

Article

Zero-FVeinNet: Optimizing Finger Vein Recognition with Shallow CNNs and Zero-Shuffle Attention for Low-Computational Devices

Nghi C. Tran ¹, Bach-Tung Pham ¹, Vivian Ching-Mei Chu ², Kuo-Chen Li ^{3,*}, Phuong Thi Le ^{4,*},
Shih-Lun Chen ⁵, Aufaclav Zatu Kusuma Frisky ⁶, Yung-Hui Li ⁷ and Jia-Ching Wang ¹

¹ Department of Computer Science and Information Engineering, National Central University, Taoyuan 320317, Taiwan; tcnghivn@gmail.com (N.C.T.); pbtung@g.ncu.edu.tw (B.-T.P.); jcw@csie.ncu.edu.tw (J.-C.W.)

² Department of Drama and Theatre, National Taiwan University, Taipei City 10617, Taiwan; vivchu@ntu.edu.tw

³ Department of Information Management, Chung Yuan Christian University, Taoyuan 32023, Taiwan

⁴ Department of Computer Science and Information Engineering, Fu Jen Catholic University, New Taipei City 242062, Taiwan

⁵ Department of Electronic Engineering, Chung Yuan Christian University, Taoyuan 32023, Taiwan; chischen@cycu.edu.tw

⁶ Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia; aufaclav@ugm.ac.id

⁷ AI Research Center, Hon Hai Research Institute, Taipei 114699, Taiwan; yunghui.li@foxconn.com

* Correspondence: kuochen@cycu.edu.tw (K.-C.L.); ptle@csie.fju.edu.tw (P.T.L.)



Citation: Tran, N.C.; Pham, B.-T.; Chu, V.C.-M.; Li, K.-C.; Le, P.T.; Chen, S.-L.; Frisky, A.Z.K.; Li, Y.-H.; Wang, J.-C. Zero-FVeinNet: Optimizing Finger Vein Recognition with Shallow CNNs and Zero-Shuffle Attention for Low-Computational Devices. *Electronics* **2024**, *13*, 1751.

<https://doi.org/10.3390/electronics13091751>

Academic Editor: Praveen Kumar Donta

Received: 16 April 2024

Revised: 28 April 2024

Accepted: 30 April 2024

Published: 1 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, deep learning has experienced tremendous growth and has been proven to be effective across various domains including image classification [1,2], speech recognition [3], object detection [4], semantic segmentation [5], and natural language processing (NLP) [6]. This rapid advancement coincides with the increased adoption of mobile and embedded devices such as smartphones, smart glasses, and smartwatches. These devices, however, face significant limitations due to their restricted computational power, memory capacity, bandwidth, and battery life, which are often inadequate for supporting the complex requirements of modern deep learning models. As the reliance on mobile

technology grows and the importance of personal data security becomes increasingly paramount, the need for robust protection methods has intensified. Biometric recognition technologies, which utilize unique physiological or behavioral characteristics for identity verification, are becoming integral to enhancing security on mobile devices. This approach not only offers a more reliable means of identification compared to traditional passwords but also integrates seamlessly with the capabilities of current mobile devices. Chen et al. [7] categorized biometric traits into two distinct groups: (1) external biological characteristics, encompassing fingerprints, irises, and facial features, and (2) internal biological characteristics, including finger and palm veins. Developments in the use of external biological characteristics have recently made substantial advancements, leading to their application in mobile payments, smartphones, and access control systems [8]. However, external characteristics are generally more vulnerable to environmental influences. For instance, the presence of wounds, oil, or stains on fingers can impede the accurate capture of biometric data by sensors, thus diminishing the reliability of biometric recognition models.

Amidst these challenges, the development of finger vein recognition technology represents a significant leap in biometric identification, offering a promising solution particularly suited for mobile and embedded devices. Finger vein recognition leverages the unique patterns of blood vessels inside the finger, which are nearly impossible to replicate or forge [9], thereby providing a secure method of identification compared to more surface-level traits such as fingerprints or facial features. The appeal of this technology lies in its use of near-infrared light to distinctly capture vein patterns, which are not affected by external skin conditions and thus remain stable over time.

The integration of finger vein recognition with deep learning technologies has opened new avenues for enhancing the security features on mobile devices. By utilizing lightweight deep learning models specifically adapted for mobile use, it is possible to deploy advanced biometric systems that operate effectively even on devices with stringent power and memory constraints. These models streamline the computational process, reducing the number of necessary operations and the overall size of the model without compromising the accuracy of vein pattern recognition. This is particularly advantageous in mobile and embedded applications where efficiency and quick processing are priorities.

Despite the advantages, vein-based recognition systems are not without issues. The quality of vein images can be compromised by poor NIR camera performance or improper finger positioning during scanning, which might alter the infrared illumination on the skin and thus the appearance of the vein patterns. These factors can introduce errors into identity verification processes, underscoring the need for robust image acquisition techniques.

Addressing these limitations, this paper introduces a novel, lightweight Convolutional Neural Network (CNN) model, named Zero-FVeinNet. This model integrates only 0.3 million parameters, which is significantly fewer than those found in typical deep learning models, thereby reducing computational demands and enhancing processing speed without sacrificing accuracy. Key innovations of this model include the following:

- Shallow CNN network combined with a re-parameterization mechanism: by adjusting the number of layers to a minimal yet sufficient level and re-parameterizing the system, this approach not only reduces the amount of redundancy in parameter usage but also retains the model's learning capability, which is crucial for achieving a high accuracy.
- Integrating a blur pool layer into a lightweight model: this modification maintains feature extraction consistency across translated images, thereby stabilizing recognition accuracy.
- Zero parameter channel shuffle coordinate attention block (ZSCA): this attention mechanism helps to reduce the computational costs associated with traditional attention models and maintains the ability to learn important features. It is particularly effective in extracting subtle vein characteristics.

