Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11107-4

Cross-view gait recognition based on residual long short-term memory

Junqin Wen[1] · Xiuhui Wang[2]

Received: 16 August 2020 / Revised: 21 May 2021 / Accepted: 25 May 2021 /
Published online: 12 June 2021

Abstract
As a promising biometric recognition technology, gait recognition has many advantages, such as non-invasive, easy to implement in a long distance, but it is very sensitive to the change of video acquisition angles. In this paper, we propose a novel cross-view gait recognition framework based on residual long short-term memory, namely, CVGR-RLSTM, to extract intrinsic gait features and carry out gait recognition. The proposed framework captures dependencies of human postures in time dimension during walking by inputting randomly sampling frame-by-frame gait energy images. The frame-by-frame gait energy images are generated by merging adjacent gait silhouette images sequentially, which integrates gait features of temporal and spatial dimensions to a certain extent. In the CVGR-RLSTM framework, the embedded residual module is used to further refine the spatial gait features, and the LSTM module is utilized to optimize the temporal gait features. To evaluate the proposed framework, we carried out a series of comparative experiments on the CASIA Dataset B and OU-ISIR LP Dataset. Experimental results show that the proposed method reaches the state-of-the-art level.

Keywords Gait classification · Deep learning · Long short-term memory · Residual network

1 Introduction

As a promising human identification approach, video-based gait recognition has many advantages, such as suitable for recognition at a distance and requiring no contact [1, 20]. Nowadays, gait recognition methods have wide application prospects in intelligent surveillance systems. However, although video-based gait recognition has attracted exten-

✉ Xiuhui Wang
wangxiuhui@cjlu.edu.cn

Junqin Wen
wenjq@zjtie.edu.cn

1 Zhejiang Technical Institute of Economics, Hangzhou 310058, China

2 Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University, Hangzhou 310018, China

sive attention from institutes and researchers, it is still in its early stage. There are still many challenges in applying gait recognition methods to practical environments because gait status of one person may be dramatically altered when the view angles are changed.

Fortunately, with successful applications of deep learning technologies in moving object detection [12, 13, 16, 35], visual segmentation and recognition [5, 7, 8, 10, 19] from images and videos, gait recognition is expected to make new breakthroughs with the help of deep learning technology. In this paper, by integrating the strong feature expression and extraction ability of residual network and long short-term memory (LSTM), we propose a novel gait recognition framework based on residual long short-term memory, namely, CVGR-RLSTM, to improve the accuracy of cross-view gait recognition. The proposed framework consists of two modules, i.e. an embedded residual module and a LSTM module. Based on the input frame-by-frame gait energy images (ff-GEIs), we use the embedded residual module to refine the spatial gait features, and utilize the LSTM module to optimize the temporal gait features. Besides, we conduct a thorough comparison and analysis of the proposed method and several other methods on the CASIA Dataset B and OU-ISIR LP Dataset, which demonstrates that our method works well comparing to several existing technologies.

The remaining sections of this paper are organized as follows. In Section 2, related work on gait recognition will be discussed. In Section 3, the proposed CVGR-RLSTM framework will be described on details including its construction, training and testing. Our experiments were carried out based on the CASIA gait dataset and the OU-ISIR gait database to evaluate the proposed algorithms in Section 4. Section 5 will summarize the proposed framework and discuss future work.

2 Related works

According to whether Deep Learning (DL) technologies are used, existing gait recognition approaches can be roughly grouped into two categories: DL-free approaches and DL-based approaches.

DL-free approaches usually utilize well-designed artificial gait features and achieve gait classification using traditional classifiers. They mainly focused on designing efficient gait feature representation or constructing some view transform models. In [9], the novel gait representation, named GEI, was first proposed. As a widely used gait representation, GEI has been employed in many existing gait recognition approaches. Wang et al. [26] proposed a gait recognition algorithm based on Gabor wavelets, which employs a two-dimensional principal component analysis (2D2PCA) method to reduce the dimension of gait features. Besides, the multiclass support vector machine (SVM) was adopted to classify different gaits. Muramatsu et al. [17] presented the VTM-based method to resolve the problem of cross-view gait recognition. The view transformation model was obtained with a set of multiple individuals from multiple views. In the recognition, the model transforms gallery gaits into the same view as that of related input gait, and the gaits can well match each other. In [27], the continuous density hidden Markov models were employed to perform gait recognition. A feature extraction algorithm is firstly proposed based on natural gait cycles and the observation vector set is constructed using the extracted features. Then, the gait vector set extracted from the training sample set is used to estimate the parameters of the proposed gait models. In addition, an adaptive algorithm is introduced based on Cox regression analysis to adaptively adjust parameters of the trained gait model. Boulgouris et al. [3] proposed a gait recognition method based on combining holistic and model-based features. In this method, the holistic features are employed for capturing general gait dynamics, while the model-based features are utilized for capturing more detailed sub-dynamics by refining upon the preceding general dynamics. Muramatsu et al. [17] presented an arbitrary VTM model that can match a pair of gaits from an arbitrary view. In [18], a VTM incorporating a score normalization framework was proposed, which calculates two quality measures from a pair of gait features, and then uses those to compute the posterior probability that both gait features originate from the same subjects together with the biased dissimilarity. In [25], a cross-view gait recognition method based on ensemble learning was proposed for gait recognition. In this approach, utilizes a well-designed gait feature based on area average distance to preserve the proximity relationships among instance triplets. Deng et al. [4] presented a gait recognition algorithm combining spatial-temporal and kinematic gait features, in which the binary silhouettes are characterized with four time-varying spatial-temporal parameters. Furthermore, by using deterministic learning, spatial-temporal gait features can be represented as the gait dynamics underlying the trajectories of lower limbs silhouette widths and holistic silhouette area. Tang et al. [23] proposed a method for gait partial similarity matching that assumes a 3D object shares common view surfaces in significantly different views. In short, DL-free gait recognition methods usually used traditional machine learning technologies to extract artificially view-independent gait features and achieve gait classification. This type of method generally does not require a lot of training data to estimate the model parameters, so it is applied in many practical applications. However, in the case where there is huge amount of training data, the recognition performance of this method is not as good as those of deep learning-based methods.

