electronics

MDPI

Article

Deep Residual Vector Encoding for Vein Recognition

Fuqiang Li [1], Tongzhuang Zhang [2], Yong Liu [1] and Feiqi Long [3,*]

[1] Xuhai College, China University of Mining and Technology, Xuzhou 221116, China
[2] The School of Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China
[3] The School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610056, China
* Correspondence: 202122080113@std.uestc.edu.cn

Abstract: Vein recognition has been drawing more attention recently because it is highly secure and reliable for practical biometric applications. However, underlying issues such as uneven illumination, low contrast, and sparse patterns with high inter-class similarities make the traditional vein recognition systems based on hand-engineered features unreliable. Recent successes of convolutional neural networks (CNNs) for large-scale image recognition tasks motivate us to replace the traditional hand-engineered features with the superior CNN to design a robust and discriminative vein recognition system. To address the difficulty of direct training or fine-tuning of a CNN with existing small-scale vein databases, a new knowledge transfer approach is formulated using pre-trained CNN models together with a training dataset (e.g., ImageNet) as a robust descriptor generation machine. With the generated deep residual descriptors, a very discriminative model, namely deep residual vector encoding (DRVE), is proposed by a hierarchical design of dictionary learning, coding, and classifier training procedures. Rigorous experiments are conducted with a high-quality hand-dorsa vein database, and superior recognition results compared with state-of-the-art models fully demonstrate the effectiveness of the proposed models. An additional experiment with the PolyU multispectral palmprint database is designed to illustrate the generalization ability.

Keywords: CNN; deep residual descriptors; sparse coding; vein recognition

Citation: Li, F.; Zhang, T.; Liu, Y.; Long, F. Deep Residual Vector Encoding for Vein Recognition. *Electronics* **2022**, *11*, 3300. https://doi.org/10.3390/electronics11203300

Academic Editor: Stefanos Kollias

Received: 5 September 2022
Accepted: 3 October 2022
Published: 13 October 2022

1. Introduction

Vein pattern, an intrinsic biometric pattern imaged under near-infrared (NIR) light, has emerged as a promising alternative for person identification. Compared with extrinsic biometric features such as face, fingerprint, palmprint, and iris, vein patterns including finger–vein, dorsa–vein, and palm–vein are highly secure, private, and convenient. These properties are the basic requirements for practical applications and are attracting more attention [1–5]. Despite the advantages of adopting vein patterns for person identification, there still exist some inherent issues (e.g., unavoidable environmental illuminations [5–7], ambient temperature effects [2,7–9], uncontrollable user behaviours [9–11], and NIR device degradation [12–14]), which make the design of robust and accurate vein recognition systems a challenging task. To alleviate the inherent influence of these issues, as shown in Figure 1, researchers [15,16] have proposed different algorithms targeted at a specific step of the traditional vein recognition framework, such as designing restoration methods to recover the details, proposing better feature extraction methods, and developing more robust matching strategies.

Other work [15–18] focuses on the improvement of vein image acquisition. Among these studies, Wang et al. [15] proposed incorporating a hierarchical vein image quality assessment index to control the lighting intensity of NIR so as to improve the contrast of captured vein images. Based on this framework, the accuracy of the index in describing vein patterns was improved [16] by designing a local image contrast estimator that is fast to compute and consumes negligible time for image capture. A similar illuminance

control strategy was proposed [17] by calculating the block intensity, where each block (vein and non-vein region) is treated equally, thus obtaining over-exposed/under-exposed vein images with low contrast. Another common problem for works focusing on the image acquisition step is that such strategies will inevitably shorten the life expectancy of an NIR system and make it inapplicable for biometric products.

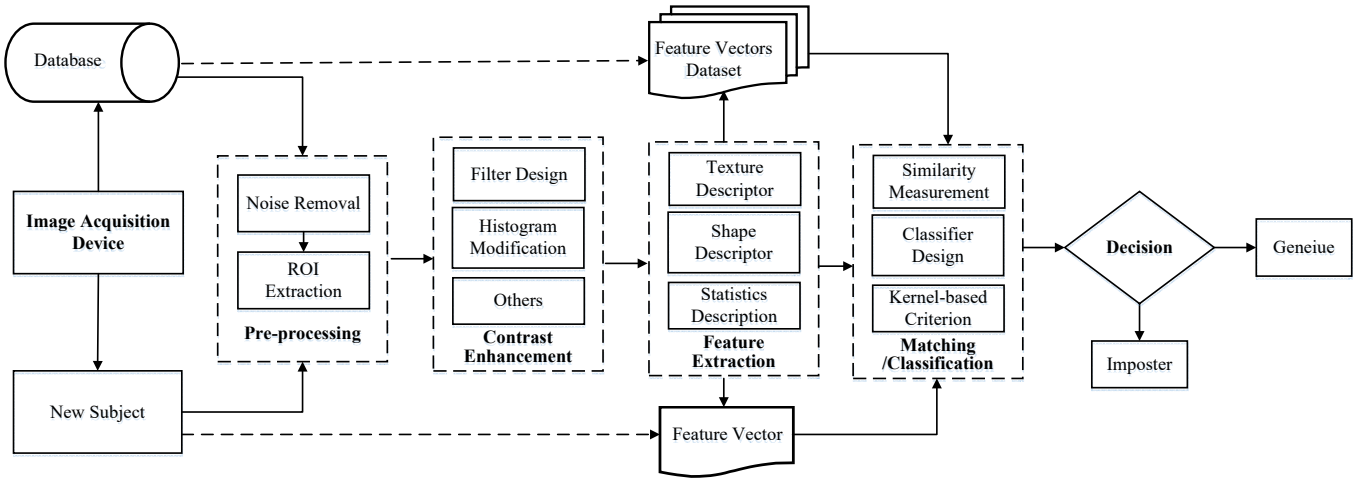

Figure 1. General framework of hand vein recognition system.

Different from the illuminance control strategies, more effective methods for handling the low-contrast vein pattern encoding and matching problems have been proposed by designing effective algorithms for vein pattern restoration [13,19], contrast enhancement [20–22], and better feature extraction [23,24]. Following this direction, methods that introduce contrast enhancement before conducting feature extraction have been researched and deployed to obtain competitive performance. In the first group, the repeated line tracking [8] approach was devised to restore the line-like vein structure and match sample-specific structures for identification. To make up for the unreliable pattern finding results and information loss with line structure detection, the approach of obtaining the maximum response along eight directions at the central pixel was proposed [25] to extract the concave region followed by a region growing method for pattern definition. Gabor based directional or structure information encoding methods were also proposed for specific vein pattern restoration and representation, and state-of-the-art recognition results [2,8,25–27] were obtained. However, the structure information mining and coding methods are sensitive to noise, illumination variation, and contactless acquisition, making them unreliable for practical vein recognition systems. Another group strictly follows the procedure shown in Figure 1, and representative hand-engineered feature models are adopted for designing robust vein recognition systems. To cope with the inherent difficulty of generating discriminative and robust representation for vein patterns that are shown as low-contrast and sparse structures with high inter-class similarities [28], scale invariant feature transform (SIFT) [29] and local binary pattern (LBP) [16] are the most commonly used feature descriptors due to their robustness and discrimination capabilities. However, some work [16,29] reveals that the vein recognition performance will be impaired by the contrast enhancement step. Specifically, the widely adopted contrast enhancement step usually introduces mismatches in SIFT and increases intra-class dissimilarity in LBP.

