Design of a Lightweight Palm-Vein Authentication System Based on Model Compression

ZIH-CHING CHEN[1], SIN-YE JHONG[2] AND CHIH-HSIEN HSIA[2,3]
¹Department of Economics
National Taiwan University
Taipei, 106 Taiwan
²Department of Engineering Science
National Cheng Kung University
Tainan, 701 Taiwan
³Department of Computer Science and Information Engineering
National Ilan University
Yilan, 260 Taiwan
E-mail: chhsia625@gmail.com

Palm-vein authentication is a secure and highly accurate vein feature authentication technology that has recently gained a lot of attention. Convolutional neural networks (CNNs) provide relatively high performance in the field of image processing, computer vision, and have been adapted for feature learning of palm-vein images. However, they often require high computation that not only are infeasible for real-time vein verification but also a challenge to apply on mobile devices. To address this limitation, we proposed a lightweight MobileNet based deep learning (DL) architecture with depthwise separable convolution (DSC) and adopt a knowledge distillation (KD) method to learn the knowledge from the more complex CNN, which makes it small but effective. Through the depth of separable convolution, the number of model parameters is significantly decreased, while still remaining high accuracy and stable performance. Experiments demonstrated that the size of the proposed model is 100 times less than the Inception_v3 model, while the performance can go beyond 98% correct identification rate (CIR) for the CASIA database.

Keywords: palm-vein recognition, knowledge distillation, lightweight, depthwise separable convolution, biometrics image.

1. INTRODUCTION

Accurate and reliable identity authentication is of great importance nowadays. The thought of measures that cannot be copied or stolen resulted in the development of user authentication with the help of biometrics. The differences and uniqueness among living things make it suitable to use biometrics such as the face, voice, finger-print, finger-vein, palm-print, palm-vein, iris, and DNA to authenticate users, especially in today's society where secure personal recognition becoming more and more crucial [1].

Among the various biometric features, palm-vein has received enormous attention as a powerful biometric identifier for user authentication because if its high security and liveness-detection. Firstly, palm-vein authentication has high security as it is interior biological information of the body concealed naturally, therefore harder for intruders to copy. Second, the vein recognition ensures liveness of the samples since it can be captured by the device without the blood flowing underneath the skin. which has dense texture, large

data volume, and high anti-counterfeit degree [2-4]. Finally, the collection of palm vein is non-intrusive since it acquires palm vein pattern image with contactless devices [4]. Many biometric devices require contact-scanning, but the recent COVID-19 pandemic has prompted a host of changes in the way we live and work, including people's acceptability of touch-based devices. Biometrics that can be collected through contactless devices, including palm-vein have received enormous attention since they can establish a more safe, healthy, and hygienic identification system.

Because the palm-vein images can be obtained through contactless devices, the difference in the position of the palm can easily cause the palm vein images to be rotated or tilted. Therefore, the feature extraction must be powerful in order to solve the real life tasks. The method of vein feature extraction requires classification method to establish a complete identification system [4]. The traditional feature extraction algorithms include statistics-based and geometry-based methods. Statistical methods have been used to extract global features [6-8] during the whole palm vein imaging, while geometric methods have been compared using similarity of local geometric structures in the image [9, 10]. However, the identification technology based on single feature may not be accurate or reliable. Li *et al*. [11] is an affine invariant feature extraction method proposed by using the four local feature algorithms. This method has excellent robustness for the rotation and inclination of palm vein images. In [12-14], multiple algorithms were proposed to use feature extraction and fuse to improve the identification rate. Classification methods are often implemented by machine learning-based methods, such as principal component analysis (PCA) [15] and support vector machine (SVM) [16, 17].

A convolutional neural network (CNN) is a neural network (NN) architecture contains convolutional layers and is mainly used in the field of image processing, image classification and computer vision tasks due to its strong performance. In the field of computer vision, CNN in various applications including image recognition, has successfully been applied to biological characteristics of vein recognition [18-20] in recent years. This method although need a lot of mark samples for training and establish a model, the related research results show that CNN can get very good effect in palm-vein recognition. However, their methods are relatively complex and time consuming.

CNN's high computation complexity and the limitation of storage space and power, makes its application in the embedded platform still a big challenge, especially for large dataset. Consequently, it is challenging to deploy these networks under limited computational resources, such as in mobile devices, and unable to meet the performance requirements of mobile payment of hand vein recognition system. Thus, compressing a trained model without a significant loss in performance has become an increasingly important task. CNN model compression can be carried out by quantization, pruning, or manual design of NN architecture (convolution mode). Well-known quantization models include Deepcompression [21], Binary-Net [22], Tenary-Net, dorefa-net, SqueezeNet, MobileNet V1&V2, shuffle Net V1&V2 [23]. Quantization turns a lot of mathematical operations into Binary-Net operations, saving time and space for forward propagation. In order to reduce the model parameters, SqueezeNet replaced the convolution kernel of 3×3 with the convolution kernel of 1×1. The main work of MobileNet V1 is to use the depthwise separable convolutions (DSC) instead of the standard convolutions to solve the computing efficiency and reduce the number of parameters of the convolutional network. However, the training procedure of DSC is time consuming and require the full training set. Among these tech-

niques, knowledge distillation (KD) and the design of the NN architecture does not change the original network structure, while model pruning and model quantization such as binary networks focus on reduce the size of the original complex model as much as possible, thus resulting in a significant change of the original network structure, such transformation is often irreversible.

