

Improved MobileNetV2 For Finger Vein Recognition

Hongcheng Zhou^{1*} and Aikang Liu²

¹ School of Electronic Information Engineering, Jinling Institute of Technology, Nanjing 211169, China

² School of Software Engineering, Jinling Institute of Technology, Nanjing 211169, China

* Corresponding author. E-mail: zhouhcnj@163.com

Received: Aug. 04, 2026; Accepted: Sep. 09, 2026

Finger vein recognition is a promising biometric technology for identity authentication, but practical systems are often constrained by limited computing and storage resources. To improve the recognition performance of lightweight models under these constraints, this paper proposes PSE-MobileNetV2, a MobileNetV2-based finger vein recognition network with partially embedded Squeeze-and-Excitation (SE) channel-attention blocks. Instead of adding SE blocks to every inverted residual stage, the proposed model inserts them only into stages 4-7, where the feature maps encode higher-level vein representations. This design strengthens channel-wise feature recalibration while avoiding unnecessary early-stage computational overhead. In the reported closed-set, image-disjoint classification experiments on the FV-USM and SDUMLA public datasets, PSE-MobileNetV2 achieves accuracies of 99.49% and 97.06%, respectively, exceeding the MobileNetV2 baseline by 0.31 and 0.84 percentage points. These results suggest that selective SE placement can improve the accuracy-efficiency trade-off of MobileNetV2. However, the current study does not establish verification performance or statistical significance across repeated runs.

Keywords: Deep Learning; Convolutional Neural Network; Finger Vein Recognition; Channel Attention; MobileNetV2; Lightweight Network; Biometric Recognition

© The Author(s). This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202701_36.028

1. Introduction

In the digital age, identity authentication requires both security and convenience. Traditional methods, such as passwords, keys, and access cards, are vulnerable to loss, theft, or replication and are therefore insufficient for high-security applications. Biometric technologies, including fingerprint recognition [1], face recognition [2, 3], offline signature recognition [4], and iris recognition [5], have been widely used for personal identification. However, externally visible biometric traits can sometimes be captured or forged with advanced artificial-intelligence techniques, which may introduce security risks [6]. Finger vein recognition [7] has therefore attracted increasing attention because vein patterns are internal biological traits that provide liveness, resistance to forgery, and individual distinctiveness.

Deep learning has increasingly been applied to finger vein recognition. Yang et al. [8] used generative adversarial learning for finger vein representation, demonstrating the potential of deep generative models while also highlighting the complexity of verification-oriented pipelines. Song et al. [9] combined composite finger vein images with DenseNet-161 and shift matching to reduce image misalignment, but the large network limits deployment on resource-constrained devices. Shaheed et al. [10] adopted an Xception architecture based on depthwise separable convolution to reduce computational complexity. More recently, Li et al. [11] proposed FV-MViT, a mobile vision transformer that combines local and global feature modelling, and Yao et al. [12] proposed a visual feature-guided lightweight convolutional network. These advances show continuing interest in compact finger vein models, but they

also leave an engineering question: can a standard mobile CNN be improved through selective attention without losing its deployment advantages?

Motivated by this accuracy-efficiency gap, this work investigates a selective SE strategy for MobileNetV2 [13]. The hypothesis is that early stages should preserve inexpensive local edge and texture extraction, whereas deeper stages may benefit more from channel recalibration because their features are semantically richer. The contributions are as follows: (1) a partial SE-enhanced MobileNetV2 variant is implemented for resource-constrained closed-set finger-identity classification; (2) SE blocks are inserted only into stages 4-7, and no-SE and full-SE variants are compared to examine the engineering trade-off; and (3) the method is evaluated on two public datasets using classification metrics and hardware-independent complexity measures. This work does not claim a new attention mechanism. Its contribution is an empirical study of selective SE placement and a transparent discussion of the limitations of the current protocol.

2. Materials and methods

2.1. MobileNetV2 backbone

MobileNetV2 [13] is a lightweight convolutional neural network introduced in 2018. Compared with MobileNetV1 [14], MobileNetV2 uses inverted residual blocks and linear bottlenecks. In each inverted residual block, a 1×1 pointwise convolution first expands the channel dimension, a 3×3 depthwise convolution then extracts spatial features, and a linear 1×1 pointwise convolution projects the features back to a lower-dimensional space. A residual shortcut is used when the stride is 1 and the input and output channel dimensions are identical.

