Machine Intelligence Research
www.mi-research.net

20(5), October 2023, 707-715
DOI: 10.1007/s11633-022-1366-8

Dual-domain and Multiscale Fusion Deep Neural Network for PPG Biometric Recognition

Chun-Ying Liu[1] Gong-Ping Yang[2] Yu-Wen Huang[1] Fu-Xian Huang[1]

[1]School of computer, Heze University, Heze 274015, China

[2]School of Software, Shandong University, Jinan 250101, China

Abstract: Photoplethysmography (PPG) biometrics have received considerable attention. Although deep learning has achieved good performance for PPG biometrics, several challenges remain open: 1) How to effectively extract the feature fusion representation from time and frequency PPG signals. 2) How to effectively capture a series of PPG signal transition information. 3) How to extract time-varying information from one-dimensional time-frequency sequential data. To address these challenges, we propose a dual-domain and multiscale fusion deep neural network (DMFDNN) for PPG biometric recognition. The DMFDNN is mainly composed of a two-branch deep learning framework for PPG biometrics, which can learn the time-varying and multiscale discriminative features from the time and frequency domains. Meanwhile, we design a multiscale extraction module to capture transition information, which consists of multiple convolution layers with different receptive fields for capturing multiscale transition information. In addition, the dual-domain attention module is proposed to strengthen the domain of greater contributions from time-domain and frequency-domain data for PPG biometrics. Experiments on the four datasets demonstrate that DMFDNN outperforms the state-of-the-art methods for PPG biometrics.

Keywords: Photoplethysmography (PPG) signal , biometric recognition, multiple scale, deep neural network, dual-domain attention.

Citation: C. Y. Liu, G. P. Yang, Y. W. Huang, F. X. Huang. Dual-domain and multiscale fusion deep neural network for PPG biometric recognition. *Machine Intelligence Research*, vol.20, no.5, pp.707–715, 2023. http://doi.org/10.1007/s11633-022-1366-8

1 Introduction

Over the past decade, biometrics using some physiological signals, such as electrocardiogram (ECG), photoplethysmography (PPG), and electromyograms (EMG), have gradually caused considerable concern. Compared with widely implemented biometric traits, such as face and fingerprint, PPG signals as biometric traits, provide the following advantages[1, 2]: 1) PPG signals are acquired by attaching sensors to living persons, which can provide proof of liveness detection. 2) It is difficult to counterfeit or replicate, and PPG signals as biometric traits are highly secure. 3) PPG signals include information not only personal identity verification but also heart health and psychological states. 4) PPG signals are one-dimensional data with small size.

PPG biometrics have attracted increasing attention from researchers because of their unique advantages. PPG biometrics are a new technology, and many methods for PPG biometrics have been proposed[2, 3]. Deep learning (DL) has achieved good performance in PPG biometrics, and many related methods have been reported[1, 4–9]. However, several challenges remain open:

1) The existing deep learning methods for PPG biometrics use small-scale convolution filters to extract the amplitude features and ignore explicitly capturing a series of transition information of PPG waves. For example, the transition information of systolic and diastolic waves and dicrotic notches is crucial for recognizing PPG signals. Small-scale convolution filers can extract local amplitude information, and large-scale convolution filters provide the transition information of different morphological features. All this information is crucial for PPG biometric recognition. Therefore, a challenging problem using deep learning for PPG biometrics is how to design a model to extract the different scale transition information from PPG signals.

2) Most of the existing deep learning methods for PPG biometrics extract features from raw PPG signals. Raw PPG signals are easily influenced by acquisition equipment, body position, and various physical and psychological factors and are unstable with time variation. PPG signals contain time-frequency domain features, and some methods convert the raw PPG signals into frequency domain features as input to deep learning models for PPG biometric recognition. However, the existing deep learning methods for PPG biometric recognition only consider utilizing single-domain features, without using both the time and frequency domain fusion as input

Research Article
Manuscript received on April 26, 2022; accepted on August 8, 2022; published online on January 11, 2023
Recommended by Guest Editor Zhe-Nan Sun
Colored figures are available in the online version at https://link.springer.com/journal/11633
© Institute of Automation, Chinese Academy of Sciences and Springer-Verlag GmbH Germany, part of Springer Nature 2023

to deep learning models. Hence, it is a challenging problem to utilize the fusion of both the time and frequency information as input to deep learning models.

3) Most of existing works for PPG biometrics do not utilize the sequential relationship from one-dimensional time-frequency data. Most of existing deep learning methods for PPG biometrics only use a convolutional neural network (CNN) to extract the deep PPG features. CNN has achieved better results in higher-dimensional image data, without considering one-dimensional sequential data. Some deep learning methods only use a long short-term memory (LSTM) network to extract time-varying information from raw PPG signals. Therefore, a challenging problem for PPG biometrics is designing a robust deep model for extracting time-varying information from one-dimensional time-frequency sequential data.

