photonics

MDPI

Article

Dual-Stream Enhanced Deep Network for Transmission Near-Infrared Dorsal Hand Vein Age Estimation with Attention Mechanisms

Zhenghua Shu, Zhihua Xie * and Xiaowei Zou

Key Lab of Advanced Electronic Materials, Jiangxi Science and Technology Normal University, Nanchang 330031, China; 1020101066@jxstnu.edu.cn (Z.S.); 1020101062@jxstnu.edu.cn (X.Z.)
* Correspondence: xie_zhihua68@aliyun.com; Tel.: +86-139-7001-8047

Abstract: Dorsal hand vein recognition, with unique stable and reliable advantages, has attracted considerable attention from numerous researchers. In this case, the dorsal hand vein images captured by the means of transmission infrared imaging are clearer than those collected by other infrared methods, enabling it to be more suitable for the biometric applications. However, less attention is paid to individual age estimation based on dorsal hand veins. To this end, this paper proposes an efficient dorsal hand vein age estimation model using a deep neural network with attention mechanisms. Specifically, a convolutional neural network (CNN) is developed to extract the expressive features for age estimation. Simultaneously, another deep residual network is leveraged to strengthen the representation ability on subtle dorsal vein textures. Moreover, variable activation functions and pooling layers are integrated into the respective streams to enhance the nonlinearity modeling of the dual-stream model. Finally, a dynamic attention mechanism module is embedded into the dual-stream network to achieve multi-modal collaborative enhancement, guiding the model to concentrate on salient age-specific features. To evaluate the performance of dorsal hand vein age estimation, this work collects dorsal hand vein images using the transmission near-infrared spectrum from 300 individuals across various age groups. The experimental results show that the dual-stream enhanced network with the attention mechanism significantly improves the accuracy of dorsal hand vein age estimation in comparison with other deep learning approaches, indicating the potential of using near-infrared dorsal hand vein imaging and deep learning technology for efficient human age estimation.

Keywords: transmission near-infrared; attention mechanism; dorsal hand vein; age estimation; the dual-stream enhanced network with attention mechanism (DSEANET)

Citation: Shu, Z.; Xie, Z.; Zou, X. Dual-Stream Enhanced Deep Network for Transmission Near-Infrared Dorsal Hand Vein Age Estimation with Attention Mechanisms. *Photonics* **2024**, *11*, 1113. https://doi.org/10.3390/photonics11121113

Received: 22 October 2024
Revised: 18 November 2024
Accepted: 23 November 2024
Published: 25 November 2024

1. Introduction

Automatic age estimation, which estimates individual ages based on facial images, is a highly regarded topic in pattern recognition, machine learning, and computer vision [1]. In 1994, Kwon et al. published early research on age classification theory and practical methods for facial images based on craniofacial features and skin wrinkle analysis [1]. By 2002, Lantis et al. [2] used learning-based age transformation features to explain the impact of age on facial appearance, indicating that reliable and accurate age estimation can be performed on facial images. Subsequently, with the establishment of numerous age recognition-related datasets, researchers have conducted more extensive and in-depth research on age estimation [3].

Early physiological age estimation methods used traditional machine learning algorithms and can be generally divided into two stages. The first stage involved manually extracting age-related features, and the second procedure employed the extracted features to optimize the age estimation model through various algorithms to accomplish classification and estimation tasks. With the advancement of deep learning technology, some

works have gradually applied deep learning to facial age estimation, achieving encouraging performance. Especially when the training samples are large enough, the performance of convolution neural network (CNN) models becomes more excellent. It can learn diverse and compact global features, thus strengthening the feature representation ability of the models. Therefore, it has been successfully applied to face analysis tasks, such as face detection, face recognition [4], face verification [5], and demographic estimation [6]. Following this principle, Niu et al. [7] developed an end-to-end facial age estimation model based on CNNs. Sendik et al. [8] proposed a facial age estimation approach in combination with CNN and SVR, which has achieved good results on small-scale datasets. Chen et al. [9] designed an unconstrained age estimation method based on a depth convolutional neural network (DCNN). Zhang et al. [10] designed a multi-level residual network as an age estimation model. The DCNN first pre-trains the model on the Image Net dataset and then optimizes it on the IMDB-WIKI-101 dataset. Tao et al. [11] came up with a new facial age estimation method based on the deep forest model. Liu H et al. [12] presented a label-sensitive deep metric learning method for facial age estimation. Taheri S et al. [13] combined the multi-layer features in a CNN with a series of age-related manual features to estimate facial age. However, the visible facial imaging usually varies with the ambient illumination and disguise, which makes facial age estimation extremely challenging.

Fortunately, physiological tissues are responsive to near-infrared light, which typically ranges between 700 and 1000 nm in wavelength. When the hand is exposed to near-infrared light of a specific wavelength, the light penetrates the epidermis and enters the subcutaneous tissue, where it scatters. During the scattering process, the near-infrared light is significantly absorbed by deoxyhemoglobin in the venous blood, resulting in a dark shadow at the location of the hand vein when captured by an image sensor, while regions without veins appear brighter. The absorption of near-infrared light by deoxyhemoglobin at the vein sites ensures that the veins appear as dark shadows in the captured images, whereas non-venous areas remain brighter [14]. This unique property of dorsal hand vein images contributes to their high stability and robust resistance against spoofing attempts.

The distribution of subcutaneous veins varies among individuals, yet this distribution remains relatively stable over time, making venous patterns reliable physiological characteristics for personal identity authentication [14]. Based on this finding, reflective near-infrared methods are commonly used to capture dorsal hand veins. Nevertheless, the original vein images captured by this method may be very blurry. To address this issue, Shu et al. designed a transmissive imaging system to more clearly capture the dorsal hand vein images. They further trained and applied deep residual networks with attention mechanisms for effective feature representation and recognition of dorsal hand vein images [15].