The linear bottleneck is a key component of MobileNetV2. The final 1×1 projection does not use a ReLU activation because a nonlinear activation in a low-dimensional feature space can destroy information. This design is appropriate for preserving subtle vein textures and structural cues.

Finger vein recognition systems are often deployed in embedded access-control or attendance devices with limited computing power and storage. Compared with heavy-weight backbones such as ResNet, MobileNetV2 substantially reduces the parameter count and computational cost through depthwise separable convolution and a streamlined architecture. Finger vein images contain discriminative branch patterns, intersections, and fine texture lines. MobileNetV2 expands features into a higher-dimensional space for nonlinear processing and then uses a linear bottleneck to retain information when projecting them back to a

compact representation. Its multi-stage hierarchy therefore provides an appropriate lightweight basis for extracting both local texture and higher-level vein structure.

2.2. Proposed PSE-MobileNetV2

2.2.1. SE attention mechanism

The Squeeze-and-Excitation (SE) block proposed by Hu et al. [15] is a channel-attention module rather than a self-attention mechanism. It adaptively recalibrates channel-wise responses by modelling dependencies among channels in a convolutional feature map. Fig. 1 illustrates its structure.

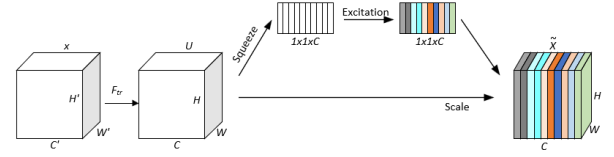


Fig. 1. SE Channel-Attention Module

The SE block comprises squeeze, excitation, and scale operations. For an input feature map U in $R^{H \times W \times C}$, global average pooling compresses the spatial dimensions and produces a C -dimensional channel descriptor z . The excitation operation applies two fully connected layers with a reduction ratio $r = 16$: the first reduces the dimension from C to C/r , and the second restores it to C . ReLU and sigmoid activations generate channel weights. Finally, the scale operation multiplies each channel of U by its corresponding weight to produce the recalibrated feature map.

The SE block thus reweights channel responses adaptively. In finger vein images, channels encoding branches, intersections, or texture contrast may receive greater weights, whereas less informative background responses may be suppressed. Because the current study does not include attention maps, Grad-CAM visualisations, or feature-map evidence, this interpretation is presented as a mechanism-based hypothesis rather than a demonstrated visual finding.

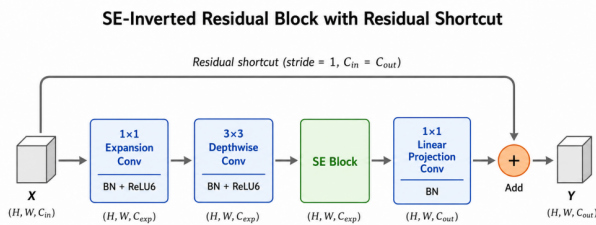
2.2.2. Partial SE-embedded inverted residual blocks

This study embeds SE channel attention into selected MobileNetV2 inverted residual blocks. In each selected block, the SE module is placed after the depthwise convolution and before the final linear 1×1 projection. This position recalibrates expanded spatial features before they are compressed by the bottleneck. When the stride is 1 and the input and output channel dimensions match, the residual shortcut is retained; otherwise, no shortcut is used. Fig. 2 shows the revised block.

Table 1. Architecture of PSE-MobileNetV2

Stage	Input Size	Block Type	Expansion Factor	Output Channels	Number of Blocks	Stride
0	224×224	Conv2d	-	32	1	2
1	112×112	Inverted Residuals	1	16	1	1
2	112×112	Inverted Residuals	6	24	2	2
3	56×56	Inverted Residuals	6	32	3	2
4	28×28	SE-Inverted Residuals	6	64	4	2
5	14×14	SE-Inverted Residuals	6	96	3	1
6	14×14	SE-Inverted Residuals	6	160	3	2
7	7×7	SE-Inverted Residuals	6	320	1	1
8	7×7	Conv2d	-	1280	1	1
9	1×1	Classifier	-	N	-	-

Note: N denotes the number of finger-identity classes (SDUMLA: N = 636; FV-USM: N = 492).

