DUAL-DOMAIN LOW-RANK FUSION DEEP METRIC LEARNING FOR OFF-THE-PERSON ECG BIOMETRICS

Guiping Zhu[1], *Mingzhu Ma*[1], *Yuwen Huang*[1,2], *Kuikui Wang*[1], *Gongping Yang*[1,2,*]

[1]School of Software,Shandong University,Jinan 250101,China
[2]School of Computer,Heze University,Heze 274015,China
*gpyang@sdu.edu.cn (G. Yang)

ABSTRACT

Electrocardiogram (ECG) biometrics has been an emerging field, and off-the-person ECG biometrics capturing the ECG from fingertips is one of the new trends in this field. However, dynamic morphological variability in the same person and low signal-to-noise ratios pose great challenges for off-the-person ECG biometrics. To reduce the dynamic morphological variability, this paper introduces deep metric learning into ECG biometrics to learn intra-individual compact features. To enforce the robust of proposed method, dual-domain features extracted from both 1D signals and 2D spectrograms are integrated by low-rank fusion. Furthermore, this method dispenses with the need for noise removal and outliers discarding completely. Experiments on two off-the-person ECG benchmark databases demonstrate that the proposed method significantly outperforms the state-of-the-art methods. Additionally, ablation experiments show the effectiveness of every part of our framework.

Index Terms— ECG biometrics, off-the-person, deep metric learning, dual-domain, low-rank fusion

1. INTRODUCTION

As the rapid demand of biometric technologies for identity security, ECG signal as an emerging biometric trait has attracted increasing research interest [1, 2]. In recent years, with the advent of the off-the-person devices that collect ECG signals with minimal skin of fingers or palms, medical devices that connect electrodes to other parts of body are gradually replaced [3, 4].

However, there are two-fold challenges on off-the person ECG biometrics: (1) dynamic morphological variability induced by emotional arousal, heart-rate variation, and posture variety [5], (2) low signal-to-noise ratios caused by power line interference, baseline drift, and electrode malposition [6]. First of all, various subspace-based methods have been attempted to reduce intra-individual ECG variability [1, 5]. Nevertheless, they are not robust to model the dynamic morphological variability of the same person in off-the-person condition. Also, several works have introduced emerging deep learning into ECG biometrics [7, 8]. However, general deep ECG methods separated different identities by a Softmax classifier which is not discriminative enough. Although the triplet loss for ECG biometrics [9] has been utilized to learn intra-class compact features, there is a combinatorial explosion in the number of ECG triplets especially for large-scale datasets. In addition, while low signal-to-noise ratio is another key problem restricting ECG performance, one common solution is to remove interference components and abnormal data [10]. However, the noise in off-the-person condition is too complex to be eliminated completely. Several deep methods transformed signals into robust feature space without any preprocessing [9], but it is still an unsolved problem to fully model the complex noise in off-the-person condition.

Inspired by supervised metric learning in face verification [11, 12], we introduce additive angular margin loss (AAM loss) into ECG biometrics. In contrast to most approaches which deemed ECG biometrics as a classification task, we view it as a metric learning task to reduce intra-individual dynamic morphological variability. Motivated by the recent success of STFT-based spectrogram in ECG arrhythmia classification [13], 2D spectrograms extracted by STFT is utilized to investigate frequency information as a complementarity of time domain. Furthermore, in contrast to input-level feature concatenation which lost out the temporal dependencies within each feature, the low-rank feature fusion [14] has been utilized for robust feature representations.

The remaining parts of this paper are organized as follows. Section 2 introduces our proposed method. The experimental results for evaluating the method will be presented in Section 3. Finally, the conclusion is given in Section 4.

2. PROPOSED METHOD

In this section, we introduce our proposed framework including feature extraction, deep metric learning, low-rank fusion

This work was supported in part by the NSFC-Xinjiang Joint Fund under Grant U1903127 and in part by the Natural Science Foundation of Shandong Province under Grant ZR2020MF052.

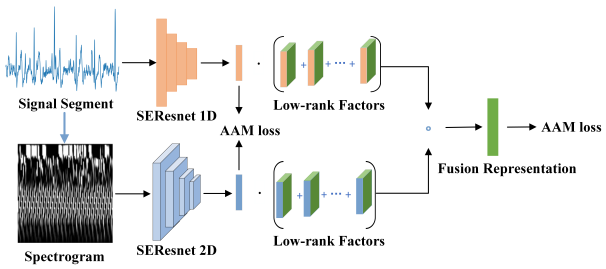


Fig. 1. **The proposed framework for ECG biometrics**

and the details of network and training. The proposed framework for ECG biometric recognition is shown in Fig.1.