More recently, hand-engineered feature representation has been significantly outperformed by CNN in almost all domains of image understanding, including object detection [30], categorization [31], and segmentation [32]. Faced by existing problems on how to design algorithms that are free of contrast enhancement and are able to generate discriminative representations to handle vein patterns with sparse structures and high inter-class similarities, researchers have turned to CNN [6,33–36]. Aimed at evaluating the effectiveness of CNN for vein discrimination tasks, Radzi et al. [34,35] proposed constructing a tiny

CNN that was trained directly or fine-tuned with finger vein databases. However, due to the insufficient model training caused by the lack of scale and diversity in the training dataset, the performance was much lower than that of the state-of-the-art methods [16], indicating the necessity of proposing new formulations for taking advantage of CNN to deal with challenging vein recognition tasks.

To discover an improved solution on utilizing powerful CNN for discriminating vein patterns, the idea of transferring the semantic knowledge encoded by both a pre-trained CNN and the training dataset (e.g., ImageNet) as the representation of vein patterns is proposed in this paper as the basis for constructing robust vein recognition systems. As shown in Figure 2, the pre-trained CNN model together with its training dataset (e.g., ImageNet) is utilized to generate a descriptor pool. Subsequently, the vein image is first processed with the same pre-trained CNN to obtain the 1×1000 probability matrix (at the softmax layer), which serves as the index for calculating the feature selection matrix by a task-specific thresholding strategy. The $K \times 4096$ sized deep residual descriptors, which are able to describe the vein pattern in a more discriminative manner by transferring knowledge from both pre-trained CNN and the ImageNet dataset, are obtained by conducting subtraction between the output of the fully connected layer from vein images and the associated reference descriptors, which are calculated with matrix multiplication between the feature selection matrix and the descriptor pool matrix. The proposed descriptors are free from the burstiness problem [37] that exists in traditional SIFT features, where some of the detected keypoints are repeated, thus degrading the discriminative ability of the descriptors.

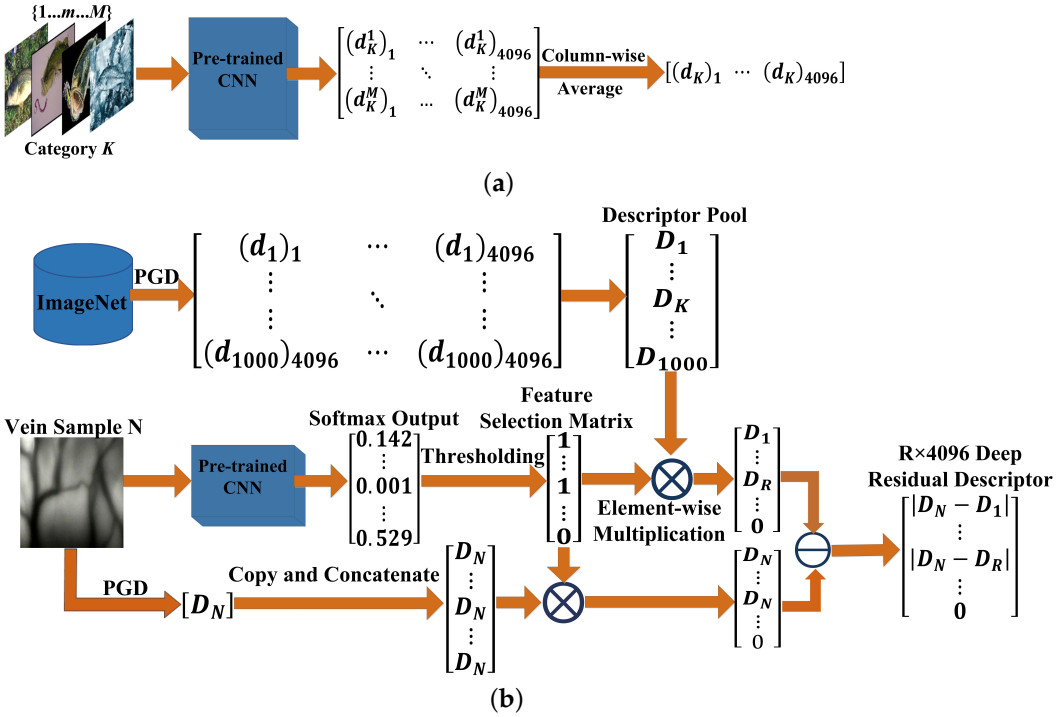

Figure 2. Overview of the proposed deep residual descriptor generation system. (**a**) Procedure of generating descriptors (PGD) from a pre-trained CNN for category K containing M samples. (**b**) Procedure of generating a descriptor pool and deep residual descriptor for a given vein sample.

After obtaining the robust and discriminative descriptors by transferring knowledge from other datasets with pre-trained CNN, high-level descriptor encoding methods [38,39] are desired for further improving the semantical discriminability of the descriptors. To achieve this, the deep residual vector encoding (DRVE) model is designed by combining sparse dictionary learning and coding algorithm with off-the-shelf deep residual descriptors, and state-of-the-art vein recognition results are obtained.

The remainder of this paper is organized as follows. In Section 2, related vein pattern encoding methods are reviewed and analyzed. In Section 3, methods for generating the deep residual descriptors are proposed, followed by introducing the DRVE model. Rigorous experiments and results analysis are given in Section 4. Finally, the paper concludes with a discussion about possible future directions in Section 5.