To address the aforementioned challenges, in this paper, we propose a dual-domain and multiscale fusion deep neural network (DMFDNN) for PPG biometric recognition. The DMFDNN consists of a two-branch deep learning framework, which can learn the time-varying and multiscale discriminative features from the time and frequency domains. Meanwhile, a multiscale extraction module is designed for capturing transition information, which is composed of multiple convolution layers with different receptive fields. Moreover, the dual-domain attention module is proposed to strengthen the domain of greater contributions from time-domain and frequency-domain data for PPG biometrics. An overview architecture of the proposed DMFDNN is shown in Fig. 1.

The main innovations of this work are the following:

1) We develop a novel two-branch deep learning framework for PPG biometrics that can learn the time-varying and multiscale discriminative features from the time and frequency domains.

2) To capture a series of transition information of PPG waves, we design a multiscale extraction module, which consists of multiple convolution layers with different receptive fields.

3) We propose a dual-domain attention module to strengthen the domain of greater contribution for PPG biometrics.

The rest of our work is organized as follows. Related work is presented in Section 2. Data preprocessing is given in Section 3. The proposed methodology is introduced in Section 4, and the experimental results and analysis are reported in Section 5. A brief conclusion and future work follow in Section 6.

2 Related work

There are fiducial and nonfiducial approaches for PPG biometrics methods[3]. Fiducial-based approaches use the peaks, downwards and upwards slopes, and the interval of extreme points as fiducial points for PPG biometrics. Yao et al.[10] extracted derivatives of PPG signals as fiducial points for PPG signal biometrics. Mancilla-Palestina et al.[2] proposed a fusion learning method by extracting fiducial points of ECG and PPG signals for biometrics of embedded system. Chakraborty and Pal[11] used 12 features of PPG signals and their derivatives for PPG biometric recognition, and achieved 100% accuracy for 15 subjects. Nadzri et al.[12] extracted the systolic peaks, diastolic peaks, and dicrotic notches of PPG signals for biometric recognition. Sancho et al.[13] extracted the fiducial points of the time domain and Karhunen-Loève transform for PPG biometric recognition. The fiducial points of the PPG waves are sensitive to various noise, so the performance of fiducial-based methods for PPG biometrics is unsatisfactory.

Nonfiducial-based approaches obtain the feature transformation of PPG signals for biometrics. Spachos et al.[14] extracted the features of PPG signals by linear discriminant analysis (LDA) and identified the subjects to use the nearest neighbor classifier. Karimian et al.[15] proposed the discrete wavelet transform (DWT) features of PPG signals for biometrics. Yadav et al.[16] fused the features of continuous wavelet transform (CWT) and direct linear discriminant analysis (DLDA) for PPG biometric recognition, and obtained an equal error rate (EER) of 0.5%–6%. Farago et al.[17] used the cross-correlation of ECG, PPG and EMG signals for biometrics. Lee et al.[18] used random forest to extract the features of wearable PPG signals for biometrics. Although nonfiducial-based approaches for PPG biometric recognition achieve better performance, the PPG signals contain redundant noise and intraclass variety information, which influence the performance of PPG biometrics.

Recently, many deep-learning-based methods have been proposed for PPG biometrics. Hwang et al.[1] developed a fusion model of CNN and LSTM for PPG bio-

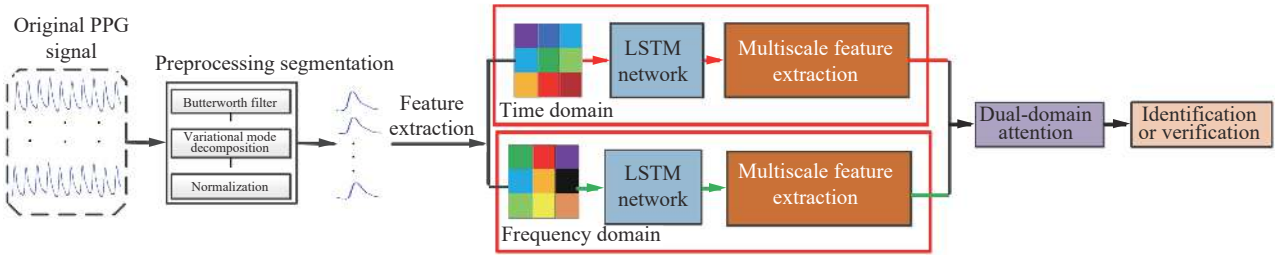


Fig. 1 An overview architecture of DMFDNN

metric recognition. Luque et al.[4] proposed convolutional networks with end-to-end architecture for PPG biometric authentication. Everson et al.[5] proposed deep-learning framework to employ two CNN layers in conjunction with two LSTM layers for PPG biometric recognition. Biswas et al.[6] proposed a deep-learning model for biometrics to use wrist-worn PPG signals for remote cardiovascular monitoring, which consisted of a four-layer deep neural network. Hwang et al.[7] proposed a PPG biometric generative adversarial network (PBGAN) by extracting the semantic representation and time stability. Ye et al.[8] investigated two CNNs and two LSTM networks to extract discriminative features, and identified the subject by the adjusted cosine similarity. Liu et al.[9] used five CNN layers, two LSTM layers and one dense output layer for PPG biometrics. Deep-learning models for PPG biometric recognition show good performance, and they are a hotspot of research. However, the existing deep learning methods do not utilize the discriminative feature from time and frequency data.