Considering the above problems, this study proposed a lightweight CNN for palm-vein recognition technology. To avoid significant performance drop at the lightweight CNN, which has a large capacity and parameter size gap compared to complex models, we distil the knowledge from the complex but high performing model. Through the DSC and KD method, the number of model parameters is significantly decreased, while still remain high accuracy and stable performance, therefore suitable for real life application. The architecture of the proposed palm-vein authentication system, as shown in Fig. 1.
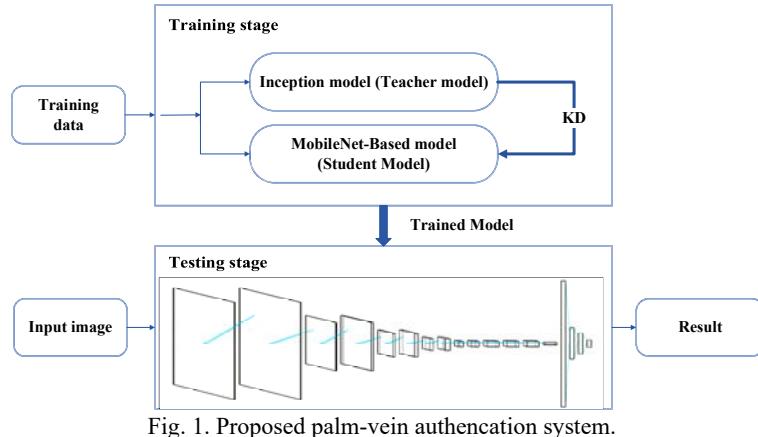

Fig. 1. Proposed palm-vein authencation system.

2. RELATED WORKS

2.1 CNN

Each module of the CNN is composed of the convolutional layer and the pooling layer, which is a DL model [24] formed by the continuous superposition of the modules and the addition of the full connection layer. The convolutional layer can extract the texture, color and other characteristics of the image through convolution operation to obtain the feature map, and then conduct dimensionless sampling through pooling layer. Sparse interactions, parameter sharing and equivariant representations are the three major concepts of CNN, to enhance the effectiveness of network training. Sparse interaction and parameter sharing reduce the number of parameters to be stored in the whole model, which can effectively reduce the computational burden and improve the computational efficiency. The mechanism of parameter sharing and appropriate pooling strategy also gives CNN the characteristics of invariant height to translation and scaling that an architectural diagram of a CNN, as shown in Fig. 2.
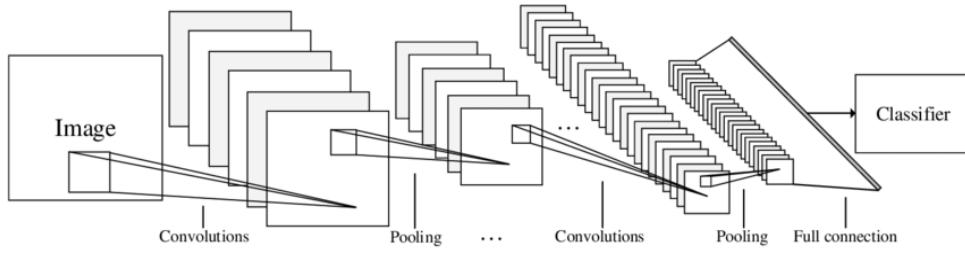
Fig. 2. Architecture of LeNet [24].

2.2 Inception_v3

In 2014, Google published a CNN architecture, called GoogleNet (Inception_v1), which proposed a parallel module called Inception, as shown in Fig. 3 (a). Inception aim to improve network performance by increasing the depth of the network as opposed to Alexnet or VGGNet, by increasing the width of parallel modules. To increase the adaptability of the model, Inception uses convolution kernel of different sizes to capture image features simultaneously, and adds a convolution kernel of 1×1 before the convolution kernel of 3×3 and 5×5 to reduce the dimension, so as to reduce the complexity of the model. Inception_v2 uses two 3×3 convolution kerns instead of one 5×5 convolution kernel, which enables CNN to learn more about features, and it contains batch normalization method to normalize each layer, which significantly reduces the training time required. In Fig. 3 (b), the Inception_v3 architecture [25] is shown in Fig. 3 (b), which splits a large two-dimensional convolution into two one-dimensional convolution, saving required parameters and reducing operation time.
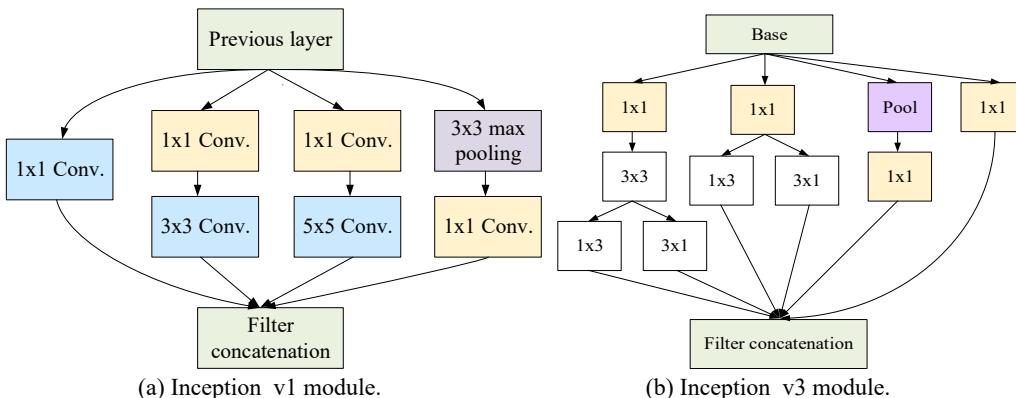

(a) Inception_v1 module.   (b) Inception_v3 module.

Fig. 3. Architecture of inception module.

