applied sciences

MDPI

Article

Finger Vein and Inner Knuckle Print Recognition Based on Multilevel Feature Fusion Network

Li Jiang [1], Xianghuan Liu [2], Haixia Wang [1,*] and Dongdong Zhao [1]

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
[2] College of Information Engineering and Technology, Zhejiang University of Technology, Hangzhou 310023, China
* Correspondence: hxwang@zjut.edu.cn

Abstract: Multimodal biometric recognition involves two critical issues: feature representation and multimodal fusion. Traditional feature representation requires complex image preprocessing and different feature-extraction methods for different modalities. Moreover, the multimodal fusion methods used in previous work simply splice the features of different modalities, resulting in an unsatisfactory feature representation. To address these two problems, we propose a Dual-Branch-Net based recognition method with finger vein (FV) and inner knuckle print (IKP). The method combines convolutional neural network (CNN), transfer learning, and triplet loss function to complete feature representation, thereby simplifying and unifying the feature-extraction process of the two modalities. Dual-Branch-Net also achieves deep multilevel fusion of the two modalities' features. We assess our method on a public FV and IKP homologous multimodal dataset named PolyU-DB. Experimental results show that the proposed method performs best and achieves an equal error rate (EER) of the recognition result of 0.422%.

Keywords: finger vein features; inner knuckle print features; multimodal recognition; convolutional neural network; feature fusion

Citation: Jiang, L.; Liu, X.; Wang, H.; Zhao, D. Finger Vein and Inner Knuckle Print Recognition Based on Multilevel Feature Fusion Network. *Appl. Sci.* **2022**, *12*, 11182. https://doi.org/10.3390/app122111182

Academic Editors: Elias N. Zois and Dimitrios Kalivas

Received: 3 October 2022
Accepted: 1 November 2022
Published: 4 November 2022

1. Introduction

Biometric recognition aims to distinguish individuals based on human physiological and behavioral characteristics and is widely applied in Internet-of-Things security [1], financial payment [2], and security systems [3], etc. A biometric system usually includes image acquisition [4], preprocessing [5], feature extraction [6], and matching [7], as shown in Figure 1. Due to the convenience of acquisition, finger-based biometric recognition methods have developed rapidly. With the increasing demand for accurate identity recognition in recent years, multimodal recognition methods have attracted wide attention [8–10]. Finger vein (FV) and inner knuckle print (IKP) are two modalities of the finger. FV is the tiny blood vessels inside the finger close to the cortex [11], and IKP is the pattern of the inner surface of the knuckle [12]. We use these two modalities to achieve identity recognition in this study.

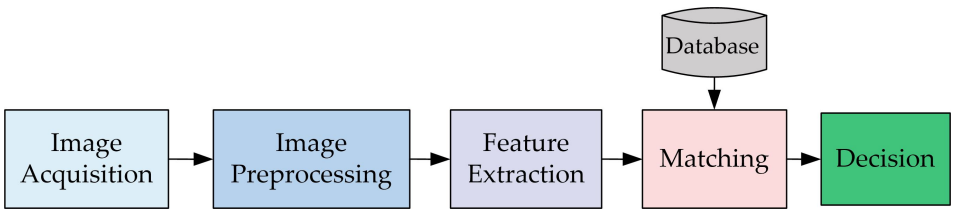


Figure 1. The architecture of the biometric recognition system.

There are two main problems to be solved in designing an method based on multimodal biometric recognition: how to extract each modality's features and how to fuse multiple modalities' information.

The traditional feature extraction of IKP and FV mainly relies on hand-designed schemes [13–17]. Although traditional feature-extraction methods can achieve good results from specific designs, they are complicated. To solve this problem, we use deep learning models for feature extraction. CNN is suitable for image feature extraction, as the model has a strong generalization ability after pretraining large-scale data. More importantly, the input images of the model do not require complicated preprocessing. We use the Inception-ResNet-v1 [18] model for feature extraction in this study. The experimental results show that the model can extract IKP and FV image features well. In addition, we introduce the triplet loss function [19] as the loss function to optimize the model. The function forces the feature representation distance of different sample images of the same person to be close and that of sample images of different people to be far away. It improves the expression ability of the fusion feature.

Feature-level fusion is the mainstream method for multimodal fusion when using deep neural networks to extract image features [20–23]. The feature dimension is large, and the operability space of fusion is strong. Some previous works only performed a simple concatenation operation on multimodal features [24,25]. Such fusion method does not consider the features of different stages in the feature-extraction process, which will affect the final recognition performance. Multilevel feature fusion can better represent features. Therefore, we propose to fuse IKP and FV features at different scales and perform multilevel feature fusion. Given the difference between the two modalities, the method takes FV as the main task modality and IKP as the auxiliary task modality. In summary, our contributions are as follows:

- We propose a Dual-Branch-Net network to extract the features of FV and IKP. The features of the FV branch at multiple levels of the network are fused with the corresponding features of the IKP branch, achieving deep fusion between the features of the FV and the IKP.
- We use transfer learning and triplet loss function to optimize the multimodal feature representation and perform identity recognition by measuring the similarity of feature vectors.
- We perform a series of experiments to verify that the recognition performance of our method is superior to both unimodal and other multimodal recognition methods.

2. Related Works

The research related to our work can be mainly divided into two categories: one is feature extraction of biometrics; the other is feature fusion of multiple modalities.