3 Data preprocessing

3.1 Preprocessing

PPG signals are easily affected by various kinds of noises, such as baseline wander, powerline interference, and motion artifacts. The process of preprocessing includes the following steps:

Butterworth filter.

We filter the PPG signals to use a fourth-order Butterworth filter, and the cutoff frequency is set as 0.5–18 Hz, which can effectively mitigate the noise of baseline wander and powerline interference.

Variational mode decomposition[19].

First, to mitigate the influence of motion artifacts, PPG signals are first decomposed into different modes by variational mode decomposition. PPG signals X can be decomposed into different modes X_k with center frequency w_k as follows:

$$\min_{\{X_k\},\{w_k\}} \sum_k \left\| \partial_t \left[\delta(t) + \frac{\mathrm{j}}{\pi t} \times X_k(t) \right] \mathrm{e}^{-\mathrm{j}w_k t} \right\|_2^2$$
$$\text{s.t.} \sum_k X_k = X \tag{1}$$

where $\{X_k\}$ is the different mode set, $\{w_k\}$ is the center frequency set, $k \in [1, 2, \cdots, K]$, K is the mode number, and δ is the Dirac distribution.

Then, we use the a reference signal to remove the motion artifacts from the remaining modes, and the PPG signals are reconstructed by the remaining modes as follows:

$$X_{recon} = \sum_{k \neq \{k_d\}} X_k \tag{2}$$

where $\{k_d\}$ denotes the modes of noise sets.

Normalization.

We use z score standard normalization for PPG signals.

The whole preprocessing process is shown in Fig. 2.

3.2 Segmentation

We segment the preprocessed PPG signals into fixed length sequences by sliding a rectangular window, and each segmented sequence is taken as a sample. To obtain sufficient training samples, we segment PPG signals by shifting the rectangular window with overlap fractions, which is a method for data augmentation. The testing samples are obtained by sliding the rectangular window without any overlap.

3.3 Feature extraction

Time-domain feature.

Similar to the work in the literature[20], we extract the waveform of PPG samples as time-domain feature, which includes the maximum and minimum amplitude, interval and standard deviation. One-dimensional local binary patters (1DLBP) can generate binary codes by comparing the amplitude of each sampling point with its neighbors[21], which can reduce noise influence, and we also use 1DLBP as a time-domain feature.

Frequency-domain feature.

We extract the discrete wavelet transform (DWT) and short-time Fourier transform (STFT) as frequency-domain features of one-dimensional PPG signals[20, 22]. PPG samples are decomposed into wavelet coefficients by wavelet packet decomposition, and the Daubechies wavelet Db8 with three-level wavelet is used to obtain the DWT feature value.

4 Model structure

The proposed DMFDNN consists of three main com-

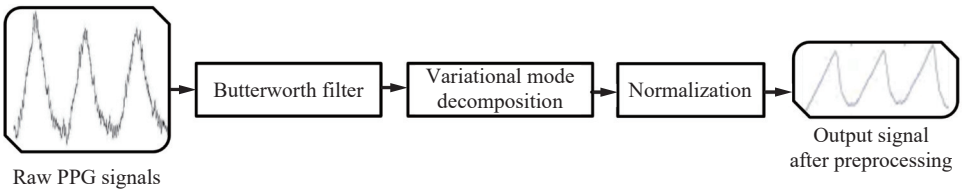


Fig. 2 Preprocessing process

ponents: 1) LSTM network for learning time-varying discriminative features, 2) multiscale feature extraction module for learning transition information, 3) dual-domain attention module for learning complementary information.

4.1 Notations

The input of the proposed model is the PPG sample matrix X, $X = [X^t, X^f]$, $X^t \in \mathbf{R}^{d_t \times n}$, $X^f \in \mathbf{R}^{d_f \times n}$, X^t and X^f are the time-domain and frequency-domain feature sets, respectively, and d_t and d_f are the dimensionality values.

4.2 LSTM network

The time-frequency features of PPG signals are long-term dependent, and we use the LSTM network to extract the time-varying discriminative features[23]. The architecture of the LSTM network is show in Fig. 3.

The time-frequency feature matrix X is processed by the LSTM network, and we can obtain the output H of the LSTM network:

$$H = LSTM(X). \tag{3}$$

4.3 Multiscale feature extraction module

Transition information is crucial for PPG biometric recognition. To capture the transition information, we utilize different scale convolution filers to extract discriminative features, and the multiscale feature extraction (MSFE) module is shown in Fig. 4.