The efficacy and versatility of Zero-FVeinNet have been rigorously tested across multiple prominent public finger vein datasets, including SDUMLA-FV [10], FV-USM [11] from the University Sains Malaysia, SCUT-FVD [12], and THU-FVFDT [13]. The findings

confirm that Zero-FVeinNet not only surpasses conventional methods in speed and efficiency but also sets new benchmarks in biometric identity recognition, promising a new direction for the use of biometric security technology in mobile and embedded systems. This study's contributions offer substantial improvements to the operational performance of deep learning models for mobile devices, particularly in the context of personal security and identity verification.

2. Literature Review

Significant advancements have been made in the field of finger vein recognition in recent years, categorized primarily into conventional and deep learning approaches.

Conventional finger vein recognition involves two primary stages: pattern extraction and identification. Initially, the pattern is extracted from an input image, followed by identification of the vein's identity. Miura introduced algorithms for pattern detection such as Repeated Line Tracking and Maximum Curvature [14,15]. Later, Huang developed the Wide Line Detector for enhanced pattern extraction [16]. These algorithms utilize the cross-sectional profiles of finger veins, in which the veins are visibly darker than the surrounding tissues. The Repeated Line Tracking algorithm randomly initializes positions and iteratively tracks vein lines, whereas the Maximum Curvature algorithm identifies vein patterns by detecting peaks in curvature across these profiles. Unlike these, the Wide Line Detector focuses on the width of vein lines. Subsequently, Ma et al. combined oriented gradient and local phase quantization techniques with pyramid histograms to refine the texture features of veins for better identification [17]. Identification is then performed by matching these extracted patterns against a database using Structural Matching or Template Matching [18,19]. Meng et al. introduced a zone-based method that segments the matching process into smaller areas, which speeds up the process while increasing the accuracy of pinpointing authentic matches [20]. These traditional methods often require manual calibration and can yield inconsistent results. They are also difficult to develop into end-to-end systems and are highly susceptible to variations in image quality such as contrast, scale, and orientation.

Conversely, deep learning techniques, which utilize extensive training datasets, are adept at capturing high-dimensional features and global contextual information from images. This approach is generally more resilient to quality variations in input images. Remarkable successes have been noted in biometric recognition using deep learning, particularly in finger vein recognition [21]. For example, the adoption of the VGG-16 architecture, which has outperformed traditional algorithms [22]. Zeng et al. proposed a fully convolutional neural network that extends U-Net with an embedded conditional random field, creating an efficient end-to-end system for accurate vein pattern segmentation [23]. R.S. Kuzu enhanced the DenseNet-161 architecture with a custom embedder module, leading to promising outcomes on standard finger vein datasets [24]. Furthermore, Generative Adversarial Networks (GANs) have been effectively used to address issues of low image quality and limited data availability. These networks employ convolutional layers over fully connected layers, reducing computational demands and optimizing feature extraction [25]. Hou and Yan implemented a triplet-classifier GAN that generates synthetic data to bolster the training process of triplet loss-based CNN classifiers, thus mitigating overfitting and improving recognition accuracy [26]. Departing from traditional single biometric identification methods, reference [27] presents an innovative biometric approach combining finger vein and facial features using a CNN enhanced by a self-attention mechanism and a ResNet structure. Extensive tests with AlexNet and VGG-19 show that this multimodal method significantly improves identification accuracy, which has exceeded 98.4% in complex scenarios.

Recently, with advancements in embedded AI devices [28], research has pivoted towards AI models optimized for mobile and embedded applications. Zhao introduced a lightweight CNN that incorporates a center loss function and dynamic regularization to tackle image quality issues and expedite convergence, demonstrating decreased error rates

and enhanced computational efficiency [29]. Furthermore, Hsia's [30] improved lightweight CNN (ILCNN) addresses translation-induced accuracy problems and enhances parameter efficiency using diverse branch blocks (DBB) [31], adaptive polyphase sampling (APS) [32], and a coordinate attention mechanism (CoAM) [33]. This model not only achieves high accuracy but does so with a minimal parameter count of just 1.23 million.

3. Methodology

Enhancing a model's capabilities to encapsulate complex features and intricate relationships within datasets constitutes a fundamental strategy to augment performance metrics in deep learning. Specifically, in the domain of finger vein recognition, research delineated in reference [22] employs the VGG16 architecture, which encompasses approximately 138 million parameters, to achieve elevated performance levels. Concurrently, another study cited in references [24,34] implements the DenseNet-161 architecture as a feature extraction model, boasting nearly 20 million parameters. Despite the promising outcomes demonstrated by these models in finger vein recognition tasks, their substantial size poses significant challenges for their deployment on mobile or embedded devices.

The objective of our research is to develop a lightweight convolutional neural network (CNN) for a finger vein recognition system. Recent advancements in the design and deployment of efficient deep learning architectures for mobile platforms [35–38] have consistently reduced both computational costs and parameter sizes, thereby enhancing model efficiency. Nonetheless, a minimal parameter count does not solely define a model's lightweight nature; for instance, parameter sharing can reduce model size but increase computational load. Additionally, operations that do not involve parameters, such as skip connections [39] and branching [40], may lead to significant costs that are associated with memory access. These challenges are further complicated by the presence of custom accelerators in efficient architectures.

To address these challenges, we propose Zero-FVeinNet, a novel architecture incorporating a shallow CNN equipped with a re-parameterization mechanism to minimize model size. Moreover, the proposed the ZeroBlur-DBB, a modified version of DBB [31] with an integrated blur pool layer and ZSCA mechanism.

The proposed finger vein recognition system illustrated in Figure 1 includes two primary phases: training and inference. In the training phase, a lightweight model architecture is employed. The optimal weight parameters are iteratively refined to enhance the model's performance. The culmination of this phase is the selection and storage of the best-performing weight parameters, which represent the trained model. In the inference phase, the pretrained model from the training phase applies a re-parameterization mechanism to generate a lightweight model, facilitating its deployment in finger vein recognition tasks. The model processes an input finger vein image by comparing it with pre-registered finger vein features stored in a database. A match is determined based on the highest similarity score. If this score surpasses a predefined threshold, the model confirms the identity as correct. Conversely, scores below this threshold are deemed misclassifications.

