

A Lightweight Neural Network for Palm Vein Recognition

Sifeng Luo, Xi Huang*

College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou, 350007, China

* Corresponding author: Xi Huang

Abstract: Due to the advantages of lightweight networks such as efficiency, portability, and fast inference, we propose a palm vein recognition network based on lightweight neural networks. The network consists of three stages combining Blocks of ShuffleNetV2, Blocks of MobileNetV3, and MBConv of EfficientNet. The number of each stage in the network is chosen based on the baseline parameters and computational effort, and the expansion factors are re-analyzed and selected. The performance of the proposed network is compared to the baseline with a 16.78% reduction in error rate and a 10.91% compression of parameters. The experimental results show that the proposed palm vein recognition network is effective, which can potentially contribute to the development of more accurate, portable and reliable biometric systems.

Keywords: Plam vein; Lightweight neural network; Attention mechanism.

1. Introduction

With the rapid development of technology, there are not only traditional identification technologies in the field of recognition, but also some more convenient biometric technologies. Palm vein recognition is a modern biometric-based technology that uses the unique vein pattern in the palm of the hand as a means of personal identification. Unlike most biometric technologies, palm veins are stable, unique, not particularly sensitive to external conditions, and have high user acceptance and high anti-counterfeiting capabilities. Palm vein recognition has a wide range of application scenarios, such as banking, finance, healthcare, government agencies, etc. In the financial industry, palm vein recognition can be used in ATM machines, mobile payment and other scenarios. In the medical field, palm vein recognition can be used for the management of electronic medical records and the control of medical equipment. In government agencies, palm vein recognition can be used for ID card processing and public security and other fields.

The traditional designs of palm vein feature extraction and recognition are summarized into 3 ways. The first one is based on structural features, where geometric features (angular features and point features) of veins are extracted. Such as Ladoux P O et al.[1] extracted point features by scale invariant feature transform (SIFT) for recognition. The second one is based on texture features, where the palm vein images are used as texture images from which texture features are extracted. For example, G K O Michael et al.[2] extracted the main texture features of veins by Radom transform combined with discrete cosine transformation (DCT) and matched them with Euclidean distance. The third is a subspace-based approach, which views the vein image as a high-dimensional matrix and uses a downscaling into a low-dimensional one for recognition in the low-dimensional space. Such as J G Wang et al.[3] used PCA algorithm to downscale the image and then used local hold projection (LPP) algorithm to extract the palm vein image features. Since the traditional method requires the researcher to design the corresponding algorithm for feature extraction based on experience and by studying the image's own features. With the proposed

Convolutional Neural Network (CNN) with extremely powerful image feature extraction, the research no longer needs to design features manually. Tarawneh et al.[4] used pre-trained CNNs such as AlexNet, VGG-16 and VGG-19 for palm vein image feature extraction. Jiashu Lou et al.[5] used the CNN with VGG-16 fused attention mechanism as a feature extraction network on infrared palm vein dataset.

Palm vein recognition using deep learning process may lead to longer computation time and higher storage footprint for resource limited devices. So the classical convolutional neural network is no longer suitable for this scenario. It is crucial to choose a more suitable lightweight neural network. In this paper, we re-analyze the principles of each block structure of the classical lightweight model and design a lightweight network model suitable for the scenario of palm vein recognition applications by combining various effective modules, further optimizing each parameter of the network to obtain better results than the baseline model on the CASIA dataset.

2. Related theory

2.1. Palm vein recognition

The current basic process of palm vein recognition: image preprocessing, image feature extraction and recognition. Since the hemoglobin in the palm veins absorbs NIR light better than the surrounding tissue, the light reflected back from the palm is received by the image sensor and the palm veins can be observed by an optical imaging system. Most palm vein images are obtained by illuminating the palm with near-infrared (NIR) light from 720 to 1100 nm. In order to improve the accuracy and stability of recognition, the acquired images are pre-processed. The main purpose of preprocessing is to remove noise and interference from the image while extracting the features of the palm vein. Feature extraction is the core step of palm vein recognition, which aims to extract the feature vector of palm vein from the preprocessed image. The features of palm veins include distribution, density, shape, length and width of blood vessels. After feature extraction, the extracted feature vectors need to be matched with the existing feature library to determine the

identity of the input image. The matching methods include Euclidean distance, cosine similarity, support vector machine and CNN. Among them, convolutional neural networks are increasingly used in palm vein recognition.