2. Related Work

As discussed earlier, the design of feature extraction algorithms is the most crucial and challenging aspect of vein recognition as a result of the inherent properties of vein images. Numerous studies in this area concentrate on creating more reliable and discriminative feature representations. The most recent methods noted in the literature can be summarized as follows:

Local invariant feature-based methods: Scale-invariant feature description models are the most often used hand-engineered feature extraction model for resolving the requirement of unconstrained geometric transformation, including SIFT [40] and its variants such as SURF [41], RootSIFT [42], and ASIFT [43], and the feature vector generation procedures of these models share the same configuration, including the use of scale-space, followed by extremal point detection and descriptor generation. Pierre et al. [44] used SIFT for hand vein description following simple preprocessing, which refers to the denoising and contrast enhancement procedure. However, it has been argued that preprocessing would result in vein loss and low contrast distribution, causing great difficulty in generating a sufficient number of descriptive keypoints. To reduce or remove the negative effects of contrast enhancement, SIFT or SURF [45–47] features were extracted directly on the input images without any pre-processing and binarization procedures, resulting in considerable performance improvement. However, the existence of matching pairs between two different samples usually increases the false acceptance rate (FAR) and equal error rate (EER), and this is unacceptable for vein recognition. Inspired by conclusions elsewhere [44–47], Kang et al. [48] attempted to improve SIFT performance through complementary feature fusion and noisy keypoint removal, as well as DoG-HE (histogram equalization) for improved contrast enhancement. LBP is used to describe the region distribution of the matching keypoints during the mismatch removal step, and the mismatching points are deleted using LBP difference, increasing accuracy overall. The keypoint fusion method's drawback is that the detected non-vein keypoints describing the palmprint or other regions tend to increase the FAR.

Local binary pattern based methods: Despite the robust and discriminative property of local invariant feature based models, the existence of mismatches [29] due to contrast enhancement makes the final results unreliable. In addition, the complexity of numerous descriptor detection and matching methods degrades their efficiency. Aimed at overcoming such disadvantages, and to take full advantage of local image region based feature extraction models, attention has been paid to local geometrical distribution coding methods, such as LBP [49], local derivative pattern (LDP) [50], local ternary pattern (LTP) [51], and local line binary pattern (LLBP) [52]. These approaches' main goal is to represent the global gray distribution histogram based on local coding calculation results; however, their discriminative power is hindered by sparse vein texture distribution and the exclusion of contrast information from feature coding.

Other task-specific methods: Vein recognition, as a classical image processing and pattern recognition problem, may be realized by transferring knowledge from systems established for solving other image recognition tasks, which share some common characteristics with vein structure. Inspired by such ideas, superpixel-based features [53], vein textons map [54,55], hyperinformation based features [56], and personalized and selected features [57,58] have been successively introduced for obtaining state-of-the-art recognition results. Despite high accuracy, the drawbacks of high computation time and uncertainty in the selection of optimal parameters prevent these methods from being widely used and

applied in real scenarios and also make it impossible to reproduce the reported results when applying the algorithms to other vein databases.

Furthermore, we believe that the proposed residual descriptor encoding method may also be applied to other image recognition tasks realized by defining a pre-trained CNN as feature extractor (i.e., detecting the diabetic retinopathy lesion subtypes in ultra-wide field images [59]).

3. Sparse Dictionary Learning with Deep Residual Descriptors

In Figure 2, a systematic diagram describing the procedure of generating the deep residual descriptors is given. Figure 2a demonstrates the specific procedure of generating descriptors (PGD) from a pre-trained CNN on some sample images: suppose that there are M samples for category K ($K \in \{1, \ldots, 1000\}$), a feed-forward operation of convolution and pooling with a pre-trained CNN is conducted for generating an $M \times 4096$ feature matrix, on which the column-wise averaging is performed to obtain the final 1×4096 descriptor for category K. By performing the simple PGD operations, the descriptor for each category of a dataset (e.g., 1000 categories for ImageNet) can be obtained and then combined to form the final descriptor pool, and we obtain the 1000×4096 descriptor pool with a CNN pre-trained on the ImageNet dataset. With the generated discriminative descriptor pool, the deep residual descriptor for one vein image can be constructed in the following manner: by performing the feed-forward operation with the same pre-trained CNN on the vein image, the output of the second last fully connected layer is obtained and denoted as D_N and is then transformed to a 1000×4096 matrix by a 'copy and concatenate' operation. At the same time, the output of the softmax layer is used for generating a feature selection matrix by thresholding, and the threshold is determined by cross-validation (see Section 4 for details). The feature selection matrix is then used to select the vein sample-specific descriptors from the descriptor pool (e.g., selecting $R \times 4096$ descriptors out of the 1000×4096 descriptor pool in Figure 2), and such a selection can be interpreted as transferring vein image related knowledge from a pre-trained CNN and its training dataset to obtain a discriminative representation for depicting vein patterns, thus alleviating the disadvantage of training a CNN with a limited dataset. To obtain the deep residual descriptors, the 1000×4096 sized descriptor (obtained by 'copy and concatenate') of the vein sample is also transformed into a $R \times 4096$ descriptor by the same feature selection matrix. Finally, the residual descriptor defined as the representation for the vein sample is obtained by absolute substraction between the selected $R \times 4096$ descriptor and the $R \times 4096$ vein sample descriptor. Based on the generated deep residual descriptors, the sparse dictionary coding based model is formulated to construct the DRVE model, and the issues on defining suitable codebook generation methods for the proposed residual descriptors are investigated. This step is followed by the introduction of a linear support vector machine (SVM) for efficient vein recognition.

Based on the selected robust and discriminative deep residual descriptors, the bag of features (BoF) model is analyzed and deployed for final feature generation and classification. Considering that the final descriptors for each sample should be sparse and by setting the threshold relatively high for generating the feature selection matrix in Stage II, as shown in Figure 2, an overcomplete dictionary and soft membership indicators [60] are desired for obtaining good recognition results. Based on theoretical and experimental analyses, sparse coding methods are capable of handling deep residual descriptors, and the state-of-the-art performance demonstrates the effectiveness of the proposed DRVE. In addition, by utilizing the simple multi-class linear SVM, computational efficiency is also improved.

3.1. Deep Residual Descriptors

Driven by the great success of CNN in image understanding tasks, four different CNN-based vein recognition models are investigated to test their abilities in describing and discriminating sparse and blurry vein patterns. Unexpected low recognition rate and deep analyses into the underlying reasons indicate the necessity for deep residual descriptors,

a more discriminative feature representation method. To take better advantage of its discriminability, sparse dictionary based high-level encoding and classification methods are also introduced. In the proposed model, CNN together with its training dataset is defined as a descriptor generation machine, and the output of the defined machine is a discriminative feature/descriptor pool.

When feeding a vein image into the pre-trained CNN of the machine, as shown in Figure 2, the value of the softmax layer output, which indicates the relative similarity of input to a certain class in the training data, is obtained. Subsequently, by setting a suitable threshold, the feature selection (FS) matrix can be calculated as given in Equation (1):

$$\begin{aligned} FS_{\text{top-}K} &= [F_1, F_2, \ldots, F_{1000}] \\ F_i &= \begin{cases} 1 & \text{if } i \leq k \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{1}$$

With the matrix, the deep residual descriptors may be obtained by element-wise matrix multiplication between FS and the original descriptor pool, as described in Equation (2):

$$\mathbf{D} = (FS_{\text{top-}K})^T [D_1, D_2, \ldots, D_{1000}] \tag{2}$$

where D_i is $(\sum_{j=1}^{M} d_i^j)/M$ with d_i^j being a 1×4096-D feature vector from the previous 4096-D fully-connected layer, and M is the total number of samples in each training category. With such feature vectors describing one vein image, high-level feature encoding methods with sparse dictionary learning algorithms are realized for linear SVM training. Note that PCA is adopted to reduce the dimension of the descriptors from 4096 to 1000.