2.1. Feature Extraction

Existing feature-extraction methods are classified into two groups: non-training-based [13,26–28] and training-based [22,24,29,30] methods. Most traditional feature-extraction methods do not require training. For example, Kumar and Zhou [13] used the Gabor filter to extract the feature maps of FV images on the preprocessed images and then used the morphological operation to enhance the definition of the feature maps. They also used the local Radon transformation to extract the features of IKP images. Evangelin and Fred [26] used the gray-level co-occurrence matrix (GLCM) to extract the second-order texture features of finger knuckle print (FKP) images. Li et al. [27] proposed a new feature expression method based on local encoding using the generalized symmetric local graph structure (GSLG) to express the position and orientation relationship of domain pixels. Veluchamy and Karlmarx [28] extracted the features of FKP and FV using a repeated line tracking method. Although these traditional feature-extraction methods do not require training of feature extractors, the quality of images affects their extraction methods. Thus, it is necessary to perform complicated preprocessing processes such as region of interest

extraction, orientation correction, and image enhancement. In addition, the characteristics of each modality are different, requiring special feature-extraction methods.

Although feature extraction through CNN requires pretraining of the network model, it has become a popular feature-extraction method in biometrics because of its less complex image preprocessing procedure and effective feature extraction. For example, Daas et al. [22] extracted the features of FV and FKP through CNN. Wang et al. [29] first used a five-layer CNN to extract each modal feature and then used principal component analysis (PCA) to standardize the features. Ren et al. [30] proposed a multimodal finger recognition network—namely, FPV-Net—to extract the features of FVs and fingerprints. Kim et al. [24] fine-tuned VGGNet [31] and ResNet [32] and used these pretrained models for the feature extraction of FV and finger shape. Regarding these works, this study utilizes a large-scale pretrained model and triplet loss function to obtain feature representations of multimodal images.

2.2. Multimodal Fusion Method

Multimodal fusion usually includes three levels: pixel level [33], score level [13,20,34,35], and feature level [22,23,30,36]. For example, Kumar and Zhou [13] proposed a score-level fusion function called Holistic Fusion, which considers that the matching score of one modality may be more stable than another. The function results in a similar trend between the total matching score and the matching score of the more stable modality. Noh et al. [20] generated composite images of FV texture and composite images of FV shape to be used as input for the DenseNet [37]. Then, matching was performed using the obtained output matching score. Finally, they performed score fusion under different rules such as weighted product, weighted sum, perceptron, and Bayesian. Alay and Heyam [34] used five convolution layers and two fully connected layers to extract the features of each modality and then inputted them into their softmax classifier to obtain the similarity score. They used two score methods, namely, the arithmetic mean rule and the product rule. Walia et al. [35] achieved optimal score-level fusion by boosting and suppressing concurrent classifiers and resolving conflict among discordant classifiers.

Feature-level fusion can achieve better recognition results than pixel-level and score-level fusion because feature-level fusion preserves more original information [38]. Daas et al. [22] first used the network to extract the features of the FV and FKP images, obtained two 1D feature vectors, and then concatenated the two feature vectors into a full feature vector. Wang et al. [36] proposed a two-channel CNN feature fusion framework. The face image feature and the FV image feature were taken as the input to the self-attention mechanism to obtain their respective attention weights. The vein and face features were combined with their respective attention weights and concatenated to form splicing features. Then the splicing features were convolved to obtain the fused features. Ren et al. [30] used a weighted average to fuse the features of the two modalities through the attention module. They also further fused the shallow feature and deep feature of the network. Wen et al. [23] used a convolutional network to extract the 3D feature tensors of each modality. Next, feature tensors were concatenated in series according to the dimensions of the channels into a total 3D feature tensor, which was then sent to the subsequent network layers for further processing. Compared with these works, we adopt a feature-level fusion strategy fusing features at a deeper level.

3. Proposed Method

This section first introduces the overall framework of fusion recognition based on FV and IKP; then, it introduces the designed Dual-Branch-Net and the loss function used in detail. Finally, it introduces using transfer learning to initialize the network parameters.

3.1. General Framework

Figure 2 shows the framework of the proposed FV and IKP fusion network. The input of the network is a batch of image pairs (image of the FV and its corresponding IKP). After the batch goes through architecture and L2 regularization, the feature representation

of N image pairs is obtained. The triplet loss function is adapted in this study to calculate the loss and then optimize the network parameters through backpropagation. Finally, the mapping space corresponding to the network satisfies the following characteristics. In this space, the feature representation distance of different image pairs belonging to the same finger is relatively close. By contrast, the feature representation distance of different image pairs belonging to different fingers is far apart.

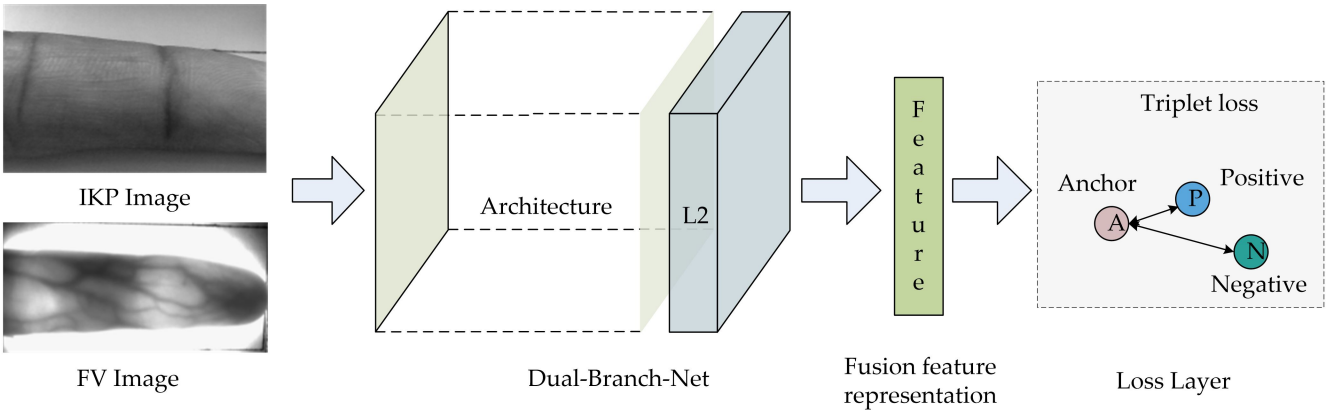


Figure 2. FV and IKP fusion network framework.