Despite the significant advancements in biometric research based on dorsal hand vein images, most scholars have concentrated on identity recognition, with relatively little attention given to individual age estimation. However, it is statistically significant that dorsal hand vein images reveal changes as individuals age. Specifically, younger individuals tend to have thinner dorsal hand veins, which become thicker and more pronounced with age. Meanwhile, as people age, their veins become more prominent and curved. More importantly, unlike the facial images, the dorsal hand veins remain relatively constant and cannot be easily disguised. In a word, near-infrared dorsal hand vein imagery presents a promising solution for age estimation.

Currently, the deep learning networks are widely applied in biometric fields and have shown good performance. For instance, He et al. [16] proposed a deep residual network construction method based on the residual learning algorithm. Zhu et al. introduced a convolutional attention module [17], which is used for facial recognition. Lee et al. [18] designed a lightweight multitasking convolution algorithm to reduce model size. Although this algorithm has good real-time performance, its accuracy is not satisfactory. Lin et al. [19] developed the method of data enhancement and improved loss function to solve the problem of face age and gender prediction under occlusion, but the network structure of

this algorithm is extremely complex. Wang et al. came up with a facial age estimation algorithm with a dual-stream enhanced shallow network under unrestricted conditions [20]. Given the strong learning capability of deep models, it is worthwhile to explore suitable models for dorsal hand vein age estimation. Currently, according to the public literature, most methods for dorsal hand vein recognition are CNNs or ResNets [21]. Specifically, Kuzu et al. explored the similarity of vein patterns in the finger vein, palm vein, and dorsal vein of a person's left and right hands based on deep learning-based feature [22]. Cimen et al. developed an advanced and complex fractal technique for extracting features from the images of hand vessel patterns [23].

Dorsal hand vein age estimation provides an alternative solution besides facial age estimation. The main motivation of our work is to explore an efficient deep learning model for dorsal hand vein age prediction equipped with a suitable vein image capturing system. Specifically, this paper proposes a dorsal hand age estimation system based on transmission near-infrared and the dual-stream enhancement network with diverse attention mechanisms, which can effectively extract discriminant age-aware features. The main contribution of this work indicates the potential of near-infrared dorsal hand vein imaging and deep learning techniques for efficient human age estimation.

2. Collection and Characteristics of Transmission Near-Infrared Dorsal Hand Vein Images

This study constructs a transmission-based capturing system for collecting near-infrared dorsal hand vein images, as shown in Figure 1 [15].

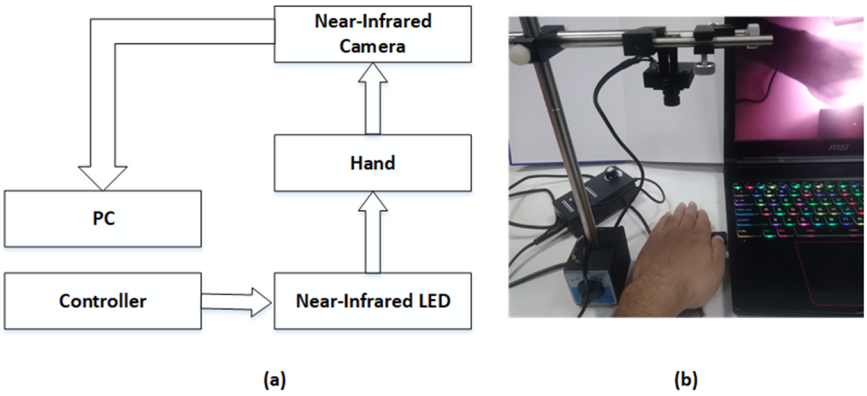

Figure 1. The schematic diagram and three-dimensional layout of near-infrared dorsal hand vein capturing system [15]. (**a**) the component diagram; and (**b**) the physical image.

Specifically, we designed a near-infrared dorsal hand vein capture system, which is composed of a light source component, a collection component, and a peripheral component. In the transmission near-infrared dorsal hand vein image collection system (Figure 1), an 850 nm LED with a 7 cm × 7 cm area light source was positioned beneath the palm. The wide motion camera (DAHUA, 3200_1080p), sensitive to wavelengths between 700 and 1100 nm, was arranged above the palms. This setup highlights the veins through near-infrared transmission, enabling the camera to capture clear dorsal hand vein images. Based on this configuration, a database of near-infrared dorsal hand vein images based on the transmission near-infrared spectrum was established for different ages.

Figure 2 provides partial near-infrared dorsal hand vein images based on transmittance across different age groups. The experiment primarily focused on adults, collecting images from 300 individuals. The age of the samples is roughly evenly distributed from 10 to 85 years old. In total, 1800 samples were obtained, with each participant providing six samples, thus forming a benchmark database of near-infrared hand vein images. As reported in reference [3], compared to the original near-infrared hand vein images obtained by reflection [15], the collected near-infrared hand vein images based on transmission are clearer. From Figure 2, it is evident that the veins on the dorsal hand change as people

grow and age. In younger individuals, the dorsal hand veins are thinner. As people mature, these veins become more pronounced. With further aging, the dorsal hand veins become even more prominent and curved. Additionally, aging leads to an increase in the number of wrinkles on the skin of the dorsal hand, as shown in Figure 2s,t,u. The veins tend to become more curved with age, as illustrated in Figure 2q–u.