3.2. Learning Discriminative Representation with DRVE

Considering the difference between the residual descriptors and the traditional SIFT descriptors, suitable dictionary generation methods by modification to the selected basis are analyzed.

Let $\mathbf{X}$ be a set of deep residual descriptors lying in a D-dimensional feature space, i.e., $\mathbf{X} = [X_1, \ldots, X_N]^T \in \mathbb{R}^{N \times D}$. The most widely used codebook optimization model, the vector quantization (VQ) algorithm [61], is adopted by applying K-means to solve the following objective function (3):

$$\min_{C} \sum_{n=1}^{N} \min_{k=1,\ldots,K} \|X_n - C_k\|_2^2 \tag{3}$$

where $C = [C_1, \ldots, C_K]^T$ is the codebook to be optimized, and $\|\cdot\|_2^2$ represents the ℓ_2-norm of the reconstruction residual matrix. The encoding problem can be solved by updating C_k with $A_k C$, and $A = [A_1, \ldots, A_K]^T$ represents the reconstruction coefficients. The overall optimization problem for solving the codebook and coefficients simultaneously can be realized by re-formulating the objective function (3) as:

$$\begin{aligned} &\min_{C, A_k} \sum_{n=1}^{N} \|X_n - A_k C\|_2^2 \\ &\text{subject to } Card(A_k) = 1, \|A_k\|_1 = 1, A_k \succeq 0, \forall k \end{aligned} \tag{4}$$

where $Card(A_k = 1)$ requires that only one element of A_k is nonzero, and $A_k \succeq 0$ restricts that all elements of A_k are nonnegative; $\|A_k\|_1$ is the ℓ_1-norm operation, which defines the summation of the absolute value of each element in A_k. Based on these constraints, it can be concluded that the hard assignment [61] with $Card(A_k) = 1$ in VQ will undoubtedly result in a coarse reconstruction with large residual errors, thus degrading the discriminability of the coefficient representation. When adopting VQ based dictionary generation and coding

methods for the residual descriptors, the performance will be worsened if N in $\mathbf{X}$ is small, which is the prerequisite to ensure that the generated residual descriptors are separable without intersection. To fully utilize the discriminability of the proposed deep residual descriptors, different constraints are added to the dictionary training and representation learning problems, resulting in functions (5)–(6), with which the off-line dictionary learning is carried out:

$$\begin{aligned} &\min_{C,A_k} \frac{1}{2}\sum_{n=1}^{N}\|X_n - A_k C\|_2^2 + \alpha\sum_{n=1}^{N}\|A_k\|_1 \\ &\text{subject to } \|C_j\|_2^2 \le 1 \quad \forall j \in \{1,\ldots,K\} \end{aligned} \tag{5}$$

where α is a tradeoff parameter between reconstruction error and sparsity constraint, and the ℓ_2-norm on C_j is used to prevent the arbitrarily small values of A_k. After learning the overcomplete dictionary $\boldsymbol{C}$ [61], a similar model by adding another constraint on the obtained coefficients is proposed for obtaining the final representation:

$$\min_{A_k} \frac{1}{2}\|X_n - A_k\boldsymbol{C}\|_2^2 + \beta_1\|A_k\|_1 + \frac{\beta_2}{2}\|A_k\|_2^2 \tag{6}$$

By optimizing function (6) under the ℓ_1 and ℓ_2 sparse coding scheme [61], a discriminative and stable representation with the learned dictionary is obtained for later classifier training and vein recognition.

With effective sparse coding on the constructed deep residual descriptors, robust and discriminative features are obtained to represent the hand–dorsa vein images, which show the sparse and high inter-class similarity structure, thus alleviating the demand for complex kernel-based classifiers for recognition tasks. Following this pipeline, a simple implementation of linear-SVM [60] is introduced to improve the efficiency of the DRVE model.

To solve the recognition problem with multiple classes, the one-against-all strategy is adopted to train multiple binary linear-SVMs. Given the training dataset $\{(F_i, y_i)\}_{i=1}^{L}$, $y_i \in \mathbf{Y} = \{1,\ldots,L\}$, the classifier aims at learning L linear and binary functions $\{w_c^T Fc \in \mathbf{Y}\}$, and the label for a certain input F can be predicted by Equation (7):

$$y = \min_{c\in\mathbf{Y}} w_c^T F \tag{7}$$

The parameters w_c of the kernel can be obtained by solving the following unconstrained convex function $J(w_c)$ with w_c as a variable:

$$\min_{w_c}\{J(w_c) = \|w_c\|_2^2 + \boldsymbol{C}\sum_{i=1}^{L} Ł(w_c, y_i^c, F_i)\} \tag{8}$$

where y_i^c equals 1 if the class label y_i is 1, otherwise y_i equals -1, and $Ł(\cdot)$ is an improved hinge loss function defined as:

$$Ł(w_c, y_i^c, F_i) = [\max(0, w_c^T F y_i^c - 1)]^2 \tag{9}$$

The replaced loss function is a differentiable quadratic hinge loss, which enables the training procedure to be conducted efficiently with a simple gradient-based optimization method (e.g., LBFGS [60]). In contrast, the original loss function in the traditional SVM is not differentiable everywhere, thus hampering the utilization of fast gradient methods for optimization. Such improvement can also be applied to other linear classification tasks for greatly improving the efficiency.

4. Experiments and Discussion

As a new approach to utilizing CNN for vein recognition tasks, comprehensive experiments were designed to verify the effectiveness of the proposed models.

4.1. Database and Baseline Model Setup

Dataset setup: Experiments for baseline model setup and performance comparison were conducted with the hand–dorsa vein database from China University of Mining and Technology [16], which was constructed with data from 98 females and 102 males with ages varying from 19 to 62. For each subject, 10 hand–dorsa vein images were acquired in two specifically set sessions separated by a time interval of more than 10 days, and each time, 5 images were acquired from each subject at the wavelength of 850 nm. The size of each image is 460 $\times$ 680 pixels. Similar to other work [16], the region of interest (ROI) for each image in this database was extracted and normalized to a size of 224 $\times$ 224 pixels. When carrying out experiments, half of the examples are randomly selected as training data, and the remaining images are utilized for testing. All results are obtained by repeating the experiments 10 times with random partitions, and the average recognition rate (RR) or equal error rate (EER) are recorded for model evaluation.