2.2. Lightweight convolution neural network

Lightweight Convolutional Neural Networks (LCNNs) have a relatively small number of parameters and low computational complexity compared to traditional deep neural networks. This type of network is designed to obtain fast and efficient performance with limited hardware resources. Their low computational requirements make them ideal for devices with limited resources, such as smartphones, embedded systems, and Internet of Things (IoT) devices. And they also have low memory and storage requirements, making them easier to deploy and maintain in edge computing or other IoT scenarios again.

However, one of the main drawbacks of lightweight neural networks compared to traditional deep neural networks is their lower accuracy. Therefore, the trade-off between model size, complexity, and accuracy must be carefully considered when choosing the right model for a particular task. Despite these limitations, lightweight neural networks have many applications, including image classification, object detection, and semantic segmentation. Low power consumption, real-time performance and small model sizes are critical in specific scenarios. They can also be used in collaborative learning systems, where multiple lightweight models can be trained and combined to achieve higher accuracy.

To achieve high performance with fewer parameters, LCNNs use a variety of techniques. The MobileNets[7] network family introduces depthwise-separable convolution as a module to significantly reduce computational costs and model complexity. The depthwise convolution within the struct operates independently on each channel of the input image, with no parameter sharing between channels, and substantially reducing the total number of model parameters. The subsequent point convolution subsequently allows for increased information exchange and combination between channels, enabling enhanced network expressivity while mitigating overfitting concerns. Due to its superior performance and efficiency, depthwise separable convolution has been widely adopted as a standard for the development of subsequent lightweight neural network models. ShuffleNetV2 introduces the Shuffle-Block module, which leverages group convolution and channel shuffling techniques to enhance the feature representation while reducing the model complexity. Specifically, the Shuffle-Block first applies group convolution to partition the input feature map into multiple subgroups, and then conducts channel blending to mix the outputs of different subgroups for facilitating inter-group information interaction and alleviating the problem of information silos. Finally, point convolution is performed to fuse the mixed channel features and obtain the final output. This module exhibits robustness against noise and outliers by promoting cross-group information exchange via channel shuffling. The MBConv module proposed by EfficientNet uses the swish function as the activation function and the SE module after deep convolution to obtain better network performance, unlike the Block of MobileNetV3. The experimental results are obtained by reprogramming and integrating these powerful modules to obtain better experimental results in palm vein recognition applications.

2.3. Attention mechanism

Attention mechanism has become a powerful technique in deep learning to improve model performance by selectively focusing on different parts of the input data, effectively capturing the most relevant information and ignoring the irrelevant information. In palm vein recognition, the attention mechanism can be used to capture the most informative features in the palm vein image, thus improving recognition accuracy. The attention mechanism allows the model to learn which parts of the input data are most relevant to the task at hand and to selectively weigh these parts when making predictions. This approach is particularly useful in cases where the input data contains a lot of noise or irrelevant information, as is often the case with palm vein images. The module designed based on the attention mechanism can be widely used in various neural networks to optimize the network feature extraction capability and increase the model performance with a small number of model parameters yet.

There are several different ways to implement attention mechanisms in deep learning models, but the most common approach is to use a soft attention mechanism, which computes a set of attention weights for each feature in the input data. These attention weights are then used to weight the feature representations before making predictions. The Squeeze-and-Excitation (SE)[9] Module mentioned in this paper is a module for enhanced feature representation in convolutional neural networks. As shown in the right part of Figure 3, the SE Module consists of two parts: Squeeze and Excitation. In the Squeeze phase, the SE Module compresses the features of each channel individually by a global pooling operation to extract the global statistics of the channel features. In the Excitation phase, unlike the original SE which uses a fully connected layer operation, MobileNetV3's SE uses two 1x1 convolutional layers instead, with the latter layer generating a channel attention vector via an h-sigmoid activation function. By introducing the SE module, the network can more effectively exploit the interdependencies and correlations between channels, thus improving the discriminative and expressive power of the features, and thereby improving the performance of the network.