We use the feature representation outputted by the network for authentication. Specifically, given image pair A of the subject and image pair B of the target object, if the similarity (expressed by Euclidean distance) between the feature representation of A and that of B is greater than the threshold, the subject and the target object are considered to belong to the same person; otherwise, they belong to different people. In addition, if the similarity between the test subject and multiple target objects is greater than the threshold, the target object category corresponding to the maximum similarity is selected as the final category of the test subject.

3.2. Model

The network architecture in Figure 2 is our proposed fusion model Dual-Branch-Net, and its specific architecture is shown in Figure 3. Dual-Branch-Net uses two branches to extract features from IKP images and FV images, respectively. The network adopts a multilevel deep fusion method to make full use of the two types of modal information. We detail the structure of the branch network and the feature fusion method below.

Dual-Branch-Net uses the Inception-ResNet-v1 [18] network as the base network for the two branches. The Inception-ResNet-v1 network includes three types of basic network modules as follows: (a) Stem—preliminary feature extraction for the original image; (b) Inception-Resnet-A/B/C—for further feature extraction, in which the input features of this module have the same number and size of dimensions as the output features; and (c) Reduction-A/B/C—further feature extraction. With this module, the height and width of the output feature will become smaller than the input feature. Meanwhile, the number of output channels will increase. Considering the recognition performance and computational efficiency, we do not directly use the original Inception-ResNet-v1 network structure but make some adjustments. The details are as follows: the number of Inception-Resnet-A was reduced from 5 to 3, the number of Inception-Resnet-B was reduced from 10 to 6, and the number of Inception-Resnet-C was reduced to 3.

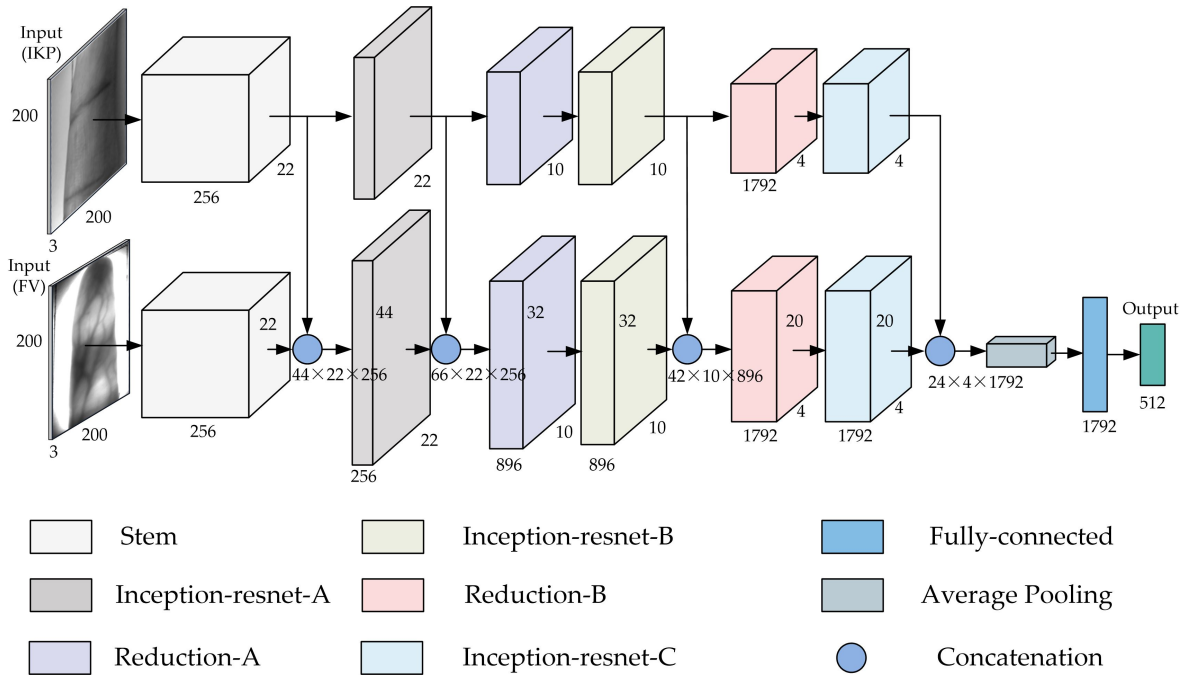


Figure 3. Dual-Branch-Net network structure.

Unlike some works [22,23] mentioned in the introduction, which directly concatenate the features of the two modalities only once, we fuse the features of the two modalities at a deeper level. The network layers at different depths have different receptive fields, and the output features of different network layers represent the features of the original image at different levels. Thus, we perform feature fusion operations at multiple levels of the network. Specifically, we splice the output features of both branch after the Stem modules, the last Inception-Resnet-A module, the last Inception-Resnet-B module, and the last Inception-Resnet-C module, respectively. The spliced features are used as the input of the subsequent network layer of the corresponding branch of the FV image, as shown in Figure 3. We have marked the number of channels, height, and width of the output of each module in the figure. Suppose we use the number of channels (C), height (H), and width (W) to represent the shape of a 3D feature; for two features of shape (C, H, W), we concatenate them into a full feature of shape (C, 2H, W) tensor. Considering that when only single-modality is used for recognition, the recognition effect of the method based on the FV is better than that based on the IKP; so, we fuse the feature of the IKP into the branch corresponding to the FV.

