followed by a batch normalization layer and a Relu activation layer to increase the stability of ear feature learning and mitigate the network gradient disappearance and overfitting phenomena.

The attention module processes the low-level detail ear feature maps output from all convolutional layers in the network to obtain further detailed ear feature maps. The first convolutional layer with a step size of 4 pixels performs convolutional operations on the input ear image of size 224 × 224 using 64 kernels of size 11 × 11. The spatial dimensions of the feature map are reduced to half of the original size using a maximum pooling layer with a window size of 3 × 3 and a step size of two pixels. The output of the first pooling layer is used as the input of the second convolutional layer, filtered using 192 kernels of size 5 × 5. The third and fourth convolutional layers both have 384 kernels of size 3 × 3. The fifth convolutional layer has 256 kernels of size 3 × 3. The last max-pooling layer outputs the high-level distinguishing feature information necessary for ear classification and then generates drop masks with adaptive shapes via an AccDrop. ReLU nonlinearity is applied to the output of the fully connected layers, and the neurons in each fully connected layer are connected to all neurons in the previous layer. In the following sections, we present the AccDrop mentioned in the network and the details of the attention module.

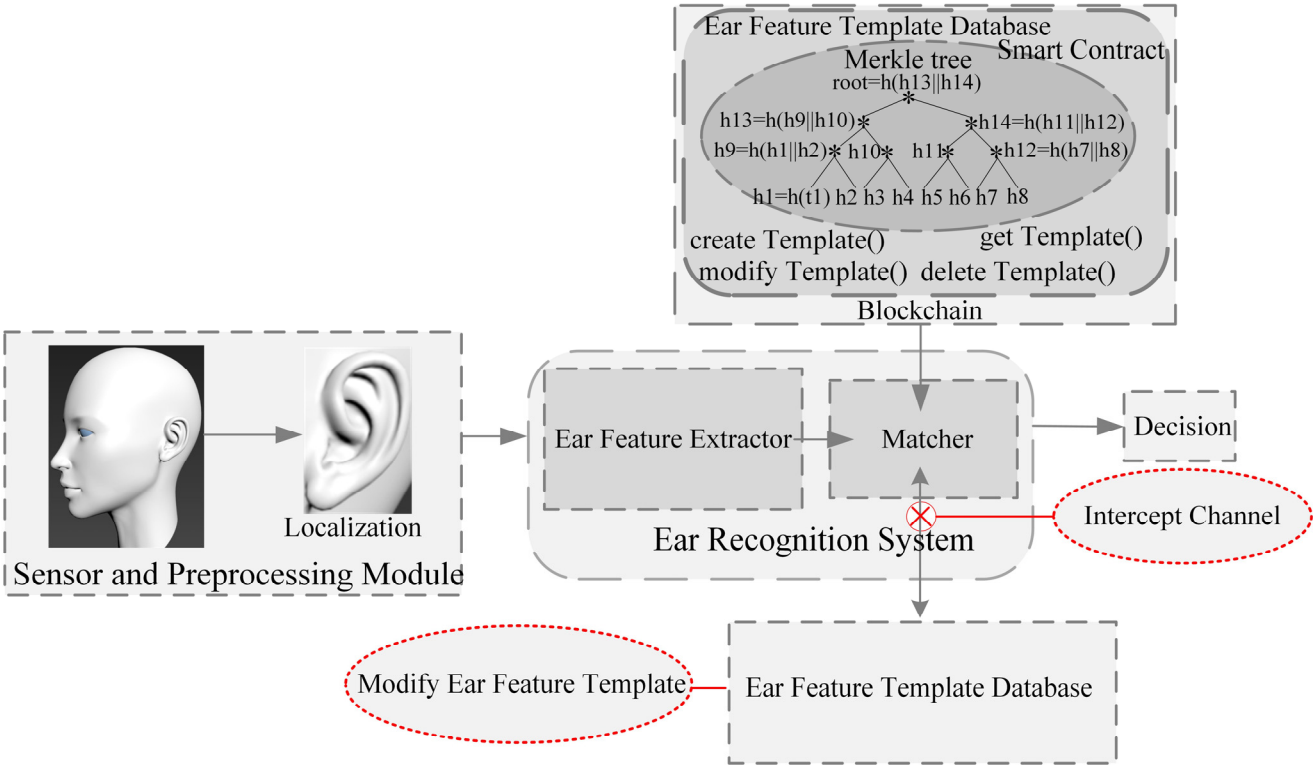


Figure 1. Proposed anti-software attack ear recognition system.

3.1. Accommodative DropBlock

Overparameterization of network models often leads to overfitting problems, and regularization methods can alleviate this problem. Dropout [47] is a commonly used regularization method that is very effective for fully connected layers. DropBlock is a structural form of regularization method that drops cells together in adjacent regions of the feature map. However, it will affect the feature learning ability of the network. In order to improve the recognition performance, our proposed AccDrop is a regularization method that can enhance the spatial dimensional feature representation to learn discriminative ear features effectively. DropBlock is a standard regularization method that tends to remove some practical feature information for basic training, resulting in poor recognition results, as shown in Figure 3. Our proposed AccDrop is a structured regularization method that can learn discriminative ear features, which can effectively mitigate the drawbacks of DropBlock. AccDrop first randomly selects the feature blocks in the ear feature map. Subsequently, a drop operation is performed on the top-zth percentile elements, and the selected feature

blocks generate drop masks with adaptive shapes. The top-zth percentile elements for the discard operation are selected based on the values in the ear feature map. The image pixel values in the feature map are continuous and neighboring pixels have similar values. Therefore, AccDrop tends to encourage the network to consider inconspicuous valid ear features when the top-zth percentile elements are removed. Figure 4 shows the three stages of AccDrop. The AccDrop algorithm is shown in Algorithm 1. It has four parameters: block_size, γ, drop_prob, and z. Block_size is the size of the mask block. γ controls the number of activation units to be deleted, and the calculation of γ can be found in [18]. Drop_prob is the probability of retaining a unit in a traditional dropout. Z indicates a discard operation for the top-zth percentile element in the block to be deleted.

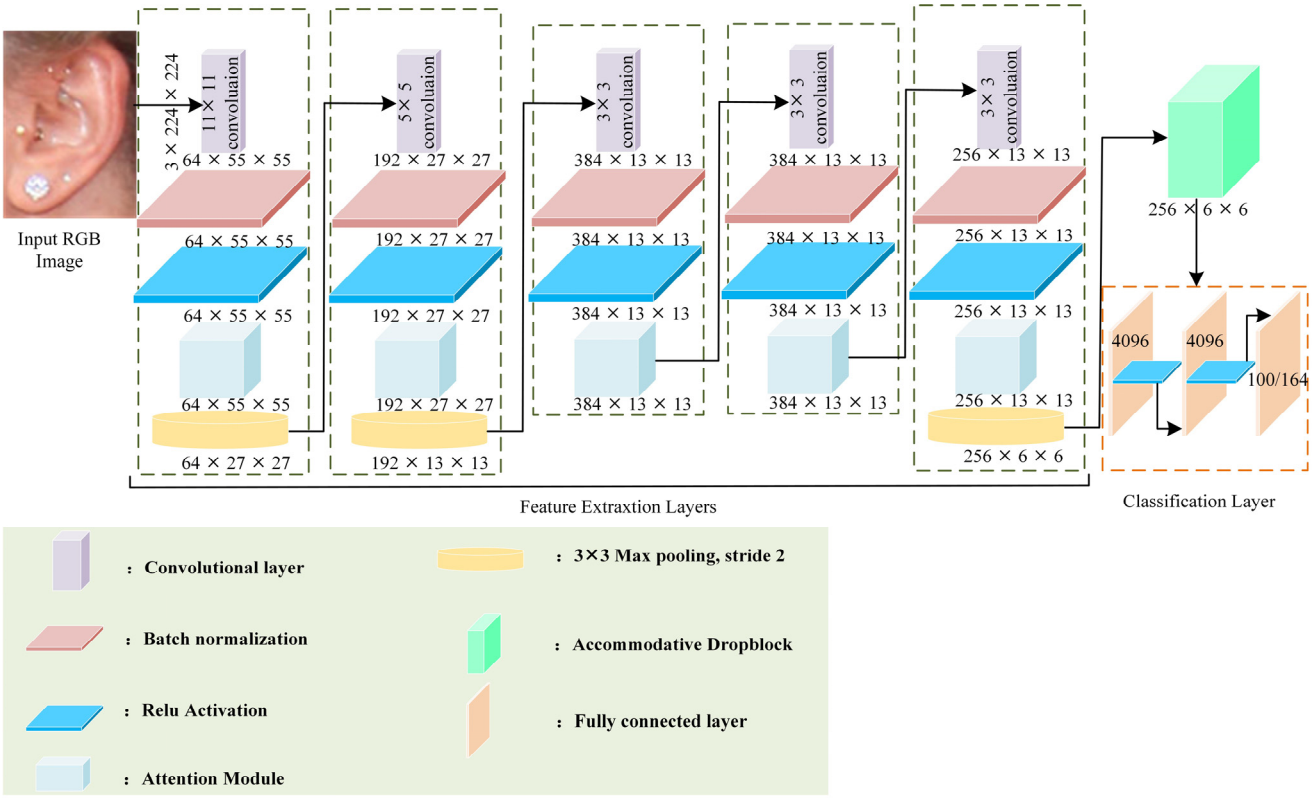


Figure 2. Architectural details of the proposed deep-feature-learning-based ear feature extraction model.