First, we employ 3 branches of dilated convolutions with dilation rates from 1 to 3 to transform the output of the LSTM network. Then, each branch is set three convolutional layers. The number of convolution kernels of each branch decreases step by step from 512 to 128, and it can continue to extract high-level semantic features and reduce dimensions at the same time, which is defined as

$$\begin{aligned} X_i^1 &= Conv_i\left(H, (3, 512, i)\right) \\ X_i^2 &= Conv_i\left(X_i^1, (3, 256, i)\right) \\ X_i^3 &= Conv_i\left(X_i^2, (3, 128, i)\right) \\ i &\in [1,\ 2,\ 3] \end{aligned} \tag{4}$$

where H is the input features mapped from LSTM, $Conv_i$ is the convolution layer with dilation rate i, $1 \le i \le 3$, and X_i^j is the output of the j-th convolution layer. Finally, the features mapped from different branches are concatenated, which is defined as

$$X_{con} = Concat\left(X_1^3, X_2^3, X_3^3\right) \tag{5}$$

where X_{con} is the multiscale concatenated feature.

Moreover, we utilize a bottleneck layer to reduce the parameters of MSFE, which can lighten our proposed model. The bottleneck is defined as

$$X_b = Bl\left(X_{con}\right) \tag{6}$$

where Bl denotes the bottleneck operation, and X_b is the final output result of MSFE.

4.4 Dual-domain attention module

Different domains have different distinctive information for personal identification. To strengthen the domain of greater contribution for PPG biometrics, we propose a dual-domain attention module (DDAM). DDAM fuses the feature maps of two domains and obtains the channel-wise attention component.

First, we obtain the dual-domain fusion component, which is defined as

$$X_{fuse} = X_b^t + X_b^f + X_b^t \odot X_b^f \tag{7}$$

where X_{fuse} denotes the fused feature map, and $\oplus$ is elementwise product operation. X_b^t and X_b^f are the outputs of multiscale feature extraction modules from time and frequency domains, respectively.

Then, similar to the work in [24], we input X_{fuse} into

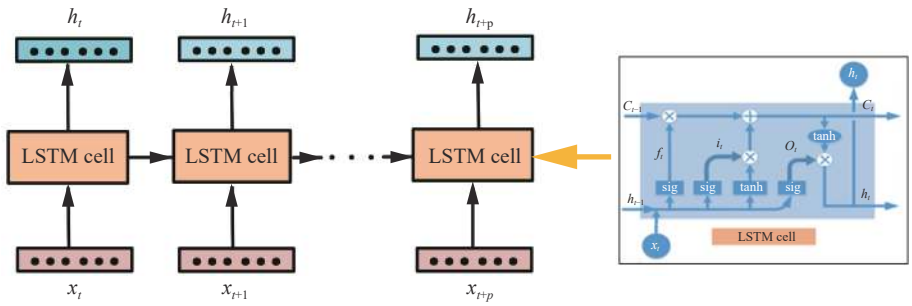


Fig. 3 Architecture of the LSTM network. x_t and h_t denote the input vector and output vector in state t, respectively. p denotes the number of segments of a sample.

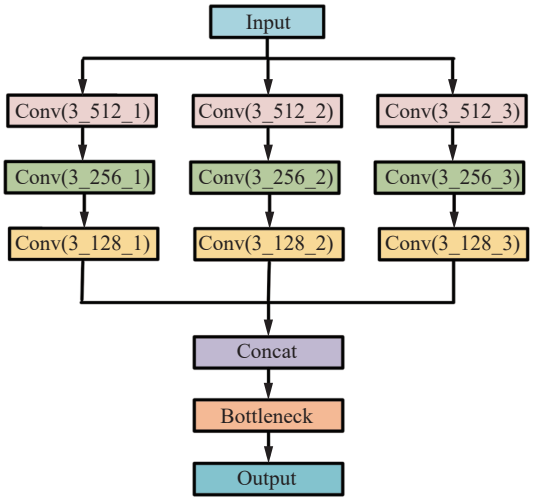


Fig. 4 Module of multiscale feature extraction. ((the kernel size)_(the number of filters)_(the dilation rate)) denotes the convolutional parameters.

the channelwise attention component to emphasize the important channels of the fused information, which is defined as

$$\begin{aligned} X_g &= Ava\left(X_{fuse}\right) \\ X_{att} &= \sigma\left(\delta\left(Full\_con\left(X_g\right)\right)\right) \\ X_{out} &= X_{fuse} \odot X_{att} \end{aligned} \tag{8}$$

where Ava is the global average pooling operation, $Full\_con$ is the fully connected operation, δ denotes the ReLu activation operation, σ is the Sigmoid activation function, and X_{out} denotes the output of DDAM.

4.5 Recognition

The proposed model turns a PPG sample into a context vector, which is further used for identification and verification. In the identification mode, we design the fully connected layer with a softmax function to predict the label of a context vector X_{out}^i, which is define as follows:

$$\hat{P}\left(S_k \| X_{out}^i\right) = softmax\left(W X_{out}^i + b\right), \quad 1 \le k \le n \tag{9}$$

where $\hat{P}\left(S_k \| X_{out}^i\right)$ denotes the posterior probability of belonging to a subject S_k, and n is the number of all subjects. The maximum $\hat{P}\left(S_k \| X_{out}^i\right)$ is identified as the candidate subject of X_{out}^i.