2.3. ResNet

As the number of layers of CNN gets deeper and deeper, gradient loss and degradation problem are easy to occur during training. The degradation problem is caused by the fact that the gradient cannot be reversed and error accumulates, so that the accuracy of the deep

network becomes saturated and even performance degrades. ResNet [26] refers to and modifies the practice of VGG19, and establishes residual unit through short-circuit connection mechanism, as shown in Fig. 4. Residual learning mechanism established by ResNet can solve the problem of difficult training in deep CNN model. For a residual unit structure composed of several hidden layers, when the input is x, the learned feature is marked as $H(x)$, and the learned residual $F(x) = H(x) - x$. When the residual error is 0, only the identity mapping is done for the accumulation layer at this time. Although the depth increases, at least the network performance will not decline. In fact, the residual error will not be zero, which means that the accumulation layer can learn some new characteristics, thus improving the performance of the network.
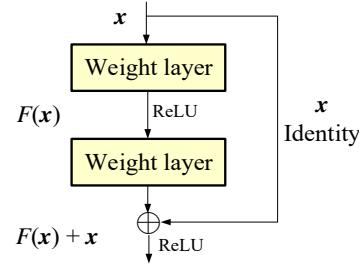

Fig. 4. Residual block.

2.4 Knowledge Distillation

The idea of training a shallower or a cheaper model under the supervision of the larger ones. Trained NN produces peaky probability distributions, which may be less informative. The soft probabilities output by a trained teacher model contains more information of the data than just the class label itself. If multiple classes are assigned high probabilities for a vein image, then that is likely that the image lie close to a decision boundary between those classes.

In Hinton *et al.* [27], knowledge is defined as the teacher model's soft outputs after the final layer, which carries more information than the one-hot encoded labels since there are extra signals of the intern-class similarities learned by the teacher model. Since the soft probabilities output by a trained teacher network contains more information than only one class label, forcing a student to mimic these probabilities can thus make the student network learn more knowledge than just the training labels alone. The basic idea of KD is described as follows.

Concretely, for an image data set $\{x_i, y_i\}$, $i = 1, 2, ..., n$, where x_i is the input image and y_i is the labeled category. If t is set as teacher model, $P_t = \text{Softmax}(z_t/T)$ is its predicted output probability, while z_t is the input of Softmax layer. T denotes the temperature parameter. Similarly, we can define $P_s = \text{Softmax}(z_s/T)$ for the Student model s. The function definition of Softmax in Eq. (1).

$$P_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \tag{1}$$

where P_i represents the output probability of the ith category, and z_i and z_j are inputs to the

Softmax layer. When $T = 1$, it is the general Softmax conversion, and the probability distribution of the predicted results among different categories is the probability with extreme value, that is, the probability of the correct category is quite close to 1, while the probability of all other categories is very close to 0. When $T > 1$, the probability distribution generated by Softmax function will become flatter and softer, thus provides more implicit information such as the degree of similarity between different classes, which can more judgment conditions for the training process of the model. As for the student model, it learns according to the loss function of Eq. (2).

$$L = \alpha \cdot L_{hard} + (1 - \alpha) \cdot L_{soft} \tag{2}$$

where L_{hard} denotes a typical cross entropy loss function in the classification problem (*i.e.*, $T = 1$), and L_{soft} is a soft cross entropy loss function predicted by teacher model. The parameter α is the weighting factor that balances the two cross entropy loss functions. Fig. 5 is a schematic of KD.
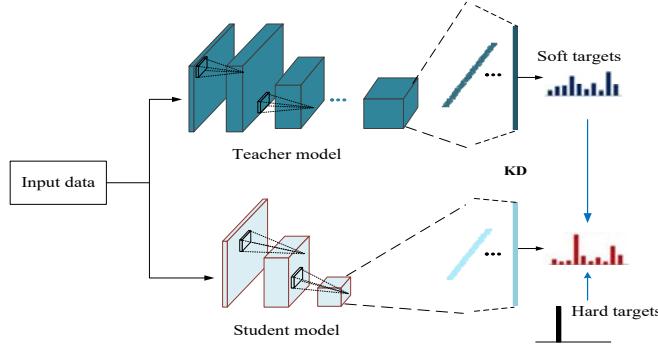

Fig. 5. The schematic overview of KD.

2.5 Depthwise Separable Convolution

In recent years, DSC has been widely used in many DNN with high performance, such as MobileNet and Xception [28], which replaces the traditional convolutional layers to reduce CNN computational cost and memory usage. The DSC, which has shown great efficiency in network design, consists of depthwise convolution and pointwise convolution, which mainly performs channel-wise feature extraction and the task of combining the separated features to generate new features. The operation is described in the form of data flow, as shown in Fig. 6.
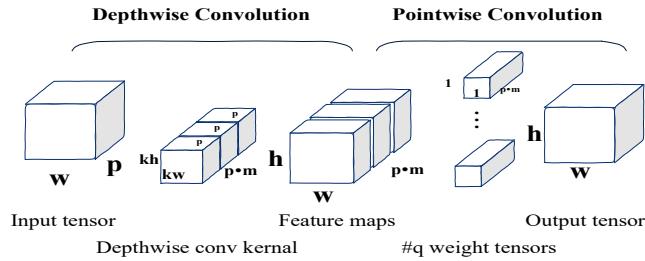

Fig. 6. The schematic overview of DSC.

3. PROPOSED A LIGHTWEIGHT MOBILENET-BASED PALM-VEIN AUTHENTICATION SYSTEM

In this study, we implement the vein recognition technology to build a palm-based system. The system is developed and extended based on deep CNN technology. Firstly, the image of palm-vein will extract the vein characteristic region to be identified through region of interest (ROI) algorithm, the ROI image will then be used as the training data of the model.