Algorithm 1: Accommodative DropBlock

Input: Feature map $F^{(n)}$ for the current layer, block_size, γ, drop_prob, z
Output: The next layer of features $F^{(n+1)}$
1: Randomly sample mask M: $M_{p,q} \sim$ Bernoulli(γ)
2: For each $M_{p,q}$, we create a square block centered at $M_{p,q}$ and of size block_size × block_size. Set the top-zth percentile elements of each square block to zero and the rest to one.
3: The mean $\overline{F^{(n)}}$ and Variance σ^2 of feature values are: $\overline{F^{(n)}} = \frac{1}{N}\sum_{i=1}^{N} F_i^{(n)}$, $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(F_i^{(n)} - \overline{F^{(n)}}\right)^2$. Z—Score Normalization: $F'^{(n)}_i = \frac{F_i^{(n)} - \overline{F^{(n)}}}{\sigma}$
4: Apply the mask: $F'^{(n)}_i = F'^{(n)}_i \times \mathrm{M}$
5: Scaling of output features: $F^{(n+1)} = F'^{(n)}_i \times count(M)/count\_ones(M)$

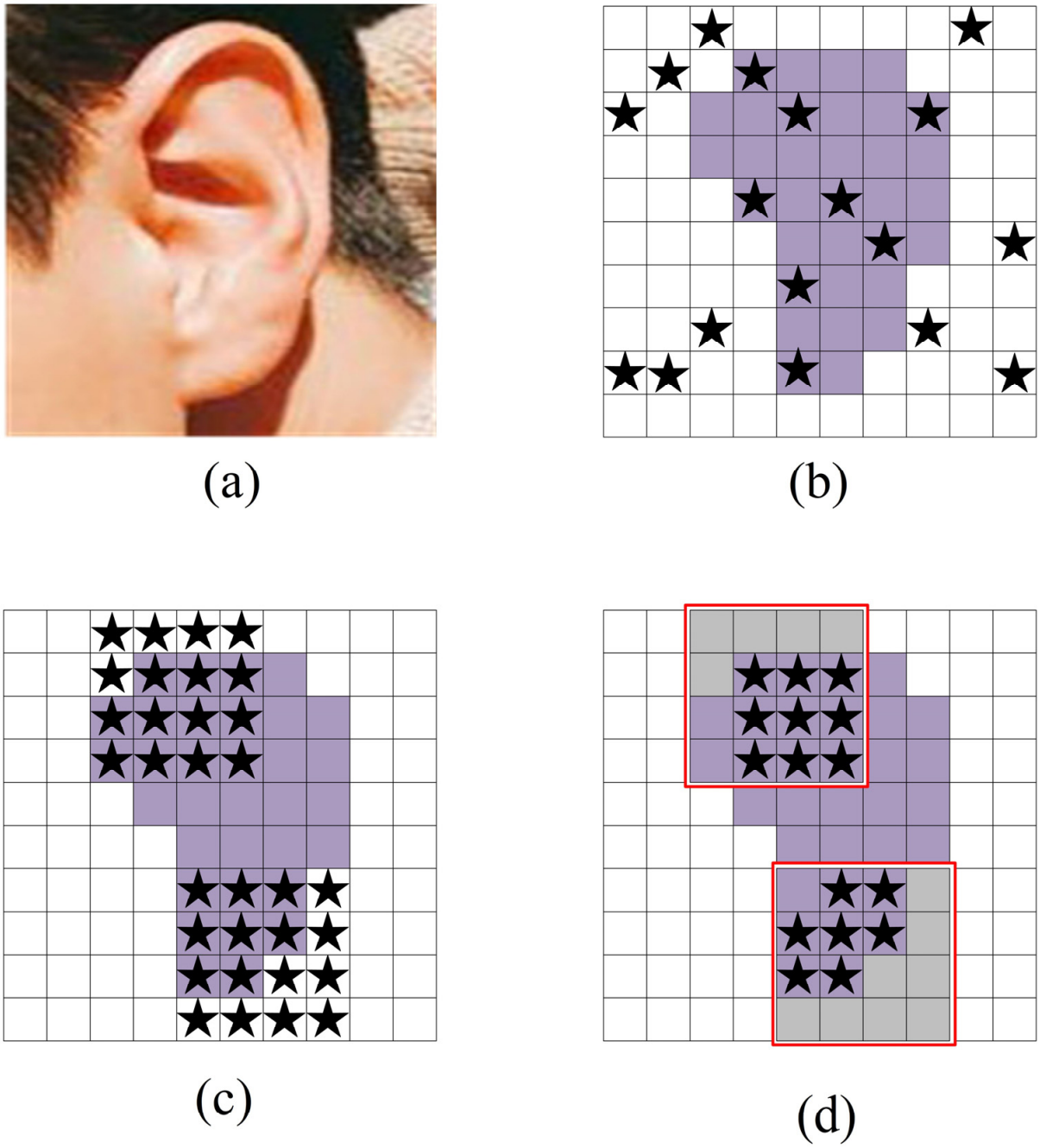


Figure 3. (**a**) Image of the input network; (**b**) Dropout; (**c**) DropBlock; and (**d**) proposed AccDrop. The regions marked using purple squares are activation units with semantic information, and black pentagram markers represent dropout operations. The elements at adjacent locations in the ear feature map share semantic information spatially. The elements adjacent to the dropped activation units still retain the semantic information at that location, causing the dropout to easily ignore spatial features. DropBlock is a structured regularization method that puts the units in adjacent regions of the feature map together to drop out. However, it will affect the feature learning ability of the network, resulting in some meaningful feature information being lost. The proposed AccDrop generates drop masks with adaptive shapes, which makes the model pay more attention to spatial information and allows it to effectively learn discriminative ear features in the face of an unconstrained ear database in realistic scenarios.

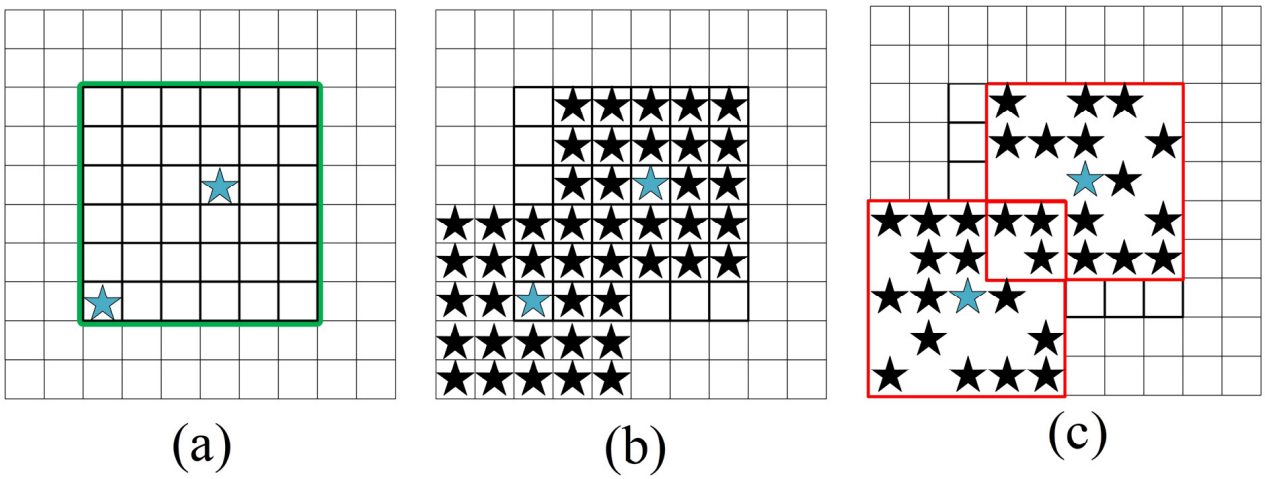


Figure 4. (**a**) We sample the mask M on each feature map in an operation similar to DropBlock. (**b**) The sampled elements are expanded into square blocks of size block_size × block_size. (**c**) The top-z-th percentile element of each square block is subjected to the drop operation.

3.2. Attention Module

The channel-based and spatial domains are currently the two most dominant attention mechanisms in computer vision. It is worth noting that the combination of the channel domain and spatial domain facilitates feature selection during information processing. Since the ear images in the dataset used in this paper have large inter- and intra-class differences, invalid features such as occlusion, background, etc., in the ear images can interfere with recognition performance. Therefore, we introduce a simple and parameter-free attention mechanism at the back end of the convolutional layer of the feature extraction network to refine the ear features output from the convolutional layer using 3D weights, so that the ear recognition system can learn the ear neurons of interest. The structure diagram is shown in Figure 5. The importance of each neuron is different, and the assigned weights should be unique. Among them, the neurons with higher importance show significant spatial inhibition effects. The linear difference between other neurons and the target neuron is derived by defining the energy function for each neuron, as shown in Equation (1).

$$e_t(w_t, \mathrm{b}_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 + (y_t - \hat{t})^2 \tag{1}$$

All values in Equation (1) are scalars. $\hat{t} = w_t t + \mathrm{b}_t$ and $\hat{x}_i = w_t x_i + \mathrm{b}_t$ are linear transformations of t and x_i, respectively, where t is the target neuron in a single channel of the input feature $X \in \mathbb{R}^{C \times H \times W}$ and x_i is the other neurons in a single channel of the input feature $X \in \mathbb{R}^{C \times H \times W}$. w_t and b_t are weight and bias transformations. $M = H \times W$ is the number of neurons on that channel and i is the index in the spatial dimension. Equation (1) reaches a minimum when all other $\hat{x}_i$ is equal to y_0 and $\hat{t}$ is equal to y_t, where y_0 and y_t do not have the same value. Minimizing this equation is equivalent to finding the linear separability between all other neurons in the same channel and the target neuron t. For simplicity, we add a regularizer to Equation (1) and use the binary notation for y_t and y_0 (i.e., 1 and -1). The energy function is finally expressed as Equation (2).

$$e_t(w_t, \mathrm{b}_t, y, x_i) = (1 - (w_t t + \mathrm{b}_t))^2 + \lambda {w_t}^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_t x_i + \mathrm{b}_t))^2 \tag{2}$$

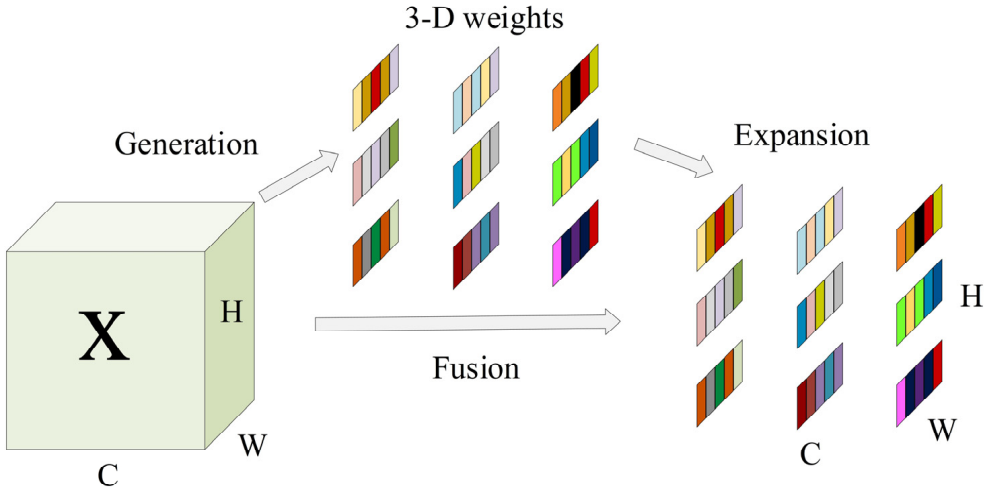


Figure 5. This attention mechanism allows the direct estimation of 3D weights to refine the ear features further. The same color indicates using a single scalar for each point on the ear feature. X is the input feature, H is the height of the input data, C is the number of channels, and W is the input data width.