The objective function of our model is trained by the multiclass cross entropy loss L, which is defined as

$$L = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{C} Q\left\{S_k = X_{out}^i\right\} \log\left(\hat{P}\left(S_k \| X_{out}^i\right)\right) \tag{10}$$

where N is the number of all samples, and k is the number of subjects enrolled. $Q\left\{S_k = X_{out}^i\right\}$ denotes the true posterior probability.

In the verification mode, we calculate the cosine similarity value between the context vector and the enrolled vector, and decide whether the context vector is a genuine or an impostor by a preset threshold value.

5 Experiment

5.1 Databases

To evaluate the performance of the proposed method, we employ four public databases: Beth Israel Deaconess Medical Center (BIDMC)[25, 26], Multi-parameter Intelligent Monitoring for Intensive Care (MIMIC)[27], CapnoBase[28] and Biosec1[16]. Table 1 gives the number of subjects, frequency and length of PPG signals on the four datasets.

Table 1 Details of the three adopted datasets

Dataset	No. of subjects	Frequency	Length (min)
BIDMC	53	125	8
MIMIC	65	125	10
CapnoBase	42	300	8
Biosec1	31	100	3

The BIDMC database was acquired from 53 hospitalized patients, and each recording contained an 8-minute duration with a sampling frequency 125 Hz. The MIMIC database was acquired from ICUs, and we chose 65 subjects in this work. The CapnoBase database has PPG, ECG, and other signals for 42 cases of 8-minute duration with a sampling frequency 300 Hz.

In all databases, 3-min-long recordings are used from each subject, and we extract the first 60% of each recording as training sets, the next 30% for validation sets, and the last 10% for testing sets. The testing, validation and training samples do not overlap, and the average testing values are obtained as experimental results. To evaluate our method on two-sessions scenario, we choose 20 subjects with two recordings in the MIMIC database, which are named as M1 and M2, and the two-sessions database is called two-sessions MIMIC. The Biosec1 database consists of two-sessions of PPG signals, and the number of subjects is 31 with a sampling rate of 100 Hz. The PPG signals from Biosec1 are collected in the leisure state, and the average time gap between the two sessions is 36 days. Similar to [7], we exclude 30 s in the first and last parts of each session, which is time consuming of wearing and adjusting the device.

5.2 Performance metrics

In the identification mode, the subject recognition rate

is used as an evaluation criterion, which is defined as

$$\text{Subject recognition rate} = \frac{N\_correct\_sample}{N\_test\_sample} \tag{11}$$

where N_test_sample is the total number of testing samples and $N\_correct\_sample$ is the number of correctly identified testing samples.

In authentication mode, the similarity is calculated by comparing a test sample and all enrolled samples. The EER is the equivalent value of both the false acceptance rate (FAR) and false rejection rate (FRR), which is defined as

$$FAR = \frac{NFA}{NIRA} \times 100\% \tag{12}$$

$$FRR = \frac{NFR}{NGRA} \times 100\% \tag{13}$$

where NFA is the number of false acceptance samples, $NIRA$ is the number of impostor attempt samples, NFR is the number of false rejection samples, and $NGRA$ is the number of genuine attempt samples.

5.3 Parameter settings

The input dimension varies along with the number of sample points of a rectangular window, and we set the 1.5 second rectangular window with an overlap of 0.8 fractions. DMFDNN is implemented using the TensorFlow framework. The training epochs of the model are 180 with a batch size of 80. In the LSTM model, we use a single layer LSTM with 150 nerve cells in the hidden layer, and the input dimension is the length of the rectangular window. Each CNN layer consists of a one-dimensional CNN operation, and the pooling window is set as 3 with stride 3.

5.4 Performance of the proposed method

To validate the performance of DMFDNN, we compared DMFDNN on three databases with baseline methods that include the single-feature and fusion-feature methods, LSTM method, and CNN method. We extracted the shape, one-dimensional local binary pattern, discrete wavelet transform and short-term Fourier transform of time and frequency features, which are named the shape, 1DLBP, DWT and STFT methods, respectively. Similar to DMFDNN, we also design the fully connected layer with a softmax function to predict the label of all extracted features. Table 2 lists the subject recognition rates of DMFDNN compared with baseline methods.

As shown in Table 2, the performance of DMFDNN has greater improvements than the single-feature and fusion-feature methods on the three databases, demonstrating that DMFDNN with dual-domain and multiscale fusion has high discrimination for PPG biometrics. Moreover, because of capturing transition and multiscale information, the performance of DMFDNN is superior to LSTM and CNN methods on the three databases. The total parameter numbers of LSTM, CNN and DMFDNN are 276 645, 298 342 and 303 794, respectively. The experimental networks require approximately 645 K, 668 K and 683 K floating point operations on LSTM, CNN and DMFDNN, respectively. It is worth noting that, as DMFDNN is the fusion approach of CNN and LSTM, DMFDNN does not increase the number of parameters and floating-point operations. This is mainly because the input data of DMFDNN is time-frequency data of PPG signals, the input of CNN and LSTM are sequence segments of original PPG signals, and the dimension of time-frequency data is less than the sequence segments.