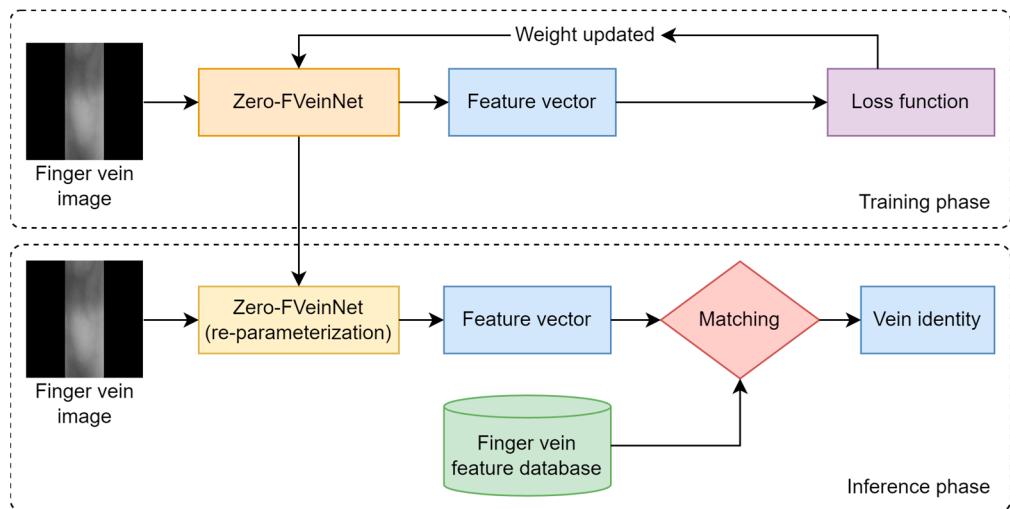


Figure 1. The finger vein recognition system proposed in this research.

3.1. Shallow CNN Network with a Re-Parameterization Mechanism

Convolutional neural networks (CNNs) are highly effective for image classification tasks due to their ability to extract features hierarchically, mimicking human perceptual processes. These networks start by identifying simple, low-level features such as edges and textures and progressively learn to recognize complex, high-level representations. This capability makes them particularly adept in applications ranging from medical imaging to autonomous driving. In finger vein recognition, the structure comprises the blood vessels in a finger, which form vascular patterns visible on the skin's surface. These patterns appear as lines and points against a black and white background in images, presenting a simpler structure compared to more complex objects like human faces or vehicles, which include intricate features such as ears, eyes, noses, mouths, lights, doors, windows, and tires. Consequently, vein detection does not require the high-level processing necessary for recognizing faces or vehicles.

Figure 2 illustrates the varying layer depths required for different objects within CNNs. It shows that, while complex structures like human faces and vehicles necessitate deeper layers for effective feature extraction and classification, simpler structures such as finger veins can be effectively recognized with shallower layers.

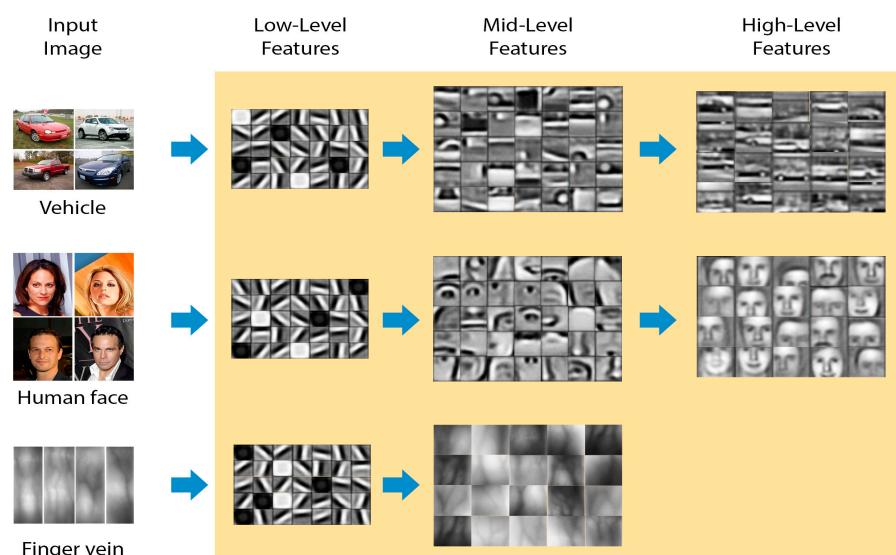


Figure 2. Different levels of features extracted by CNNs for human faces, vehicles, and finger veins.

To further contribute to the model's efficiency, we integrated a re-parameterization mechanism. Reparameterization, a technique outlined in the research in [31], utilizes the DBB to perform structural reparameterization, which effectively reduces the network's parameters while maintaining the model's expressive power. Re-parameterization recalibrates the parameter space to optimize the network's representational capacity, thereby enabling a more compact model with reduced computational demands. Consequently, the integration of a shallow CNN architecture with re-parameterization promotes a lightweight yet potent model for finger-vein recognition, achieving the dual goals of minimizing parameter count and computational expense.

This paper proposes adopting a shallow CNN architecture combined with a re-parameterization mechanism for finger vein recognition tasks. Characterized by fewer convolutional layers, this architecture is designed to maintain high accuracy while ensuring computational efficiency. By reducing the network's depth, this approach not only conserves computational resources but also optimizes parameter usage, meeting the demands of real-world applications in which efficiency is crucial. Figure 3 displays the initial model structure of the shallow CNN integrated with DBB for reparameterization. In the subsequent section, we will explore a modified version of DBB, termed ZeroBlur-DBB, which is intended to enhance model performance without increasing the number of model parameters.