The energy functions are M on each channel. Using an iterative solver, such as SGD, to compute all these functions can be complicated. It is worth noting that Equation (2) has fast closed-form solutions for w_t and b_t, which can be obtained from Equations (3) and (4) as follows:

$$w_t = -\frac{2(t-u_t)}{2{\sigma_t}^2 + 2\lambda + (t-u_t)^2} \tag{3}$$

$$b_t = -\frac{1}{2} w_t (u_t + t) \tag{4}$$

${\sigma_t}^2 = \frac{1}{M-1}\sum_i^{M-1}(x_i - \mu_t)^2$ and $\mu_t = \frac{1}{M-1}\sum_{i=1}^{M-1} x_i$ denote the calculation of the variance and mean of all neurons excluding t in that channel, respectively, which can significantly reduce the computational cost by avoiding the repeated calculation of μ and σ at each location. We calculate the minimum energy using Equation (5).

$$e_t^* = \frac{4(\lambda + \hat{\sigma}^2)}{2\lambda + 2\hat{\sigma}^2 + (t-\hat{u})^2} \tag{5}$$

where $\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i$ and $\hat{\sigma}_t{}^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i - \hat{\mu})^2$. From Equation (5), if the energy e_t^* is lower, it indicates that the difference between the neuron t and the surrounding neurons is greater, and the neuron is more important. The importance of each neuron can be expressed using $\frac{1}{e_t^*}$. We use the scaling operator for feature refinement and finally derive an energy function as shown in Equation (6).

$$\widetilde{X} = X \odot sigmoid\left(\frac{1}{E}\right) \tag{6}$$

where E groups all e_t^* in spatial dimensions and channels. Excessively large values in E can be restricted by *sigmoid*.

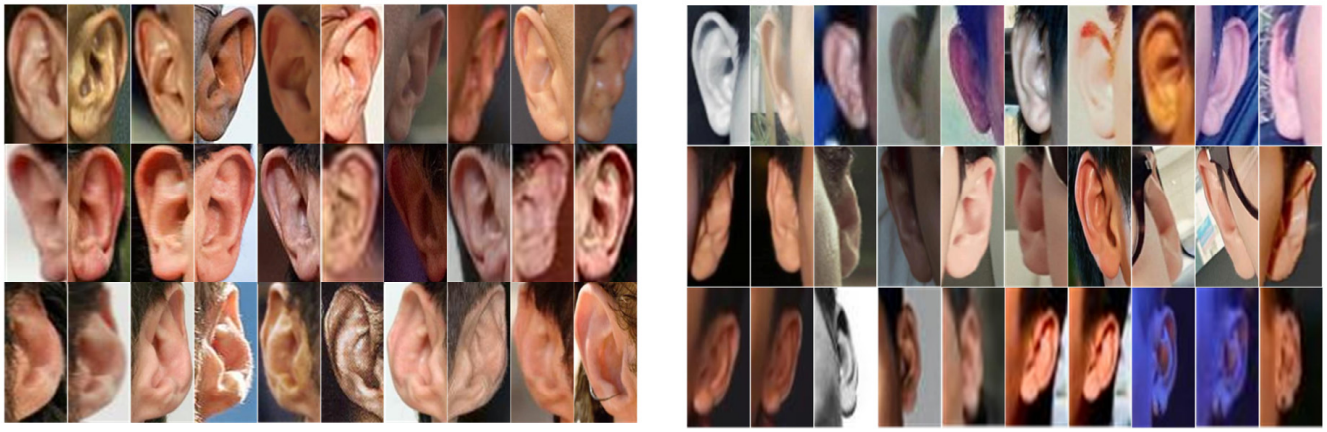
4. Experiments and Results

4.1. Datasets and Experiment Setup

4.1.1. Dataset Introduction

The Annotated Web Ears (AWE) [43–45] ear database consists of 100 subjects with 10 ear images per subject, for a total of 1000 ear images, and was provided by Ljubljana

University. The ear images are challenged by occlusions caused by hair and ear ornaments as well as by angular variations. The EARVN1.0 [46] unconstrained ear database contains 28,412 images from 164 subjects and is a large unconstrained ear database. These images are highly variable in terms of angle, scale, resolution, and lighting. Most of the ear images face challenges due to hair, background, and ornament occlusion. We randomly selected three subjects from each database and selected 10 ear images from each presentation subject, as shown in Figure 6.



(a)AWE dataset (b)EARVN1.0 dataset

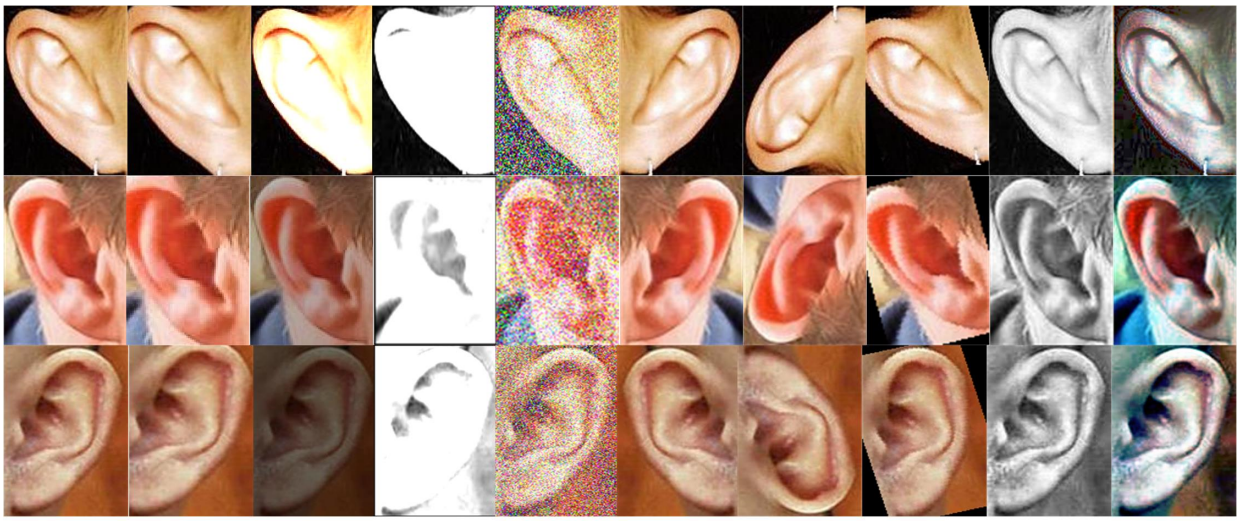
Figure 6. Ear images of three subjects were randomly selected from two ear databases, AWE and EARVN1.0, for presentation, with ten images presented for each subject. These ear images had large variations in angle, resolution, and brightness, and had the challenge of being obscured by jewelry and hair.

4.1.2. Data Augmentation

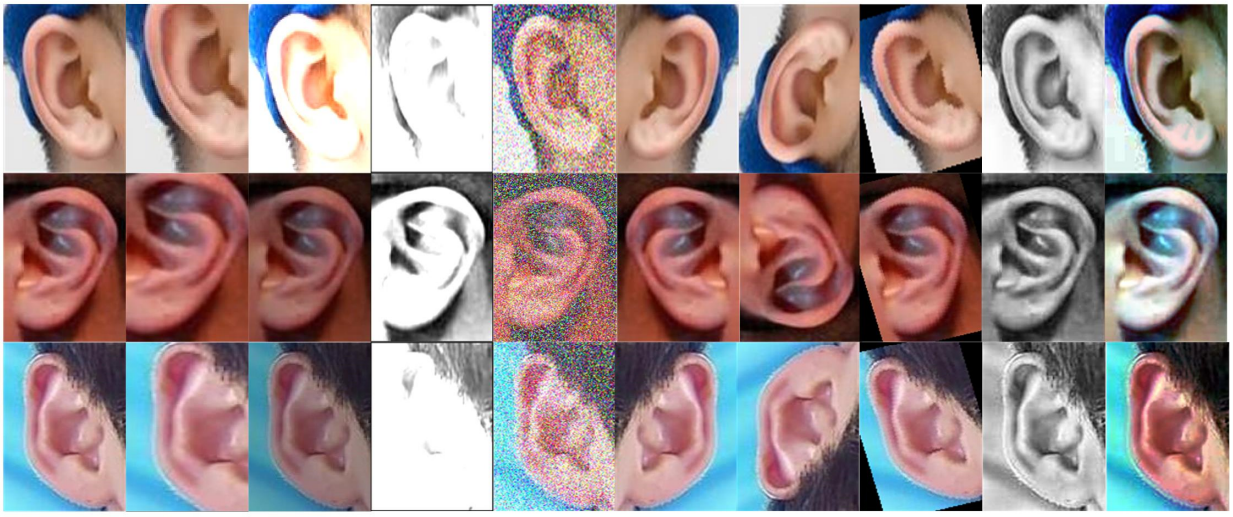
In deep learning network training, the more sufficient samples are trained, the stronger the generalization of the network model and the higher the robustness. Data augmentation can solve the sample imbalance problem, increase the noise data, improve the robustness of the model, increase the amount of data for training, improve the generalization ability of the model, and to some extent, solve the overfitting problem. We enhanced the training data with image processing techniques such as histogram equalization, cropping, and brightness increase. The augmented ear images are shown in Figure 7.