Table 2 Comparison of DMFDNN with baseline methods

Feature	BIDMC	MIMIC	CapnoBase
Shape	93.35%	92.37%	94.08%
1DLBP	94.56%	93.72%	94.93%
STFT	95.23%	94.41%	95.62%
DWT	95.68%	94.75%	96.03%
Shape + DWT	96.63%	95.72%	97.24%
Shape + STFT	96.72%	95.87%	97.36%
1DLBP + STFT	97.34%	96.17%	97.98%
1DLBP + DWT	97.62%	96.34%	98.14%
LSTM	98.42%	97.35%	99.12%
CNN	98.64%	97.54%	99.24%
Ours	**99.16%**	**98.42%**	**99.87%**

Furthermore, we perform the multi-session analysis in the MIMIC database. The subject-recognition-rate performance of DMFDNN and other baseline methods in multi-session scenarios is given in Table 3.

Table 3 Comparison of results with multi-session scenarios in MIMIC

Method	Training	Testing	Subject recognition rate (%)
Shape	M1	M2	91.63
	M2	M1	91.42
DWT	M1	M2	92.62
	M2	M1	92.21
Shape + DWT	M1	M2	94.52
	M2	M1	94.36
Ours	M1	M2	**98.05**
	M2	M1	**97.74**

From Table 3, compared to the baseline methods, DMFDNN can achieve the best results, demonstrating that dual-domain and multiscale fusion can improve the performance for multi-session data.

5.5 Comparison with state-of-the-art methods

To further validate the performance of DMFDNN, we compare it with several state-of-the-art methods, including the nondeep learning methods and deep learning methods. In two-session MIMIC, M1 recordings are collected as a training set, and M2 recordings are equally divided into validation and testing sets. The performance results are shown in Table 4.

Table 4 Comparison with state-of-the-art methods

Database	Method	Subject recognition rate (%)	EER (%)
BIDMC	[20]	98.06	1.94
	[27]	98.65	–
	[3]	98.78	1.43
	[7]	–	1.34
	Ours	**99.16**	**0.92**
MIMIC	[20]	97.62	2.36
	[3]	97.86	2.23
	[7]	–	2.15
	Ours	**98.42**	**1.43**
CapnoBase	[18]	99.00	–
	[13]	–	1.0
	[3]	99.82	0.34
	[7]	–	0.3[1]
	[1]	**100**	**0.1**
	Ours	99.87	0.18
Two-session MIMIC	[20]	96.43	3.87
	[3]	96.74[2]	–
	[7]	–	2.43
	Ours	**98.05**	**2.26**
Biosec1	[3]	87.24	13.15
	[7]	88.5[3]	–
	Ours	**89.26**	**11.42**

[1]Cited from [7] with the structure of wide-shallow model.
[2]Cited from [3].
[3]Cited from [7] with the structure of PBGAN-LS (W).

From Table 4, it can be observed that our method outperforms other methods on BIDMC, MIMIC and Biosec1, and it is evident that DMFDNN with dual-domain and multiscale fusion representation can obtain more robust and discriminative recognition performance than other methods. On all databases, DMFDNN can achieve competitive PPG biometric recognition compared with state-of-the-art methods.

5.6 Ablation experiment

To evaluate the contributions of each part of our model, we removed the different parts of our model to obtain the following different models:

TMDNN and FMDNN. We remove the frequency-domain branch and retain only time-domain input, and obtain a time-domain and multiscale deep neural network (TMDNN). We remove the time-domain branch and retain only frequency-domain input, and obtain a frequency-domain and multiscale deep neural network (FMDNN).

DMFDNN without LSTM. We remove the LSTM network, and obtain the DMFDNN model without LSTM.

DFDNN. We remove the multiscale feature extraction model, and obtain the model of the dual-domain fusion deep neural network (DFDNN).

DMFDNN without attention. We replace the dual-domain attention module, and concatenate only time frequency features.

Table 5 gives the recognition performance results using the above different models.

Table 5 Comparison of DMFDNN with baseline methods

Feature	BIDMC	MIMIC	CapnoBase
TMDNN	97.98%	96.78%	99.12%
FMDNN	98.14%	97.31%	99.26%
DMFDNN without LSTM	98.16%	97.26%	99.14%
DFDNN	98.77%	97.84%	99.34%
DMFDNN without attention	99.04%	98.23%	99.44%
Ours	**99.16%**	**98.42%**	**99.87%**

As shown in Table 5, we can see that DMFDNN has better performance than all other methods on the three databases, which clearly shows that different parts of DMFDNN are important to improve the performance. Compared with the TMDNN and FMDNN, the recognition performance increases by 1.18% and 1.02% on BIDMC, respectively, which can prove the importance of dual-domain fusion. LSTM can learn the time-varying discriminative information, and DMFDNN without LSTM has lower performance than our proposed method. Transition information is crucial for PPG biometric recognition, and multiscale extraction can capture transition information, so DMFDNN has better performance than DFDNN. In addition, the attention can strengthen the contribution of different domains for PPG biometrics, so DMFDNN without attention has lower performance than our method. Therefore, it is important to exploit the information of dual-domain fusion, multiscale feature extraction and attention to improve the recognition performance.