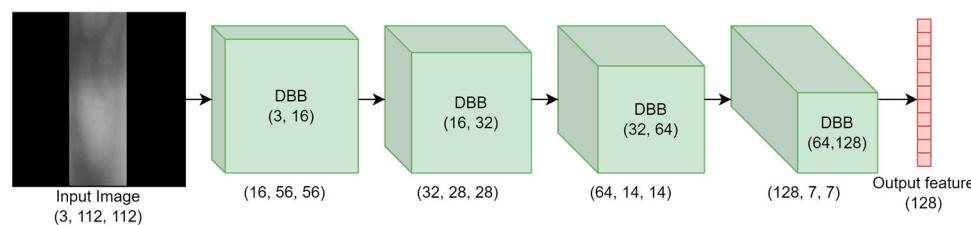


Figure 3. The Shallow CNN network with diverse branch block (DBB).

3.2. ZeroBlur-DBB Module: Diverse Branch Block with Blur Pool and Zero Shuffle Coordinate Attention

To further enhance model performance, we have integrated the DBB with a blur pool and zero parameter channel shuffle coordinate attention, creating the ZeroBlur-DBB. The detailed structure of the ZeroBlur-DBB is depicted in Figure 4.

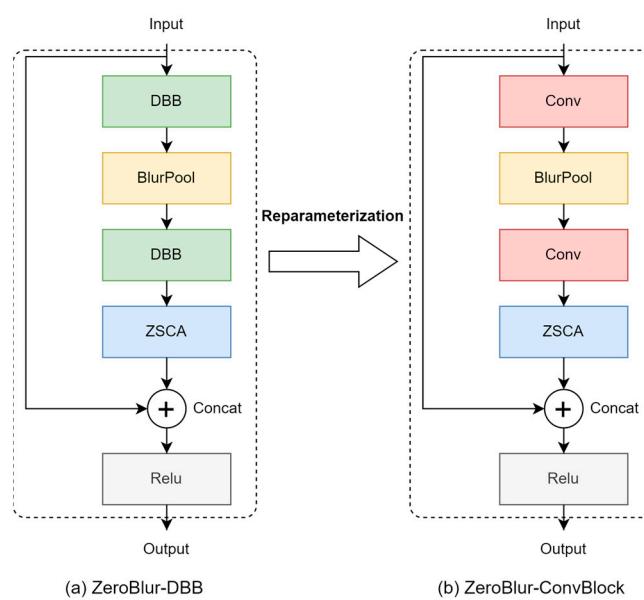


Figure 4. The detailed structure of the ZeroBlur-DBB and ZeroBlur-ConvBlock. (a) The ZeroBlur-DBB architecture; (b) the ZeroBlur-ConvBlock architecture.

Figure 4a illustrates the detailed structure of ZeroBlur-DBB itself, highlighting the additions of the blur pool and ZSCA, whereas Figure 4b presents the ZeroBlur-ConvBlock, which executes the re-parameterization by replacing the multi-branch convolutional layer in the DBB block with a single-layer convolutional layer. This adjustment reduces the number of parameters while maintaining performance.

To elucidate how ZeroBlur-DBB contributes to improved performance, we will further explore the functionalities of the blur pool layer and ZSCA.

3.2.1. Blur Pool Layer

The blur pool, as proposed in [41], enhances the shift-equivariance—or alternatively, promotes shift-invariance—of convolutional neural networks (CNNs). In CNNs, a slight positional shift of features within input images can variably alter the resulting output feature maps. These feature maps are crucial for classification tasks, as CNNs extract these maps from input images to identify and classify objects.

Traditional pooling layers aim to downsample these feature maps to diminish their sensitivity to positional changes in features. However, the variance issue persists due to the downsampling process itself. For example, a pooling layer with a stride of 2 handles even-pixel shifts effectively, yet it inaccurately processes odd-pixel shifts. This condition implies that, if an input shift impacts the output feature map, the CNN exhibits shift variance; conversely, if there is no effect, it demonstrates shift invariance. Equations (1) and (2) [41] illustrate these cases of shift variance and shift invariance, respectively.

$$\text{Shift}_{\Delta x, \Delta y} \left(\tilde{G}(I) \right) = \tilde{G} \left(\text{Shift}_{\Delta x, \Delta y}(I) \right) \forall (\Delta x, \Delta y) \quad (1)$$

$$\tilde{G}(I) = \tilde{G} \left(\text{Shift}_{\Delta x, \Delta y}(I) \right) \forall (\Delta x, \Delta y) \quad (2)$$

In this representation, $I \in \mathbb{R}^{H \times W \times 3}$ denotes the input image. The feature map produced by the CNN model, denoted as $G(I) \in \mathbb{R}^{H_i \times W_i \times C_i}$, has a spatial resolution of $H_i \times W_i$ and contains C_i channels. The original resolution $\tilde{G}(I) \in \mathbb{R}^{H \times W \times C_i}$ is obtained by upsampling this feature map. The terms Δx and Δy represent the horizontal and vertical shift distances, respectively.

In our proposed model, blur pool layers are utilized to replace the traditional pooling layers. These layers blur the input features before the downsampling step, significantly reducing the impact of positional shifts on the input features. This blurring technique smooths out the details in the feature map, thereby improving shift-equivariance and enhancing the model's overall robustness against positional variations in the input data.

This strategic integration of blur pool layers into our model ensures greater consistency in feature map extraction, which is essential for maintaining high accuracy in object classification regardless of minor shifts in input image positioning.

3.2.2. Zero Parameter Channel Shuffle Coordinate Attention (ZSCA)

To enhance the model's ability to extract effective features for finger vein recognition, we incorporated an attention mechanism into our lightweight model architecture. However, conventional attention mechanisms like SENet [42] and CBAM [43] are unsuitable for mobile networks due to their substantial computational demands and large model sizes. Instead, Hou et al. [33] developed the Coordinate Attention Mechanism (CoAM), which embeds positional information into channel attention to cover larger regions efficiently. This mechanism compresses spatial information using two directionally oriented average pooling layers, and the resulting vectors are concatenated and processed through a convolutional layer to capture channel-specific information. The process ensures shared convolutional weights between the two vectors, facilitating feature encoding. The vectors are then independently transformed through additional convolutional layers, while a Sigmoid function normalizes these features, which are multiplied back with the original input

to apply the attention mechanism effectively. This structure allows CoAM to detect the precise location of relevant features within the map, enhancing the model's finger vein recognition capabilities over SENet and CBAM.