2.1. Feature extraction

The feature extraction stage consists of signal segmentation and spectral feature extraction. On the contrary to most of the work which needs fiducial points, we adopt a constant duration blind segmentation to get ECG segments. We assume that there is a collection of N samples obtained by blind segmentation from C individuals, denoted as $X = [X_1, X_2, \ldots, X_C] = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N}]$, where $\boldsymbol{x_t}$ is the ECG sample obtained by blind segmentation and X_t is the set of segments from the t^{th} individual.

Afterward, we use STFT to get spectrogram of each sample to obtain a compact representation with frequency information, denoted as $I = [I_1, I_2, \ldots, I_C] = [\boldsymbol{i_1}, \boldsymbol{i_2}, \ldots, \boldsymbol{i_N}]$, where $\boldsymbol{i_t}$ is the spectrogram of ECG sample and I_t is the set of spectrograms from the t^{th} individual. STFT converts the 1D ECG signal $\boldsymbol{x_t}$ into 2D spectrogram $\boldsymbol{i_t}$ as follows:

$$\boldsymbol{i_t}[m,n] = \sum_{k=0}^{L-1} \boldsymbol{x_t}[k] g[k-m] e^{-j2\pi nk/L} \tag{1}$$

where $g[k-m]$ is the window function to divide the signals into shorter segments, $\boldsymbol{x_t}[k]g[k- m\ ]$ is the shorter segments of input ECG segment $\boldsymbol{x_t}$, and L is the window length.

2.2. Deep metric learning

Deep metric learning aims to project samples into a low-dimensional metric space where the samples from the same identity should be similar and the samples from the different identities should be dissimilar by training a deep forward network. It focuses on the comparison of features, while classification is to find the classification boundaries. In this paper, we learn two branch functions by deep metric learning: $\boldsymbol{u} = f_1(\boldsymbol{x}; \Upsilon) \in R^{d_u}$ and $\boldsymbol{v} = f_2(\boldsymbol{i}; \Phi) \in R^{d_v}$ to represent the signal segments and spectrograms in new metric spaces U and V, where d_u, d_v are the dimensionalities of the representations in two spaces, and Υ, Φ are the trainable parameters. Considering the complex dynamic morphological variability of ECG signals, AAM loss is utilized to enhance the discriminant performance of the representation spaces.

The AAM is proposed based on Softmax loss which is widely used to separate different identities. We assume that the extracted ECG feature of the i-th sample is $\boldsymbol{z_i} \in R^d$ and its label is y_i. The batch size and the identity number are denoted as N and n. Softmax loss is presented as follows:

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{W_{y_i}^T \boldsymbol{z_i} + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T \boldsymbol{z_i} + b_j}} \tag{2}$$

where $W_j \in R^d$ denotes the j-th column of the weight $W \in R^{d \times n}$ and $b_j \in R^n$ is the bias term. To enforce higher similarity for intra-class samples, AAM loss transforms the logit as follows:

$$\mathrm{W}_j^T \boldsymbol{z_i} = \|W_j\| \, \|\boldsymbol{z_i}\| \cos\theta_j \tag{3}$$

where θ_j is the angle between the weight W_j and the feature $\boldsymbol{z_i}$. When the $\|W_j\|$ and $\|\boldsymbol{z_i}\|$ are fixed to 1 by l_2 normalization, we get the

$$W_j^T \boldsymbol{z_i} = \cos\theta_j \tag{4}$$

The angle θ between the feature $\boldsymbol{z_i}$ and the label weight W_{y_i} can be calculated by $\arccos\theta_{y_i}$. Then we add the margin m on θ by $\cos(\theta_{y_i} + m)$ to obtain discriminative features. Subsequently, the AAM loss is presented as follows:

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)} + \sum_{j=1, j\neq 1}^{n} e^{s\cos\theta_j}} \tag{5}$$

While the Softmax loss provides separable representation space, the AAM loss reduces the distance between features and centre of each class to enhance the intra-class compactness.