6 Conclusions

In this paper, a dual-domain and multiscale fusion deep neural network is proposed for PPG biometric recognition. By exploring dual-domain fusion and multiscale feature extraction, the proposed method can learn the time-varying and multiscale discriminative features from the time and frequency domains. Furthermore, to strengthen the domain of greater contribution from time-domain and frequency-domain data, we design a dual-domain attention module. In future work, we plan to explore a new deep learning approach that can incorporate the novel loss function to further improve PPG biometric recognition performance.

Acknowledgements

This work was supported by National Nature Science Foundation of China (No. 62276093) and in part by Natural Science Foundation of Shandong Province, China (No. 2022MF86).

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] D. Y. Hwang, B. Taha, D. S. Lee, D. Hatzinakos. Evaluation of the time stability and uniqueness in PPG-based biometric system. *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 116–130, 2021. DOI: 10.1109/TIFS.2020.3006313.

[2] D. E. Mancilla-Palestina, J. A. Jimenez-Duarte, J. M. Ramirez-Cortes, A. Hernandez, P. Gomez-Gil, J. Rangel-Magdaleno. Embedded system for bimodal biometrics with fiducial feature extraction on ECG and PPG signals. In *Proceedings of IEEE International Instrumentation and Measurement Technology Conference*, Dubrovnik, Croatia, 2020. DOI: 10.1109/I2MTC43012.2020.9128394.

[3] C. Y. Liu, J. J. Yu, Y. W. Huang, F. X. Huang. Time–frequency fusion learning for photoplethysmography biometric recognition. *IET Biometrics*, vol. 11, no. 3, pp. 187–198, 2022. DOI: 10.1049/bme2.12070.

[4] J. Luque, G. Cortès, C. Segura, A. Maravilla, J. Esteban, J. Fabregat. End-to-end Photoplethysmography (PPG) based biometric authentication by using convolutional neural networks. In *Proceedings of the 26th European Signal Processing Conference*, IEEE, Rome, Italy, pp. 538–542, 2018. DOI: 10.23919/EUSIPCO.2018.8553585.

[5] L. Everson, D. Biswas, M. Panwar, D. Rodopoulos, A. Acharyya, C. H. Kim, C. Van Hoof, M. Konijnenburg, N. Van Helleputte. BiometricNet: Deep learning based biometric identification using wrist-worn PPG. In *Proceedings of International Symposium on Circuits and Systems*, IEEE, Florence, Italy, 2018. DOI: 10.1109/ISCAS.2018.8350983.

[6] D. Biswas, L. Everson, M. Q. Liu, M. Panwar, B. E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. Van Hoof, M. Konijnenburg, N. Van Helleputte. CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment. *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 2, pp. 282–291, 2019. DOI: 10.1109/TBCAS.2019.2892297.

[7] D. Y. Hwang, B. Taha, D. Hatzinakos. PBGAN: Learning PPG representations from GAN for time-stable and unique verification system. *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 5124–5137, 2021. DOI: 10.1109/TIFS.2021.3122817.

[8] Y. L. Ye, G. C. Xiong, Z. Y. Wan, T. J. Pan, Z. W. Huang. PPG-based biometric identification: Discovering and identifying a new user. In *Proceedings of the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, Mexico, pp. 1145–1148, 2021. DOI: 10.1109/EMBC46164.2021.9630883.

[9] X. C. Liu, Z. H. Yuan, D. L. Ma. Deep learning framework for biometric identification from wrist-worn PPG with acceleration signals. In *Proceedings of the 6th International Conference on Signal and Image Processing*, IEEE, Nanjing, China, pp. 1–5, 2021. DOI: 10.1109/ICSIP52628.2021.9688605.

[10] J. C. Yao, X. D. Sun, Y. B. Wan. A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, France, pp. 4576–4579, 2007. DOI: 10.1109/IEMBS.2007.4353358.

[11] S. Chakraborty, S. Pal. Photoplethysmogram signal based biometric recognition using linear discriminant classifier. In *Proceedings of the 2nd International Conference on Control, Instrumentation, Energy & Communication*, IEEE, Kolkata, India, pp. 183–187, 2016. DOI: 10.1109/CIEC.2016.7513792.

[12] N. I. M. Nadzri, K. A. Sidek, A. F. Ismail. Biometric recognition for twins inconsideration of age variability using PPG signals. *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 10, no. 1–5, pp. 97–100, 2018.

[13] J. Sancho, Á. Alesanco, J. García. Biometric authentication using the PPG: A long-term feasibility study. *Sensors*, vol. 18, no. 5, Article number 1525, 2018. DOI: 10.3390/s18051525.

[14] P. Spachos, J. X. Gao, D. Hatzinakos. Feasibility study of photoplethysmographic signals for biometric identification. In *Proceedings of the 17th International Conference on Digital Signal Processing*, IEEE, Corfu, Greece, 2011. DOI: 10.1109/ICDSP.2011.6004938.