3.3. Loss Function

In the authentication phase, the Euclidean distance between two feature vectors determines whether the test subject matches the target object. Therefore, the selected loss function should make the samples of the same category close in the feature space and the samples of different categories far away in the feature space. The triplet loss function [19] is a function that can make the optimization direction of the network parameters meet this characteristic. Therefore, this paper will introduce the main idea of the triplet loss function.

When using the triplet loss function to optimize the network, some triples need to be constructed for each batch of FV and IKP image pairs (referred to as "samples" hereafter) and corresponding labels from them first. A triplet consists of an anchor sample, a positive sample, and a negative sample. Among them, anchor samples can be selected arbitrarily, samples of the same category as the anchor samples can be regarded as positive samples, and samples of different categories from the anchor samples can be regarded as negative samples. For a batch of image pairs, multiple triples can be constructed.

For each triplet, we first use Dual-Branch-Net to extract features to obtain the corresponding three feature vectors, as shown in Figure 4, and then calculate the loss of the network on the triplet according to formula (1):

$$L(A, P, N) = max(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0) \quad (1)$$

where A is the anchor sample, P is the positive sample, and N is the negative sample. $f()$ represents the Dual-Branch-Net network, and $f(x)$ represents the feature vector corresponding to the sample x. α is a fixed distance threshold. Only if the distance between the negative sample and the anchor sample and the distance between the positive sample and the anchor sample is not less than α will the loss of the network on the corresponding triplet be 0; otherwise, the loss will be greater than 0. Then, the model will then optimize the network parameters to adjust the feature representation of the sample (the arrow roughly shows the adjustment direction in the Euclidean space in Figure 4).

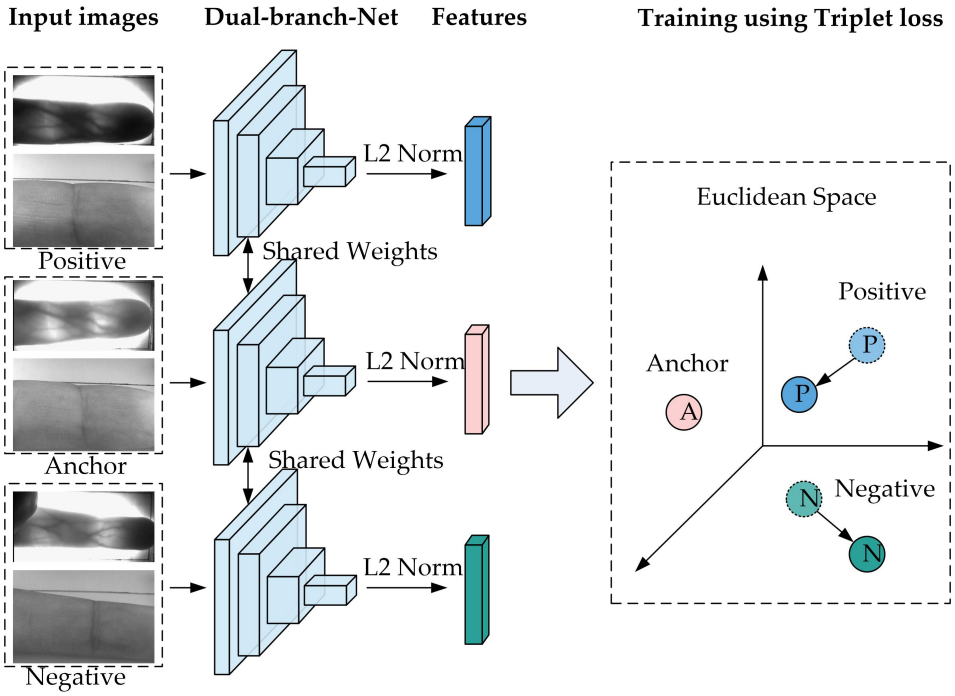


Figure 4. Training a network using triplet loss function.

3.4. Initialize Network Weights

Since the network contains many trainable parameters, the dataset containing FV and IKP images is relatively tiny. If the network parameters are randomly initialized, it is difficult for them to converge to the optimal value. Therefore, we use transfer learning to load the parameter values of the pretrained model. We initialize the Dual-Branch-Net with the weights of the Inception-ResNet-v1 network pretrained on the large-scale dataset CASIA-WebFace [39].

4. Experiments

This section first introduces the experimental environment, dataset, and evaluation metrics. Next, we introduce the content and results of multiple experiments in detail in several subsections: (a) In the comparative experiment in Section 4.2, the proposed method is compared with other multimodal recognition algorithms to verify the effectiveness of the proposed method. (b) In the ablation experiment in Section 4.3, we compare the proposed fusion method with other fusion methods and compare the recognition performance of the multimodal and unimodal methods. Through these experiments, we explore whether the proposed fusion method is appropriate and whether it is necessary to use multimodal

information for recognition. (c) In Section 4.4, by comparing different initialization methods of parameters, we explore whether it is necessary to use the parameter values of the pretrained model to initialize the proposed model.