2.3. Low-rank fusion

To jointly use features extracted from signal segments and spectrograms, low-rank fusion is introduced. This is a noise-robust approach which computes tensor fusion model with fewer parameters. Tensor fusion is an effective method to investigate shared and feature-specific interactions. It first creates a tensor by outer product between different features:

$$Q = \boldsymbol{q}_1 \otimes \boldsymbol{q}_2 \tag{6}$$

where $\otimes$ denotes the outer product over a set of vectors, and $\boldsymbol{q}_1, \boldsymbol{q}_2$ are obtained by appending 1 to the end of $\boldsymbol{u}_i, \boldsymbol{v}_i$. The tensor $Q \in R^{d_1 \times d_2}$ is then passed through a linear layer $g(\cdot)$ with a weight W to produce a vector representation:

$$h = g(Q; W) = W \cdot Q, \quad h \in R^{d_h} \tag{7}$$

In the linear layer product $W \cdot Q$, the weight tensor can be then viewed as d_h order-2 tensors $W_k \in R^{d_1 \times d_2}$ whose dimension is equal to Q. Each element-wise product on W_k and

Q contributes to one dimension of the output vector h. However, the weight W will be a tensor of order-3 in $R^{d_1 \times d_2 \times d_h}$. To reduce the number of parameters, low-rank fusion is utilized to decompose the weights into low-rank factors. We use a fixed rank r as the rank of W. Then W can be displaced by its rank r decomposition factors $\left\{w_1^i, w_2^i\right\}_{i=1}^r$. W can be recovered by:

$$W = \sum_{i=1}^{r} w_1^i \otimes w_2^i \tag{8}$$

Using the Equation 6 and 8, we can simplify Equation 7:

$$\begin{aligned} h &= W \cdot Q \\ &= \left(\sum_{i=1}^{r} w_1^i \otimes w_2^i\right) \cdot (\boldsymbol{q}_1 \otimes \boldsymbol{q}_2) \\ &= \left(\sum_{i=1}^{r} w_1^i \cdot \boldsymbol{q}_1\right) \circ \left(\sum_{i=1}^{r} w_2^i \cdot \boldsymbol{q}_2\right) \end{aligned} \tag{9}$$

where $\circ$ denotes the element-wise product. This simplification exploits the decomposition of both Q and W, so that we can compute h with decomposition factors $\left\{w_1^i, w_2^i\right\}_{i=1}^r$ instead of tensor Q. In this way, low-rank fusion models interactions between dual-domain features and temporal interactions in each feature. Moreover, low-rank fusion obtains a robust performance for off-the-person biometrics by reducing the number of parameters in tensor fusion.

2.4. Network training

As is shown in Fig. 1, the framework of our model includes two streams for signal segments and spectral images and a low-rank fusion at the end. For the two streams, 1D and 2D Resnet 18 are selected as backbones. The representations $\boldsymbol{u}$ and $\boldsymbol{v}$ are extracted before the last fully connected layer of the ResNet 18. Squeeze and excitation (SE) [15] attention which models channel-wise relationships is added to the end of all 1D residual blocks and 2D residual blocks to improve the representation power of stream networks in an efficient manner.

We train the entire network using a two-stage approach. First, we train the two stream networks (c.f., the blue and orange streams in Fig. 1) with trainable parameters Υ and Φ initialized by Kaiming Initialization [16]. AAM loss is applied to reduce the intra-individual variability of features extracted from two individual branch networks. Once the stream networks are trained, we fix their parameters and feed the representations $\boldsymbol{u}$ and $\boldsymbol{v}$ extracted before the last fully connected layer of the ResNet 18 into the fusion stream. The trainable low-rank factors $\left\{w_1^i, w_2^i\right\}_{i=1}^r$ of the fusion network can therefore be optimized by using AAM loss to minimize the distance between the fusion representations and their corresponding class centres.

3. EXPERIMENTS

3.1. Database and evaluation metrics

Two public off-the-person databases, i.e., the Check Your Biosignals Here initiative Database (CYBHiDB) [6] and the University of Toronto Database (UOFTDB) [17] are selected to evaluate the performance of the proposed method. The CYBHiDB contains short-term and long-term datasets, where the short-term dataset was collected from 65 healthy participants and the long-term dataset was collected from 63 healthy participants. We adopt all the 63 subjects of the first session in long-term dataset, randomly choosing ninety percent of the samples of every subject as training data and the rest as testing data. The UOFTDB contains 1020 identities of 398 males and 622 females captured from five postures. The first session data from 1012 subjects is selected as experimental data. Half of the samples randomly selected from each subject are used as training data and the other samples are used as testing data.