Inspired by the creation of lightweight CNNs that employ an attention module without extra parameters, we have innovated the Coordinate Attention block into a zero-parameter version by removing the convolutional layers and retaining the activation function. This adjustment ensures the attention mechanism still learns significant features. Additionally, we introduced a channel shuffle [44] at the block's end to improve information flow across feature channels, reducing computational costs while preserving model accuracy. The architectural differences between the traditional Coordinate Attention and our ZSCA are depicted in Figure 5.

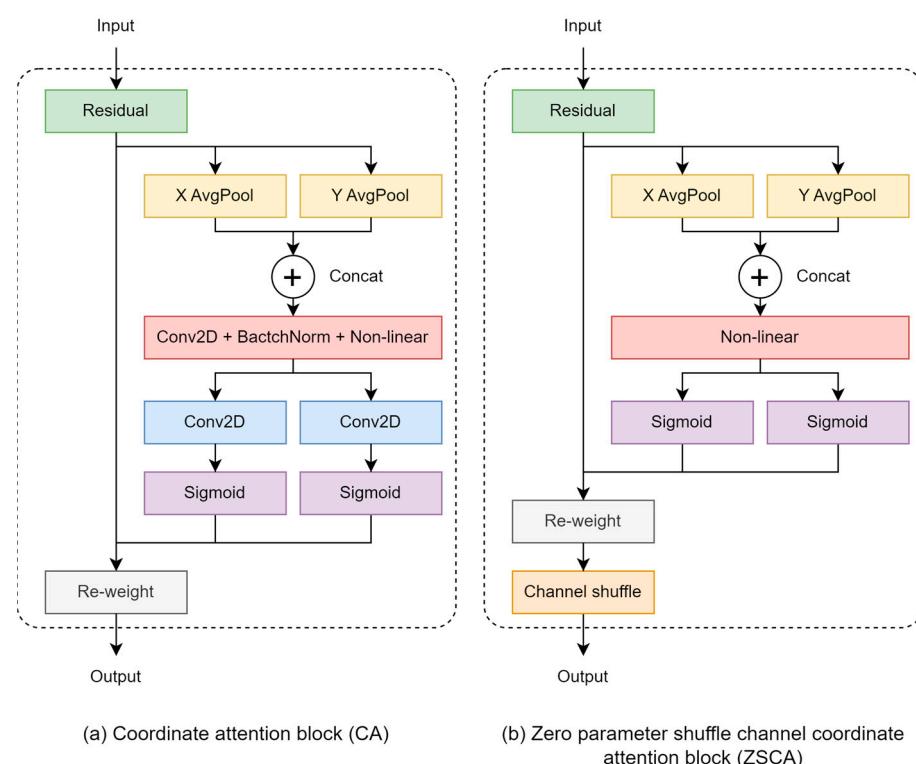


Figure 5. Architecture of different attention mechanism. (a) Coordinate attention block (CA) architecture. (b) ZSCA block architecture.

3.3. Evaluation Metrics and Loss Function

3.3.1. Evaluation Metrics

To thoroughly evaluate the lightweight model, we assess not only its performance in terms of accuracy but also its complexity, which includes model size, speed, and computational cost.

- **Model Complexity:** We calculate the model size by counting the number of parameters. We measure the model's speed in terms of inference time (millisecond), and we assess computational costs (FLOPs) associated with the model's multiply operations.
- **Model Performance:** We utilize the Correct Identity Rate (CIR) as the evaluation metric. The CIR measures the model's security efficacy; a higher CIR value signifies greater security and improved recognition performance, making it ideal for one-to-many finger vein recognition tasks. The method for calculating the CIR is detailed in Equation (3).

$$CIR = \frac{\text{Correct identity prediction}}{\text{Total number of identity}} \quad (3)$$

3.3.2. Loss Function

In our system, a CNN model is employed to derive embedding vectors from input images for classification purposes. This model is designed to reduce intra-class distances and increase inter-class distances. To boost the accuracy of identity recognition, we implemented Elastic Margin Loss (EML) [45]. EML enhances class separability by using margin values randomly selected from a normal distribution in each training cycle. This dynamic adjustment of the decision boundaries allows for more adaptable learning of class distinctions. Demonstrably, EML outperforms other margin losses like ArcFace and CosFace [45]. The computational formula used for EML is detailed in Equation (4).

$$L = -\frac{1}{N} \sum_{i \in N} \log \frac{e^{s(\cos(\theta_{y_i} + E(m, \sigma)))}}{e^{s(\cos(\theta_{y_i} + E(m, \sigma)))} + \sum_{j=1, j \neq y_i}^n e^{s(\cos(\theta_{y_j}))}} \quad (4)$$

where $E(m, \sigma)$ is a normal function that returns a random value from a Gaussian distribution with the mean m and the standard deviation σ .

4. Experiments

4.1. Finger-Vein Public Datasets

To assess our methodology, we utilized four publicly available datasets specifically focused on finger veins: SDUMLA-FV [10], FV-USM [11], SCUT-FVD [12], and THU-FVFDT [13]. The detailed properties of these datasets are discussed in the subsequent section.

- The SDUMLA-FV dataset [10]: This dataset has been compiled by Shandong University and includes data from 106 participants. Each participant provided images of their index, middle, and ring fingers from both hands, constituting 636 unique finger classes in total. In a single collection session, each finger was imaged six times, summing up to 3816 images. The original images have a resolution of 320×240 pixels. For our study, we implemented basic image processing techniques, including edge detection combined with image contrast enhancement, to extract Regions of Interest (ROI) from finger vein images with dimensions of 300×150 pixels, as the dataset does not originally include ROI images.
- The FV-USM dataset [11]: Developed by the University Sains Malaysia, this dataset involves images from 123 individuals, capturing the index and middle fingers of both left and right hands. The cohort includes 83 males and 40 females, aged 20 to 52 years. Each finger was photographed six times across two sessions, and the dataset provides pre-extracted ROI images with a resolution of 100×300 pixels that are suitable for finger vein recognition.
- The SCUT-FVD dataset [12]: Introduced by the South China University of Technology (SCUT), this dataset is designed for both finger vein recognition and spoof detection tasks. It comprises over 7000 images, evenly split between genuine and spoof images. For the purposes of this research, only the genuine images were utilized, involving 101 subjects each with six distinct finger vein identities. Each identity was documented in six separate data samples, resulting in a total of 3636 genuine samples.
- The THU-FVFDT dataset [13]: The THUFV2 dataset, released by Tsinghua University in 2014, includes ROIs of finger veins and dorsal textures. It features 610 subjects, with each providing one image of each type, captured in two sessions. The ROIs have been standardized to a resolution of 200×100 pixels. The participants were predominantly students and staff from the Tsinghua University's Graduate School at Shenzhen.