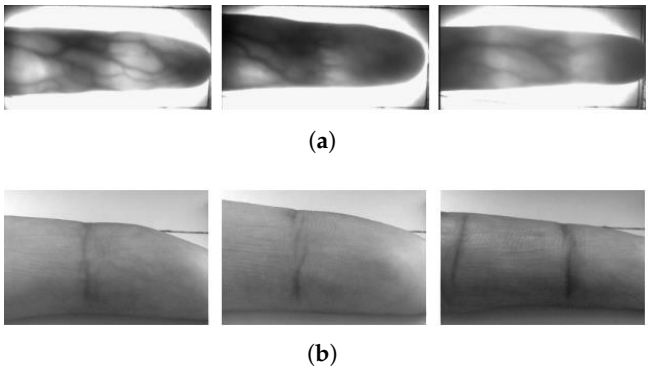
4.1. Experimental Setting

The memory of the experiment environment in this study is 32 GB, the CPU is an Intel Core i9-9940X, the GPU is an Nvidia GeForce RTX 2080 Ti, the programming language is Python, and the deep learning framework is TensorFlow.

During training, network parameters are optimized with stochastic gradient descent (SGD) using an adaptive gradient algorithm (AdaGrad). We set the learning rate to 0.01 and the training number to 500 epochs. We set the batch size to 30 due to GPU memory constraints and the margin value α in the triplet loss value to 0.2.

4.1.1. Dataset

Our experiments use the Hong Kong Polytechnic University public finger image dataset (version 1) [13], referred to as PolyU-DB. Kumar and Zhou [13] divided the collection process of this dataset into two sessions: the data collected in the first stage consist of 6 images of the index finger and middle finger of one hand of 156 people, a total of 1872 images; in the second session, for 105 of the 156 people in the stage, 6 images each of the index and middle fingers of the same hand were collected again, for a total of 1260 images. The final dataset contains 3132 FV and IKP each, belonging to 312 categories. Figure 5 shows some of the images in the PolyU-DB dataset.



(**a**)

(**b**)

Figure 5. Sample images of PolyU-DB. (**a**) FV images and their corresponding (**b**) IKP images.

During the experiment, we divide 312 categories of the PolyU-DB dataset into training and test sets. Among them, the training set contained 210 categories (2520 FV and IKP image pairs), while the test set contained the remaining 102 categories (612 image pairs), as shown in Table 1. To alleviate the problem that the model may not converge due to the small training set, we adopt image augmentation techniques such as translation clipping and conversion on the training set images. Thus, the data volume of the final training set is five times that of the original training set (12,600 image pairs).

Table 1. Dataset division and description.

Dataset	Subject Numbers	Finger Numbers	Image Numbers of each Finger	Total Image Number
Training Set	105	2	12	2520
Test Set	51	2	6	612

4.1.2. Evaluation Metrics

The experiments use the receiver operating characteristic (ROC) [40] curve and the equal error rate (EER) to evaluate the performance of the fusion recognition system of FV

and IKP. We draw the ROC curve according to the value of the false positive rate (FPR) and the value of the true positive rate (TPR) based on different thresholds, which can reflect the balance between the false reject rate (FRR) and the false accept rate (FAR). When the value of FRR is equal to that of FAR in the ROC curve graph, the value of EER is that of FAR. The smaller the value of EER, the fewer instances of matching and recognition errors, and the higher the recognition accuracy.

4.2. Comparative Experiments

To verify the effectiveness of the proposed method, we compare multiple methods, as shown in Table 2. We divide the methods in the table into the following four categories:

(a) Non-training-based unimodal methods: Including the wide line detector method [15], mean curvature method [16], and repeated line tracking method [17]. These three algorithms were common methods in FV recognition before deep neural networks became the mainstream method for feature extraction.
(b) Non-training-based multimodal methods: Holistic [13] and nonlinear [13] are two methods using score-level fusion.
(c) Training-based unimodal methods: Light CNN [41], ResNet-50 [32], and SqueezeNet [42] are all general networks based on CNN. Wimmer et al. [43] used them to realize FV recognition by combining triplet loss.
(d) Training-based multimodal methods: VGGNet-16 [31], Resnet-50, Resnet-101 [32], ResNet50-Softmax-Add [22], and AsymmetricNet [23]. The former three are also general convolutional networks. Kim et al. [24] combined them with score fusion methods to complete identity recognition. The latter two, like our method, are based on feature-level fusion. We reproduced them for multimodal recognition based on IKP and FV.

Table 2. Results of comparative experiments. Non-training-based represents traditional methods without training, and training-based represents methods based on CNNs. The unimodal methods use only FV, while the multimodal methods use FV and IKP (methods with * use FV and hand shape).

	Non-Training-Based Method	**EER/%**	**Training-Based Method**	**EER/%**
Unimodal	Wide line detector [44]	9.46	SqueezeNet [43]	3.7
	Mean curvature [44]	6.49	Light CNN [43]	10
	Repeated line tracking [44]	4.45	ResNet-50 [43]	5.6
Multimodal	Holistic [13]	2.72	VGGNet-16 * [24]	2.4433
	Nonlinear [13]	2.45	ResNet-50 * [24]	1.0235
			ResNet-101 * [24]	0.7859
			AsymmetricNet [23]	3.393
			ResNet50-Softmax-Add [22]	1.109
			Dual-Branch-Net (ours)	0.422

We found several phenomena from Table 2. First, the EERs of all multimodal methods are lower than those of the unimodal methods. The main reason behind this phenomenon is that multimodal recognition methods can use more biological information to distinguish individuals from more perspectives. Second, among multimodal recognition methods, most training-based methods can achieve better recognition results than non-training-based methods. This verifies the effectiveness of CNN in this field. Third, our proposed method achieves the lowest EER. Although AsymmetricNet and ResNet50-Softmax-Add both use feature-level fusion methods, their effects are not as good as our proposed Dual-Branch-Net. This is probably because the fusion methods they designed only fuse deep features of multiple modalities. In contrast, Dual-Branch-Net fuses multilevel multimodal features, making the fusion more thorough.