3.1 Teacher Model Training Based on Inception_v3

Inception_v3, which is widely-used image recognition model, was used for the training of palm-vein image model. The model, combining many ideas proposed by different studies over the years, is based on the idea proposed by Szegedy *et al*. [25]. In Fig. 7, the model is made up of symmetric and asymmetric building blocks, including convolutions, average pooling, max pooling, contacts, dropouts, and fully connected layers, with Batchnorm layer extensively used throughout the model and applied to activation inputs and loss computed via the Softmax layer. Here, the main purpose was to achieve high accuracy rate, and factors such as complexity and power consumption were not needed to be considered temporarily.
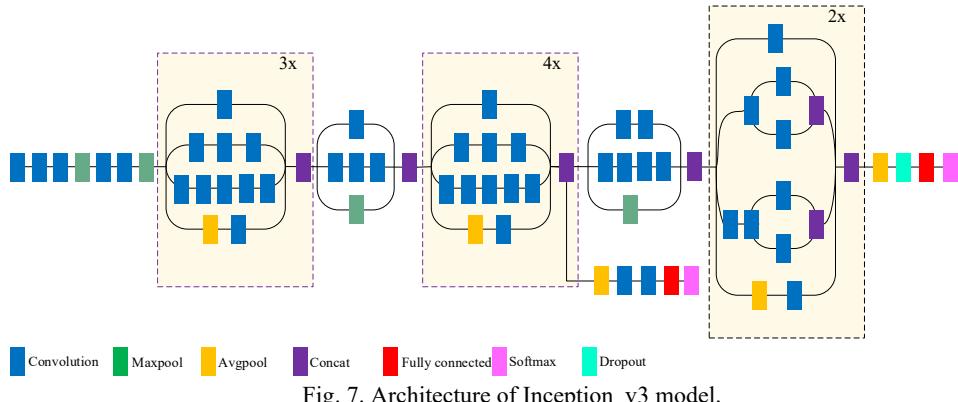

Fig. 7. Architecture of Inception_v3 model.

3.2 Construct a MobileNet-based Student Model

The CNN architecture is composed of 8 modules and 68 convolutional layers. In order to achieve a lightweight model contains fewer parameters, the depthwise and pointwise convolutional layers are used to extract features by referring to the classical model architecture of MobileNet proposed by Google [28]. Such architecture is often found in lightweight networks because it has less computing costs than traditional CNN. The architecture uses 256,779 parameters, 40 times less than the 11 million parameters used by highly accurate models such, as ResNet18, and 100 times less than Inception_v3. The following is an introduction based on input, intermediate and output, and the detailed architecture diagram is shown in Fig. 8.

A. Input flow:

The input process is composed of two modules. The input layer of this study is 256×256×3. After the operation of this process, the input image will be generated by 256×256×3 operation to generate 64×8×8 feature images.

B. Middle flow:

After the feature extraction and sampling of the previous layer, the group was added into the convolution layer for depthwise convolution, and the pointwise convolution was combined to extract image information. This process extracts information by maximizing the features of the module, and the dimension of the feature graph is reduced from 64×8×8 to 128×4×4.

C. Output flow:

For the final process, the module composed of DSC is first used for the final feature extraction. The dimension of the feature graph is increased from 128×4×4 to 512×2×2. Here, since the feature graph has been downsample for many times, max pooling is no longer used. Finally, project the outputs to 400 dimensions.
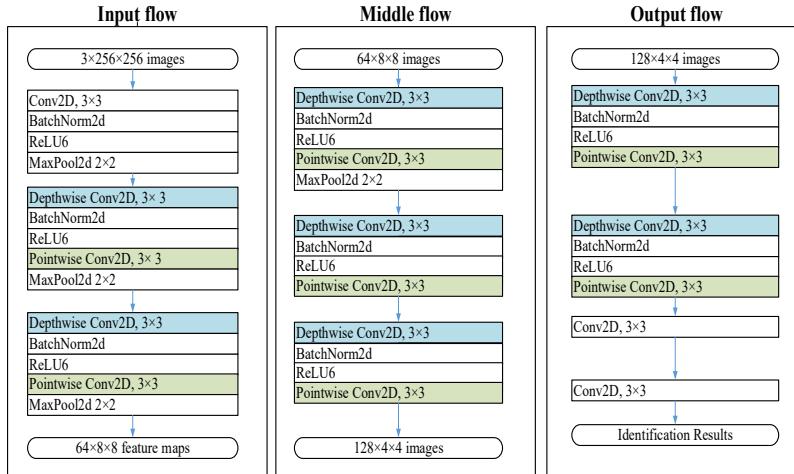

Fig. 8. The proposed MobileNet-based CNN structure.

3.3 Student Model Training Using Knowledge Distillation

In order to have the student model obtain similar performance as our teacher model, which is Inception_v3, we want to make the student model mimic the teacher model's output from the last layer. We examine the Kullback-Leibler (KL) divergence [32] between the teacher's and the student's outputs as Eq. (3), in order to determine if the student model is learning properly from the teacher model and the cross-entropy loss between the outputs of the two models.

$$Loss = \alpha T^2 \times KL\left(\frac{Teacher's\ Logits}{T} \mid \frac{Student\ Logits}{T}\right) + (1-\alpha) \times (Orignial\ Loss) \tag{3}$$