**Fig. 2.** SE-Inverted Residual Block

PSE-MobileNetV2 integrates SE-inverted residual blocks into stages 4-7 of MobileNetV2 (Table 1). Stages 1-3 retain standard inverted residual blocks because their higher-resolution maps mainly encode local edges and fine textures; applying channel recalibration at every shallow block would add overhead before stable high-level vein representations have formed. Stages 4-7 operate on lower-resolution, semantically richer maps and are therefore selected for channel recalibration. This placement is a motivated engineering choice rather than an exhaustive optimum: the current ablation compares no-SE, full-SE, and stages 4-7, but does not test every alternative range or reduction ratio.

2.3. Evaluation metrics

The proposed PSE-MobileNetV2 is evaluated on the FV-USM [16] and SDUMLA [17] public finger vein datasets as a closed-set, multiclass finger-identity classifier. Performance is reported using accuracy, macro-averaged precision, macro-averaged recall, and an aggregate F1 value calculated as the harmonic mean of the displayed macro-precision and macro-recall. This aggregate harmonic F1 is not a micro-F1 score and is not the mean of the per-class F1 values.

For each class, TP, TN, FP, and FN are defined in a one-

versus-rest manner. Class-level precision and recall are calculated first and then averaged equally across all finger-identity classes. Accuracy is calculated over all test images. The definitions are given below.

Accuracy is the proportion of correctly classified test images among all test images. It is calculated using Eq. (1):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

For a selected class, TP is the number of images from that class predicted correctly; TN is the number of images from all other classes rejected correctly; FP is the number of images from other classes assigned incorrectly to the selected class; and FN is the number of images from the selected class assigned to another class. In multiclass evaluation, these quantities are computed for each class using a one-versus-rest formulation.

Precision is the proportion of predictions assigned to a class that are correct. It is calculated using Eq. (2):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Higher precision indicates fewer false-positive assignments to a class. Precision is particularly informative when false acceptances or incorrect identity assignments are costly.

Recall is the proportion of images from a class that are identified correctly. It is calculated using Eq. (3):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Higher recall indicates fewer false negatives and therefore fewer missed samples from the target class.

The F1 score is the harmonic mean of precision and recall. It is calculated using Eq. (4):

$$\text{F1 Score} = \frac{2PR}{P + R} \quad (4)$$

A higher F1 score indicates a better balance between precision and recall. In Tables 4 and 5, the displayed F1 values are recomputed from the displayed macro-precision and macro-recall values using Eq. (4) so that the table entries are internally consistent.

Together, accuracy, macro-precision, macro-recall, and the aggregate harmonic F1 describe complementary aspects of closed-set classification. They should not be interpreted as biometric verification measures such as equal error rate (EER), false acceptance rate (FAR), or false rejection rate (FRR).

2.4. Datasets

Model evaluation used the public FV-USM and SDUMLA finger vein datasets, which differ in acquisition devices, participant populations, and image quality. These two datasets provide a useful but limited academic benchmark; broader generalisation requires cross-device, cross-session, and open-set verification experiments on additional datasets.

The FV-USM dataset contains 5,904 images collected from 123 volunteers aged between 20 and 35 years. Images were collected from the index and middle fingers of both hands, and each finger was captured 12 times across two sessions. Representative samples from this dataset are shown in Fig. 3.

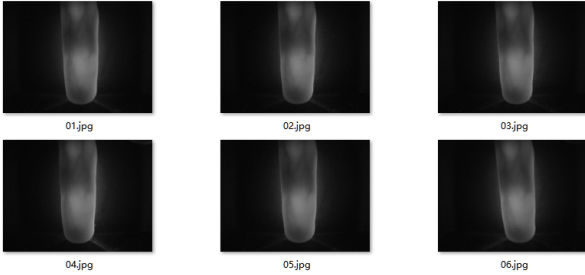


Fig. 3. Representative Images from the FV-USM Dataset

The SDUMLA finger vein database contains 3,816 images from 106 volunteers aged between 18 and 40 years. For each participant, images were captured from the index, middle, and ring fingers of both hands. Each finger was sampled six times. Representative samples from this dataset are shown in Fig. 4.