On the other hand, DL-based approaches generally use deep learning techniques in the gait classification process. A comprehensive review was presented in [32] with respective to a cross-view and cross-walking condition, also with various pre-processing methods and CNN network architectures. Besides, three network architectures, namely, LB, MT (Mid-Level @ Top), and GT (Global @ Top) were presented. The MT network is similar with the LB network, except that two extra nonlinear projections are applied before comparing the differences between the GEI pairs. There is no significant difference found between the LB and MT networks in our actual tests. The GT network suffers from severe overfitting and has less satisfactory performance. Takemura et al. [22] proposed an input and output architecture for cross-view gait recognition based on convolution neural network, which discussed the verification and identification problems with different subjects and views. The Siamese network with a pair of inputs and contrastive loss is used for verification, a triplet network with a triplet of inputs and triplet ranking loss is used for identification. Wolf et al. [31] presented a deep CNN network for gait recognition in multiple views capturing spatial-temporal features. Shiraga et al. [21] proposed a CNN-based gait recognition approach, which uses gait energy image (GEI) as its input. Wang et al. [28] proposed a gait recognition scheme by utilizing the strong expression of convolution neural networks, which constructs a multichannel CNN network to tackle a set of sequential images in parallel. Thapar and Nigam [24] proposed a 3D Convolution Deep Neural Network for person identification using cross-view gait data. Feng et al. [6] presented a LSTM-based gait feature learning method, which employs heatmaps extracted by a CNN-based pose estimate method to describe gait information. Yu et al. [34] proposed a gait feature extraction method based on generative adversarial networks (GAN). This method takes each GAN model as a regressor to create invariant gait images under changing conditions. Yu et al. [36] employed one deep model based on auto-encoder for invariant gait extraction. The obtained model can combine gait features in a progressive way by stacked multi-layer auto-encoders, which can extract invariant gait feature using only one model. Watanabe and Kimura [30] applied RNN and LSTM to 3-axis accelerations of walking acquired by a smartphone for gait identification and authentication. Wang and Yan [29] developed a specialized deep RNN architecture and a variation of gait energy images for cross-view gait recognition, which can deal with relatively small datasets without using any augmentation or fine-tuning techniques. Battistone and Petrosino [2] proposed a graph-based learning approach, called TGLSTM network, which can dynamically learn graphs when they may change during time. In summary, existing DL-based gait recognition methods are mainly based on convolutional neural networks (CNNs), and are less involved with recurrent neural networks (RNNs). These methods make full use of the ability of the convolution operation to extract and represent gait features, but ignore the temporality between adjacent gaits.

The algorithm in this paper belongs to the second category. Different from existing DL-based methods, by utilizing the powerful feature express capacity of ffGEIs [29], we propose a novel cross-view gait recognition framework based on residual network and long short-term memory to extract intrinsic gait features and carry out gait recognition.

3 Methodology

Gait feature extraction and classifier design are the key components in process of gait recognition. In this paper, on the basis of previous work, we use ffGEI [29] as the representation of gait features, and focus on the design of gait classifier.

As shown in Fig. 1, the proposed framework mainly includes five steps, 1) Generate primary gait features: Through segmentation, normalization and Combination, ffGEIs are generated from the input gait silhouette image sequences; 2) Enhance spatial gait features: The residual module is mainly used to enhance spatial gait features.The shortcut connection mechanism is used to make the gradient propagation lossless when the network is deepened, and the space size of the feature map is reduced by half by the down sampling operation; 3) Extract spatiotemporal gait features: The LSTM module is used to optimize the temporal components of ffGEIs and extracts comprehensive spatiotemporal gait features; 4) Train residual LSTM (Res-LSTM) model: In this step, the extracted spatiotemporal gait features are used to train the Res-LSTM network model, and the trained Res-LSTM model is obtained; 5) Gait classification and recognition: The trained Res-LSTM model is used to classify the test data, and we obtain the average classification accuracy. Since the construction process of ffGEI has been described in related work [29], and the traditional back