By KL Divergence loss (KLDivLos), here it requires inputs to be probability distributions and log-probability distributions, and that is why we are using Softmax and Log-Softmax on teacher/student outputs, which were raw scores. During the retraining process, the student can query the teacher for knowledge, which are input-output pairs in this case. During the training process, we horizontally flipped the images for data augmentation to increase the robustness of the system.
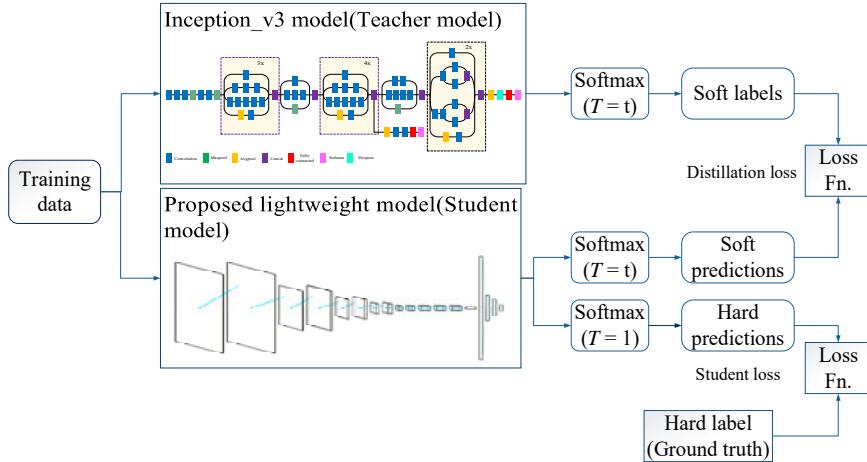

Fig. 9. The proposed method using KD method.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This study uses Pytorch to train and test the model. CASIA dataset was used for training. For the experiments on CASIA, we randomly divide each dataset into a training set, validation and test set, each contains 1,600, 100, and 700 images, respectively. We first run each model for 150 epochs using Adam optimizer with *learning rate* = 0.001, then we run the models for 50 epochs with SGD with the *learning rate* = 0.001 and the *momentum* = 0.002. Standard data augmentation was applied to the palm vein dataset. Compared with other works, the model can achieve stable convergence with relatively few training times. For KD, we use Inception_v3 as our teacher model, a high-performance network, and the proposed CNN network as our StudentNet, a lightweight network.

4.1 Public Database

To evaluate the proposed method, the most representative multi-spectral palm-vein image database V1.0 (CASIA dataset) was used in this study. The CASIA database used contactless devices to obtain a total of 7,200 palm-vein images from 100 different peoples, all of which were 8-bit gray-scale JPEG images with a pixel size of 768×576. These palm images using six different wavelengths of light to capture at the same time, corresponding to six wavelengths 460 nm, 630 nm, 700 nm, respectively, 850 nm and 940 nm and visible light, and for each person's palm captured images are from two times, the time interval of

more than a month, each time contains three samples. Between the images has a certain degree of gestures change, this material has a considerable diversity of samples, very close to the actual use. To verify the adaptability of the system to multi-spectral data, 850 nm and 940 nm images with characteristic images of the vein were used as the experimental efficacy evaluation in this work.

4.2 Performance of Classic Models

The proposed method results are reported in Table 1. We tested three different network architectures and four methods. The first 11 experiments are with classic models, including VGG16, ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, DenseNet121, DenseNet201, Inception_v3, and MobileNet, which are widely used in image classifying. MobileNet is also known as a classic lightweight model. The last two experiments are with the StudentNet designed with DSC layers. In the last experiment, we use Inception_v3 as TeacherNet for KD since it has the best performance of all the models. For the inference time, we compared the time needed for the model to predict the 700 images in test set. Overall, our work achieves the best result, same accuracy as it is TeacherNet, Inception_v3, but costs way less memory.

Table 1. The comparison of methods to accuracy and cost.

Models	Validation	Memory cost (KB)	Total size (MB)	Used parameter	Inference time (sec)
VGG16 [28]	68.72%	138,000	902.32	135,079,944	4.5591
ResNet18 [23]	93.32%	43,000	158.31	11,279,112	1.7421
ResNet34 [23]	90.12%	82,000	258.31	21,387,272	2.3341
ResNet50 [23]	91.37%	92,000	615.88	23,917,832	3.4403
ResNet101 [23]	89.00%	164,000	952.03	42,909,960	4.8761
ResNet152 [23]	91.50%	224,000	1337.45	58,553,608	6.6181
DenseNet121 [29]	90.87%	28,000	89.88	7,158,856	3.4053
DenseNet201 [29]	92.25%	72,000	239.616	18,477,128	4.8762
Inception_v3 [24]	93.75%	85,000	324.68	27,161,264	2.6270
MobileNet [26]	93.5%	10,000	289.47	2,480,072	1.7000
StudentNet train from scratch	74.63%	1,023	83.56	256,779	1.1200
StudentNet train with KD	93.87%	1,023	83.56	256,779	1.0792

The training process of the classic models is shown in Fig. 10. Although most training accuracies and training losses reach the stable state after training for 100 epochs, the accuracies and losses of validation set are unstable, and can drop sharply even after the training process achieve stable convergence. It takes longer for Inception_v3 model to train, but the accuracy of the model is the best among all models, and therefore it is been selected as our TeacherNet.