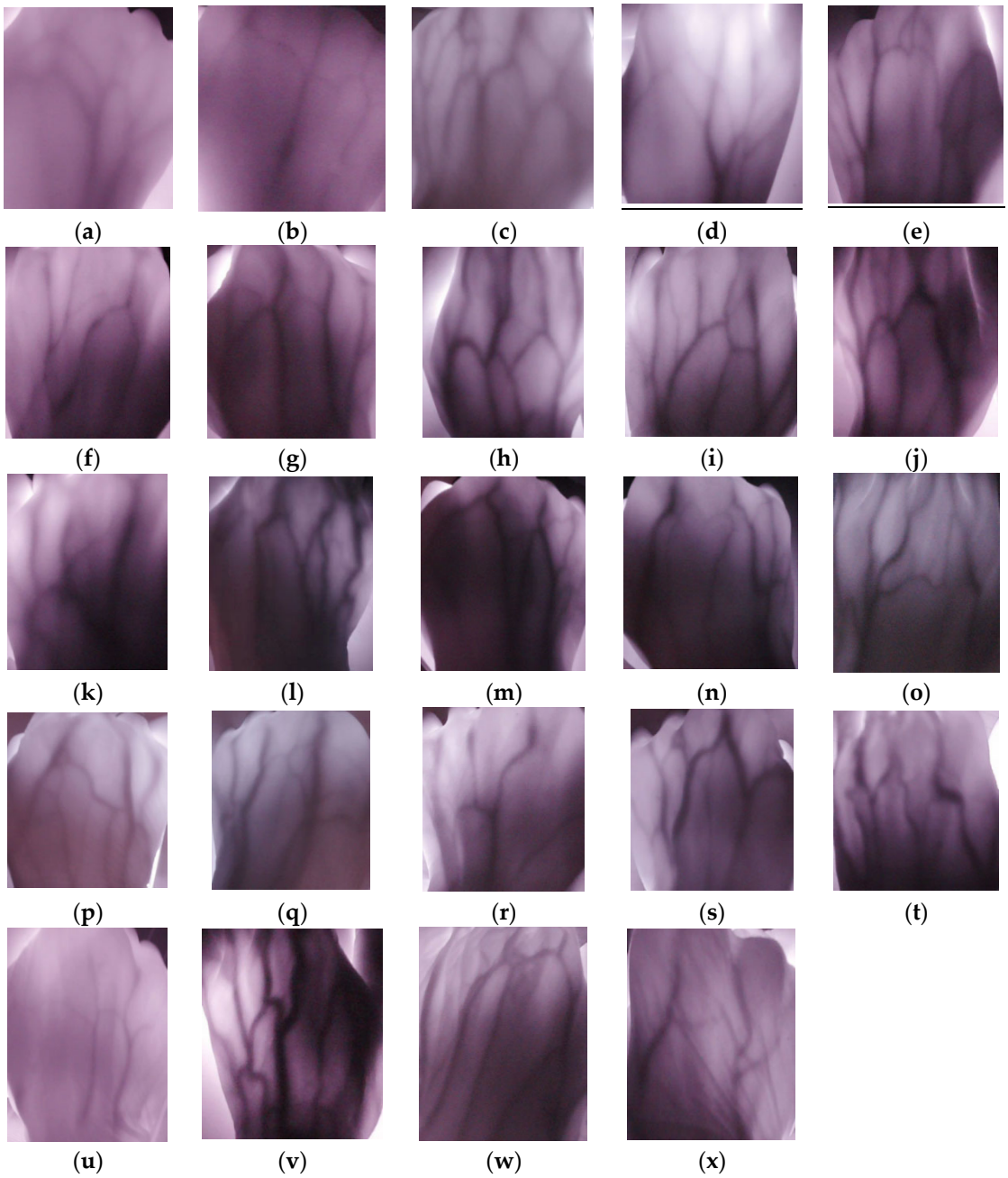

Figure 2. Transmission near-infrared dorsal hand vein at some ages: (**a**) 10 years old; (**b**) 15 years old; (**c**) 17 years old; (**d**) 20 years old; (**e**) 24 years old; (**f**) 26 years old; (**g**) 31 years old; (**h**) 35 years old; (**i**) 39 years old; (**j**) 41 years old; (**k**) 43 years old; (**l**) 48 years old; (**m**) 52 years old; (**n**) 54 years old; (**o**) 55 years old; (**p**) 56 years old; (**q**) 60 years old; (**r**) 62 years old; (**s**) 64 years old; (**t**) 66 years old; (**u**) 69 years old; (**v**) 72 years old; (**w**) 77 years old; and (**x**) 85 years old.

Figure 3 shows the Local Binary Pattern (LBP) image [24] and LBP histogram [24] of hand dorsal vein images for different ages. It can be seen that the LBP histogram of the dorsal hand veins varies with age. In younger individuals, the LBP histogram

distribution of the dorsal hand veins is relatively uniform. As individuals mature, the distribution of LBP histograms becomes more uneven. However, the LBP histogram distribution tends to return to a more uniform pattern in older individuals. Therefore, it can be inferred that the statistical distribution of texture features in different ages tends to have significant discrepancies.