4.1.3. Parameter Initialization and Evaluation Metrics

The experiments in this paper are conducted with NVIDIA Tesla V100 SXM2 16G servers as hardware support and are based on the Pytorch open-source framework. This server was released by NVIDIA Corporation in Santa Clara, CA, USA. We set up a learning rate descent method with cosine annealing. The optimizer for this experiment is stochastic gradient descent (SGD), and the momentum, weight recession, and batch size are set to 0.9, 0.0001, and 32, respectively. The number of training iterations for all experiments is set to 500 rounds. Finally, we evaluate the performance of the proposed ear recognition system using Rank-1 (R1) recognition rate, Rank-5 (R5) recognition rate, and cumulative matching feature (CMC) curves.



(a)AWE



(b)EarVN1.0

Figure 7. We augmented the two ear databases with data, and the image processing techniques used were vertical crop, luminance increase, Gaussian blur, Gaussian noise, horizontal flip, vertical flip, rotation, limited contrast adapted histogram equalization, and color histogram equalization.

4.2. Analysis of block_size

Block_size is an essential parameter in AccDrop that affects recognition performance. In this experiment, we refer to the base network without inserting the AccDrop and attention module as EARNet, and the network with only the AccDrop inserted as AD-EARNet. Neighborhood noise can easily interfere with contextual information and thus affect recognition performance. Figures 8 and 9 show the recognition performance of the AWE and EARVN1.0 ear databases under different block_size. To evaluate the recognition performance, we set the block_size to 3, 5, 7, 9, and 11. It can be observed that the recognition performance of AD-EARNet is better than that of EARNet. The best recognition performance is achieved when block_size = 7. Therefore, we set the block_size to 7 in all subsequent experiments. As block_size increases from 3 to 7, more contextual information can be taken into account, and the recognition performance improves. However, too large a mask block can lead to continuous blank areas, making the network training less stable.

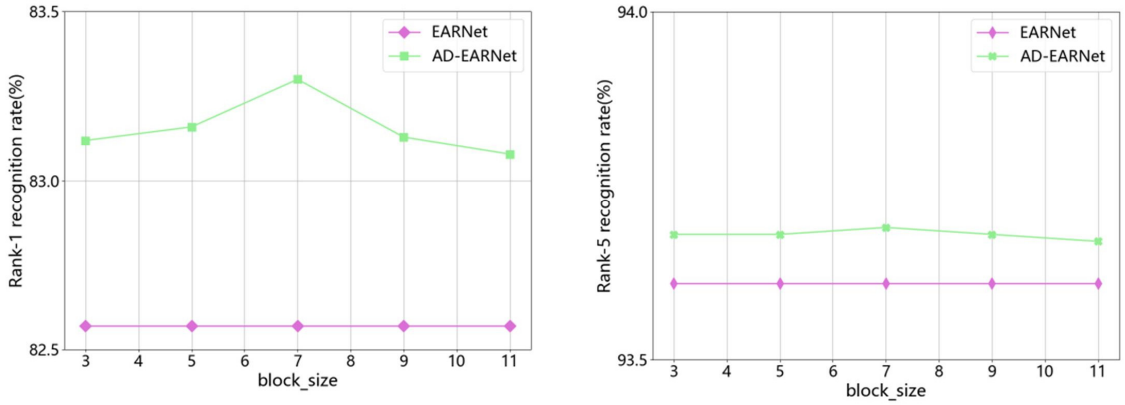


Figure 8. Relationship between R1, R5 recognition rate, and block_size of the AWE database.

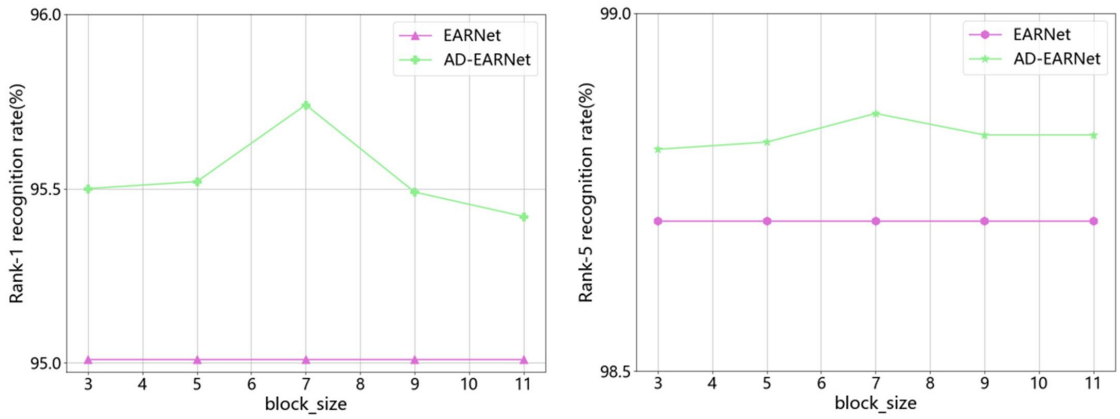


Figure 9. Relationship between R1, R5 recognition rate, and block_size of the EARVN1.0 database.

4.3. The Impact of drop_prob

The parameter drop_prob indicates the probability of retaining a unit in dropout, which is also an essential parameter in AccDrop. In the experiments, drop_prob varies from 0.75 to 0.95. Figures 10 and 11 illustrate the effect of drop_prob on the recognition performance of the two ear databases, AWE and EARVN1.0. It can be seen that the recognition performance of AD-EARNet is consistently better than that of EARNet. When drop_prob = 0.9, the best recognition rates are achieved for R1 and R5. When drop_prob is small, it will impact the feature learning process. When drop_prob is too large, it will destroy the stability of network learning. Therefore, drop_prob was chosen to be 0.9 in the later experiments.

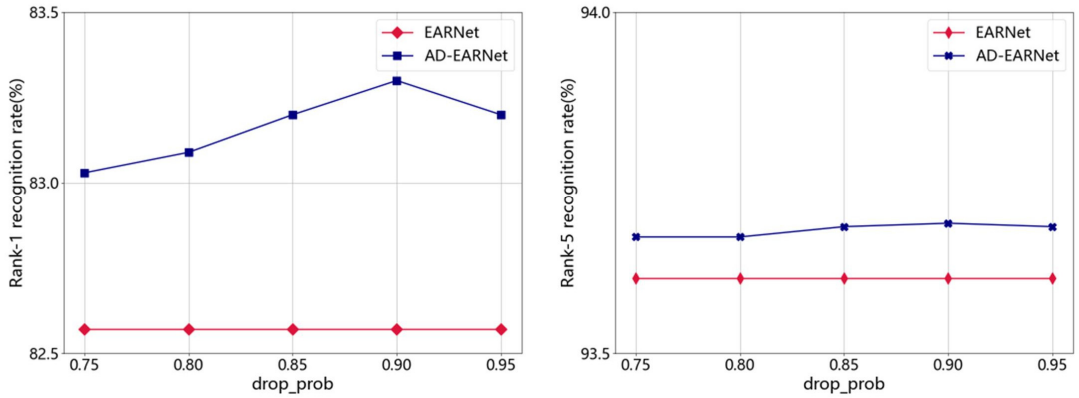


Figure 10. Relationship between R1, R5 recognition rate, and drop_prob of the AWE database.

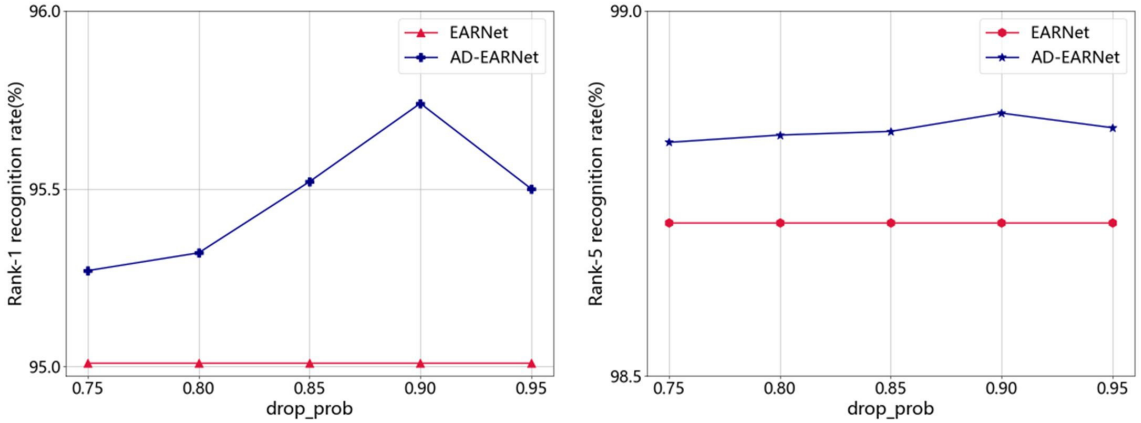


Figure 11. Relationship between R1, R5 recognition rate, and drop_prob of the EARVN1.0 database.

4.4. Analysis of z

The parameter z is crucial in AccDrop, indicating that the mask block's top zth percentile elements are dropped. Figures 12 and 13 illustrate the effect of z on the recognition performance of the AWE and EARVN1.0 ear databases. We set z to 35, 40, 45, and 50 to evaluate the recognition performance. It can be observed that when z is set to 40, the best recognition rates are achieved for R1 and R5. When setting a largerz, too many valuable features are discarded, causing the network to learn interference features such as background. When z is too small, it means that the network discards a small number of features, which can lead to the occurrence of network overfitting. So, in the following experiments, we set z = 40.