3. Proposed approach

In this study, the basic process of palm vein recognition system is implemented to extract the captured palm vein images with ROI and then perform image enhancement. For feature recognition, a palm vein recognition method based on a lightweight neural network is proposed, which is a recombination of classical and efficient lightweight modules. This subsection focuses on the image enhancement part and the principles of each lightweight module needed to design the network.

3.1. Image preprocessing

The training data used for the experiments in this paper are derived from the CASIA dataset. The CASIA dataset is a global database, collected and published by the Institute of Automation, Chinese Academy of Sciences, containing 7200 palm images from 100 volunteers, with 6 images per palm sample, take simultaneously using 6 electromagnetic spectra. The goal of this dataset is to increase sample diversity within categories and to improve accuracy by capturing palm vein images using a stable device. Five spectra in the 460 nm, 630 nm, 700 nm, 850 nm, and 940 nm bands are used to capture

in the dataset, where the palm images obtained by capturing in the 850 nm and 940 nm spectra contain relatively clear vein images that can be used as a palm vein library.

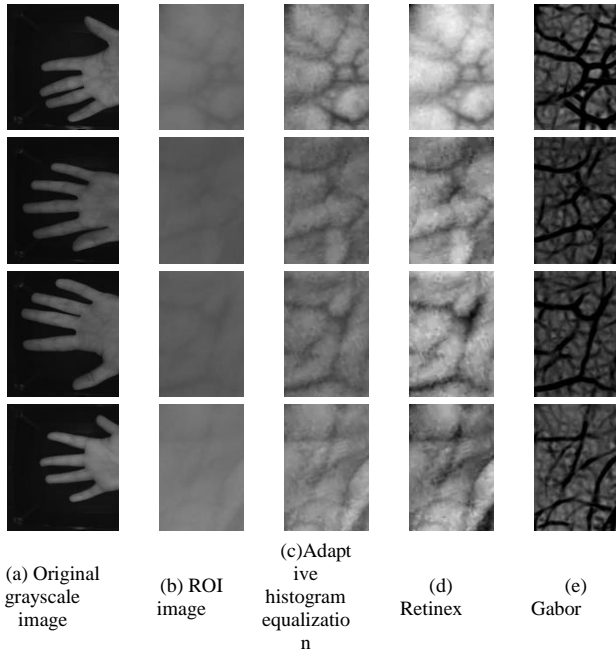


Fig. 1 ROI and different enhancement images

In this paper, the widely used tangent positioning method is used in selecting the ROI positioning method to correct the palm deflection direction through the gap between the fingers, and the obtained graph is shown in Fig. 1(b). When enhancing ROI images, as shown in Fig. 1(c) for the image after adaptive histogram equalization enhancement, the image is divided into local regions and then the histogram equalization is applied to each local region. Fig. 1(d) shows the Retinex enhanced image by decomposing the image at multiple scales and then performing brightness enhancement at different scales. Fig. 1(e) shows the Gabor enhancement map obtained by filtering in different directions and frequencies through Gabor filters. Since the adaptive enhancement and Retinex enhancement methods may over-enhance the brightness and contrast of the image, resulting in increased image distortion and noise. In contrast, the Gabor filter performs well in texture analysis and feature extraction, capturing texture information and edge information in the image and generating an image with high contrast and sharpness, which can better present the features of the palm vein image. Finally, the Gabor-enhanced image is selected as the input image for palm vein recognition.

3.2. Design Network

The purpose of designing the network in this paper is to be able to extract useful features when processing palm vein images, while keeping the network lightweight and efficient. In order to design a suitable lightweight network and to minimize repetitive work, the optimal network is selected as the baseline by filtering among the classical lightweight networks. The final network, called PalmNet, is obtained by optimizing again on the basis of the baseline.