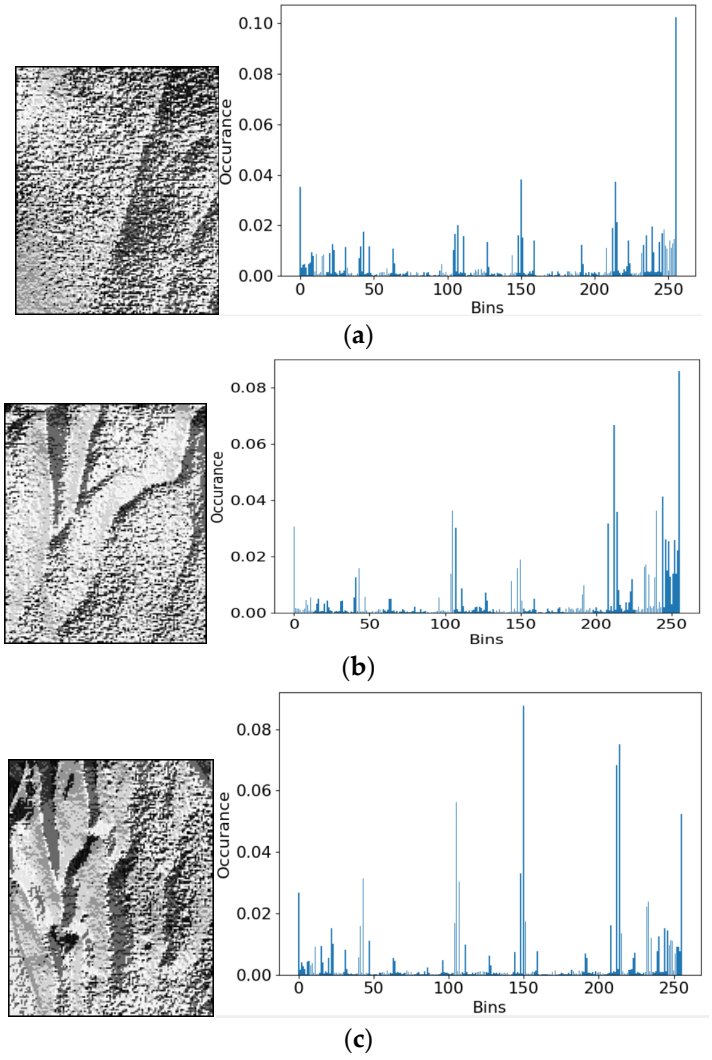

Figure 3. LBP and LBP histogram. (**a**) The LBP image (**left**) and LBP histogram (**right**) of the 15 year old; (**b**) the LBP image (**left**) and LBP histogram (**right**) of the 26 year old; and (**c**) the LBP image (**left**) and LBP histogram of the 72 year old.

3. The Proposed Approach

Deep learning networks can automatically extract high-dimensional features from near-infrared dorsal hand vein data. However, due to the extensive number of convolution and pooling operations, CNNs can also lose some important features of near-infrared dorsal hand veins, leading to an insufficient representation of the data features extracted by the network. To cope with this problem, this paper proposes a dual-stream enhancement network with an attention mechanism for dorsal hand vein age estimation. The network model comprises a CNN and a residual network. Additionally, an adaptive fusion module based on attention mechanisms is incorporated into the proposed model to achieve the dual enhancement effect. This integration operator enables the model to capture the subtle aging features of near-infrared dorsal hand veins more effectively.

3.1. The Dual-Stream Enhanced Network with Attention Mechanism

To address the issue of insufficient feature extraction ability in a single neural network model, a dual-stream enhancement network model with attention for estimating the dorsal hand veins age is proposed. The dual-stream enhancement network model is composed of a convolutional neural network embedded with an attention mechanism and a residual network with an attention mechanism. Each single network model independently extracts the features of the dorsal hand vein information and then separately predicts the dorsal hand vein age. For the first stream, the basic structure of the convolutional neural network with attention comprises a convolution module, an attention module, batch normalization (BN), a nonlinear hyperbolic tangent function (Tanh) activation function, and the maximum pooling layer. With regards to the second stream, the structure of the residual network with attention includes a residual network, ReLU activation function, average pooling, and the attention module.

The primary difference between the two single-stream networks lies in their activation functions and pooling strategies. Due to the good performance of the Tanh in distinguishing feature differences, making full use of the nonlinearity of the Tanh function is particularly beneficial for handling the dorsal hand vein characteristics with obvious differences in datasets, such as those between children and adults. In contrast, the residual network model with attention employs the ReLU activation function, which helps each single-stream network learn distinct features, thereby achieving complementary advantages. The detail framework of the dual-stream enhanced network model with attention mechanisms is illustrated in Figure 4. Initially, the output of the first fully connected layer in two single-stream shallow convolutional neural networks is concatenated to perform feature fusion. Then, it goes through two fully connected layers containing 512 neurons. In the end, the SoftMax classifier from the fused stream is used to make the final age prediction, which outputs the probability of belonging to a certain person in the age stage label.

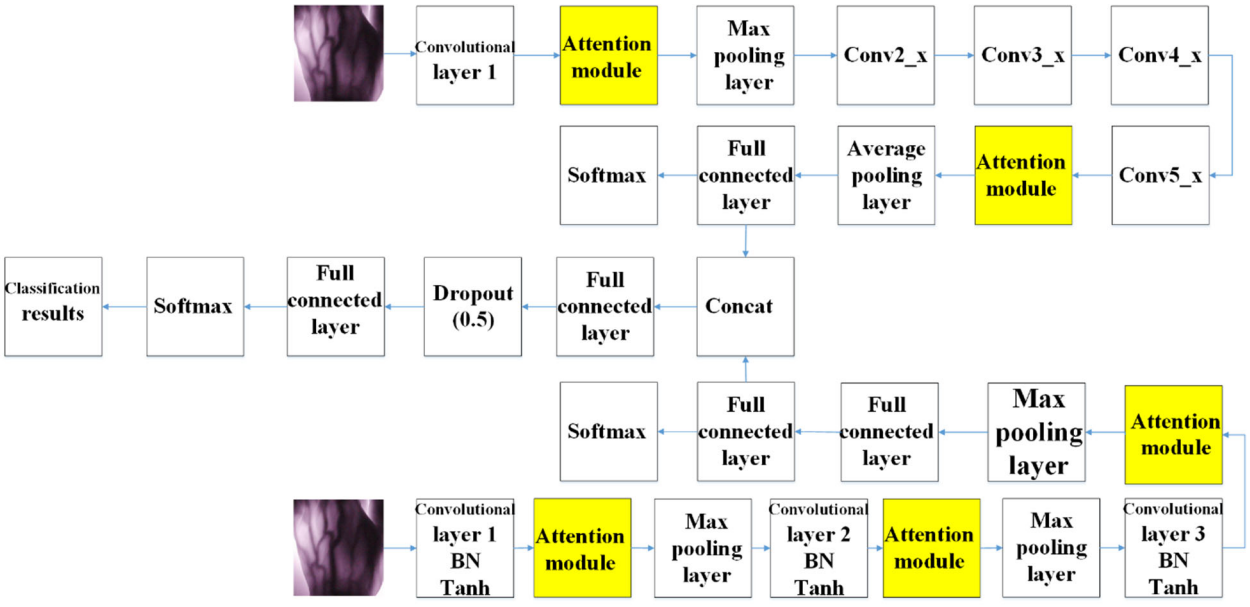

Figure 4. The framework of the dual-stream enhanced network with an attention mechanism (DSEANET).