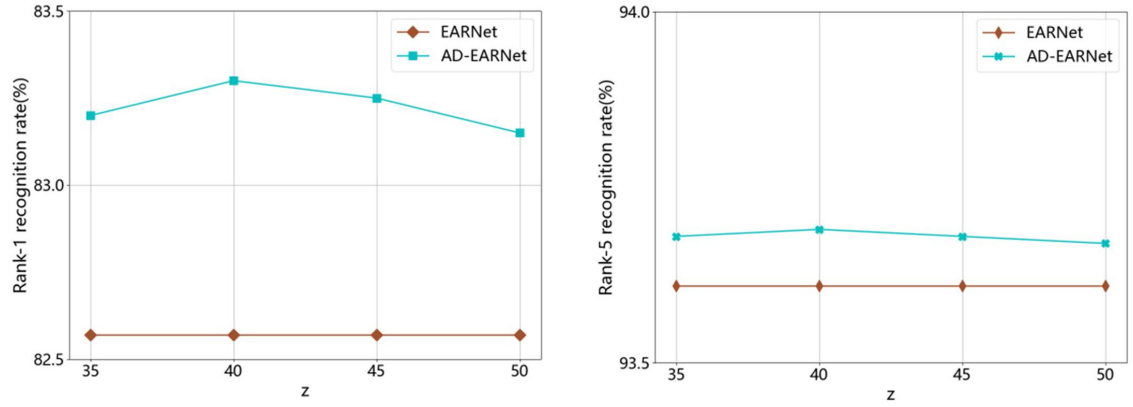


Figure 12. Relationship between R1, R5 recognition rate, and z of the AWE database.

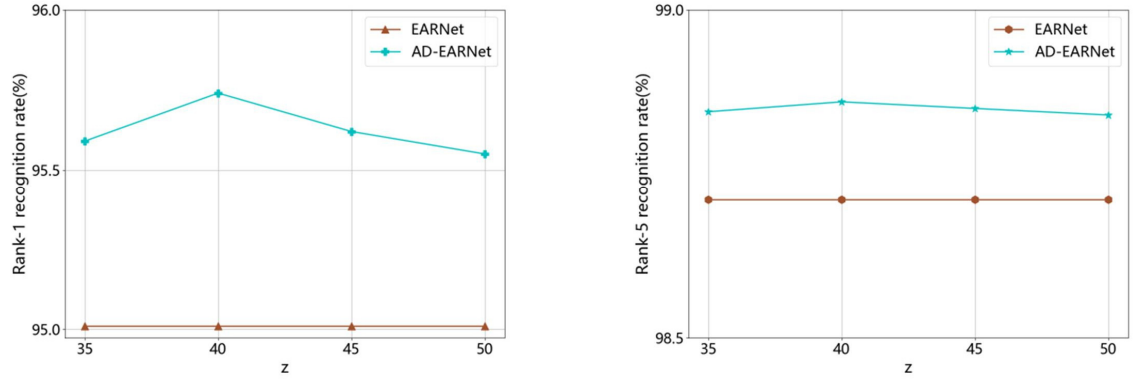


Figure 13. Relationship between R1, R5 recognition rate, and z of the EARVN1.0 database.

4.5. Comparison with Other Commonly Used Regularization Methods

In this experiment, we refer to the network applying the DropBlock regularization method as DB-EARNet. Experimental results show that our proposed AccDrop facilitates ear recognition. We chose Dropout and DropBlock to compare with AccDrop. As shown in Figure 14 and Table 1, the recognition performance of our proposed AccDrop is superior to both Dropout and DropBlock. Among the three regularization methods, Dropout is the least effective. Since the adjacent position elements in the ear feature map share semantic information spatially, although a cell is dropped, the elements adjacent to it can still retain the semantic information of that position, and the information can still circulate in the convolutional network. DropBlock is a structural form of regularization method that drops the cells together in adjacent regions of the feature map. However, it will affect the feature learning ability of the network to some extent. AccDrop generates drop masks with adaptive shapes, which can effectively learn discriminative ear features in the face of an unconstrained ear database in realistic scenarios.

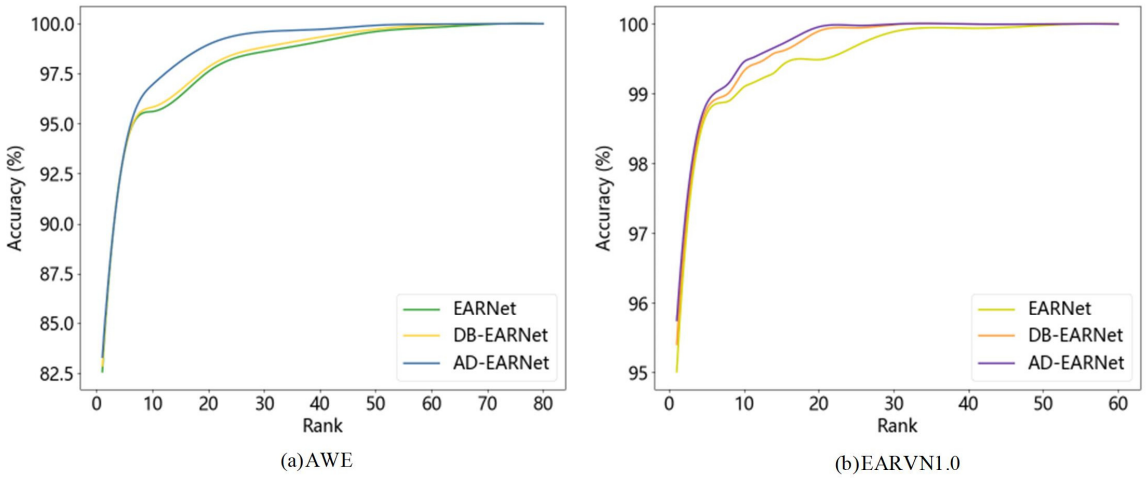


Figure 14. The CMC curves compare the effects of different regularization methods on the recognition performance of the system.

Table 1. We compare the quantitative performance metrics (R1, R5) when different regularization methods are used.

Model	AWE		EARVN1.0	
	R1 (%)	R5 (%)	R1 (%)	R5 (%)
EARNet	82.57	93.61	95.01	98.71
DB-EARNet	82.87	93.63	95.40	98.78
AD-EARNet	83.30	93.69	95.74	98.86

4.6. The Impact of λ

The parameter λ is the critical parameter in SimAM, and its details can be found in [19]. In the experiments described in this paper, we refer to the network with only the attention module inserted as AM-EARNet. We set the parameter λ to vary from 10^{-6} to 10^{-1}. Figures 15 and 16 show the effect of λ on the recognition performance of both AWE and EARVN1.0 ear databases. From the figures, we can conclude that SimAM can significantly improve the recognition performance by using a wide range of λ (ranging from 10^{-6} to 10^{-1}). When λ is set to 10^{-4}, the best recognition rates are achieved for R1 as well as R5. Therefore, we set $\lambda = 10^{-4}$ in the later experiments.

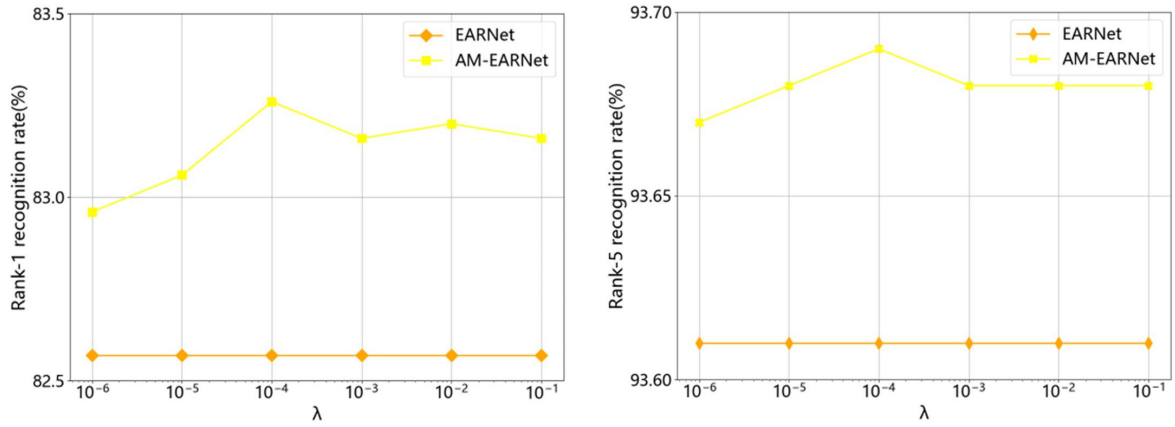


Figure 15. Relationship between R1, R5 recognition rate, and λ of the AWE database.

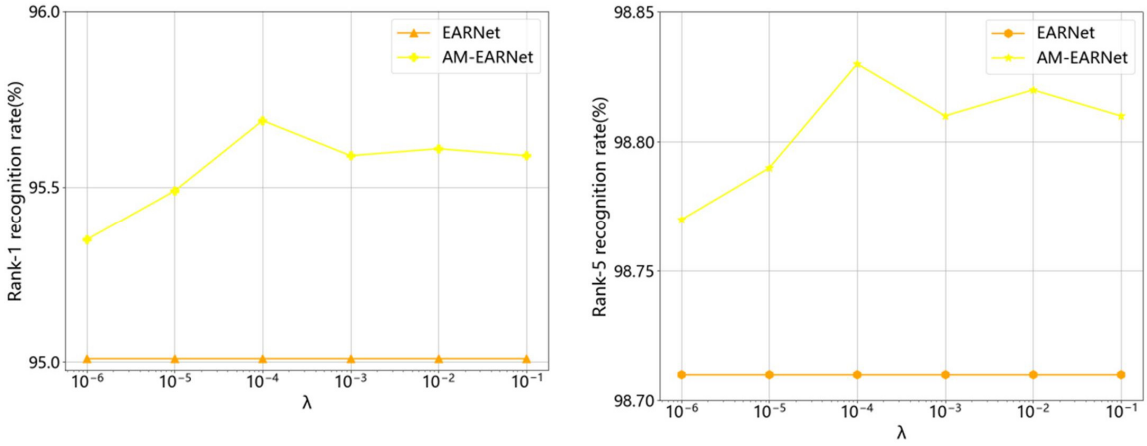


Figure 16. Relationship between R1, R5 recognition rate, and λ of the EARVN1.0 database.