Specifically, this network uses several classical convolutional blocks, such as the MobileNetV3 block, ShuffleNetV2 block and EfficientNet-lite single layer in Fig. 2, to improve the feature extraction capability of the network and reduce the computational effort. The block of ShuffleNet

in Fig. 2(a) splits the input feature map into two parts with adjacent channels and performs the operation separately. One part uses depth-separable convolution for feature acquisition, and the output is stitched with the other part to obtain a new feature map that is the same size as the input. In order to achieve information exchange between the two groups after segmentation, the new output feature map is subjected to channel rearrangement. This design of rearrangement after segmentation allows the information in the network to be more fully exchanged and transmitted, improving the diversity of features and the expressiveness of the network. As in Fig. 2(b) the block of MobileNetV3 is constructed using a module of deeply separable convolutional layers, placing the SE module at the last point of the module and then using residual connections, which improves the network and alleviates the problem of gradient disappearance while keeping the computational effort low. The MBConv module is shown in Figure 2(c), which is proposed and adopted by EfficientNe. Unlike the MobileNetV3 module which is placed in point convolution, where the number of channels is reduced and then the attention module is used after the number of channels is reduced, the MBConv module places the SE module after the deep convolution and uses a special activation function internally to improve the network performance while maintaining high efficiency.

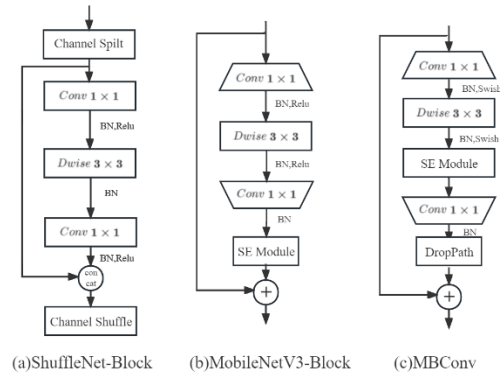


Fig. 2 Different network blocks

In addition, a global average pooling layer and a fully connected layer are added to the network to integrate and classify features. The global average pooling layer averages the features of each channel to obtain a feature vector with a fixed size. The fully connected layer then maps this feature vector to a 1024-dimensional feature space and finally outputs a probability value used to determine whether the palm veins match or not.

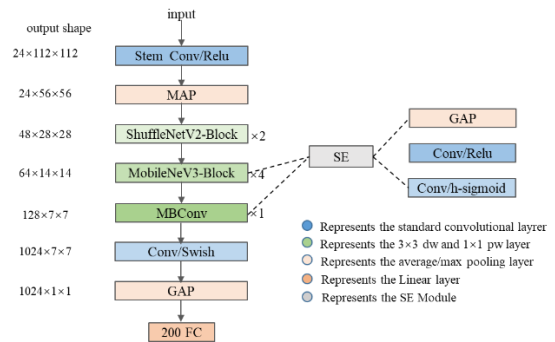


Fig. 3 A detailed view of PalmNet

By reprogramming those three important lightweight modules, the PalmNet is designed as in Fig. 3. Input the palm

vein grayscale map into the network, scale the feature map size slowly and gradually increase the number of channels. This processing helps to improve the perceptual field of the network, and also reduces the number of parameters and improves the computational efficiency of the network. As the feature map is gradually scaled down, the network can better capture the details and texture information of the palm vein image, while the increase in the number of channels helps to improve the expressiveness of the network. stage2 and stage3 contain SE Modules that can further improve the robustness and accuracy of the network. Most of the hyperparameters in the network are re-explored to be more lightweight than the baseline network and obtain a better performance in palm vein recognition.

4. Experimental Results and Analysis

4.1. Detail of the experiment

To ensure the fairness of the experiments, the experiments are conducted based on pytorch, using RTX3060, using the CASIA dataset containing 850 nm, and the dataset is primitively augmented using luminance enhancement, contrast enhancement, and rotation, and the obtained palm vein training set contained 4750 and the test set contained 1185. The optimizer uses Adam, the loss function uses Cross entropy loss. Epoch, batchsize, learning rate and weight decay are set to 1500, 32, 0.0001 and 0.0005 respectively.

4.2. Ablation and Results of experimental

In designing a suitable network for vein recognition, the first step is to start from a classical neural network, using the same training method and data, as can be seen in Table 1, ShuffleNetV2 gives the best results. For example, PalmNet-0.5x2413 means that the PalmNet with a width factor of 0.5 contains 2 ShuffleNet-Blocks, 4 MobileNetV3 Blocks, 1 MBConv, and the expansion factor is set to 3.