3.2. Dual Enhanced Convolutional Attention Module

The attention mechanism focuses on the relative relationship between the target area and the relationship of key features in the image, making it very suitable for images containing structured objects, such as the dorsal hand veins. Based on the above characteristics and to deal with the problem of feature information loss during extraction, the convolutional attention module (CAM) proposed by Zhu et al. [17] was introduced to the dual-stream en-

hanced network, forming a dual enhancement effect. This integration enables the model to capture subtle aging features of the dorsal hand vein. The CAM module includes two parts, namely, the channel attention module and the spatial attention module. Considering the given feature map, attention weights are derived along the two dimensions of channel and space. These weights are then multiplied with the original feature map to dynamically adjust the balance between useful and irrelevant features. This process enables the network to learn global features and enhances the model's generalization ability. The main intention of the channel attention module is to highlight useful hand vein features while suppressing irrelevant information. In contrast, the spatial attention module focuses on identifying significant features of the dorsal hand veins and tracking changes in the hand vein contours.

3.3. Dorsal Hand Vein Age Estimation Based on the Enhanced Dual-Stream Network with Attention Mechanism

The DSEANET incorporates an attention mechanism to accelerate the network's feature extraction capability via the feature fusion module. The DSEANET employs diverse pooling and activation functions to address the challenge of easily losing certain features. To further facilitate the network in learning global features and balancing the confidence level of feature maps across various spatial and channel dimensions, we integrate the CAM (Class Activation Map) module into the dual-stream enhanced network model. This integration confirms the advantageousness of the DSEANET in furthering the identification of hand vein images in age estimation models. It enhances the model's nonlinearity and feature recalibration, thereby enhancing the network's classification ability and robustness.

A perfect neural network model not only excels on the training dataset but also exhibits robust performance on the testing set. To minimize errors on the test set, meet real-time requirements, and enhance the overall generalization ability, regularization of the network model is imperative. Unlike single-stream network models, the dual-stream enhanced network model expands network parameters due to an increased network width, thus elevating the risk of overfitting, particularly with small-scale data. To capitalize on the advantages of the dual-stream enhanced network model while mitigating potential drawbacks, it is crucial to employ suitable strategies to prevent overfitting. Given the basic network model's three-layer architecture, the fully connected layer of the DSEANET utilizes the Dropout (0.5) method to counter overfitting.

When estimating the age of an unknown dorsal hand vein image, the process involves several steps. Initially, a single-stream CNN network and a residual network are employed to simultaneously extract features from the dorsal hand veins. Subsequently, fundamental aging characteristics are derived through a sequence of convolutional, pooling, and batch normalization operations. Following this, a convolutional attention module is utilized to further extract and filter aging features from the dorsal hand veins, considering both channel and spatial dimensions. This ensures that pertinent features receive greater emphasis while disregarding irrelevant or less significant ones. Finally, the extracted features are flattened and fed into two pathways: in the first pathway, they are inputted into the last two fully connected layers of each single-stream network for age prediction, while in the second pathway, they are fused with features extracted from the two single-stream networks and then inputted into two fully connected layers for age estimation.

4. Experiments and Results

4.1. Experimental Platform and Dataset

The experimental setup utilized an Ubuntu 19.10 running on the TensorFlow framework, implemented in Python. The hardware configuration facilitated the efficient execution of the deep learning tasks, particularly the computationally intensive operations involved in training and inference processes.

The experiment utilized a database comprising dorsal hand vein images captured by a near-infrared transmission acquisition instrument. This database contains dorsal hand vein images from 300 individuals, with 3 images of the left hand and 3 images of the right

hand captured for each person, resulting in a total of 1800 images. The age of the samples is roughly evenly distributed from 10 to 80 years old: 41 people for 10~20 years old, 42 people for 21~30 years old, 43 people for 30~40 years old, 45 people for 40~50 years old, 42 people for 50~60 years old, 45 people for 60~70 years old, and 42 people for 70+ years old. An example of the dorsal hand vein images can be seen in Figure 2. For the experiment, 800 dorsal hand vein images of the left or right hands from 200 individuals were allocated for the training set, while the remaining 1000 hand vein images captured using near infrared were designated for the test set.

All the input images were uniformly processed and resized to 227 × 227 before being fed into the network model. During the training process, each iteration consisted of 800 images. To further optimize the network model, the random gradient descent algorithm, cross-entropy loss function, and Momentum optimizer were employed for training. The initial learning rate was set to 0.01, with a maximum of 2000 iterations. The momentum was fixed at 0.9. The decay factor was set to 0.95, the decay steps to 80, and a decay coefficient of 0.1 was applied. The batch size during training was set to 80. After 200 epochs of training iterations, a network model for classification was obtained. The training process of the DSEANET, the deep residual network with attention mechanism (DRNAM), and ResNet network models is illustrated in Figure 5.