4.7. Ablation Experiment

In this subsection, we verify the effects of the accommodative DropBlock (AccDrop) and attention module on the model recognition performance by ablation experiments. The specific experimental results are shown in Table 2. From the experimental results, we can see that inserting AccDrop or the attention module alone will improve the recognition performance of the model by a small margin. However, the model performance is optimized only when AccDrop and the attention module are inserted into EARNet at the same time. Figure 17 shows the CMC curves of the ablation experiment on different ear databases.

Table 2. We compared the quantitative performance metrics (R1, R5) under different ablation experiments.

Model	AWE		EARVN1.0	
	R1 (%)	R5 (%)	R1 (%)	R5 (%)
EARNet	82.57	93.61	95.01	98.71
AM-EARNet	83.26	93.69	95.69	98.83
AD-EARNet	83.30	93.69	95.74	98.86
Proposed	83.87	93.74	96.52	99.12

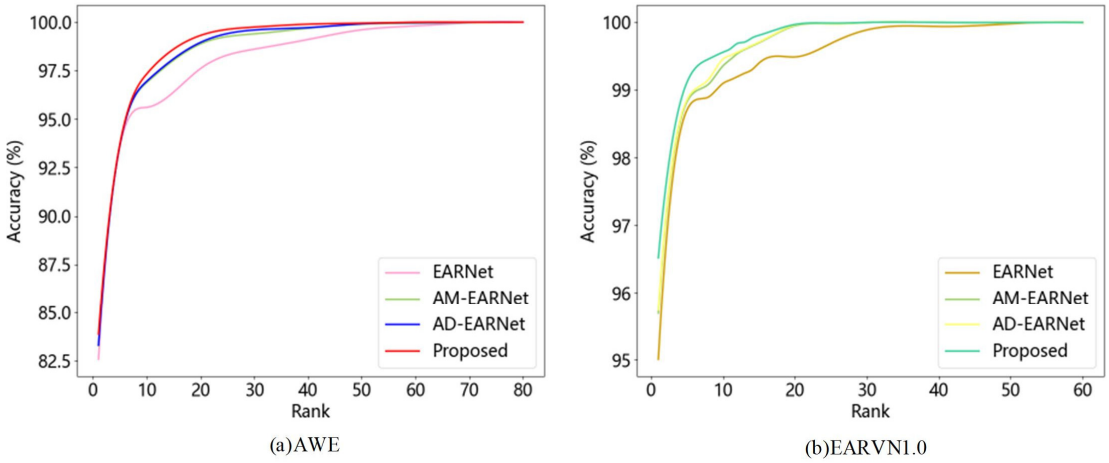


Figure 17. CMC curves comparing the identification performance of different models in ablation experiments.

4.8. Analysis of the Role of Blockchain Protection

We tampered with the ear feature template to verify the system's security. We measured the impact of tampering on the proposed network (classified as whether the blockchain protects it or not). Figure 18 shows the CMC curves of the AWE and EARVN1.0 databases before and after template tampering on the proposed network. When the proposed network is not protected by blockchain, the recognition performance is significantly degraded after template tampering. Since the blockchain prevents template tampering, the network protected by the blockchain maintains the same recognition performance as the network that has not been tampered with, fully demonstrating the combined advantages of the blockchain in the ear recognition system. Table 3 shows the Rank-1 recognition rate of two databases using the proposed ear recognition system (divided into whether they are protected by blockchain or not).

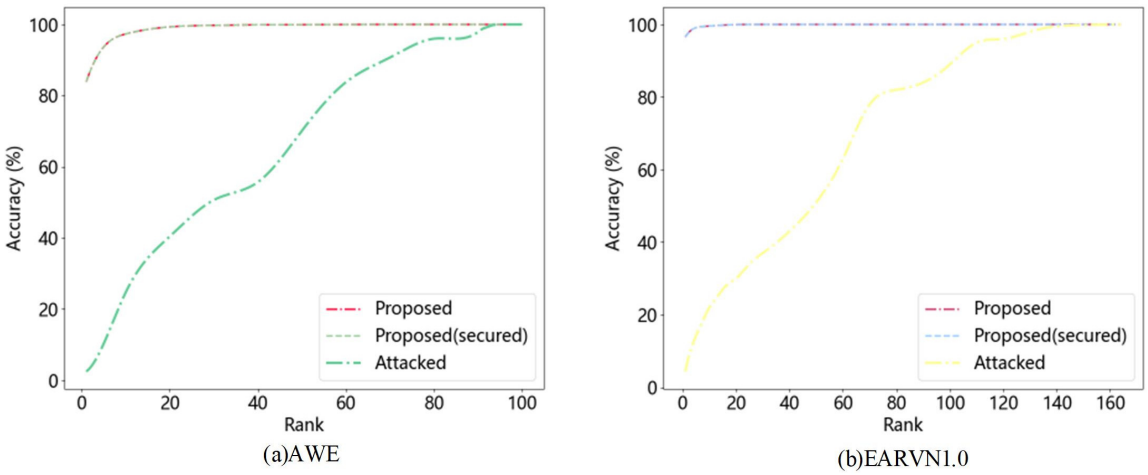


Figure 18. CMC curves of the AWE and EARVN1.0 databases before and after template tampering with two ear identification systems (divided into whether they are protected by blockchain or not).

4.9. Visual Explanations

One method often used for visual interpretation is the gradient-weighted class activation map (GradCAM) [48]. It can provide a category-distinct interpretation of ear recognition, locating regions of interest in the ear image that embody semantic information about category gradients, further helping us to understand the predictions made by different models. We list some prediction cases where the Proposed and Proposed (secured)

models made correct predictions for the subjects and the EARNet and Attacked models made incorrect predictions for the subjects. Figure 19 (AWE) and Figure 20 (EARVN1.0) show the original image of the ear, EARNet localization results, Proposed localization results, Attacked localization results, and localization results of the Proposed (secured) model. Based on the results, we can see that the geometry of the ear is the most crucial region, and ignoring invalid features such as hair, accessories, etc., will lead to correct predictions.

Table 3. Rank-1 recognition rate of the two ear recognition systems (divided into whether they are blockchain protected) before and after ear feature template tampering.

Dataset	Rank-1 Recognition Rate (%)			
	Before Template Tampering		After Template Tampering	
	Proposed	Proposed (Secured)	Proposed	Proposed (Secured)
AWE	83.87	83.87	2.46	83.87
EARVN1.0	96.52	96.52	4.37	96.52

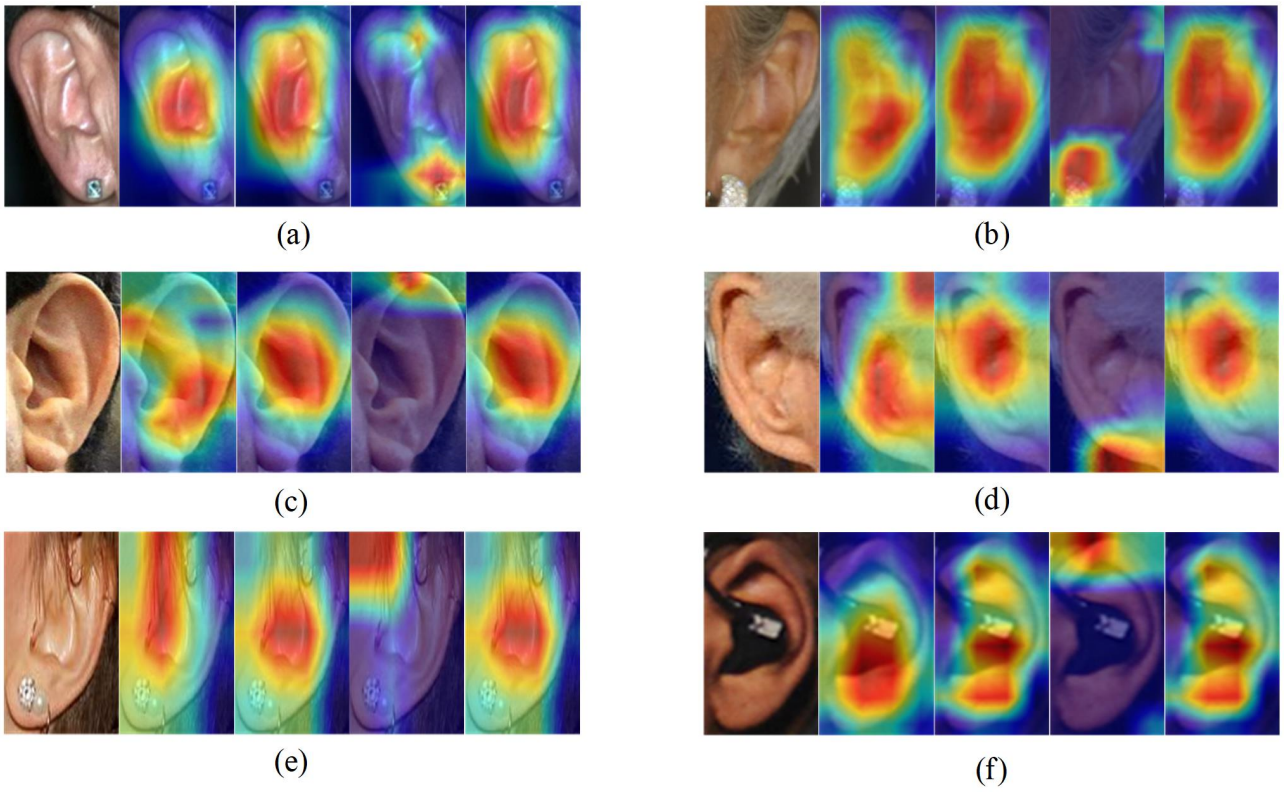


Figure 19. GradCAM visually interprets the ear category distinctions of different models on the AWE database to localize the region of interest and further help us understand the predictions made by different models. The visualization results of different models are shown in each subplot from left to right as the original image, EARNet model, Proposed model, Attacked model, and Proposed (secured) model. (**a**) EARNet only focuses on the middle part of the ear contour and ignores the top and bottom parts of the ear contour. Attacked only focuses on the earlobe and stud, which leads to incorrect predictions. (**b**) EARNet does not pay enough attention to the ear contour features in the upper part. Attacked focuses on earlobes and ear ornaments. (**c**) EARNet ignores the bulk ear contour features and only focuses on local ear features. Attacked focuses on local hair interference features. (**d**) Both EARNet and Attacked focus on hair interference features. (**e**) EARNet pays excessive attention to hair interference features. Attacked pays attention to hair and facial features. (**f**) EARNet focuses only on earplug and earlobe features. Attacked ignores the ear features in the middle and lower parts.