2.5. Experimental design and data protocol

The experiments use closed-set finger-identity classification, not a subject-independent protocol. Each finger is treated as one class (492 classes for FV-USM and 636 classes for SDUMLA), so every class must be represented in the training, validation, and test subsets. Images were assigned at the image level within each finger identity using the stated training/validation/test ratios of 8:1:1 for FV-USM

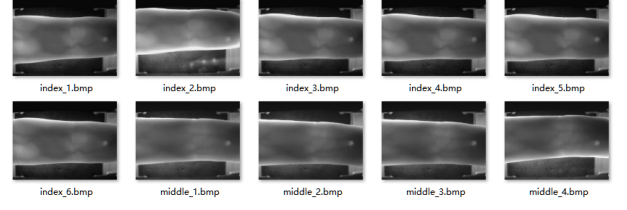


Fig. 4. Representative Images from the SDUMLA Dataset

and 4:1:1 for SDUMLA. Each source image belongs to exactly one subset; duplicate file paths were excluded, and dataset splitting was completed before model training to prevent the same image from appearing in multiple subsets. All models used the same split. Images were resized to 224×224 pixels, consistent with the network input in Table 1. The supplied manuscript and accompanying materials do not contain the original training code or sufficient records to verify the precise ROI-extraction, intensity-normalisation, or stochastic-augmentation settings. These settings are therefore not reconstructed retrospectively, and the missing implementation metadata is acknowledged as a reproducibility limitation. All models were trained for 100 epochs with a batch size of 64 using the Adam optimiser, an initial learning rate of 0.001, weight decay of 0.0001, cross-entropy loss, and a ReduceLROnPlateau scheduler that reduced the learning rate by a factor of 0.5 after 10 epochs without improvement in validation accuracy.

3. Results and discussion

PSE-MobileNetV2 is compared with ResNet50, ResNet50SE, ShuffleNetV2, ShuffleNetV3, MobileNetV2, MobileNetV2SE, and MobileNetV3-Small on FV-USM and SDUMLA. ResNet50SE denotes ResNet50 with SE channel attention, whereas MobileNetV2SE denotes a full-SE MobileNetV2 in which SE modules are inserted into all inverted residual stages. All locally evaluated models used the same closed-set image split, optimiser, batch size, and training duration. Fig. 5 shows the training curves of the validation sets of the two datasets on different models.

Table 2 shows that PSE-MobileNetV2 is competitive on both datasets, but it is not presented as a universal state-of-the-art model. On FV-USM, ResNet50SE obtains the highest local accuracy of 99.59%, while PSE-MobileNetV2 reaches 99.49%, 0.31 percentage points above MobileNetV2. On SDUMLA, PSE-MobileNetV2 obtains the highest local accuracy of 97.06%, 0.84 percentage points above MobileNetV2. Because these gains are modest and repeated-run statistics are unavailable, the main claim is an accuracy-complexity trade-off rather than a statistically established superiority claim.