Table 1 shows a summary of the four different finger vein datasets.

All of the finger vein images employed in our experiments utilize Regions of Interest (ROIs) either extracted directly from the dataset or identified using basic image processing techniques as specified in the SDUMLA-FV dataset. The ROI extraction stage is crucial in practical applications. The finger vein ROI extraction methodology can adopt the technique detailed in reference [46], which proposes a novel method for finger vein images. This

method includes a weighted horizontal Sobel operator, a contour-based edge detection method, and a gradient detection operator with a large receptive field. Evaluated against public finger vein datasets, this method has shown a substantial reduction in processing time and enhancements in robustness and accuracy.

Table 1. Summary of finger vein datasets.

Database	# of Classes	# of Samples per Class	Total Samples
SDUMLA-FV	636	6	3816
FV-USM	492	12	5904
SCUT-FVD	606	6	3636
THU-FVFDT	610	2	1220

4.2. Experimental Configuration

This study utilized the Zero-FVeinNet architectural framework for the recognition of finger veins. The preprocessing involved zero-padding the finger vein images' ROIs to convert them into square formats that were then resized to align with the prescribed input dimensions of the model. The training phase of the model incorporated EML as the loss function and employed AdamW for optimization purposes. The set hyperparameters were as follows: an image size of 112 pixels, a batch size of 64, a training duration of 300 epochs, and an initial learning rate set at 0.0001. The training operations were executed within the PyTorch deep learning environment. Computational support was provided by an Intel Core™ i7-9800X CPU and an Nvidia RTX 3060 GPU.

4.3. Experimental Results

In our system, the model processes an input finger vein image by comparing it with pre-registered finger vein features in a database, using a similarity score relative to a set threshold of 0.5. This threshold is critical for validating the authenticity of the finger vein data, where a score exceeding 0.5 confirms the identity as correct, and scores below this mark indicate misclassification. This recognition threshold was established empirically through a detailed analysis of similarity scores for both genuine and impostor matches. During the testing of vein identities not recorded in the database, it was observed that the similarity scores for these unidentified entries fell below 0.5, whereas scores for accurately identified veins were above this threshold, thereby justifying the employment of a 0.5 threshold for robust identity verification.

To evaluate the effectiveness of our proposed model, its performance was compared with other lightweight models focusing on the Correct Identity Rate (CIR) and model parameters. The comparative analysis included the following baseline models: ResNet-50 [47], which represents a deep model architecture; Inception V3 [2], exemplifying a wider model architecture; and other state-of-the-art mobile networks such as MobileNet [36,37], MobileViT [38], EfficientNet [48], and ILCNN [30]. All of the models underwent evaluation under uniform conditions to ensure a fair comparison. The evaluation criteria were designed to assess not only the accuracy (CIR) of each model but also their operational efficiency, encompassing model size (parameters), inference time, and computational cost (FLOPs). These experimental findings are detailed in Table 2.

Table 2 shows that our proposed method achieves the best Correct Identification Rate (CIR) on three different datasets (FV-USM, SCUT-FVD, and SDUMLA-FV), despite having only 0.3 million parameters. This is significantly fewer parameters: up to 82 times less compared to the largest model size in the table (ResNet50), and three times smaller than the smallest model size listed (MobileNetV3_small_050). Additionally, it also records the shortest inference time at 0.49 ms, compared to the other models which ranged from 0.7 to 1.13 ms on our testing hardware.

Table 2. The experimental comparison of the proposed model with others on finger vein datasets.

Model	CIR (%)				Params (M)	Inference Time (ms)	FLOPs (G)
	FV-USM	SCUT-FVD	SDUMLA-FV	THU-FVFDT			
resnet50 [47]	99.70	97.17	95.44	70.10	24.81	1.13	2.18
inception_v3 [2]	99.59	96.67	95.75	60.33	22.8	1.03	1.082
mobilenetv2_035 [36]	99.09	82.17	84.28	31.80	1.21	0.77	0.035
mobilenetv3_small_050 [37]	99.19	90.17	84.91	60.57	1.07	0.69	0.024
mobilevit_xxs [38]	99.29	89.33	84.28	50.98	1.16	0.92	0.115
mobilevitv2_050 [38]	99.39	83.17	82.08	28.20	1.24	0.83	0.196
efficientnet_b0 [48]	99.24	93.83	88.52	46.56	4.64	0.82	0.218
ILCNN [30]	99.80	95.50	97.48	51.31	1.23	0.88	0.198
Proposed	99.90	97.83	97.80	68.69	0.365	0.49	0.149

The bold numbers represent the best results from the experiment.

Regarding computation cost (FLOPs), MobileNetV3_small_050 has the lowest, at 0.024 GFLOPs, whereas the proposed method has a cost of 0.149 GFLOPs. Most mobile-oriented models have GFLOPs ranging from 0.024 to 0.218, in contrast to the more basic models like Inception_v3 and ResNet50, which have costs of 1.082 GFLOPs and 2.18 GFLOPs, respectively.

In the THU-FVFDT dataset, the best CIR is 70.1%, compared to 68.69% achieved by our proposed method. This indicates that larger models still perform better in extracting higher levels of features. Moreover, the THU-FVFDT is an insufficient dataset with only two samples per class, which restricts us to using just one sample per class for training and evaluation. This limitation is why the performance of the model on this dataset is relatively lower compared to its performance on the other finger vein datasets. To enhance its performance on this dataset, one could apply augmentation techniques; however, in our experiments, we wanted to evaluate the performance of the model even on datasets without any data enhancement techniques.