Table 1. Results of experimental

model	Top-1 Acc(%)	Params(M)	Flops(M)
MobileNetV1-0.5x	95.19	0.92	153.06
MobileNetV2-0.5x	95.53	0.71	97.90
MobileNetV3_Small	94.18	1.48	59.45
ShuffleNetV2-0.5x	96.96	0.55	37.46
PalmNet-0.5x	97.47	0.49	43.40
PalmNet-1.0x	97.89	1.37	192.77
palmNet-2.0x	99.07	4.47	765.12
PalmNet-0.5x2223	96.20	0.56	37.67
PalmNet-0.5x4223	97.05	0.56	40.11
PalmNet-0.5x6223	96.46	0.56	42.56
PalmNet-0.5x2323	97.63	0.58	43.09
PalmNet-0.5x2313	95.95	0.46	37.98
PalmNet-0.5x2413	97.47	0.49	43.40
PalmNet-0.5x2412	96.80	0.44	34.11
PalmNet-0.5x2411	96.20	0.39	24.83

With ShuffleNetV2 as the target, the initial network is set up to perform a trade-off between parameters and computational effort. The initial network, like the baseline,

divides the network into three stages. In order to obtain smaller parameters for the network, the options for the stages with smaller parameters are directly selected and the control variable method is used in these options.

The performance metrics obtained for the networks with subscripts [2,2,2,3] to [6,2,2,3] in Table 1 can be found that the number of first stages cannot exceed 4. Fixing the number of first stages to 2, the best two options are chosen to obtain the best performing options [2,3,2] and [2,4,1], respectively, and the second option is chosen after a trade-off between the number of parameters, the amount of computation and the accuracy. When designing the palm vein recognition network, the expansion factor of the SE module in MobileNetV3-block and MBConv, as well as the number of expansion layers used for the first time within these two modules, were set to 3. For fairness, different sizes of expansion factors are set for the last three data in Table 1, respectively. It can be seen that the expansion factor layer correlation, the larger the expansion factor the better the effect may be, due to the number of covariates. Due to the limitation of the number of parameters, 3 is finally chosen as the expansion factor of the feature extraction network.

5. Conclusion

In this paper, a palm vein recognition network with three stages is designed using a combination of Block of ShuffleNet, Block of MobileNetV3, and MBConv of EfficientNet. The number of stages and expansion factors of the network are chosen based on the parameters and computational effort of the baseline. The proposed network achieves better performance compared to the baseline. This work demonstrates the effectiveness of combining different block structures when designing palm vein recognition networks and the importance of carefully selecting network parameters to optimize feature recognition performance. The accuracy and efficiency of the proposed method for palm vein recognition is improved. Since this paper is only experimented and validated on a closed set dataset, future optimization work will be performed on an open set dataset to validate again.

References

- [1] Ladoux P O, Rosenberger C, Dorizzi B. Palm vein verification system based on SIFT matching, Advances in Biometrics: Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009. P.1290-1298.
- [2] Michael G K O, Connie T, Hoe L S, et al. Design and implementation of a contactless palm vein recognition system, The 1st Symposium on Information and Communication Technology. 2010, p. 92-99.
- [3] Wang J G, Yau W Y, Suwandy A, et al. Person recognition by fusing palmprint and palm vein images based on "Laplacianpalm" representation. Pattern recognition, 2008, 41(5), p. 1514-1527.
- [4] Tarawneh A S, Chetverikov D, Hassanat A B. Pilot comparative study of different deep features for palmprint identification in low-quality images. arXiv preprint arXiv:1804.04602, 2018.
- [5] Lou J, Wang B. Palm Vein Recognition via Multi-task Loss Function and Attention Layer. arXiv preprint arXiv:2211.05970, 2022.

- [6] Watanabe M. Palm Vein Authentication in *Advances in Biometrics*. Heidelberg, Germany: Springer-Verlag, 2008, p. 75-88.
- [7] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3, *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, p. 1314-1324.
- [8] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design, *Proceedings of the European conference on computer vision (ECCV)*. 2018, p. 116-131.
- [9] Tan M, Le Q. Efficientnetv2: Smaller models and faster training, *International conference on machine learning*. PMLR, 2021, p. 10096-10106.
- [10] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design, *Proceedings of the European conference on computer vision (ECCV)*. 2018, p. 116-131.