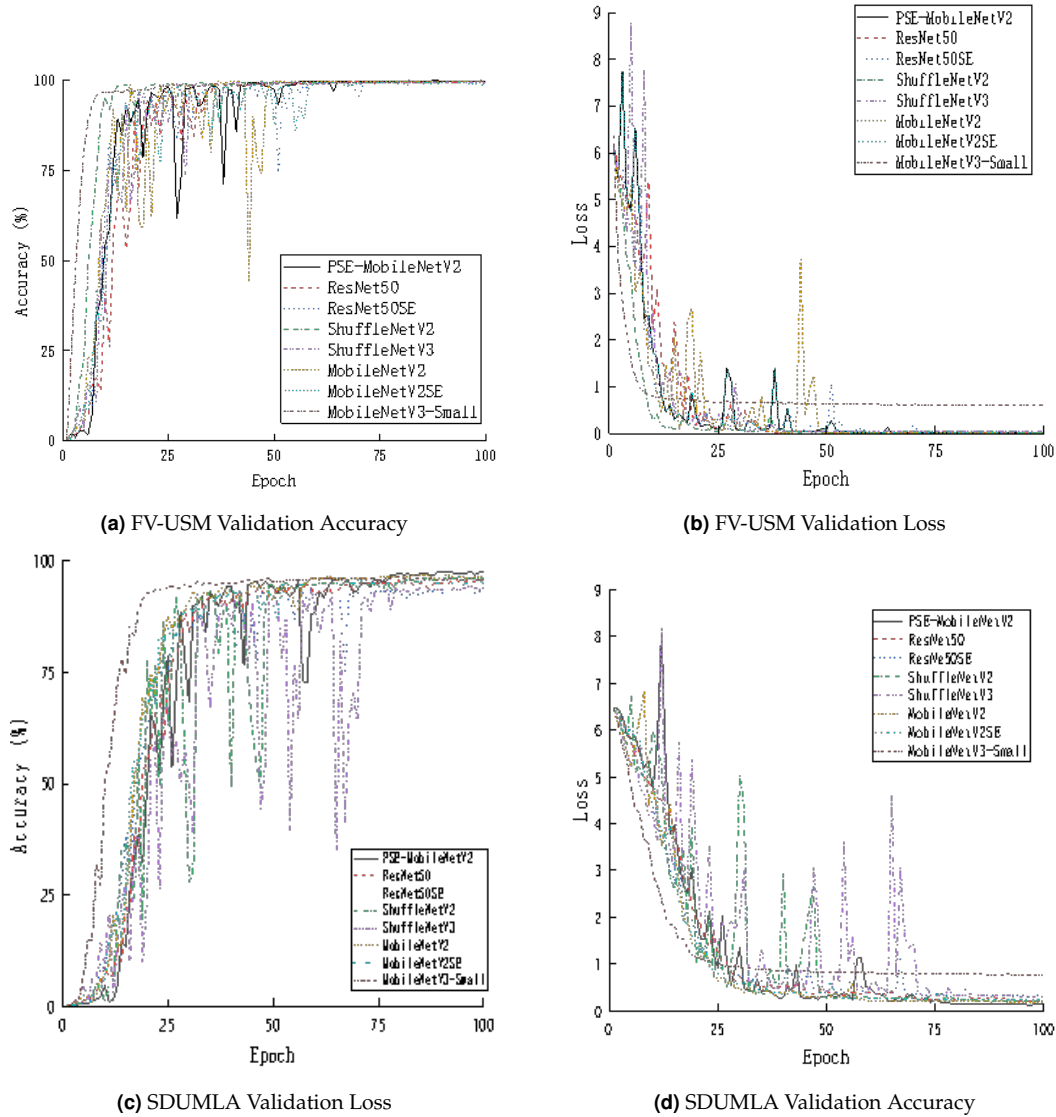


Fig. 5. Validation Accuracy and Loss Curves on FV-USM and SDUMLA.

In Table 2, rows correspond to models re-implemented in this study and columns correspond to the two benchmark datasets. Table 3 provides a qualitative comparison with representative finger-vein-specific studies. Those studies use different preprocessing, partitions, architectures, and in some cases verification metrics rather than closed-set classification accuracy. Table 3 is therefore contextual and must not be interpreted as a same-protocol ranking.

3.1. Recognition performance metrics

Tables 4 and 5 report macro-precision, macro-recall, and an aggregate harmonic F1 for FV-USM and SDUMLA. Every displayed F1 value has been recalculated from the displayed precision and recall values using $F1 = 2PR/(P + R)$. This produces internally consistent table values, but

it should not be confused with micro-F1 or the arithmetic mean of per-class F1 values.

Table 2. Test Accuracy of the Evaluated Models on FV-USM and SDUMLA

Model	FV-USM Accuracy	SDUMLA Accuracy
ResNet50	99.28%	96.85%
ResNet50SE	99.59%	96.70%
ShuffleNetV2	99.39%	95.44%
ShuffleNetV3	98.78%	94.96%
MobileNetV2	99.18%	96.22%
MobileNetV3-Small	99.39%	96.38%
PSE-MobileNetV2	99.49%	97.06%

Table 3. Qualitative Contextual Comparison with Published Finger-Vein-Specific Methods