Since our objective is to design a lightweight finger vein recognition model suitable for mobile embedded devices, our system can be converted into ONNX format, enabling full deployment on devices such as Arduino and other mobile platforms. Additionally, our approach allows for the direct extraction of image regions on the embedded device during the collection of finger vein images. Importantly, our system does not require any preprocessing steps, such as image enhancement, thus facilitating the complete deployment of the model directly on the device.

4.4. Ablation Study

To explore the effects of various techniques within the proposed Zero-FVeinNet model on finger vein recognition, an ablation study was carried out to evaluate their efficacy. The integration of the Zero Shot Channel Attention (ZSCA) mechanism and the blur pool layer in the model facilitates the preservation of original characteristics while concurrently extracting features across multiple scales and translations. This approach enhances the diversity of the features learned. The empirical findings indicate that incorporating the ZSCA into the model significantly improves its ability to capture the textural features of finger vein images.

Table 3 indicates that the proposed model in (2), without blur pool layers, will experience decreased performance across all four finger vein datasets. However, it will maintain the same model parameters, inference time, and computational cost (FLOPs). Conversely, removing the ZSCA from proposed model (3) leads to a significant decrease in its performance on the THU-FVFDT dataset. This change also results in increased model parameters, inference time, and computational cost (FLOPs). These findings demonstrate that the ZSCA module significantly enhances the model's performance, particularly on datasets with insufficient data such as THU-FVFDT, achieving a Correct Identification Rate (CIR) of 60.33%, compared to 57.05% with the blur pool layer. The proposed model with

both ZSCA and blur pool layers (1) achieves the best results in terms of model performance and complexity.

Table 3. The ablation study on the proposed model.

Model	CIR (%)				Params (M)	Inference Time (ms)	FLOPs (G)
	FV-USM	SCUT-FVD	SDUMLA-FV	THU-FVFDT			
(1) Proposed (ZSCA + Blur)	99.9	97.83	97.8	68.69	0.365	0.498	74.58
(2) Proposed (ZSCA)	99.8	96.67	97.43	60.33	0.365	0.498	74.58
(3) Proposed (Blur)	99.9	97.17	97.64	57.05	0.371	0.512	74.73

The bold numbers represent the best results from the experiment.

5. Conclusions

This study presents the Zero-FVeinNet model, developed for finger vein recognition, with an architectural design aimed at enhancing feature diversity learning during training. It proficiently captures essential global features and demonstrates shift-invariance by incorporating a blur pool layer. Furthermore, the integration of a ZSCA block in the model serves to enhance its performance while reducing complexities such as model size, inference speed, and computational expenses, thereby addressing common challenges related to feature extraction stability. Empirical evaluations of the Zero-FVeinNet model show a Correct Identification Rate (CIR) of 99.9% on the FV-USM dataset, 97.83% on SCUT-FVD, 97.8% on SDUMLA-FV, and 68.69% on THU-FVFDT datasets, surpassing the performance of recent methodologies cited in the literature. The lower performance on the THU-FVFDT dataset, attributed to data insufficiency, suggests potential enhancements through methods like data augmentation or self-supervised learning, which are considerations for future research. Additionally, the model's compatibility with the ONNX format facilitates its deployment on devices such as Arduino and other mobile platforms without the need for preprocessing steps like image enhancement. Future investigations will focus on assessing the model's performance on mobile platforms to develop a comprehensive, lightweight solution for finger vein recognition, particularly improving outcomes on datasets with limited data.

Author Contributions: Conceptualization, J.-C.W.; Methodology, N.C.T., B.-T.P.; resources, V.C.-M.C., K.-C.L., P.T.L., S.-L.C., A.Z.K.F., Y.-H.L.; writing—original draft preparation, N.C.T.; writing—review and editing, B.-T.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: a. The datasets collected and analyzed during this work are accessible from the SDUMLA dataset repository [10], available at <https://tsapps.nist.gov/BDbC/Search/Details/420> (accessed on 29 April 2024). b. The datasets collected and analyzed during this work are accessible from the THUFV2 dataset repository [11], available at <https://www.sigs.tsinghua.edu.cn/labs/vipl/thu-fvfdt.html> (accessed on 29 April 2024). c. The datasets collected and analyzed during this work are accessible from the FV-USM dataset repository [12], available at http://drfendi.com/fv_usm_database (accessed on 29 April 2024). d. The datasets collected and analyzed during this work are accessible from the SCUT-FVD dataset repository [13], available at <https://github.com/BIP-Lab/SCUT-SFVD> (accessed on 29 April 2024).

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

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
2. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