Parameters setup: As the key parameter to obtain the residual descriptor, the size for the dimension of the feature selection matrix in Figure 2 was determined by 5-fold cross validation from $\{1, \ldots, 50\}$ with the training set. After obtaining the dimension R, PCA was adopted for transforming the $R \times 4096$ descriptors to $R \times 1000$ descriptors for better dictionary generation. For DRVE, the codebook size was fixed as 512. The setup of other parameters can be referenced from our implementation.

Baseline model setup: With the availability of the numbers of pre-trained CNN models (e.g., VGG-16 [62], GoogLeNet [31], and ResNet-128 [63]) for descriptor pool generation, an experiment on finding out the most appropriate one for DRVE and SRRV was designed, and three models, including VGG-16 [62], GoogLeNet [31], and ResNet-128 [63]) pre-trained with ImageNet were involved. Following the training and testing setups described in the previous section, the average RR with the corresponding standard deviation were obtained and are shown in Table 1.

Table 1. Recognition rate (%) of DRVE with different pre-trained models.

	VGG	**GoogLeNet**	**ResNet**
DRVE	**98.83 $\pm$ 1.02**	91.25 $\pm$ 0.65	90.41 $\pm$ 0.78

As shown in Table 1, the overall recognition rate using VGG was better than the others for the DRVE model by a large margin, and the VGG based models were used as the baseline for comparative experiments. Furthermore, the experimental results also revealed that the VGG based network structure is more capable of discriminating the sparse vein structures. This may motivate other research on designing a VGG-like network structure and training it with a large-scale vein database or fine-tuning to seek the best results with model training instead of feature extraction from off-line models.

4.2. Comparison with State-of-the-Art

After obtaining the baseline formulations of DRVE, rigorous comparison experiments were designed to demonstrate the superiority of the proposed model over the current state-of-the-art. Comparisons were made with CNN-based vein recognition models as well as hand-engineered feature-based vein recognition models. Following the setup in most of the existing vein recognition models [16,29,34,35,64], the evaluation metric for the comparison is the EER rather than the direct RR. The EER is obtained by plotting the receiver operating characteristics (ROC) curve between the false acceptance rate (FAR) and the false rejection rate (FRR). After discovering the best parameter setup by cross-validation with the training dataset, the genuine matching and imposter matching on the testing set

are conducted with the trained DRVE for obtaining the FAR and FRR, respectively, with which we obtain the EER, as shown in Table 2 and Figure 3.

Table 2. EER (%) with different pre-trained models; the benchmark performance of the first four groups are highlighted. Group 1 represents direct training (DT) from scratch; group 2 represents fine-tuning (FT); group 3 represents off-line feature extractor (OFEx); group 4 represents off-line feature encoding (OFEn).

Group	**1 (DT)**		**2 (FT)**		**3 (OFEx)**		**4 (OFEn)**		**5 (Proposed)**
Methods	FingerveinNet	AlexNet	AlexNet	VGG	AlexNet	VGG	FV	VLAD	DRVE
Accuracy (%)	**2.089**	2.711	**3.104**	3.641	4.215	**2.835**	**1.028**	1.031	**0.016**

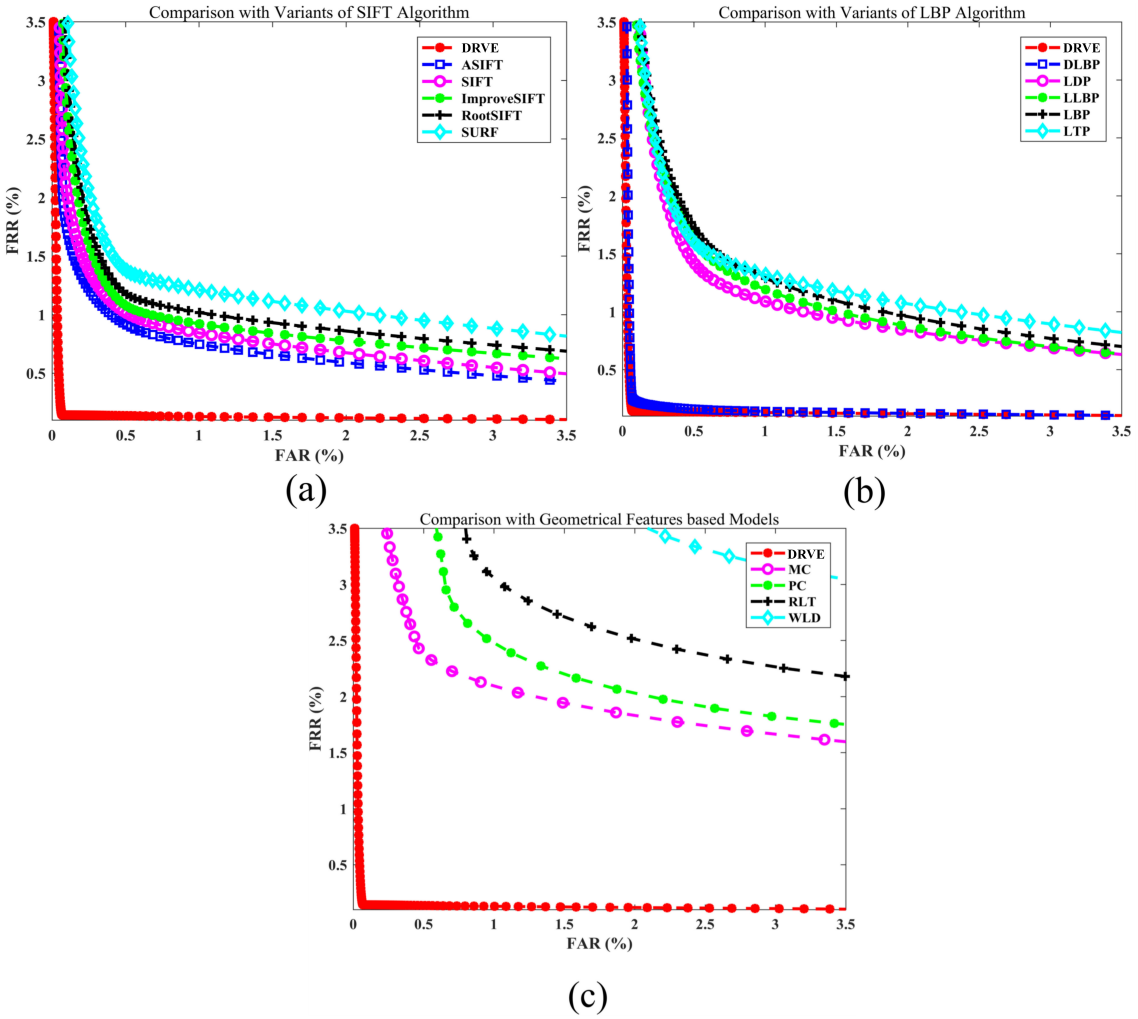

Figure 3. Comparison of ROC curves between the proposed models and representative hand-engineered methods including (**a**) SIFT based models, (**b**) LBP based models, and (**c**) geometrical feature based models.