4.3 Hyper Parameters of Knowledge Distilling Training Process

For the KD process, we first examine the different parameters of KD. The comparison of different hyper parameters of KD, as shown in Table 2.
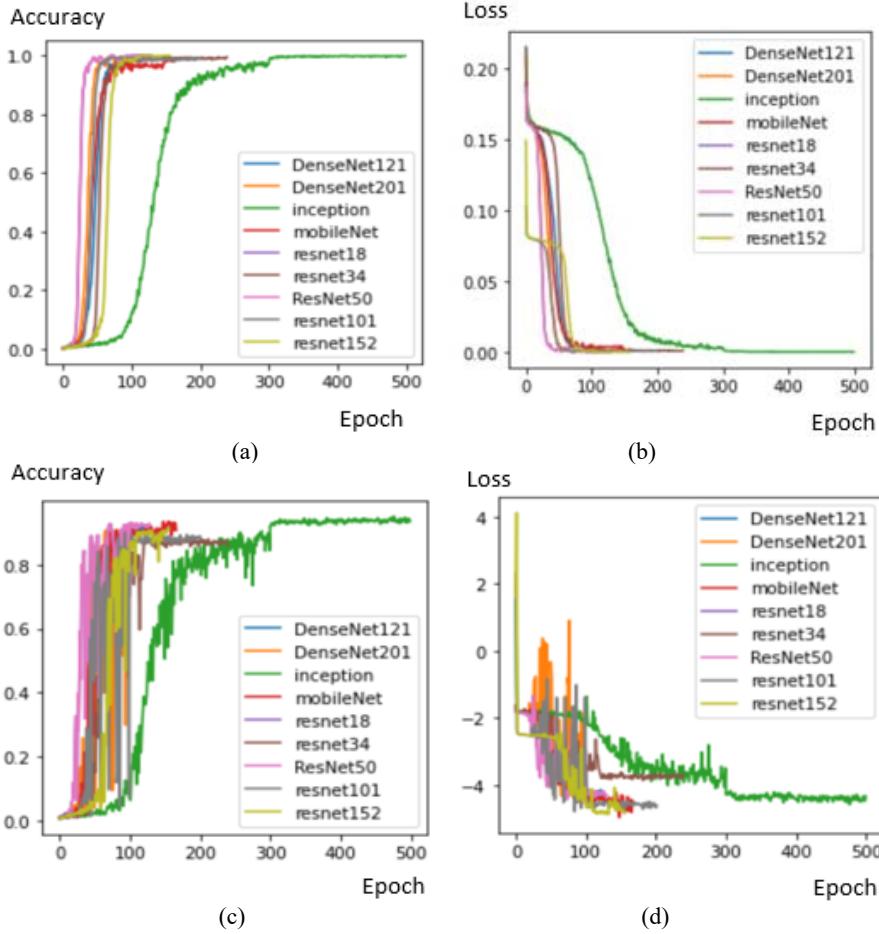
Fig. 10. The training process with different models; (a) Training accuracy; (b) Training loss; (c) Validation accuracy; (d) The log of validation loss.

Table 2. The comparison of hyper parameter α for KD.

α	Testing accuracy rate
0.1	90.62%
0.3	92.20%
0.5	93.87%
0.7	92.50%
0.9	89.88%

The StudentNet reaches its best performance when $\alpha = 0.5$. As seen in Table 3, the accuracy drops as α became bigger or smaller. We hypothesized that because our StudentNet has low capacity, that is, it may not have enough capacity to minimize both the training loss and the KD loss. It might end up minimizing only one loss, at the end of training, that is, the KD loss. We found out that it is best to set $\alpha = 0.5$ to achieve best performance of the model.

Table 3. The comparison of hyper parameter *T* for KD.

T	Testing accuracy rate
10	89.75%
20	93.87%
30	93.12%

Where *T* represents a high temperature here, can theoretically mitigate the peakiness of the teacher logits and may result better performance. The result in Table 3 shows that high temperature does increase the overall performance for the training process, compared to popular choices $T \in \{3, 4, 5\}$ [29].

4.4 Effectiveness of the Training Framework

From Fig. 12 and Table 1, it can be observed that KD reduces the variance of the StudentNet's loss and accuracy which makes the network a lot more stable with less memory cost. With the help of TeacherNet, our proposed model converge significantly better, as shown in Fig. 12, comparing to the result in Fig. 11, where the network was trained from scratch.
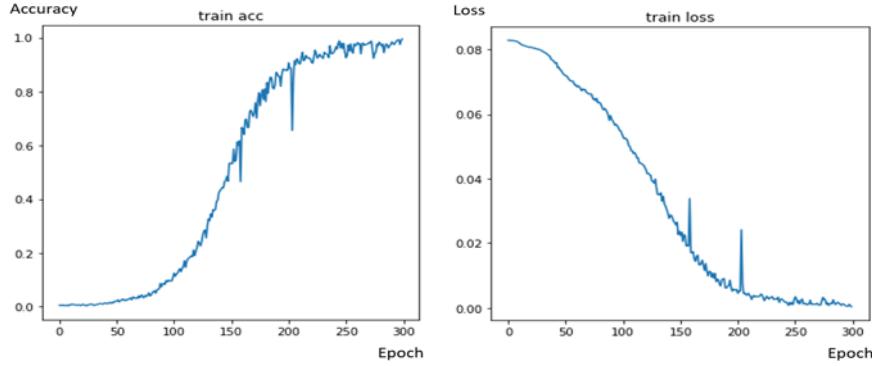

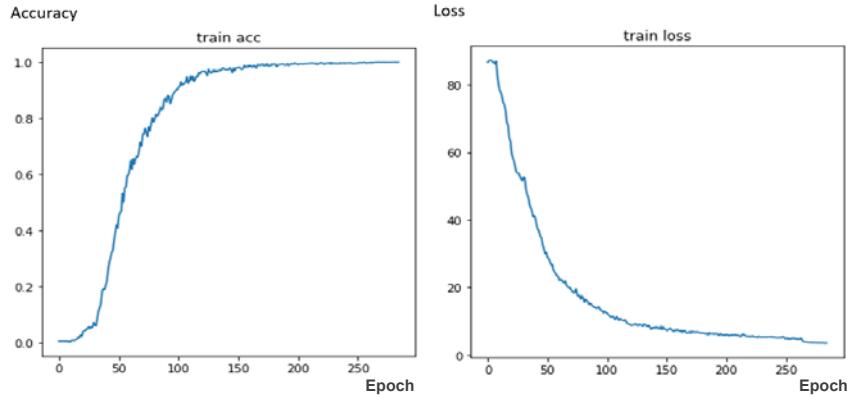Fig. 11. Learning curve of StudentNet trained from scratch.