4.3. Ablation Experiments

To verify that the proposed network model can effectively extract the fusion features of FV and IKP, we perform ablation experiments based on different modalities and different fusion methods.

First, to verify the effectiveness of the proposed fusion method, we compare the proposed fusion method with another simple splicing fusion method. Specifically, the features of the two modalities are only fused once at the output end of the last fully connected Inception-Resnet-v1 by concatenation operation. We concatenate the 512-dimensional output features of FV and the 512-dimensional output features of IKP to form 1024-dimensional features. We denote a Dual-Branch-Net with only stage 4 connections as Dual-Branch-Net (only stage 4 fusion). We present the ROC curve of the experiment in Figure 6 and the precise EER value in Table 3. The ROC curve of the Dual-Branch-Net (only stage 4 fusion) method shifts to the lower right compared with that of the Dual-Branch-Net. This indicates that the recognition performance of the simple-fusion method is worse than that of the proposed fusion method. The EER value based on the Dual-Branch-Net (only stage 4 fusion) in Table 3 indicates that the corresponding EER values vary by 0.377%. There is one possible cause for the deterioration of the performance of Dual-Branch-Net (only stage 4 fusion). The feature fusion depth of the multimodal collaborative network output is not enough, as it does not consider the multilevel information fusion of the network. However, Dual-Branch-Net fuses features at different levels. Thus, there is less information loss and more thorough fusion.

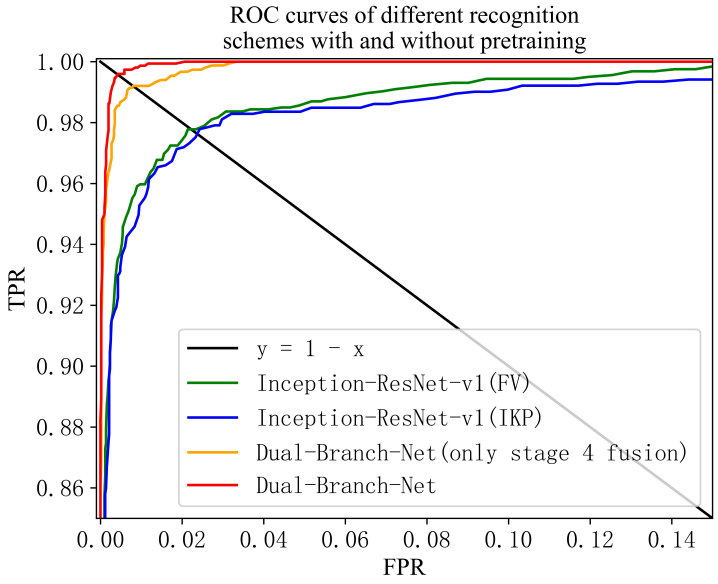


Figure 6. ROC curves of ablation experiment results.

Table 3. Results of ablation experiments.

Method	EER/%
Inception-ResNet-v1 (IKP)	2.373
Inception-ResNet-v1 (FV)	2.222
Dual-Branch-Net (only stage 4 fusion)	0.799
Dual-Branch-Net	0.422

Second, we compare our multimodal recognition method with a unimodal recognition method based on IKP or FV to study the necessity of recognition based on multimodal features. In the unimodal experiment, we first use Inception-ResNet-v1 to extract the features of the unimodal image and perform L2 regularization on output features. We then combine the triplet loss function to optimize the network. After that, the trained network model is used to extract the feature of the image pairs. Finally, according to

the similarity between the feature vector of the test image and the target image, we can determine whether the experimenter corresponding to the test image is a real matcher or an impostor. From the experimental results in Figure 6 and Table 3, we find that the performance of multimodal recognition of FV and IKP is significantly better than that of the unimodal recognition of FV or IKP. Therefore, the result confirms the feasibility of the proposed multimodal recognition system.

In addition, from the experimental results in Figure 6, the performance of FV recognition is better than that of IKP recognition. Therefore, we use FV as the primary task modality of Dual-Branch-Net, and IKP as the auxiliary task modality.

4.4. Discussion of Pretraining

We conduct experiments with and without pretraining parameters loaded to explore the necessity to initialize the Dual-Branch-Net model. We present the ROC curves of the experiments in Figure 7 and the EER values in Table 4. The methods with the "with pretraining" suffix in Figure 7 represent using the pretrained network weights to initialize the weights. The pretrained network weights are generated by Inception-Resnet training on the large-scale dataset CASIA-WebFace. Meanwhile, the methods with the "without pretraining" suffix randomly initialize the network weights.

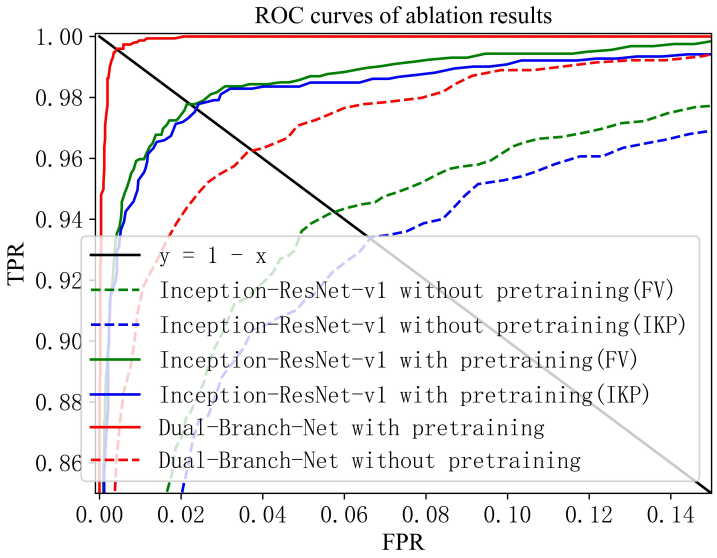


Figure 7. ROC curves of different recognition schemes with and without pretraining.