In this section, identification and verification experiments are performed to provide further comprehensive understanding of our framework. In all evaluation experiments, we select one sample per person to form the enroll set and the rest are used as the probe samples. For the identification task, the recognition rate (RR) is the criteria, which is the percentage of correctly identified probe samples. For the verification task, the equal error rate (EER) is the criteria, which is the value where false acceptance rate (FAR) and false rejection rate (FRR) are equal, defined as follows:

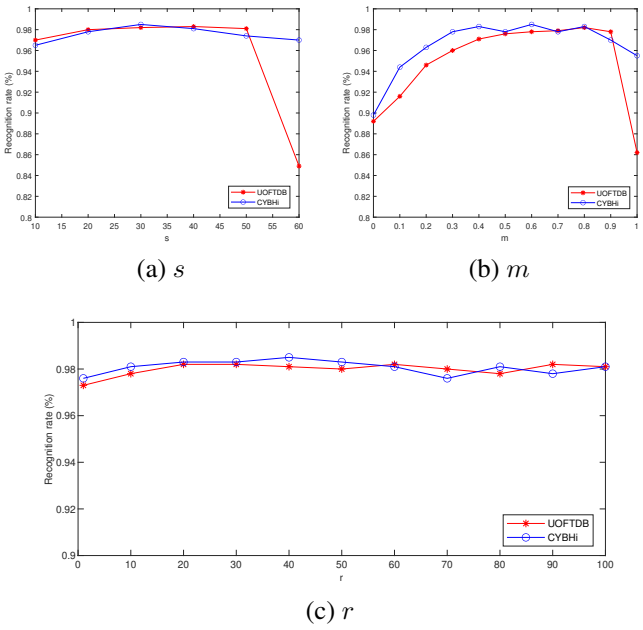
$$FRR = FP/(TP + FP) \tag{10}$$

$$FAR = FN/(TN + FN) \tag{11}$$

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

3.2. Effects of parameters settings and performance of proposed method

In this section, we analyze the impact of four hyper-parameters, including the scale factor s and the relaxation factor m for loss function in Equation 5, the rank r for feature fusion in Equation 9, and the time t for blind segmentation. The scale factor s determines the largest scale of each similarity score. As illustrated in Fig. 2(a), $s = 30$ is appropriate to achieve fairly competent performance. The relaxation factor m determines the margin penalty between features and their class centres. From the influence of m illustrated in Fig. 2(b), we can observe that it achieves the best performance when m reaches 0.6 and 0.8 for CYBHiDB and UOFTDB respectively. We explore the performance of our model with rank settings in low-rank fusion varying from 1 to 100. The results presented in Fig. 2(c) demonstrate that our model is robust for a wide range of rank settings. In addition, to



(a) s (b) m

(c) r

Fig. 2. **Influence of parameters s, m and r.**

balance the time and recognition rate, the t is set to 1.4 s for the CYBHiDB and 7 s for UOFTDB.

3.3. Ablation study

In this experiment, we provide ablation study on UOFTDB and CYBHiDB to verify the effectiveness of each component of our model. We first compare AAM loss with Softmax loss and other metric learning losses include triplet loss [18] which was a weak-supervised metric learning loss and Circle loss [12] which proposed a unified perspective of weak-supervised and supervised metric learning. We can see from Table 1 that AAM loss is much better than Softmax and it also has better performance than the Circle loss. The results prove that AAM loss can produce more intra-compact features. Also, as shown in Table 1, Resnet with SE block which models channel-wise dependencies performs better than Resnet. Additionally, the low-rank fusion not only combines frequency and time domain information but also leads to noise robust by using fewer parameters. The comparison reported in the last four rows of Table 1 demonstrates that low-rank model outperforms concentrated fusion and methods using single feature. These competitive performances emphasize the advantage of AAM loss, SE Resnet and low-rank feature fusion in our framework for off-the-person biometrics.

3.4. Comparison with state-of-the-art methods

Table 2 shows the comparison between the proposed method and other state-of-the-arts on off-the-person databases, including unified sparse representation [3], CN fusion [19] and

Table 1. **RR (%) in ablation study of loss functions, attention and features**

Ablation Study	UOFTDB	CYBHiDB
Softmax	61.0	67.4
Triplet loss	76.1	78.4
Circle loss	98.1	97.2
AAM loss	**98.3**	**98.5**
Resnet	97.0	97.9
Resnet+SE	**98.3**	**98.5**
Signal segments	93.4	97.0
Spectrograms	97.3	94.0
Concentrated fusion	97.7	97.0
Low-rank fusion	**98.3**	**98.5**

Table 2. **Comparison with state-of-the-arts**

Method	Databases (Subject No.)	RR (%)	EER (%)
[3]	CYBHiDB (63)	97.4	1.26
	UOFTDB (1012)	97.4	2.86
[19]	CYBHiDB (63)	97.1	1.85
	UOFTDB (1012)	97.9	2.34
[20]	CYBHiDB (63)	96.1	2.52
	UOFTDB (1012)	92.6	5.48
Ours	CYBHiDB (63)	**98.5**	**0.86**
	UOFTDB (1011)	**98.3**	**0.82**

hybridized approach [20]. As shown in the table, it can be observed that our method has much better performance on RR and EER compared to other state-of-the-art methods. It validates that our method can well capture distinctive and robust features from the off-the-person ECG samples. Moreover, our model does not need any noise removing or outliers discarding step in contrast to [3], [19], [20]. Therefore, not only the performance of proposed model is better than other methods on two off-the-person databases, but also our algorithm has much better robustness for noise and outliers.