Fig. 12. Learning curve of StudentNet trained with KD.

4.5 Network Performance Evaluation

Fig. 13 shows the parameters used by the different classic models. We can see that our proposed lightweight model's size has been significantly decreased relative to the classic CNN models. Our proposed architecture, also known as the StudentNet in the KD process, only contains 256,779 parameters, which is approximately 100 times less than TeacherNet, Inception_v3. To be applied in embedded systems and solve real life tasks, inference time is also a crucial factor. For the experiment, we examine the inference time for the models to predict 800 images. Fig. 14 shows that it only costs our proposed method 1 second to get the result, while other models need 2 to 6 seconds to predict the label of 800 vein images. The inference time reduces to one half of the TeacherNet model, which is considerable. We compare the performance of recognition, as shown in Table 1. We can see that the lightweight CNN's performance only drops slightly compared with that of the TeacherNet in our dataset. Considering the reduction of resources, the performance of our proposed lightweight CNN is satisfactory.
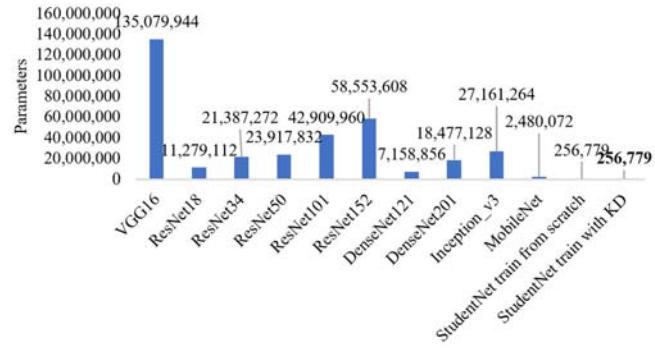

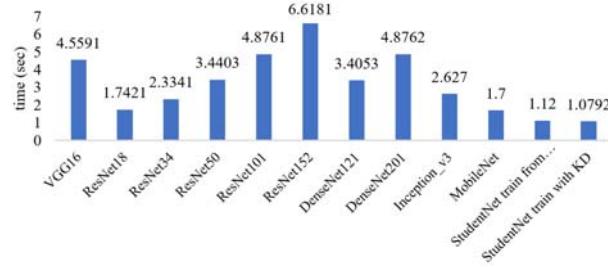Fig. 13. Parameters comparison of different networks.

Fig. 14. Inference time comparison of different networks.

4.6 Evaluation of Palm-Vein Identification Efficiency

In this study, two wavelengths of near-infrared (NIR) light from the CASIA dataset were used for analysis. However, in order to accurately evaluate the performance of the method proposed in this work, palm-vein images were extracted from different periods of time in the experimental analysis. The second time period was used for testing. In this

study, the left and right hands of 100 people were regarded as different people, and a total of 200 categories could be obtained. Thus, the number of amplified samples could be compared to make the experimental analysis more complete.

Compared with the CNN model applied in the vein field in recent years, the correct identification rate (CIR) of the proposed method is higher than other methods, as shown in Table IV, which makes this method not limited by hardware devices, but can be applied in handheld platforms or embedded systems, which is more in line with the application requirements in real life.

Table 4. CIR evaluation.

Methods	CIR
Wirayuda *et al*. [30]	90.87%
Mirmohamdsadeghi *et al*. [31]	93.20%
Zhou *et al*. [10]	97.50%
This work	98.12%

5. CONCLUSION

A new mobile vein recognition model is presented as a lightweight MobileNet network designed with DSC. This can not only significantly decrease the memory cost of the network used in biometric image recognition, but at the same time stabilize the convergence of the network and reduce the inference time of the network significantly, which makes it suitable for implementing on mobile devices. The experimental results in the most representative public CASIA database in order to analyze the performance have shown that the proposed method is capable of achieving great recognition rate and the instantaneous, and can be effectively used to mobile application in the future, which brings the power of artificial intelligence (AI) driven application directly into the practitioner's hand.

REFERENCES

1. C.-H. Hsia, J.-M. Guo, and C.-S. Wu, "Finger-vein recognition based on parametric-oriented corrections," *Multimedia Tools and Applications*, Vol. 76, 2017, pp. 25179-25196.
2. L. M. Dinca and G. P. Hancke, "The fall of one, the rise of many: a survey on multi-biometric fusion methods," *IEEE Access*, Vol. 5, 2017, pp. 6247-6289.
3. E. Turki, R. Alabboodi, and M. Mahmood, "A proposed hybrid biometric technique for patterns distinguishing," *Journal of Information Science and Engineering*, Vol. 36, 2020, pp. 337-345.
4. W. Wu, S. J. Elliott, S. Lin, S. Sun, and Y. Tang, "Review of palm vein recognition," *IET Biometrics*, Vol. 9, 2020, pp. 1-10.
5. K. F. H. Holle, J. Y. Sari, and Y. P. Pasrun, "Local line binary pattern and fuzzy K-NN for palm vein recognition," *Journal of Theoretical and Applied Information Technology*, Vol. 95, 2017, pp. 2906-2912.
6. J.-C. Lee, "A novel biometric system based on palm vein image," *Pattern Recognition Letters*, Vol. 33, 2012, pp. 1520-1528.

7. W.-Y. Han and J.-C. Lee, "Palm vein recognition using adaptive Gabor filter," *Expert Systems with Applications*, Vol. 39, 2012, pp. 133225-13234.

8. D. Zhang, Z.-H. Guo, and G.-M. Lu, "Online joint palmprint and palmvein verification," *Expert Systems with Applications*, Vol. 38, 2011, pp. 2621-2631.