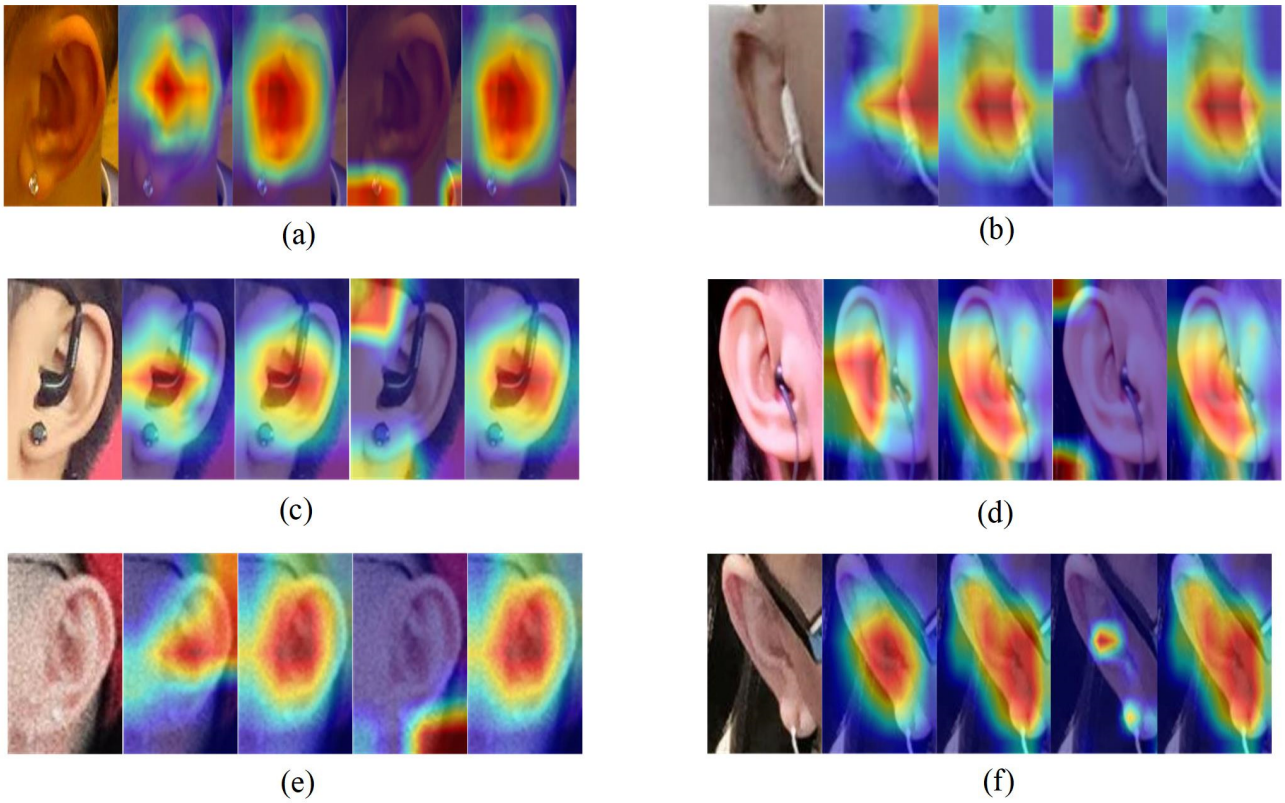


Figure 20. GradCAM visually interprets the ear category distinction of the different models on the EARVN1.0 database to localize the region of interest and further help us understand the predictions made by the different models. The visualization results of different models are shown in each subplot from left to right as the original image, EARNet model, Proposed model, Attacked model, and Proposed (secured) model. (**a**) EARNet focuses only on local ear contour features. Attacked focuses on invalid features such as earpieces and backgrounds. (**b**) EARNet focuses on earphones and face features. Attacked only focuses on background interference features. (**c**) EARNet is disturbed by earphone features. Attacked is disturbed by background features. (**d**) EARNet ignores the upper and lower ear contour features. Attacked is disturbed by features such as hair as well as background. (**e**) Both EARNet and Attacked are disturbed by invalid features such as the background. EARNet even ignores the upper and lower parts of ear features. (**f**) Both EARNet and Attacked ignore some ear features.

4.10. Compared with Other Methods

We compared our method with existing ear recognition techniques based on Rank-1 and Rank-5 recognition accuracy in Table 4. From the comparison results in Table 4, it is clear that our proposed method has the best recognition performance.

Table 4. The proposed method is compared with other representative methods based on the quantitative performance metrics Rank-1 and Rank-5 recognition accuracy.

Dataset	Method	R1 (%)	R5 (%)
AWE	Hassaballah et al. [49]	49.60	-
	Emersic et al. [44]	49.60	-
	Dodge et al. [50]	56.35	74.80
	Dodge et al. [50]	68.50	83.00
	Dodge et al. [50]	80.03	93.48
	Zhang et al. [51]	50.00	70.00
	Emersic et al. [52]	62.00	80.35
	Khaldi et al. [53]	50.53	76.35

Table 4. *Cont.*

Dataset	Method	R1 (%)	R5 (%)
AWE	Hassaballah et al. [54]	54.10	-
	Khaldi et al. [55]	48.48	-
	Khaldi et al. [56]	51.25	-
	Alshazly et al. [57]	67.25	84.00
	Omara et al. [58]	78.13	-
	Kacar et al. [59]	47.80	72.10
	Chowdhury et al. [60]	50.50	70.00
	Hansley et al. [61]	75.60	90.60
	Aiadi et al. [62]	82.50	-
	Xue bin et al. [63]	83.52	93.71
	Proposed (secured)	83.87	93.74
EARVN1.0	Ramos-Cooper et al. [64]	92.58	97.88
	Alshazly et al. [65]	93.45	98.42
	Xue bin et al. [63]	96.10	99.28
	Proposed (secured)	96.52	99.12

5. Conclusions

This paper proposes an ear recognition system against software attacks using deep feature learning and blockchain protection. Our proposed AccDrop generates drop masks with adaptive shapes for ear images and does not discard fixed-sized ear regions, effectively improving the recognition performance of the ear recognition system in the face of unconstrained ear databases in realistic scenarios. We use the attention mechanism to infer the 3D attention weights of the output feature maps of the convolutional layer, which not only does not increase the parameters and computation of the network but also enables the network to learn more discriminative neurons. We introduce a public blockchain when exchanging the ear feature template database. The use of Merkle tree blockchain storage technology to enhance the protection of ear feature templates dramatically improves the security of the ear recognition system and protects users' privacy. We evaluated the model on representative AWE and EARVN 1.0 unconstrained ear databases and achieved Rank-1 (R1) recognition accuracies of 83.87% and 96.52%, respectively, significantly better than most existing ear recognition systems. Finally, the Grad-CAM technique was used to visualize and interpret our model. From the visualization results, our model can effectively learn more discriminative ear features. In the future, we will continue to optimize the ear feature extraction stage method to further improve the system's recognition performance on unconstrained ear databases, which will significantly help surveillance security and financial security.

Author Contributions: All of the authors made significant contributions to this work. X.X., Y.L. and L.L. devised the approach and analyzed the data; Y.L., C.L. and L.L. helped design the experiments and provided advice for the preparation and revision of the work; Y.L. and C.L. performed the experiments; X.X. and Y.L. wrote this manuscript. 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); the Scientific Research Project of the Education Department of Shaanxi Province (21JK0921); the Key Research and Development Program of Shaanxi Province, under grant 2017GY-071; the Technical Innovation Guidance Special Project of Shaanxi Province, under grant 2017XT-005; and the research program of Xian Yang City under grant 2017K01-25-3.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: AWE can be found at http://awe.fri.uni-lj.si/ (accessed on 12 November 2021). EARVN1.0 can be found at https://data.mendeley.com/datasets/yws3v3mwx3/3 (accessed on 12 November 2021).