Table 4. Results of experiments with and without pretraining.

Method	EER/%	
	Without Pretraining	With Pretraining
Inception-ResNet-v1 (IKP)	6.634	2.373
Inception-ResNet-v1 (FV)	5.764	2.222
Dual-Branch-Net	3.750	0.422

Compared with the method of loading pretrained weights, the recognition performance of randomly initializing the network parameters dropped significantly. Numerically, the variation of the EER value is more than 3%. The recognition performance of randomly initializing the network parameters is not even as good as some traditional methods mentioned in Table 2. The main reason for this phenomenon is the large number of model parameters. Specifically, the combined space of model parameter values is large, and the training data used are less, being insufficient to make the model weights converge to the optimal value. Therefore, when using a large-scale network model in the multimodal

recognition of FV and IKP, it is necessary to use transfer learning to load the weights of the pretrained model.

5. Conclusions

We proposed a network named Dual-Branch-Net, which can perform identity recognition based on dual modalities of IKP and FV. Unlike previous work that fuses multimodal features using simple splicing, our method combined two features at multiple levels. Moreover, we used both transfer learning and triplet loss to optimize the feature representation of the model. The experimental results on the public dataset PolyU-DB showed that the proposed method can reduce the EER of the recognition result to 0.422%. In the future, we will try to use some model distillation methods to reduce the computational complexity of the proposed method so that low-cost embedded devices can also run the method.

Author Contributions: Conceptualization, L.J. and D.Z.; methodology, X.L.; software, X.L.; validation, X.L., H.W. and D.Z.; formal analysis, L.J.; investigation, L.J.; resources, X.L.; writing—original draft preparation, X.L.; writing—review and editing, H.W.; supervision, H.W.; funding acquisition, L.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (No. 61976189, No. 62001418), Leading Innovation Team of Zhejiang Province under Grant (2021R01002), and Natural Science Foundation of Zhejiang Province (LQ21F010011).

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from The Hong Kong Polytechnic University and are available at http://www4.comp.polyu.edu.hk/~csajaykr/fvdatabase.htm (accessed on 3 November 2022) with the permission of The Hong Kong Polytechnic University.

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

References

1. Yang, W.; Wang, S.; Sahri, N.M.; Karie, N.M.; Ahmed, M.; Valli, C. Biometrics for Internet-of-Things security: A review. *Sensors* **2021**, *21*, 6163. [CrossRef] [PubMed]
2. Ramya, S.; Sheeba, R.; Aravind, P.; Gnanaprakasam, S.; Gokul, M.; Santhish, S. Face Biometric Authentication System for ATM using Deep Learning. In Proceedings of the 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 25–27 May 2022; pp. 1446–1451. [CrossRef]
3. Srinivas, K.K.; Vijitha, U.; Chandra, G.A.; Kumar, K.S.; Peddi, A.; Uppala, B.S. Artificial Intelligence based Optimal Biometric Security System Using Palm Veins. In Proceedings of the 2022 International Mobile and Embedded Technology Conference (MECON), Noida, India, 10–11 March 2022; pp. 287–291. [CrossRef]
4. Carvalho, J.M.; Brãs, S.; Ferreira, J.; Soares, S.C.; Pinho, A.J. Impact of the acquisition time on ECG compression-based biometric identification systems. In *Iberian Conference on Pattern Recognition and Image Analysis*; Springer: Cham, Switzerland, 2017; pp. 169–176. [CrossRef]
5. Tolosana, R.; Vera-Rodriguez, R.; Ortega-Garcia, J.; Fierrez, J. Preprocessing and feature selection for improved sensor interoperability in online biometric signature verification. *IEEE Access* **2015**, *3*, 478–489. [CrossRef]
6. Shende, P.; Dandawate, Y. Convolutional neural network-based feature extraction using multimodal for high security application. *Evol. Intell.* **2021**, *14*, 1023–1033. [CrossRef]
7. Basheer, S.; Nagwanshi, K.K.; Bhatia, S.; Dubey, S.; Sinha, G.R. FESD: An approach for biometric human footprint matching using fuzzy ensemble learning. *IEEE Access* **2021**, *9*, 26641–26663. [CrossRef]
8. Vijayakumar, T. Synthesis of palm print in feature fusion techniques for multimodal biometric recognition system online signature. *J. Innov. Image Process. (JIIP)* **2021**, *3*, 131–143. [CrossRef]
9. Ryu, R.; Yeom, S.; Kim, S.H.; Herbert, D. Continuous multimodal biometric authentication schemes: A systematic review. *IEEE Access* **2021**, *9*, 34541–34557. [CrossRef]
10. Purohit, H.; Ajmera, P.K. Optimal feature level fusion for secured human authentication in multimodal biometric system. *Mach. Vis. Appl.* **2021**, *32*, 1–12. [CrossRef]
11. Shaheed, K.; Liu, H.; Yang, G.; Qureshi, I.; Gou, J.; Yin, Y. A systematic review of finger vein recognition techniques. *Information* **2018**, *9*, 213. [CrossRef]