4. CONCLUSION

In this paper, we provide a dual-domain low-rank fusion deep metric learning framework to reduce the noise and dynamic morphological variability. We first view ECG biometrics as a task of supervised metric learning and introduce AAM loss to obtain discriminative features. Furthermore, we perform low-rank feature fusion to combine frequency and time domain information for noise robust. More specifically, our model does not need any noise removing or outliers discarding step. Our model achieves competitive results on two databases compared to the state-of-the-arts. Future work could explore the applications of deep metric leaning for multiple sessions situation, and design specific metric learning loss function for ECG biometrics.

5. REFERENCES

[1] Shun-Chi Wu, Peng-Tzu Chen, A Lee Swindlehurst, and Pei-Lun Hung, "Cancelable biometric recognition with ecgs: subspace-based approaches," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1323–1336, 2018.

[2] Rui Li, Gongping Yang, Kuikui Wang, Yuwen Huang, Feng Yuan, and Yilong Yin, "Robust ecg biometrics using gnmf and sparse representation," *Pattern Recognition Letters*, vol. 129, pp. 70–76, 2020.

[3] Yuwen Huang, Gongping Yang, Kuikui Wang, Haiying Liu, and Yilong Yin, "Learning joint and specific patterns: a unified sparse representation for off-the-person ecg biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 147–160, 2020.

[4] Gokhan Guven, Hakan Gürkan, and Umit Guz, "Biometric identification using fingertip electrocardiogram signals," *Signal, Image and Video Processing*, vol. 12, no. 5, pp. 933–940, 2018.

[5] Jeehoon Kim, Dongsuk Sung, MyungJun Koh, Jason Kim, and Kwang Suk Park, "Electrocardiogram authentication method robust to dynamic morphological conditions," *IET Biometrics*, vol. 8, no. 6, pp. 401–410, 2019.

[6] Hugo Plácido Da Silva, André Lourenço, Ana Fred, Nuno Raposo, and Marta Aires-de Sousa, "Check your biosignals here: A new dataset for off-the-person ecg biometrics," *Computer methods and programs in biomedicine*, vol. 113, no. 2, pp. 503–514, 2014.

[7] Sara S Abdeldayem and Thirimachos Bourlai, "A novel approach for ecg-based human identification using spectral correlation and deep learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 1, pp. 1–14, 2019.

[8] João Ribeiro Pinto, Jaime S Cardoso, and André Lourenço, "Deep neural networks for biometric identification based on non-intrusive ecg acquisitions," in *The Biometric Computing*, pp. 217–234. Chapman and Hall/CRC, 2019.

[9] João Ribeiro Pinto, Miguel V Correia, and Jaime S Cardoso, "Secure triplet loss: Achieving cancelability and non-linkability in end-to-end deep biometrics," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 2, pp. 180–189, 2020.

[10] Kuikui Wang, Gongping Yang, Yuwen Huang, and Yilong Yin, "Multi-scale differential feature for ecg biometrics with collective matrix factorization," *Pattern Recognition*, vol. 102, pp. 107211, 2020.

[11] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.

[12] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei, "Circle loss: A unified perspective of pair similarity optimization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6398–6407.

[13] Jingshan Huang, Binqiang Chen, Bin Yao, and Wangpeng He, "Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.

[14] Zhun Liu, Ying Shen, Varun Bharadhwaj Lakshminarasimhan, Paul Pu Liang, Amirali Bagher Zadeh, and Louis-Philippe Morency, "Efficient low-rank multimodal fusion with modality-specific factors," in *ACL (1)*, 2018.

[15] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *ICCV*. 2015, pp. 1026–1034, IEEE Computer Society.

[17] Saeid Wahabi, Shahrzad Pouryayevali, Siddarth Hari, and Dimitrios Hatzinakos, "On evaluating ecg biometric systems: Session-dependence and body posture," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, 2014.

[18] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[19] Eduardo Jose da Silva Luz, Gladston JP Moreira, Luiz S Oliveira, William Robson Schwartz, and David Menotti, "Learning deep off-the-person heart biometrics representations," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1258–1270, 2017.

[20] Ubaid ur Rehman, Khurram Kamal, Javaid Iqbal, and Muhammad Fahad Sheikh, "Biometric identification through ecg signal using a hybridized approach," in *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence*, 2019, pp. 226–230.