Method	Reference	Study focus	Qualitative finding	Relevance and limitation
FV-GAN	[8]	GAN-based finger-vein representation for verification	Uses a fully convolutional generative framework to learn vein representations	Verification-oriented protocol; not directly comparable with closed-set classification
Deep DenseNet	[9]	Composite-image DenseNet for finger-vein recognition	Uses composite images and DenseNet to improve robustness to image displacement	Uses a different architecture and protocol; included as finger-vein-specific reference
DS-CNN	[10]	Xception/depth-wise separable CNN for finger-vein recognition	Uses depthwise separable Xception features for compact finger-vein recognition	A stronger task-specific baseline, but protocol details differ from this study
FV-MViT	[11]	Mobile Vision Transformer for finger-vein recognition	Combines lightweight CNN and transformer components for local-global modelling	Transformer-based method with different computation profile
PSE-MobileNetV2	This work	Selective SE placement in MobileNetV2 under this study's closed-set protocol	Tests an accuracy-complexity trade-off rather than a new attention mechanism	Same protocol as Tables 2, 4, 5, and 7; not a cross-paper SOTA claim

Note: The studies in Table 3 use different protocols. The table summarises methodological relevance and does not claim direct numerical superiority.

On FV-USM, PSE-MobileNetV2 achieves a macro-precision of 0.9942, a macro-recall of 0.9949, and a recalculated aggregate harmonic F1 of 0.9945. ResNet50SE has the highest accuracy, whereas PSE-MobileNetV2 provides balanced classification metrics with the smaller MobileNetV2-based backbone.

On SDUMLA, PSE-MobileNetV2 achieves the highest locally evaluated accuracy, macro-precision, macro-recall, and recalculated aggregate harmonic F1, with values of 97.06%, 0.9575, 0.9707, and 0.9641, respectively. The improvement is modest and should be interpreted together with model complexity and the absence of repeated-run uncertainty estimates.

Table 4. Classification Metrics on FV-USM

Model	Precision	Recall	F1 Score (harmonic of displayed macro-P/R)
ResNet50	0.9953	0.9929	0.9941
ResNet50SE	0.9936	0.9929	0.9932
ShuffleNetV2	0.9939	0.9939	0.9939
ShuffleNetV3	0.9898	0.9878	0.9888
MobileNetV2	0.9946	0.9919	0.9932
MobileNetV3-Small	0.9939	0.9939	0.9939
PSE-MobileNetV2	0.9942	0.9949	0.9945

Table 5. Classification Metrics on SDUMLA

Model	Precision	Recall	F1 Score (harmonic of displayed macro-P/R)
ResNet50	0.9528	0.9686	0.9606
ResNet50SE	0.9597	0.9670	0.9633
ShuffleNetV2	0.9334	0.9544	0.9438
ShuffleNetV3	0.9269	0.9497	0.9382
MobileNetV2	0.9464	0.9623	0.9543
MobileNetV3-Small	0.9481	0.9638	0.9559
PSE-MobileNetV2	0.9575	0.9707	0.9641

3.2. Model complexity

Table 6 compares the parameter count, floating-point operation count, and model size of the evaluated networks. ResNet50 and ResNet50SE exceed 24 M parameters and 8 GFLOPs, whereas the MobileNet- and ShuffleNet-based models are substantially smaller. Relative to MobileNetV2, PSE-MobileNetV2 increases the parameter count from 3.04 M to 3.59 M, FLOPs from 0.654 G to 0.657 G, and model size from 11.83 MB to 13.95 MB. The original latency and FPS values have been removed because the supplied materials do not identify the hardware platform, software environment, batch size used for timing, warm-up pro-

cedure, or measurement method. Retaining unverifiable timing values would imply a degree of reproducibility that the available records do not support.

Table 6. Hardware-Independent Model Complexity

Model	Parameters /M	FLOPs /G	Size /MB
ResNet50	24.81	8.266	94.96
ResNet50SE	27.33	8.282	104.59
ShuffleNetV2	1.91	0.305	7.45
ShuffleNetV3	0.72	0.082	2.99
MobileNetV2	3.04	0.654	11.83
MobileNetV2SE	3.60	0.660	14.02
MobileNetV3-Small	2.54	0.066	9.8
PSE-MobileNetV2	3.59	0.657	13.95

3.3. Ablation study

The SE-placement ablation compares three configurations: the original MobileNetV2 without SE modules, MobileNetV2SE with SE modules in all inverted residual stages, and PSE-MobileNetV2 with SE modules in stages 4-7. Table 7 reports the results.