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

References

1. Appati, J.K.; Nartey, P.K.; Yaokumah, W.; Abdulai, J.-D. A systematic review of fingerprint recognition system development. *Int. J. Softw. Sci. Comput. Intell. IJSSCI* **2022**, *14*, 1–17. [CrossRef]
2. Babu, G.; Khayum, P.A. Elephant herding with whale optimization enabled ORB features and CNN for Iris recognition. *Multimed. Tools Appl.* **2022**, *81*, 5761–5794. [CrossRef]
3. Kaur, P.; Krishan, K.; Sharma, S.K.; Kanchan, T. Facial-recognition algorithms: A literature review. *Med. Sci. Law* **2020**, *60*, 131–139. [CrossRef] [PubMed]
4. Manley, G.A. An evolutionary perspective on middle ears. *Hear. Res.* **2010**, *263*, 3–8. [CrossRef] [PubMed]
5. Abate, A.F.; Nappi, M.; Riccio, D.; Ricciardi, S. Ear recognition by means of a rotation invariant descriptor. In Proceedings of the 18th International Conference On Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; pp. 437–440.
6. Alkababji, A.M.; Mohammed, O.H. Real time ear recognition using deep learning. *Telkomnika Telecommun. Comput. Electron. Control* **2021**, *19*, 523–530. [CrossRef]
7. Galbally, J.; Haraksim, R.; Beslay, L. A study of age and ageing in fingerprint biometrics. *IEEE Trans. Inf. Secur.* **2018**, *14*, 1351–1365. [CrossRef]
8. Lanitis, A. A survey of the effects of aging on biometric identity verification. *Int. J. Biom.* **2010**, *2*, 34–52. [CrossRef]
9. Alexander, K.S.; Stott, D.J.; Sivakumar, B.; Kang, N. A morphometric study of the human ear. *J. Plast. Reconstr. Aesthetic Surg.* **2011**, *64*, 41–47. [CrossRef]
10. Krishan, K.; Kanchan, T.; Thakur, S. A study of morphological variations of the human ear for its applications in personal identification. *Egypt. J. Forensic Sci.* **2019**, *9*, 6. [CrossRef]
11. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [CrossRef]
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
13. Wang, J.; Gao, F.; Dong, J.; Du, Q. Adaptive dropblock-enhanced generative adversarial networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5040–5053. [CrossRef]
14. Jain, A.K.; Nandakumar, K.; Ross, A. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognit. Lett.* **2016**, *79*, 80–105. [CrossRef]
15. Goel, A.; Agarwal, A.; Vatsa, M.; Singh, R.; Ratha, N. Securing CNN model and biometric template using blockchain. In Proceedings of the 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), Tampa, FL, USA, 23–26 September 2019; pp. 1–7.
16. Delgado-Mohatar, O.; Fierrez, J.; Tolosana, R.; Vera-Rodriguez, R. Blockchain meets biometrics: Concepts, application to template protection, and trends. *arXiv* **2020**. [CrossRef]
17. Zhang, W.; Yuan, Y.; Hu, Y.; Nandakumar, K.; Chopra, A.; Sim, S.; De Caro, A. Blockchain-Based Distributed Compliance in Multinational Corporations' Cross-Border Intercompany Transactions: A New Model for Distributed Compliance Across Subsidiaries in Different Jurisdictions. In Proceedings of the Advances in Information and Communication Networks: Proceedings of the 2018 Future of Information and Communication Conference (FICC), San Francisco, CA, USA, 14–15 March 2019; Volume 2, pp. 304–320.
18. Ghiasi, G.; Lin, T.-Y.; Le, Q.V. Dropblock: A regularization method for convolutional networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1–11.
19. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Online, 8–24 July 2021; pp. 11863–11874.
20. Merkle, R.C. A digital signature based on a conventional encryption function. In Proceedings of the Conference on the Theory and Application Of Cryptographic Techniques, Santa Barbara, CA, USA, 16–20 August 1987; pp. 369–378.
21. Moreno, B.; Sanchez, A.; Vélez, J.F. On the use of outer ear images for personal identification in security applications. In Proceedings of the IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology (Cat. No. 99ch36303), Madrid, Spain, 5–7 October 1999; pp. 469–476.
22. Choras, M.; Choras, R.S. Geometrical algorithms of ear contour shape representation and feature extraction. In Proceedings of the Sixth international Conference on Intelligent Systems Design and Applications, Jian, China, 16–18 October 2006; pp. 451–456.
23. Choraś, M. Perspective methods of human identification: Ear biometrics. *Opto-Electron. Rev.* **2008**, *16*, 85–96. [CrossRef]
24. Dong, J.; Mu, Z. Multi-pose ear recognition based on force field transformation. In Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application, Shanghai, China, 20–22 December 2008; pp. 771–775.
25. Chang, K.; Bowyer, K.W.; Sarkar, S.; Victor, B. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1160–1165. [CrossRef]
26. Alaraj, M.; Hou, J.; Fukami, T. A neural network based human identification framework using ear images. In Proceedings of the TENCON 2010-2010 IEEE Region 10 Conference, Fukuoka, Japan, 21–24 November 2010; pp. 1595–1600.
27. Xie, Z.; Mu, Z. Ear recognition using LLE and IDLLE algorithm. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.

28. Yuan, L.; Mu, Z.-C.; Zhang, Y.; Liu, K. Ear recognition using improved non-negative matrix factorization. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; pp. 501–504.
29. Dewi, K.; Yahagi, T. Ear photo recognition using scale invariant keypoints. In Proceedings of the Computational Intelligence, San Francisco, CA, USA, 20–22 November 2006; pp. 253–258.
30. Kumar, A.; Wu, C. Automated human identification using ear imaging. *Pattern Recognit.* **2012**, *45*, 956–968. [CrossRef]
31. Nosrati, M.S.; Faez, K.; Faradji, F. Using 2D wavelet and principal component analysis for personal identification based on 2D ear structure. In Proceedings of the 2007 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, 25–28 November 2007; pp. 616–620.
32. Benzaoui, A.; Kheider, A.; Boukrouche, A. Ear description and recognition using ELBP and wavelets. In Proceedings of the 2015 International Conference on Applied Research In Computer Science And Engineering (Icar), Beiriut, Lebanon, 8–9 October 2015; pp. 1–6.
33. Jacob, L.; Raju, G. Ear recognition using texture features-a novel approach. In Proceedings of the Advances in Signal Processing and Intelligent Recognition Systems, Trivandrum, India, 13–15 March 2014; pp. 1–12.
34. Ahila Priyadharshini, R.; Arivazhagan, S.; Arun, M. A deep learning approach for person identification using ear biometrics. *Appl. Intell.* **2021**, *51*, 2161–2172. [CrossRef]
35. Štepec, D.; Emeršič, Ž.; Peer, P.; Štruc, V. Constellation-based deep ear recognition. In *Deep Biometrics*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 161–190.
36. Radhika, K.; Devika, K.; Aswathi, T.; Sreevidya, P.; Sowmya, V.; Soman, K. Performance analysis of NASNet on unconstrained ear recognition. In *Nature Inspired Computing for Data Science*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 57–82.
37. Ganapathi, I.I.; Ali, S.S.; Prakash, S.; Vu, N.-S.; Werghi, N. A survey of 3d ear recognition techniques. *ACM Comput. Surv.* **2023**, *55*, 1–36. [CrossRef]
38. Benzaoui, A.; Khaldi, Y.; Bouaouina, R.; Amrouni, N.; Alshazly, H.; Ouahabi, A. A comprehensive survey on ear recognition: Databases, approaches, comparative analysis, and open challenges. *Neurocomputing* **2023**, *537*, 236–270. [CrossRef]
39. Hadid, A.; Evans, N.; Marcel, S.; Fierrez, J. Biometrics systems under spoofing attack: An evaluation methodology and lessons learned. *IEEE Signal Process. Mag.* **2015**, *32*, 20–30. [CrossRef]
40. Nourmohammadi-Khiarak, J.; Pacut, A. An ear anti-spoofing database with various attacks. In Proceedings of the 2018 International Carnahan Conference on Security Technology (ICCST), Montreal, QC, Canada, 22–25 October 2018; pp. 1–5.
41. Toprak, I.; Toygar, Ö. Ear anti-spoofing against print attacks using three-level fusion of image quality measures. *Signal Image Video Process.* **2020**, *14*, 417–424. [CrossRef]
42. Sepas-Moghaddam, A.; Pereira, F.; Correia, P.L. Ear presentation attack detection: Benchmarking study with first lenslet light field database. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 2355–2359.
43. 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]
44. Emeršič, Ž.; Štruc, V.; Peer, P. Ear recognition: More than a survey. *Neurocomputing* **2017**, *255*, 26–39. [CrossRef]
45. 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]
46. Hoang, V.T. EarVN1. 0: A new large-scale ear images dataset in the wild. *Data Brief* **2019**, *27*, 104630. [CrossRef]
47. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
48. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
49. 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]
50. Dodge, S.; Mounsef, J.; Karam, L. Unconstrained ear recognition using deep neural networks. *IET Biom.* **2018**, *7*, 207–214. [CrossRef]
51. Zhang, Y.; Mu, Z.; Yuan, L.; Yu, C. Ear verification under uncontrolled conditions with convolutional neural networks. *IET Biom.* **2018**, *7*, 185–198. [CrossRef]
52. Emeršič, Ž.; Štepec, D.; Štruc, V.; Peer, P. Training convolutional neural networks with limited training data for ear recognition in the wild. *arXiv* **2017**. [CrossRef]
53. 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]
54. Hassaballah, M.; Alshazly, H.A.; Ali, A.A. Robust local oriented patterns for ear recognition. *Multimed. Tools Appl.* **2020**, *79*, 31183–31204. [CrossRef]
55. 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.

56. 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]
57. 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]
58. 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]
59. Kacar, U.; Kirci, M. ScoreNet: Deep cascade score level fusion for unconstrained ear recognition. *IET Biom.* **2019**, *8*, 109–120. [CrossRef]
60. Chowdhury, D.P.; Bakshi, S.; Pero, C.; Olague, G.; Sa, P.K. Privacy preserving ear recognition system using transfer learning in industry 4.0. *IEEE Trans. Ind. Inform.* **2022**, *19*, 6408–6417. [CrossRef]
61. Hansley, E.E.; Segundo, M.P.; Sarkar, S. Employing fusion of learned and handcrafted features for unconstrained ear recognition. *IET Biom.* **2018**, *7*, 215–223. [CrossRef]
62. Aiadi, O.; Khaldi, B.; Saadeddine, C. MDFNet: An unsupervised lightweight network for ear print recognition. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *14*, 13773–13786. [CrossRef] [PubMed]
63. Xu, X.; Liu, Y.; Cao, S.; Lu, L. An efficient and lightweight method for human ear recognition based on MobileNet. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 9069007. [CrossRef]
64. Ramos-Cooper, S.; Gomez-Nieto, E.; Camara-Chavez, G. VGGFace-Ear: An extended dataset for unconstrained ear recognition. *Sensors* **2022**, *22*, 1752. [CrossRef] [PubMed]
65. Alshazly, H.; Linse, C.; Barth, E.; Martinetz, T. Deep convolutional neural networks for unconstrained ear recognition. *IEEE Access* **2020**, *8*, 170295–170310. [CrossRef]