Comparison with CNN based models: To demonstrate the superiority of the proposed encoding and recognition method over other types of CNN based models on the hand–dorsa vein dataset, four different kinds of formulations for utilizing CNN for vein recognition tasks are involved, including training a task-specific network from scratch [64], pre-trained model fine-tuning [64], direct off-line feature extraction with pre-trained VGG [31], and the combination of direct feature extraction with high-level feature encoding. In the case of training a new model from scratch, the models introduced in [34]

and [64] were realized. In the case of fine-tuning for the domain in focus (here, the hand–dorsa vein), again two models were realized for comparison. Apart from one model [64] that fine-tunes the AlexNet trained with ImageNet, VGG trained with ImageNet was also fine-tuned, as VGG is the most widely cited model for fine-tuning. In the third case of using the pre-trained model directly as a feature extractor, structure-growing CNN [64] and the VGG model were used, and the classifier is set the same as the one used in [64]. In the fourth case of encoding the directly extracted features, two representative methods [61], namely, the Fisher vector (FV) and vector of locally aggregated descriptors (VLAD), were used as the baseline for performance comparison. The last group is the proposed DRVE model, and only the baseline obtained from Table 1 was selected for comparison. It should be noted that the proposed vein recognition model was defined as a union but not as the combination of deep descriptors and feature encoding, with the result being that the performance of deep residual descriptors and the sparse coding models were not compared with other similar baseline models. The experimental results of these models with the hand–dorsa vein database are listed in Table 2.

The following conclusions can be drawn directly from Table 2: (1) When adopting CNN for vein pattern encoding and recognition, the proposed deep residual descriptor generation and encoding methods are the best, outperforming the existing four methods by a large margin ranging from 1.012% to 4.199%; (2) In the method of training a new model directly with vein database (Group 1), a task-specific structure [64] is probably a better choice than the generic models (e.g., VGG or ResNet); (3) In the method of fine-tuning (Group 2), the overall inferior performance compared with other groups indicates that the strategy of fine-tuning is not preferred when the target images differ greatly from the original training dataset, especially when the target dataset has a high inter-class similarity; (4) In the method of utilizing pre-trained CNN as feature extractor (Group 3), acceptable results confirm the transferability of the pre-trained model.

Further analysis of the experimental results yields: (5) For the feature encoding models (Group 4), the highest accuracy is achieved when compared with the previous three methods, which not only confirms the conclusion in (4) but also indicates that direct feature extraction followed by high-level feature encoding should be the very first step when adopting pre-trained CNN models for different tasks; (6) As the deep descriptor generation step is theoretically similar to those in Group 4, the large difference in the final results also demonstrates the superiority of the proposed feature extraction strategy over the traditional ones by transferring knowledge from other images to increase the discriminability of the extracted features. (7) Theoretically, the last three groups are somewhat similar, and the performance gain between neighbouring groups also shows that incorporating high-level feature encoding methods on the extracted pre-trained CNN features will greatly improve the performance. It also indicates that modifying the methods of feature extraction (e.g., generating the deep residual descriptors) and feature encoding methods simultaneously may result in state-of-the-art performance. In addition, the proposed generic approach of adopting CNN together with the training dataset for deep descriptor generation may also be applicable to other image classification tasks.

Comparison with hand-engineered feature based models: As shown in [16], hand-engineered features continue to dominate vein recognition tasks. Following a comparison of the proposed method with all possible CNN-based frameworks, a performance comparison with state-of-the-art hand-engineered features is presented here to demonstrate the superiority of the proposed DRVE models. Note that only the LBP and SIFT based models are selected here for comparison, because they are commonly used and capable of dealing with different kinds of vein databases, and the topology/geometry or task-specific representation based methods are not generally applicable to different kinds of databases. Statistical methods (e.g., PCA) are preferably adopted for feature post-processing but not directly for feature extraction in vein recognition tasks, and they are not included for comparison either. Specifically, two kinds of representative hand-engineered feature extraction algorithms are used as references: one is the local invariant feature model, including SIFT [40], SURF [41],

ASIFT [43], RootSIFT [42], and improved SIFT [29]; this is the best hand-engineered algorithm as it has the advantages of being invariant to rotation, translation, scale uncertainty, and uniform illumination. The other one is LBP [49] and its variants, including LDP [51], LTP [51], LLBP [52], and DLBP [16]; because of its efficiency, such a model is widely used for vein-based identification applications, and it also provides competitive recognition results. The specific EER results of different models are shown in Figure 3.

Judging from the EER result of identification with the hand–dorsa vein database, it can be concluded that the proposed deep residual descriptor encoding models perform better than the local invariant feature (LIF) models with an EER of 0.016%, whereas the best of LIF is 0.820% with ASIFT and the best of LBPs is 0.058% with DLBP. The new state-of-the-art vein recognition results fully demonstrate the ability of the proposed model in discriminating the sparse vein patterns with high inter-class similarity.

Comparison with geometrical feature based models: Vein patterns can be used for authentication due to differences in topological patterns between images. Driven by this, a series of feature extraction methods [8,11,65–69] are designed for discriminating vein patterns by geometrical difference. In this section, four representative methods, including the maximum curvature (MC) from [8], the wide line detector (WLD) [11], the principal curvature (PC) [67], and the repeated line tracking (RLT) method from [66], were selected for performance comparison. The specific EER results of different models are shown in Figure 3.

As shown in Figure 3, similar to the results from models based on hand-crafted features, large performance improvements using the proposed model over the other representative methods were shown as ranging from 1.841% to 3.005% on the hand–dorsa vein dataset. In addition, by comparing the results in Figure 3, the performance discrepancy between the best geometrical feature (MEC) and the best hand-crafted feature (DLBP) shows that geometrical features have limitations, because they can only be used effectively on a limited number of patterns and are sensitive to small modifications. However, benefiting from the proposed discriminative knowledge transfer mechanism and the hierarchical descriptor encoding methods (e.g., the sparse coding in this paper), the proposed model performs well across different biometrical patterns, as evidenced by the superior results with the palmprint datasets described in the following section.