12. Bahmed, F.; Ould Mammar, M. Basic finger inner-knuckle print: A new hand biometric modality. *IET Biom.* **2021**, *10*, 65–73. [CrossRef]
13. Kumar, A.; Zhou, Y. Human identification using finger images. *IEEE Trans. Image Process.* **2011**, *21*, 2228–2244. [CrossRef]
14. Liu, C. A new finger vein feature extraction algorithm. In Proceedings of the 2013 6th International Congress on Image and Signal Processing (CISP), Hangzhou, China, 16–18 December 2013; Volume 1, pp. 395–399. [CrossRef]
15. Huang, B.; Dai, Y.; Li, R.; Tang, D.; Li, W. Finger-vein authentication based on wide line detector and pattern normalization. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 1269–1272. [CrossRef]
16. Song, W.; Kim, T.; Kim, H.C.; Choi, J.H.; Kong, H.J.; Lee, S.R. A finger-vein verification system using mean curvature. *Pattern Recognit. Lett.* **2011**, *32*, 1541–1547. [CrossRef]
17. Miura, N.; Nagasaka, A.; Miyatake, T. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Mach. Vis. Appl.* **2004**, *15*, 194–203. [CrossRef]
18. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Association for the Advancement of Artificial Intelligence (AAAI): San Francisco, CA, USA, 2017. [CrossRef]
19. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823. [CrossRef]
20. Noh, K.J.; Choi, J.; Hong, J.S.; Park, K.R. Finger-vein recognition based on densely connected convolutional network using score-level fusion with shape and texture images. *IEEE Access* **2020**, *8*, 96748–96766. [CrossRef]
21. Song, J.M.; Kim, W.; Park, K.R. Finger-vein recognition based on deep DenseNet using composite image. *IEEE Access* **2019**, *7*, 66845–66863. [CrossRef]
22. Daas, S.; Yahi, A.; Bakir, T.; Sedhane, M.; Boughazi, M.; Bourennane, E.B. Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion. *IET Image Process.* **2020**, *14*, 3859–3868. [CrossRef]
23. Wen, M.; Zhang, H.; Yang, J. End-To-End Finger Trimodal Features Fusion and Recognition Model Based on CNN. In *Chinese Conference on Biometric Recognition*; Springer: Shanghai, China, 2021; pp. 39–48. [CrossRef]
24. Kim, W.; Song, J.M.; Park, K.R. Multimodal biometric recognition based on convolutional neural network by the fusion of finger-vein and finger shape using near-infrared (NIR) camera sensor. *Sensors* **2018**, *18*, 2296. [CrossRef]
25. Goswami, G.; Mittal, P.; Majumdar, A.; Vatsa, M.; Singh, R. Group sparse representation based classification for multi-feature multimodal biometrics. *Inf. Fusion* **2016**, *32*, 3–12. [CrossRef]
26. Evangelin, L.N.; Fred, A.L. Feature level fusion approach for personal authentication in multimodal biometrics. In Proceedings of the 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM), Chennai, India, 23–24 March 2017; pp. 148–151. [CrossRef]
27. Li, S.; Zhang, H.; Shi, Y.; Yang, J. Novel local coding algorithm for finger multimodal feature description and recognition. *Sensors* **2019**, *19*, 2213. [CrossRef]
28. Veluchamy, S.; Karlmarx, L. System for multimodal biometric recognition based on finger knuckle and finger vein using feature-level fusion and k-support vector machine classifier. *IET Biom.* **2017**, *6*, 232–242. [CrossRef]
29. Wang, L.; Zhang, H.; Yang, J. Finger multimodal features fusion and recognition based on CNN. In Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, 6–9 December 2019; pp. 3183–3188. [CrossRef]
30. Ren, H.; Sun, L.; Guo, J.; Han, C. A Dataset and Benchmark for Multimodal Biometric Recognition Based on Fingerprint and Finger Vein. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 2030–2043. [CrossRef]
31. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Available online: https://arxiv.org/abs/1409.1556 (accessed on 22 April 2022).
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
33. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [CrossRef]
34. Alay, N.; Al-Baity, H.H. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors* **2020**, *20*, 5523. [CrossRef] [PubMed]
35. Walia, G.S.; Singh, T.; Singh, K.; Verma, N. Robust multimodal biometric system based on optimal score level fusion model. *Expert Syst. Appl.* **2019**, *116*, 364–376. [CrossRef]
36. Wang, Y.; Shi, D.; Zhou, W. Convolutional Neural Network Approach Based on Multimodal Biometric System with Fusion of Face and Finger Vein Features. *Sensors* **2022**, *22*, 6039. [CrossRef]
37. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. [CrossRef]
38. Soleymani, S.; Dabouei, A.; Kazemi, H.; Dawson, J.; Nasrabadi, N.M. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 3469–3476. [CrossRef]

39. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Learning Face Representation from Scratch. Available online: https://arxiv.org/abs/1411.7923 (accessed on 16 May 2022).
40. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
41. Wu, X.; He, R.; Sun, Z.; Tan, T. A light CNN for deep face representation with noisy labels. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2884–2896. [CrossRef]
42. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size. Available online: https://arxiv.org/abs/1602.07360 (accessed on 15 June 2022).
43. Wimmer, G.; Prommegger, B.; Uhl, A. Finger vein recognition and intra-subject similarity evaluation of finger veins using the cnn triplet loss. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 400–406. [CrossRef]
44. Joseph, R.B.; Ezhilmaran, D. An efficient approach to finger vein pattern extraction using fuzzy rule-based system. In *Innovations in Computer Science and Engineering*; Springer: Hyderabad, India, 2019; pp. 435–443. [CrossRef]