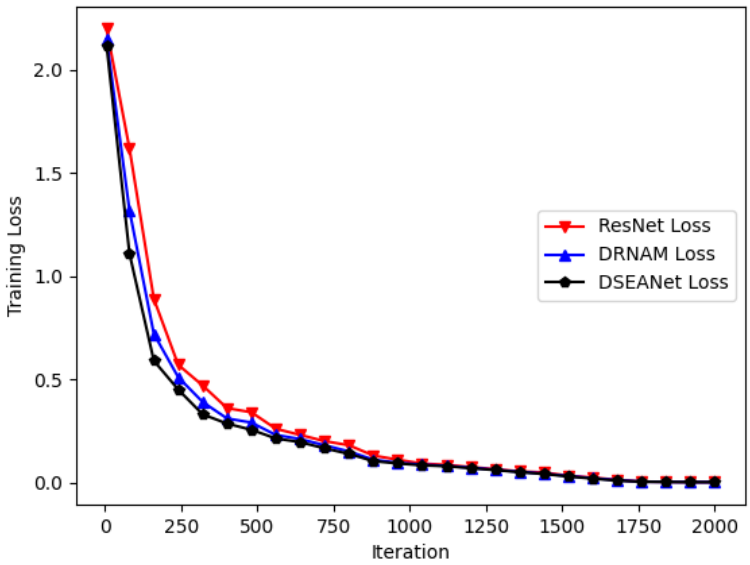

Figure 5. The training loss curves of different models.

This visualization provides insights into the progression of the training process and the convergence of the models over iterations. Figure 5 clearly illustrates that during the initial training period, although both the DSEANET and ResNet exhibit similar convergence trends, the DSEANET method demonstrates a favorable advantage in achieving lower loss values. Specifically, the loss value of the DSEANET method after training is notably smaller compared to both ResNet and the DRNAM. This observation underscores the effectiveness and rationality of the DSEANET algorithm in optimizing the model and enhancing its performance. By prioritizing smaller loss values early in the training process, the DSEANET showcases its capability to efficiently learn and extract relevant features from the input data, thus contributing to improved model convergence and overall performance.

Figure 6 reveals significant fluctuations in the recognition rate of the ResNet algorithm with changes in iterations. Conversely, the DSEANET algorithm consistently demonstrates superior estimation performance on the dataset of dorsal hand vein images captured using transmission near infrared. Particularly noteworthy is that during the early training iterations, the estimation performance of the DSEANET algorithm surpasses that of ResNet. Furthermore, after stabilizing around 200 training iterations, the recognition rate of the

DSEANET remains consistently higher than that of ResNet as the training period progresses. This trend indicates the robustness and effectiveness of the DSEANET algorithm in achieving superior recognition performance on dorsal hand vein images, especially in comparison to ResNet, across various stages of the training process.

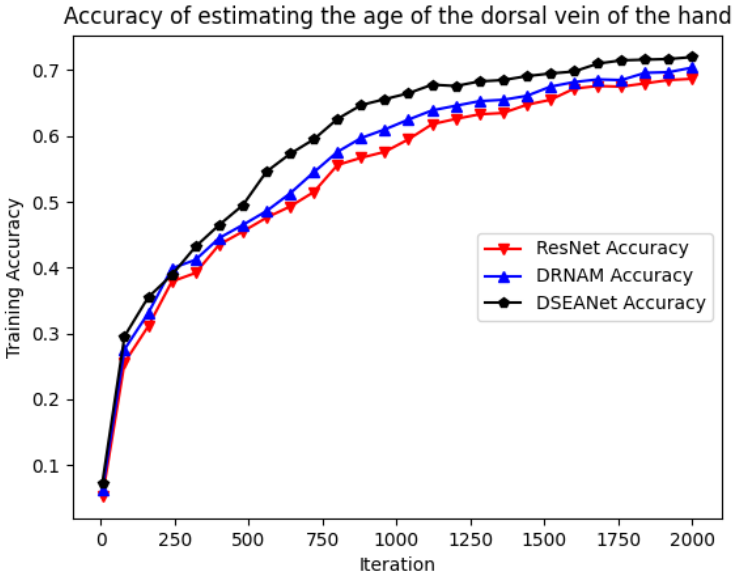

Figure 6. Training accuracy of different models.

4.2. Experimental Results and Analysis

The experiments utilized a database comprising near-infrared dorsal hand vein images captured using the transmission near-infrared spectrum to investigate and compare the age estimation capabilities of different algorithms.

To evaluate the effectiveness of the DSEANET model in estimating the age of individuals across different age groups, samples ranging from 10 to 70 years old were segmented into intervals of 10 years each. Consequently, the final dataset of near-infrared dorsal hand vein images was divided into 7 age groups: 10–20, 21–30, 31–40, 41–50, 51–60, 61–70, and 70+. The estimated results obtained from the DSEANET model, ResNet, and the deep residual network with attention mechanism (DRNAM) model for the seven age groups in the dorsal hand vein images based on the reflection-type near-infrared spectrum are presented in Table 1.

Table 1. Accuracy (%) of estimating the age of the dorsal vein of the hand on different CNNs.

Age Groups	Neural Network Models		
	DSEANET	**DRNAM**	**ResNet**
10~20 years old	83	81	79
21~30 years old	70	68	66
30~40 years old	65	63	61
40~50 years old	63	61	59
50~60 years old	72	70	69
60~70 years old	71	69	68
70+ years old	82	81	79
Average accuracy	72.0	70.4	68.7

From Table 1, it is obvious that the age estimation accuracy of the DSEANET is higher than that of the other two deep networks on the near-infrared dorsal hand vein

image database. The integration of the attention mechanism empowers the DSEANET to effectively emphasize salient age characteristics within near-infrared dorsal hand vein images, thus boosting the accuracy of age estimation. Upon further comparison of the results in Table 1, it is observed that the overall accuracy of age estimation for middle-aged individuals in the dorsal hand vein dataset based on reflection-type near infrared has decreased. This decline is attributed to the gradual and subtle changes in age characteristics of dorsal hand veins throughout adulthood (21–50 years old). The minimal variations in age characteristics among different adult age groups hinder effective feature learning and subsequently impact the accuracy of age estimation.