[15] N. Karimian, M. Tehranipoor, D. Forte. Non-fiducial PPG-based authentication for healthcare application. In *Proceedings of the IEEE/EMBS International Conference on Biomedical & Health Informatics*, IEEE, Orlando, USA, pp. 429–432, 2017. DOI: 10.1109/BHI.2017.7897297.

[16] U. Yadav, S. N. Abbas, D. Hatzinakos. Evaluation of PPG biometrics for authentication in different states. In *Proceedings of International Conference on Biometrics*, IEEE, Gold Coast, Australia, pp. 277–282, 2018. DOI: 10.1109/ICB2018.2018.00049.

[17] P. Farago, R. Groza, L. Ivanciu, S. Hintea. A correlation-based biometric identification technique for ECG, PPG and EMG. In *Proceedings of the 42nd International Conference on Telecommunications and Signal Processing*, IEEE, Budapest, Hungary, pp. 716–719, 2019. DOI: 10.

1109/TSP.2019.8768810.

[18] S. W. Lee, D. K. Woo, Y. K. Son, P. S. Mah. Wearable biosignal (PPG)-based personal authentication method using random forest and period setting considering the feature of PPG signals. *Journal of Computers*, vol. 14, no. 4, pp. 283–294, 2019. DOI: 10.17706/jcp.14.4.283-294.

[19] K. Dragomiretskiy, D. Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014. DOI: 10.1109/TSP.2013.2288675.

[20] J. F. Yang, Y. W. Huang, R. L. Zhang, F. X. Huang, Q. G. Meng, S. X. Feng. Study on PPG biometric recognition based on multifeature extraction and naive Bayes classifier. *Scientific Programming*, vol. 2021, Article number 5597624, 2021. DOI: 10.1155/2021/5597624.

[21] S. Nikan, F. Gwadry-Sridhar, M. Bauer. Pattern recognition application in ECG arrhythmia classification. In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies*, Porto, Portugal, pp. 48–56, 2017. DOI: 10.5220/0006116300480056.

[22] R. Raj, J. Selvakumar, M. Anburajan. Evaluation of hypotension using wavelet and time frequency analysis of photoplethysmography (PPG) signal. In *Proceedings of International Conference on Advances in Computational Intelligence in Communication*, Puducherry, India, vol. 14, pp. 57–61, 2016.

[23] M. Y. Bian, B. Peng, W. Wang, J. Dong. An accurate LSTM based video heart rate estimation method. In *Proceedings of the 2nd Pattern Recognition and Computer Vision*. Springer, Xi'an, China, pp. 409–417, 2019. DOI: 10.1007/978-3-030-31726-3_35.

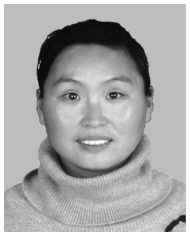
[24] Z. Y. Jia, Y. F. Lin, J. Wang, X. H. Wang, P. Y. Xie, Y. B. Zhang. SalientSleepNet: Multimodal salient wave detection network for sleep staging. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, Montreal, Canada, pp. 2614–2620, 2021. DOI: 10.24963/ijcai.2021/360.

[25] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, D. A. Clifton. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017. DOI: 10.1109/TBME.2016.2613124.

[26] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000. DOI: 10.1161/01.CIR.101.23.e215.

[27] J. F. Yang, Y. W. Huang, F. X. Huang, G. P. Yang. Photoplethysmography biometric recognition model based on sparse softmax vector and *k*-nearest neighbor. *Journal of Electrical and Computer Engineering*, vol. 2020, Article number 9653470, 2020. DOI: 10.1155/2020/9653470.

[28] W. Karlen, S. Raman, J. M. Ansermino, G. A. Dumont. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, 2013. DOI: 10.1109/TBME.2013.2246160.

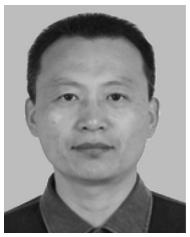


Chun-Ying Liu received the M. Eng. degree in computer science from Shandong University of Science and Technology, China in 2008. She is now an associate professor at School of Computer, Heze University, China.

Her research interests include biometrics, data-mining and pattern recognition.

E-mail: lcy810204@163.com

ORCID iD: 0000-0002-6062-4224



Gong-Ping Yang received the Ph. D. degree in computer software and theory from Shandong University, China in 2007. He is currently a professor at School of Software Engineering, Shandong University, and an adjunct professor at School of Computer, Heze University, China.

His research interests include pattern recognition, image processing and biometrics.

E-mail: gpyang@sdu.edu.cn

ORCID iD: 0000-0001-7637-2749

Yu-Wen Huang received the Ph. D. degree in computer science and technology from Shandong University, China in 2021. Now he is an associate professor at School of Computer, Heze University, China.

His research interests include ECG recognition, biometrics and machine learning.

E-mail: hzxy_hyw@163.com

Fu-Xian Huang received the M. Eng. degree in computer science from Shandong University of Science and Technology, China in 2005. He is now a professor at School of Computer, Heze University, China.

His research interests include biometrics and machine learning.

E-mail: huangfuxian@126.com (Corresponding author)

ORCID iD: 0000-0002-2838-9973