5. Conclusions

The main contribution of this paper is a completely new knowledge transfer concept of defining a pre-trained CNN together with its training dataset (e.g., ImageNet) as a robust descriptor pool, from which representation knowledge is transferred to enrich the domain-specific image representation and to increase the final classification results. With the deep residual descriptor generation mechanism, discriminative representations for describing the sparse vein patterns with high inter-class similarity were obtained by the proposed DRVE model, which incorporated the sparse dictionary learning scheme for improving the semantic properties of the representation. VGG was adopted as the basis for generating the baseline model, and rigorous comparative experiments with both CNN based models and other state-of-the-art algorithms demonstrate the representation and generalization ability of the proposed model. In addition, by realizing different kinds of CNN based vein recognition baseline models, some conclusions on how to utilize CNN appropriately for challenging vein recognition tasks were obtained. Furthermore, research on utilizing the proposed model for other domain-specific image recognition tasks is ongoing.

Author Contributions: Conceptualization, F.L. (Fuqiang Li) and T.Z.; methodology, F.L. (Fuqiang Li); validation, F.L. (Fuqiang Li); formal analysis, T.Z.; writing—original draft preparation, F.L. (Fuqiang Li); writing—review and editing, F.L. (Feiqi Long); visualization, Y.L.; supervision; project administration, Y.L.; funding acquisition, F.L. (Fuqiang Li). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under grant 20KJB510051, and the Sichuan Science and Technology Program under grant 2022YFG0032.

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

References

1. Khellat, S.; Abrishambaf, R.; Monteiro, J.; Benyettou, M. Multimodal fusion of the finger vein, fingerprint and the finger-knuckle-print using Kernel Fisher analysis. *Appl. Soft Comput.* **2016**, *42*, 439–447. [CrossRef]
2. Kumar, A.; Zhou, Y. Human identification using finger images. *IEEE Trans. Image Process.* **2012**, *21*, 2228–2244. [CrossRef] [PubMed]
3. Xie, S.; Lu, Y.; Yoon, S.; Yang, J.; Park, D. Intensity variation normalization for finger vein recognition using guided filter based singe scale retinex. *Sensors* **2015**, *15*, 17089–17105. [CrossRef] [PubMed]
4. 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]
5. Jain, T.; Kumar, R. A study of vein recognition system. *Acta Inform. Malays.* **2019**, *3*, 13–15. [CrossRef]
6. 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]
7. Song, W.; Kim, T.; Kim, H.; Choi, J.; Kong, H.; Lee, S. A finger-vein verification system using mean curvature. *Pattern Recognit. Lett.* **2011**, *32*, 1541–1547. [CrossRef]
8. Miura, N.; Nagasaka, A.; Miyatake, T. Extraction of finger-vein patterns using maximum curvature points in image profiles. *IEICE Trans. Inf. Syst.* **2007**, *90*, 1185–1194. [CrossRef]
9. Mulyono, D.; Jinn, H. A study of finger vein biometric for personal identification. In Proceedings of the International Symposium on Biometrics and Security Technologies, Isalambad, Pakistan, 23–24 April 2008; pp. 1–8.
10. Hashimoto, J. Finger vein authentication technology and its future. In Proceedings of the 2006 Symposium on VLSI Circuits, Honolulu, HI, USA, 15–17 June 2006; pp. 5–8.
11. 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 20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 23–26 August 2010; pp. 1269–1272.
12. Cheong, W.; Prahl, S.; Welch, A. A review of the optical properties of biological tissues. *IEEE J. Quantum Electron.* **1990**, *26*, 2166–2185. [CrossRef]
13. Lee, E.; Park, K. Image restoration of skin scattering and optical blurring for finger vein recognition. *Opt. Lasers Eng.* **2011**, *49*, 816–828. [CrossRef]
14. Yang, Y.; Shi, Y. Towards finger-vein image restoration and enhancement for finger-vein recognition. *Inf. Sci.* **2014**, *268*, 33–52. [CrossRef]
15. Wang, J.; Wang, G.; Li, M.; Yu, W.; Tian, H. An improved hand vein image acquisition method based on the proposed image quality evaluation system. *Comput. Model. New Technol.* **2014**, *18*, 1204–1208.
16. Wang, J.; Wang, G. Quality-Specific Hand Vein Recognition System. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2599–2610. [CrossRef]
17. Chen, L.; Wang, J.; Yang, S.; He, H. A Finger Vein Image-Based Personal Identification System With Self-Adaptive Illuminance Control. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 294–304. [CrossRef]
18. Abd Rahman, A.B.; Juhim, F.; Bade, A.; Chee, F.P. Effect of NIR LED Power in Enhancing the Vein Acquisition. In Proceedings of the 2021 IEEE 19th Student Conference on Research and Development (SCOReD), Kota Kinabalu, Malaysia, 23–25 November 2021; pp. 456–459.
19. Lee, E.; Park, K. Restoration method of skin scattering blurred vein image for finger vein recognition. *Electron. Lett.* **2009**, *45*, 1074–1076. [CrossRef]
20. Nguyen, D.; Park, Y.; Shin, K.; Park, K. New Finger-vein Recognition Method Based on Image Quality Assessment. *KSII Trans. Internet Inf. Syst.* **2013**, *7*, 347–365.
21. Yang, L.; Yang, G.; Yin, Y.; Xiao, R. Finger vein image quality evaluation using support vector machines. *Opt. Eng.* **2013**, *52*, 27003–27010. [CrossRef]
22. Nardelli, P.; Jimenez-Carretero, D.; Bermejo-Pelaez, D.; Washko, G.R.; Rahaghi, F.N.; Ledesma-Carbayo, M.J.; Estépar, R.S.J. Pulmonary artery–vein classification in CT images using deep learning. *IEEE Trans. Med. Imaging* **2018**, *37*, 2428–2440. [CrossRef]
23. Sachar, S.; Kumar, A. Survey of feature extraction and classification techniques to identify plant through leaves. *Expert Syst. Appl.* **2021**, *167*, 114181. [CrossRef]
24. Kovač, I.; Marák, P. Finger vein recognition: Utilization of adaptive gabor filters in the enhancement stage combined with SIFT/SURF-based feature extraction. *Signal Image Video Process.* **2022**, 1–7. [CrossRef]
25. Qin, H.; Qin, L.; Yu, C. Region growth–based feature extraction method for finger-vein recognition. *Opt. Eng.* **2013**, *50*, 57208–57213. [CrossRef]