Table 7. Ablation Results for SE Placement Strategies

Embedding Strategy	FV-USM Accuracy (%)	SDUMLA Accuracy (%)
No SE (MobileNetV2)	99.18%	96.22%
Full SE (MobileNetV2SE)	99.39%	96.01%
Partial SE, stages 4-7 (PSE-MobileNetV2)	99.49%	97.06%

Relative to MobileNetV2, the stages 4-7 configuration increases accuracy from 99.18% to 99.49% on FV-USM and from 96.22% to 97.06% on SDUMLA. Full SE placement reaches 99.39% on FV-USM but 96.01% on SDUMLA, indicating that attention in every stage does not consistently improve performance. Stages 4-7 were selected because deeper, lower-resolution feature maps contain more abstract channel representations and the added attention cost is smaller than at high spatial resolutions. Nevertheless, this ablation does not prove that stages 4-7 are globally optimal: stages 3-6, stages 5-7, and alternative reduction ratios were not evaluated. The reduction ratio $r = 16$ follows the standard SE configuration [15].

3.4. Discussion and limitations

The proposed method is a lightweight MobileNetV2 engineering variant rather than a fundamentally new network family. Its practical contribution is evidence that selective SE placement can improve closed-set classification accuracy while retaining most of the hardware-independent efficiency of MobileNetV2 under the reported experimental setting.

Several limitations remain. First, the evaluation is closed-set and image-disjoint, not subject-independent or open-set; consequently, classification accuracy does not establish biometric verification performance measured by EER, FAR, FRR, or ROC curves. Second, only two public academic datasets are used. Third, the current materials do not support Grad-CAM, feature-map visualisation, failure-case analysis, or independent repeated runs, so visual interpretability and statistical significance have not been demonstrated. Fourth, the ablation does not test all insertion ranges or reduction ratios. Fifth, the original code and logs needed to verify detailed ROI extraction, normalisation, augmentation, and hardware timing are unavailable. These limitations restrict reproducibility and prevent broad state-of-the-art claims.

Accordingly, PSE-MobileNetV2 is positioned as a compact accuracy-complexity trade-off for the reported closed-set benchmark, not as a validated replacement for verification-oriented finger vein systems.

4. Conclusions

1. PSE-MobileNetV2 selectively inserts Squeeze-and-Excitation channel-attention blocks into stages 4-7 of MobileNetV2. This strategy recalibrates deeper channel responses while avoiding full-stage attention overhead.
2. Under the reported closed-set, image-disjoint protocol, the model achieves classification accuracies of 99.49% on FV-USM and 97.06% on SDUMLA.
3. With 3.59 M parameters, 0.657 GFLOPs, and a model size of 13.95 MB, PSE-MobileNetV2 provides a favourable accuracy-complexity trade-off relative to MobileNetV2. The current evidence does not establish verification performance, statistical significance, or universal superiority.
4. Future work should evaluate open-set verification using EER, FAR, FRR, and ROC curves; report statistics over multiple random seeds; add Grad-CAM and failure-case analysis; test additional SE insertion

ranges and reduction ratios; release complete pre-processing and timing protocols; and validate cross-device generalisation.

Availability of data and material

Data will be made available upon reasonable request.