To verify the effectiveness of the DSEANET algorithm, it was compared with other age estimation algorithms on the near-infrared hand vein dataset. The precision and recall metrics for each algorithm are presented in Table 2.

Table 2. Age classification results of hand vein dataset.

Neural Network Model	Precision (%)	Recall (%)
ResNet	59.5	82.4
DRNAM	61.2	83.6
DSEANET	62.7	77.3

As shown in Table 2, the precision of the DSEANET is better than that of ResNet and the DRNAM, indicating a higher accuracy in correctly identifying age groups. However, the recall of the DSEANET is smaller compared to ResNet and the DRNAM, implying that it may miss some instances of certain age groups. Despite this, the overall performance of the dorsal hand vein age estimation method based on transmission near-infrared, leveraging the DSEANET, remains notably superior to other near-infrared hand vein age estimation algorithms.

The combined insights from Tables 1 and 2 provide compelling evidence of the effectiveness of the DSEANET in estimating the age of dorsal hand veins. While precision reflects the model's ability to make accurate predictions, recall highlights its capability to capture instances of each age group effectively. Although there may be slight trade-offs between precision and recall, the superior performance of the DSEAET confirms its suitability and reliability for age estimation tasks in dorsal hand vein images.

Moreover, the Mean Absolute Error (MAE) and Deviation Coefficient (ε_{error}) are two important measurement criteria to further evaluate the performance of facial age estimation [25–27]. Here, we also use the two metrics, MAE and ε_{error} [28], to present the superiority of the proposed dorsal hand vein age estimation model.

Firstly, we can calculate the difference between the predicted age and the actual age obtained by estimating the age of the dorsal hand vein in the test sample image. Then, the absolute averages of the differences between all the samples are processed to obtain the MAE. The computing formula is shown in Equation (1).

$$MAE = (1/\hat{N})\sum_{i=1}^{\hat{N}}|a_i - \hat{a}_i| \tag{1}$$

where $\hat{N}$ represents the total number of samples to be tested; a_i and $\hat{a}$ are the label age and predicted age of the i-th sample. The smaller the MAE value, the better the performance of the algorithm and the higher the accuracy of age estimation.

The specific formula for ε_{error} is shown in Equation (2) [28].

$$\varepsilon_{error} = 1 - (1/N)\sum_{i=1}^{N} e^{\frac{(\hat{b}_i - b_i)^2}{2\hat{\sigma}_i^2}} \tag{2}$$

where $\hat{\delta}_i$ is the standard deviation of the predicted age, $\hat{b}_i$ is the average predicted age, and b_i is the actual age. ε_{error} ranges from 0 (best) to 1 (worst); the closer it is to 0, the more accurate the estimation is.

Considering the network models, the pre-trained ResNet, DRNAM, and DSEANET are selected for further experimental comparison, with performance indicators of the MAE and ε_{error}, respectively. The experimental results on the dorsal hand vein dataset are shown in Table 3.

Table 3. Comparison of accuracy results of hand vein dataset. Bold fonts indicate the best results.

Neural Network Models	MAE (Year)	ε_{error}
ResNet	5.637	0.437
DRNAM	5.271	0.418
DSEANET	5.138	0.395

Compared with other models, the DSEANET achieved better experimental results on the near-infrared dorsal hand vein dataset. The main reason for the outstanding performance of the DSEANET is that this model adopts two heterogeneous single-stream neural networks to improve the learning ability. This is because the convolutional attention module (CAM) introduced by the DSEANET can refine shallow features into high-level features through attention modules. The CAM also enhances the network's ability to extract high-level features and improves the differentiation and expression of critical age information. At the same time, the combination of the residual structure and improved attention module can better express the correlation between different channel features, thereby enhancing the expressive power of refined attention features. On the other hand, there will be no network degradation due to network deepening, thus achieving better recognition performance. The serial structure of the attention module and convolutional layer first enhances the representation of key features in the feature map and then extracts features more in line with the human visual mechanism. The DSEANET has shown significant improvements compared to ResNet and the DRNAM in all aspects, proving the feasibility of the DSEANET.

5. Conclusions

To enhance the generalization ability of dorsal hand vein age estimation, this paper proposes a novel dual-stream enhanced network that integrates attention mechanism modules. This model, built upon the foundation of CNNs, comprises two streams: a convolutional neural network and a residual network, both augmented with attention mechanisms. By leveraging attention mechanisms, the dual-stream enhanced network model enhances the learning capacity of the network. To address the issue of feature information loss during feature extraction, the two constituent single-stream networks employ distinct activation and pooling functions. The experimental results demonstrate that the dual-stream enhanced network model with attention mechanisms achieves notable success in estimating age based on transmission near-infrared dorsal hand vein images. However, samples of our hand vein database are relatedly limited for deep neural networks and there is still room for good generalization on the gender and ethnicity factors. In the future, more training and testing samples can be captured to facilitate training a deeper model with the self-attention mechanism, which can further enhance the performance of age estimation. Meanwhile, the differences in dorsal hand vein patterns between males and females will be explored for the robust age estimation system.

Author Contributions: Conceptualization and methodology, Z.S.; software, Z.X.; validation, X.Z.; formal analysis, X.Z.; investigation and resources, Z.X.; data curation, X.Z.; writing—original draft preparation, Z.S.; writing—review and editing, Z.X.; visualization, Z.S.; supervision, Z.X.; project

administration, Z.X.; funding acquisition, Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Nature Science Foundation of China (No. 62362037 and No. 12264018), the Natural Science Foundation of Jiangxi Province of China (No. 20224ACB202011), and the Jiangxi Province Graduate Innovation Special Fund Project (No. YC2022-s790).

Institutional Review Board Statement: Our studies were conducted under IRB of Jiangxi Science and Technology Normal University approval, and the following is the information. Approval Code: IRB-JXSTNU-2023007, Approval Date: 2 April 2023.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. All other public datasets used are available and cited in the references.