9. C. Yang and J. Chen, and L. Wei. "Palm vein feature extraction based on median-length included angle chain," *Journal of Computer Applications*, Vol. 29, 2009, pp. 3048-3050.

10. Y.-B. Zhou and A. Kumar, "Human identification using palm-vein images," *IEEE Transactions on Information Forensics and Security*, Vol. 6, 2011, pp. 1259-1274.

11. W. Li and W.-Q. Yuan, "Comparison of four local invariant characteristics based on palm vein," in *Proceedings of IEEE International Conference on Computational Science and Engineering*, 2017, pp. 850-853.

12. H. Qin, L. Qin, L. Xue, X. He, C. Yu, and X. Liang, "Finger-vein verification based on multi-features fusion," *Sensors*, Vol. 13, 2013, pp. 15048-15067.

13. X. Yan, W. Kang, F. Deng, and Q. Wu, "Palm vein recognition based on multi-sampling and feature-level fusion," *Neurocomputing*, Vol. 151, 2015, pp. 798-807.

14. X. Meng, X. Xi, Z. Li, and Q. Zhang, "Finger vein recognition based on fusion of deformation information," *IEEE Access*, Vol. 8, 2020, pp. 50519-50530.

15. Y. Matsuda, N. Miura, A. Nagasaka, H. Kiyomizu, and T. Miyatake, "Finger-vein authentication based on deformation-tolerant feature-point matching," *Machine Vision and Applications*, Vol. 27, 2016, pp. 237-250.

16. J.-D. Wu and C.-T. Liu, "Finger-vein pattern identification using SVM and neural network technique," *Expert Systems with Applications*, Vol. 38, 2011, pp. 14284-14289.

17. C.-H. Hsia and C.-F. Lai, "Embedded vein recognition system with wavelet domain," *Sensors and Materials*, Vol. 32, 2020, pp. 3221-3234.

18. R. Das, E. Piciucco, E. Maiorana, and P. Campisi, "Convolutional neural network for finger-vein-based biometric identification," *IEEE Transactions on Information Forensics and Security*, Vol. 14, 2019, pp. 360-373.

19. M. Wulandari, Basari, and D. Gunawan, "On the performance of pretrained CNN aimed at palm vein recognition application," in *Proceedings of IEEE International Conference on Information Technology and Electrical Engineering*, 2019, pp. 1-6.

20. O. Toygar, F. O. Babalola, and Y. Bitrim, "FYO: A novel multimodal vein database with palmar, dorsal and wrist biometrics," *IEEE Access*, Vol. 8, 2020, pp. 82461-82470.

21. L. Zonglei and X. Xianhong, "Deep compression: a compression technology for apron surveillance video," *IEEE Access*, Vol. 7, 2019, pp. 129966-129974.

22. T. Do, T. Hoang, D. L. Tan, A. Doan, and N. Cheung, "Compact hash code learning with binary deep neural network," *IEEE Transactions on Multimedia*, Vol. 22, 2020, pp. 992-1004.

23. K. Nan, S. Liu, J. Du, and H. Liu, "Deep model compression for mobile platforms: a survey," *Tsinghua Science and Technology*, Vol. 24, 2019, pp. 677-693.

24. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, Vol. 86, 1998, pp. 2278-2323.

25. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826.

26. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

27. G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv Preprint*, 2015, arXiv:1503.02531.

28. F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1800-1807.

29. J.-H. Cho and B. Hariharan, "On the efficacy of dnowledge distillation," in *Proceedings of IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4793-4801.

30. T. A. B. Wirayuda, "Palm vein recognition based-on minutiae feature and feature matching," in *Proceedings of International Conference on Electrical Engineering and Informatics*, 2015, pp. 350-355.

31. L. Mirmohamadsadeghi and A. Drygajlo, "Palm vein recognition with local texture patterns," *IET Biometrics*, Vol. 3, 2014, pp. 198-206.

32. N. Passalis and A. Tefas, "Learning deep representations with probabilistic knowledge transfer," in *Proceedings of European Conference on Computer Vision*, 2018, pp. 268-284.

Zih-Ching Chen was born in New Taipei City, Taiwan, in 1999. She is currently working toward the B.S. degree in Economics at National Taiwan University, Taipei City, Taiwan. Her current research interests include computer vision, data science, and machine learning.

Sin-Ye Jhong received the B.S. degree in Electrical Engineering from Chinese Culture University in 2017, and the M.E. degree in Graduate Institute of Automation Technology from National Taipei University of Technology in 2019, both in Taipei, Taiwan. He is currently working toward the Ph.D. student at the Department of Engineering Science from National Cheng Kung University, Tainan, Taiwan. Mr. Jhong is also the Vice Chair of Young Professionals with the IET Taipei Local Network. His research interests include digital image and video processing, computer vision, and deep learning.

Chih-Hsien Hsia was born in Taipei city, Taiwan, in 1979. He received the Ph.D. degree in Electrical Engineering from Tamkang University, New Taipei, Taiwan, in 2010. In 2007, he was a Visiting Scholar with Iowa State University, Ames, IA, USA. From 2010 to 2013, he was a Postdoctoral Research Fellow with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. From 2013 to 2015, he was an Assistant Professor with the Department of Electrical Engineering at Chinese Culture University, Taiwan. He was an Associate Professor with the Department of Electrical Engineering, Chinese Culture University, Taiwan, from 2015 to 2017. He currently is a Professor with the Department of Computer Science and Information Engineering, National Ilan University, Taiwan. His research interests include DSP IC design, multimedia signal processing, and ccognitive learning.

Dr. Hsia is the Chapter Chair of IEEE Young Professionals Group, Taipei Section, and the Director of the IET Taipei Local Network. He has serves on Associate Editor of the Journal of Imaging Science and Technology and the Journal of Computers.