Funding statement

The authors declare no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

References

- [1] R. Singh, R. Singh, R. K. Tripathi, and P. Agarwal, (2025) "Fingerprint recognition using artificial neural networks" **Proceedings of the National Academy of Sciences, India Section A: Physical Sciences** 95(2): 127–135. DOI: [10.1007/s40010-025-00917-y](https://doi.org/10.1007/s40010-025-00917-y).
- [2] G. Jeevan, G. C. Zacharias, M. S. Nair, and J. Rajan, (2022) "An empirical study of the impact of masks on face recognition" **Pattern Recognition** 122: 108308. DOI: [10.1016/j.patcog.2021.108308](https://doi.org/10.1016/j.patcog.2021.108308).
- [3] K. V. Shakrani, N. M. Kanyangarara, P. T. Parowa, V. Gupta, and R. Kumar, (2022) "A deep learning model for face recognition in presence of mask" **Acta inform Malays** 6(2): 43–6. DOI: [10.26480/aim.02.2022.38.41](https://doi.org/10.26480/aim.02.2022.38.41).
- [4] N. Calik, O. C. Kurban, A. R. Yilmaz, T. Yildirim, and L. D. Ata, (2019) "Large-scale offline signature recognition via deep neural networks and feature embedding" **Neurocomputing** 359: 1–14. DOI: [10.1016/j.neucom.2019.03.027](https://doi.org/10.1016/j.neucom.2019.03.027).
- [5] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun, (2020) "Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition" **IEEE Transactions on information forensics and security** 15: 2944–2959. DOI: [10.1109/TIFS.2020.2980791](https://doi.org/10.1109/TIFS.2020.2980791).
- [6] B. Hou, H. Zhang, and R. Yan, (2022) "Finger-vein biometric recognition: A review" **IEEE Transactions on Instrumentation and Measurement** 71: 1–26. DOI: [10.1109/TIM.2022.3200087](https://doi.org/10.1109/TIM.2022.3200087).
- [7] J. Yang, Y. Shi, and G. Jia, (2017) "Finger-vein image matching based on adaptive curve transformation" **Pattern Recognition** 66: 34–43. DOI: [10.1016/j.patcog.2017.01.008](https://doi.org/10.1016/j.patcog.2017.01.008).
- [8] W. Yang, C. Hui, Z. Chen, J.-H. Xue, and Q. Liao, (2019) "FV-GAN: Finger vein representation using generative adversarial networks" **IEEE Transactions on Information Forensics and Security** 14(9): 2512–2524. DOI: [10.1109/TIFS.2019.2902819](https://doi.org/10.1109/TIFS.2019.2902819).
- [9] J. M. Song, W. Kim, and K. R. Park, (2019) "Finger-vein recognition based on deep DenseNet using composite image" **Ieee Access** 7: 66845–66863. DOI: [10.1109/ACCESS.2019.2918503](https://doi.org/10.1109/ACCESS.2019.2918503).
- [10] K. Shaheed, A. Mao, I. Qureshi, M. Kumar, S. Hussain, I. Ullah, and X. Zhang, (2022) "DS-CNN: A pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition" **Expert Systems with Applications** 191: 116288. DOI: [10.1016/j.eswa.2021.116288](https://doi.org/10.1016/j.eswa.2021.116288).
- [11] X. Li, J. Feng, J. Cai, and G. Lin, (2024) "FV-MViT: Mobile vision transformer for finger vein recognition" **Sensors** 24(4): 1331. DOI: [10.3390/s24041331](https://doi.org/10.3390/s24041331).
- [12] Q. Yao, D. Song, X. Xu, and K. Zou, (2024) "Visual feature-guided diamond convolutional network for finger vein recognition" **Sensors** 24(18): 6097. DOI: [10.3390/s24186097](https://doi.org/10.3390/s24186097).
- [13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks". In: *2018 IEEE/CVF conference on computer vision and pattern recognition*. Ieee. 2018, 4510–4520. DOI: [10.1109/CVPR.2018.00474](https://doi.org/10.1109/CVPR.2018.00474).
- [14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, (2017) "Mobilenets: Efficient convolutional neural networks for mobile vision applications" **arXiv preprint arXiv:1704.04861**: DOI: [10.48550/arXiv.1704.04861](https://doi.org/10.48550/arXiv.1704.04861).
- [15] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, (2020) "Squeeze-and-Excitation Networks" **IEEE transactions on pattern analysis and machine intelligence**: DOI: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372).
- [16] M. S. M. Asaari, S. A. Suandi, and B. A. Rosdi, (2014) "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics" **Expert Systems with Applications** 41(7): 3367–3382. DOI: [10.1016/j.eswa.2013.11.033](https://doi.org/10.1016/j.eswa.2013.11.033).
- [17] Y. Yin, L. Liu, and X. Sun. "SDUMLA-HMT: A multimodal biometric database". In: *Chinese conference on biometric recognition*. Springer. 2011, 260–268. DOI: [10.1007/978-3-642-25449-9_33](https://doi.org/10.1007/978-3-642-25449-9_33).