Acknowledgments: We would like to thank the editor and the reviewers for their valuable comments.

Conflicts of Interest: On behalf of all the authors, the corresponding author states that there are no conflicts of interest.

References

1. Kwon, Y.H.; Lobo, D.V. Age classification from facial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 762–767.
2. Lanitis, A.; Taylor, C.J.; Cootes, T.F. Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 442–455. [CrossRef]
3. Panis, G.; Lanitis, A.; Tsapatsoulis, N.; Cootes, T.F. Overview of research on facial ageing using the FG-NET ageing database. *IET Biom.* **2016**, *5*, 37–46. [CrossRef]
4. Yi, S.; Wang, X.; Tang, X. Deep Convolutional Network Cascade for Facial Point Detection. In Proceedings of the IEEE Internation Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3476–3483.
5. Taigman, Y.; Ming, Y.; Ranzato, M.; Wolf, L. Deep Face: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
6. Ming, Y.; Zhu, S.; Lv, F.; Yu, K. Correspondence driven adaptation for human profile recognition. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 505–512.
7. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal Regression with Multiple Output CNN for Age Estimation. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4920–4928.
8. Sendik, O.; Keller, Y. Deep Age: Deep Learning of face-based age estimation. *Signal Process. Image Commun.* **2019**, *78*, 368–375.
9. Chen, J.C.; Kumar, A.; Ranjan, R.; Patel, V.M.; Alavi, A.; Chellappa, R. A cascaded convolutional neural network for age estimation of unconstrained faces. In Proceedings of the International Conference on Biometrics Theory, Applications and Systems, Buffalo, NY, USA, 6–9 September 2016; pp. 20–28.
10. Zhang, K.; Gao, C.; Guo, L.; Sun, M.; Yuan, X.; Han, T.X.; Zhao, Z.; Li, B. Age Group and Gender Estimation in the Wild with Deep Ro R Architecture. *IEEE Access* **2017**, *5*, 22492–22503. [CrossRef]
11. Tao, W.; Zhao, Y.; Liu, L.; Li, H.; Xu, W.; Chen, C. A novel hierarchical regression approach for human facial age estimation based on deep forest. In Proceedings of the International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, China, 27–29 March 2018; pp. 1512–1524.
12. Liu, H.; Lu, J.; Feng, J.; Zhou, J. Label-sensitive deep metric learning for facial age estimation. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 292–305. [CrossRef]
13. Taheri, S.; Toygar, O. Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images. *IET Biom.* **2019**, *8*, 24–133. [CrossRef]
14. Li, K.; Zhang, G.; Wang, Y.; Wang, P.; Ni, C. Hand-dorsa vein recognition based on improved partition local binary patterns. In Proceedings of the 10th Chinese Conference (CCBR 2015), Tianjin, China, 13–15 November 2015; pp. 312–320.
15. Zhenghua, S.; Zhihua, X. Dorsal hand vein recognition based on transmission-type near infrared imaging and deep residual network with attention mechanism. *Opt. Rev.* **2022**, *29*, 335–342.
16. 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, Washington, DC, USA, 27–30 June 2016; pp. 770–778.
17. Zhu, Y.; Ni, X.; Yao, Y. Face Recognition Based on SVDF and Deep Learning with Attention. *J. Chin. Comput. Syst.* **2020**, *41*, 1763–1767.
18. Lee, J.H.; Chan, Y.M.; Chen, T.Y.; Chen, C.S. Joint estimation of age and gender from unconstrained face images using lightweight multi–task cnn for mobile applications. In Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 10–12 April 2018; pp. 162–165.

19. Lin, L.E.; Lin, C.H. Data augmentation with occluded facial features for age and gender estimation. *IET Biom.* **2021**, *10*, 640–653. [CrossRef]
20. Wang, X.; Yili Hamu, Y. Face age estimation with dual-stream enhanced shallow network under unrestricted conditions. *Laser J.* **2022**, *43*, 88–94.
21. Li, J.; Li, K.; Zhang, G.; Wang, J.; Li, K.; Yang, Y. Recognition of Dorsal Hand Vein in Small-Scale Sample Database Based on Fusion of ResNet and HOG Feature. *Electronics* **2022**, *11*, 2698. [CrossRef]
22. Kuzu, R.S.; Maiorana, E.; Campisi, P. On the intra-subject similarity of hand vein patterns in biometric recognition. *Expert Syst. Appl.* **2022**, *192*, 116305. [CrossRef]
23. Cimen, M.E.; Boyraz, O.F.; Yildiz, M.Z.; Boz, A.F. A new dorsal hand vein authentication system based on fractal dimension box counting method. *Optik* **2021**, *226*, 165438. [CrossRef]
24. Huang, X.; Zhao, G.; Hong, X.; Zheng, W.; Pietikäinen, M. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing* **2016**, *175*, 564–578. [CrossRef]
25. Hanaoka, K.; Ngan, M.; Yang, J.; Quinn, G.W.; Hom, A.; Grother, P. *Face Analysis Technology Evaluation: Age Estimation and Verification*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2024. [CrossRef]
26. Yeom, H.-G.; Lee, B.-D.; Lee, W.; Lee, T.; Yun, J.P. Estimating chronological age through learning local and global features of panoramic radiographs in the Korean population. *Sci. Rep.* **2023**, *13*, 21857. [CrossRef] [PubMed]
27. Bekhouche, S.E.; Benlamoudi, A.; Dornaika, F.; Telli, H.; Bounab, Y. Facial Age Estimation Using Multi-Stage Deep Neural Networks. *Electronics* **2024**, *13*, 3259. [CrossRef]
28. Rasmus, R.; Radu, T.; Gool, L.V. DEX: Deep expectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision Workshop, Boston, MA, USA, 7–12 June 2015.