3. Dahl, G.E.; Yu, D.; Deng, L.; Acero, A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *20*, 30–42. [[CrossRef](#)]
4. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Conference on International Conference on Neural Information Processing Systems, Montreal, QB, Canada, 7–12 December 2015; pp. 91–99.
5. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
6. Mikolov, T.; Karafiat, M.; Burget, L.; Cernocky, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the Conference of the International Speech Communication Association, Makuhari, Japan, 26–30 September 2010; pp. 1045–1048.
7. Chen, Y.-Y.; Hsia, C.-H.; Chen, P.-H. Contactless multispectral palm-vein recognition with lightweight convolutional neural network. *IEEE Access* **2021**, *9*, 149796–149806. [[CrossRef](#)]
8. Algarni, M. An Extra Security Measurement for Android Mobile Applications Using the Fingerprint Authentication Methodology. *J. Inf. Secur. Cybercrimes Res.* **2023**, *6*, 139–149. [[CrossRef](#)]
9. Syazana-Itqan, K.; Syafeeza, A.; Saad, N.; Hamid, N.A.; Saad, W. A review of finger-vein biometrics identification approaches. *Indian J. Sci. Technol.* **2016**, *9*, 1–9. [[CrossRef](#)]
10. Yin, Y.; Liu, L.; Sun, X. SDUMLAHMT: A multimodal biometric database. In *Chinese Conference on Biometric Recognition*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 260–268.
11. Asaari, M.S.M.; Suandi, S.A.; Rosdi, B.A. Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics. *Expert Syst. Appl.* **2014**, *41*, 3367–3382. [[CrossRef](#)]
12. Qiu, X.; Kang, W.; Tian, S.; Jia, W.; Huang, Z. Finger vein presentation attack detection using total variation decomposition. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 465–477. [[CrossRef](#)]
13. Yang, W.; Qin, C.; Liao, Q. A database with ROI extraction for studying fusion of finger vein and finger dorsal texture. In *Chinese Conference on Biometric Recognition*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 266–270.
14. Miura, N.; Nagasaka, A.; Miyatake, T. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Mach. Vis. Appl.* **2004**, *15*, 194–203. [[CrossRef](#)]
15. Miura, N.; Nagasaka, A.; Miyatake, T. Extraction of finger vein patterns using maximum curvature points in image profiles. *IEICE TRANSACTIONS Inf. Syst.* **2007**, *90*, 1185–1194. [[CrossRef](#)]
16. Huang, B.; Dai, Y.; Li, R.; Tang, D.; Li, W. Finger vein authentication based on wide line detector and pattern normalization. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1269–1272.
17. Ma, H.; Hu, N.; Fang, C. The biometric recognition system based on near-infrared finger vein image. *Infrared Phys. Technol.* **2021**, *116*, 103734. [[CrossRef](#)]
18. Zhang, W.; Wang, Y. Core based structure matching algorithm of Finger Vein verification. In *Object Recognition Supported by User Interaction for Service Robots*; IEEE: Piscataway, NJ, USA, 2002; Volume 1, pp. 70–74.
19. Nagao, M. *Methods of Image Pattern Recognition*; Corona: San Antonio, TX, USA, 1983.
20. Meng, X.; Zheng, J.; Xi, X.; Zhang, Q.; Yin, Y. Finger vein recognition based on zone-based minutia matching. *Neurocomputing* **2021**, *423*, 110–123. [[CrossRef](#)]
21. Minaee, S.; Abdolrashidi, A.; Su, H.; Bennamoun, M.; Zhang, D. Biometrics recognition using deep learning: A survey. *arXiv* **2019**, arXiv:1912.00271. [[CrossRef](#)]
22. Hong, H.G.; Lee, M.B.; Park, K.R. Convolutional neural network based finger vein recognition using nir image sensors. *Sensors* **2017**, *17*, 1297. [[CrossRef](#)] [[PubMed](#)]
23. Zeng, J.; Wang, F.; Deng, J.; Qin, C.; Zhai, Y.; Gan, J.; Piuri, V. Finger vein verification algorithm based on fully convolutional neural network and conditional random field. *IEEE Access* **2020**, *8*, 65402–65419. [[CrossRef](#)]
24. Kuzu, R.S.; Maioranay, E.; Campisi, P. Vein-based biometric verification using transfer learning. In Proceedings of the 2020 43rd International Conference on Telecommunications and Signal Processing (TSP), Milan, Italy, 7–9 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 403–409.
25. Yang, W.; Hui, C.; Chen, Z.; Xue, J.; Liao, Q. FV-GAN: Finger vein representation using generative adversarial networks. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2512–2524. [[CrossRef](#)]
26. Hou, B.; Yan, R. Triplet-classifier gan for finger-vein verification. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 212–223. [[CrossRef](#)]
27. Yang, W.; Shi, D.; Zhou, W. Convolutional neural network approach based on multimodal biometric system with fusion of face and finger vein features. *Sensors* **2022**, *22*, 6039. [[CrossRef](#)] [[PubMed](#)]
28. Lin, H.Y. Embedded Artificial Intelligence: Intelligence on Devices. *Computer* **2023**, *56*, 90–93. [[CrossRef](#)]
29. Zhao, D.; Ma, H.; Yang, Z.; Li, J.; Tian, W. Finger vein recognition based on lightweight CNN combining center loss and dynamic regularization. *Infrared Phys. Technol.* **2020**, *105*, 103221. [[CrossRef](#)]
30. Hsia, C.H.; Ke, L.Y.; Chen, S.T. Improved Lightweight Convolutional Neural Network for Finger Vein Recognition System. *Bioengineering* **2023**, *10*, 919. [[CrossRef](#)]
31. Ding, X.; Zhang, X.; Han, J.; Ding, G. Diverse branch block: Building a convolution as an inception-like unit. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.

32. Chaman, A.; Dokmanic, I. Truly shift-invariant convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
33. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
34. Noh, K.J.; Choi, J.; Hong, J.S.; Park, K.R. Finger-Vein Recognition Based on Densely Connected Convolutional Network Using Score-Level Fusion with Shape and Texture Images. *IEEE Access* **2020**, *8*, 96748–96766. [[CrossRef](#)]
35. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
36. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
37. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
38. Mehta, S.; Rastegari, M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178.
39. Wu, D.; Wang, Y.; Xia, S.T.; Bailey, J.; Ma, X. Skip connections matter: On the transferability of adversarial examples generated with resnets. *arXiv* **2020**, arXiv:2002.05990.
40. Zhu, X.; Bain, M. B-CNN: Branch convolutional neural network for hierarchical classification. *arXiv* **2017**, arXiv:1709.09890.
41. Zhang, R. Making convolutional networks shift-invariant again. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 9–15 June 2019.
42. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
43. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
44. Zhang, Q.-L.; Yang, Y.-B. Sa-net: Shuffle attention for deep convolutional neural networks. In Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; IEEE: Piscataway, NJ, USA, 2021.
45. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. ElasticFace: Elastic margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, New Orleans, LA, USA, 18–24 June 2022; pp. 1578–1587.
46. Lu, H.; Wang, Y.; Gao, R.; Zhao, C.; Li, Y. A novel roi extraction method based on the characteristics of the original finger vein image. *Sensors* **2021**, *21*, 4402. [[CrossRef](#)] [[PubMed](#)]
47. 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.
48. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 9–15 June 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.