26. Yang, J.; Yang, J. Combination of Gabor wavelets and circular Gabor filter for finger-vein extraction. In *Emerging Intelligent Computing Technology and Applications*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 346–354.
27. Yang, J.; Yang, J.; Shi, Y. Finger-vein segmentation based on multi-channel even-symmetric Gabor filters. In Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems, Shanghai, China, 20–22 November 2009; pp. 500–503.
28. Wang, G.; Sun, C.; Sowmya, A. Multi-weighted co-occurrence descriptor encoding for vein recognition. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 375–390. [CrossRef]
29. Wang, J.; Wang, G. SIFT Based Vein Recognition Models: Analysis and Improvement. *Comput. Math. Methods Med.* **2017**, *50*, 1–14. [CrossRef]
30. Mulyono, D.; Jinn, H. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
31. Christian, S.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
32. Hariharan, B.; Arbeláez, P.; Girshick, R.; Malik, J. Simultaneous detection and segmentation. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 297–312.
33. Hong, H.; Lee, M.; Park, K. Convolutional Neural Network-Based Finger-Vein Recognition Using NIR Image Sensors. *Sensors* **2017**, *17*, 1297. [CrossRef]
34. Radzi, S.; Hani, M.; Bakhteri, R. Finger-vein biometric identification using convolutional neural network. *Turk. J. Electr. Eng. Comput. Sci.* **2016**, *24*, 1863–1878. [CrossRef]
35. Qin, H.; El-Yacoubi, M. Deep Representation-Based Feature Extraction and Recovering for Finger-Vein Verification. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 1816–1829. [CrossRef]
36. Ren, H.; Sun, L.; Guo, J.; Han, C.; Wu, F. Finger vein recognition system with template protection based on convolutional neural network. *Knowl.-Based Syst.* **2021**, *227*, 107159. [CrossRef]
37. Hoang, T.; Do, T.; Tan, D.; Cheung, N. Selective Deep Convolutional Features for Image Retrieval. In Proceedings of the ACM International Conference on Multimedia, Mountain View, CA USA, 23–27 October 2017; pp. 1600–1608.
38. Jou, H.; Douze, M.; Schmid, C.; Pérez, P. Aggregating local descriptors into a compact image representation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 3304–3311.
39. Perronnin, F.; Sánchez, J.; Mensink, T. Improving the Fisher kernel for large-scale image classification. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 143–156.
40. Lowe, D. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **2004**, *60*, 91–110. [CrossRef]
41. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
42. Arandjelovi̇, R.; Zisserman, A. Three things everyone should know to improve object retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 2911–2918.
43. Morel, J.; Yu, G. ASIFT: A new framework for fully affine invariant image comparison. *SIAM J. Imaging Sci.* **2009**, *2*, 438–469. [CrossRef]
44. Ladoux, P.; Rosenberger, C.; Dorizzi, B. Palm vein verification system based on SIFT matching. In *Advances in Biometrics*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1290–1298.
45. Yuan, L.; Gen, L.; Chao, D.; Jin, H.; Xin, W. Fast recognition of hand vein with SURF descriptors. *Chin. J. Sci. Instrum.* **2011**, *32*, 831–836.
46. Wang, H.; Tao, L.; Hu, X. Novel algorithm for hand vein recognition based on retinex method and SIFT feature analysis. In *Electrical Power Systems and Computers*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 559–566.
47. Kim, H.; Lee, E.; Yoon, G.; Yang, S.; Lee, E.; Yoon, S. Illumination normalization for SIFT based finger vein authentication. In *International Symposium on Visual Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 21–30.
48. Kang, W.; Liu, Y.; Wu, Q.; Yue, X. Contact-free palm-vein recognition based on local invariant features. *PLoS ONE* **2014**, *9*, e97548. [CrossRef] [PubMed]
49. Lee, E.; Lee, H.; Park, K. Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction. *Int. J. Imaging Syst. Technol.* **2009**, *19*, 179–186. [CrossRef]
50. Kang, B.; Park, K.; Yoo, J.; Kim, J. Multimodal biometric method that combines veins, prints, and shape of a finger. *Opt. Eng.* **2011**, *50*, 201–209.
51. Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **2010**, *19*, 1635–1650.
52. Rosdi, B.; Shing, C.; Suandi, S. Finger vein recognition using local line binary pattern. *Sensors* **2010**, *11*, 11357–11371. [CrossRef]
53. Liu, F.; Yin, Y.; Yang, G.; Dong, L.; Xi, X. Finger vein recognition with superpixel-based features. In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), Clearwater, FL, USA, 29 September–2 October 2014; pp. 1–8.
54. Dong, L.; Yang, G.; Yin, Y.; Liu, F.; Xi, X. Finger vein verification based on a personalized best patches map. In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), Clearwater, FL, USA, 29 September–2 October 2014; pp. 9–16.

55. Su, K.; Yang, G.; Yang, L.; Li, D.; Su, P.; Yin, Y. Learning binary hash codes for finger vein image retrieval. *Pattern Recognit. Lett.* **2019**, *117*, 74–82. [CrossRef]
56. Xi, X.; Yang, G.; Yin, Y.; Yang, L. Finger vein recognition based on the hyperinformation feature. *Opt. Eng.* **2014**, *53*, 013108. [CrossRef]
57. Xi, X.; Yang, G.; Yin, Y.; Meng, X. Finger vein recognition with personalized feature selection. *Sensors* **2013**, *13*, 11243–11259. [CrossRef]
58. Turkoglu, M. COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble. *Appl. Intell.* **2021**, *51*, 1213–1226. [CrossRef]
59. Levenkova, A.; Sowmya, A.; Kalloniatis, M.; Ly, A.; Ho, A. Lesion detection in ultra-wide field retinal images for diabetic retinopathy diagnosis. In *Medical Imaging 2018: Computer-Aided Diagnosis*; SPIE: Bellingham, WA, USA, 2018; pp. 314–321.
60. Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 1794–1801.
61. Koniusz, P.; Yan, F.; Gosselin, P.; Mikolajczyk, K. Higher-order occurrence pooling on mid-and low-level features: Visual concept detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 313–326. [CrossRef]
62. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
63. 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 (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
64. Wang, J.; Wang, G. Hand-dorsa Vein Recognition with Structure Growing Guided CNN. *Opt.-Int. J. Light Electron Opt.* **2017**, *149*, 469–477. [CrossRef]
65. Zhang, H.; Liu, Z.; Zhao, Q.; Zhang, C.; Fan, D. Finger vein recognition based on Gabor filter. In Proceedings of the International Conference on Intelligent Science and Big Data Engineering, Nanjing, China, 18–20 October 2013; pp. 827–834.
66. 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]
67. Choi, J.; Song, W.; Kim, T.; Lee, S.; Kim, H. Finger vein extraction using gradient normalization and principal curvature. In *Image Processing: Machine Vision Applications II*; SPIE: Bellingham, WA, USA, 2009; pp. 827–834.
68. Kang, K. Vein pattern extraction based on vectorgrams of maximal intra-neighbor difference. *Pattern Recognit. Lett.* **2012**, *33*, 1916–1923. [CrossRef]
69. Zhang, Y.; Li, Q.; You, J.; Bhattacharya, P. Palm vein extraction and matching for personal authentication. *Adv. Vis. Inf. Syst.* **2007**, *4781*, 154–164.