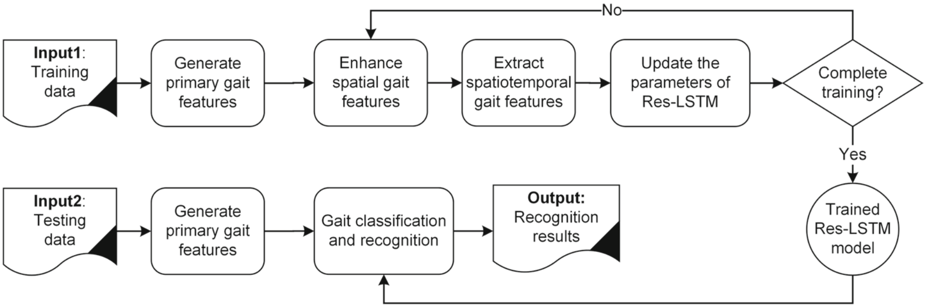

Fig. 1 The overview architecture of the proposed CVGR-RLSTM framework

propagation through time (BPTT) method [15] is used in the model training steps, the following sections will focus on how to extract spatiotemporal gait features through residual module and LSTM module.

3.1 Architecture of the Res-LSTM network

Our Res-LSTM Network consists of three modules, as shown in Fig. 2, a residual block and two LSTM blocks. After inputting ffGEIs as primary gait features, the spatial components are enhanced by the residual block. Then, two LSTM blocks are used to extract the spatiotemporal gait features in turn. By adding a shortcut connection, we can get the counterpart residual version of a general neural network, i.e. a residual network. Typically, a residual block is composed of two branches, shortcut connection and main branch, which can be defined as:

$$y = d(x) + f(x, W), \tag{1}$$

where y is the total output of the residual block, $d(x)$ is obtained by the shortcut connection branch, $f(x, W)$ is the output component of the main branch, and W is the weight coefficient of the main branch.

Compared with the existing methods, we add two special layers into the Res-LSTM network: 1×1 convolutional layers and batch normalization layers. A 1×1 convolutional layer is used to match dimensions of different paths. The essence of a 1×1 convolutional layer is to multiply each input channel by its convolutional weight and add them together, which is equivalent to "connecting" the independent channels in the input. No matter how many channels are in the input, the number of output channels is always equal to the number of convolutional kernels.

Batch normalization layers are to force the distribution of neuron input values back to the standard normal distribution with mean value of 0 and variance of 1 by certain normalization means, so that the input value of activation function falls in the region where nonlinear function is more sensitive to input [33]. In this way, small change of input will lead to large changes of loss function, which means that the gradient becomes larger and avoids the problem of gradient disappearance. At the same time, the larger the gradient means that the convergence speed of learning is fast, which can greatly speed up the training speed.

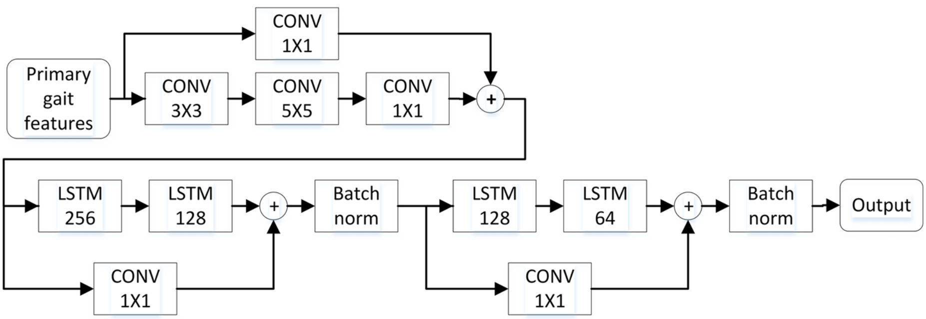

Fig. 2 Architecture of the proposed Res-LSTM model

3.2 Res-LSTM model training

Res-LSTM is the residual version of plain LSTM network. Before introducing the training process of Res-LSTM, we briefly introduce the structure of a plain LSTM network. As shown in Fig. 3, in each LSTM cell, there are three gates, the forget gate, the input gate, and the output gate.

The forget gate determines what information will be ignored, as defined by:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{2}$$

where $\sigma(\cdot)$ is the activation function, h_{t-1} is the output of the (t-1)-th time, x_t is the input of t-th time, W_f and b_f are the weight and bias of the forget gate respectively, and f_t is the output of the forget gate.

The input gate determines which new information remains in the cell state and updates the cell state, as following:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{3}$$

$$a_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * a_t, \tag{5}$$

where $\sigma(\cdot)$ and $tanh(\cdot)$ are the activation functions, W and b are the weights and biases of different units cell respectively, C_{t-1} is the status of (t-1)-th time, i_t is the output of the input gate, a_t is the input transform unit, and C_t is the status of current LSTM cell.

The output gate determines what information will be output, and is defined as:

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{6}$$

$$h_t = O_t * tanh(C_t), \tag{7}$$

where $\sigma(\cdot)$ and $tanh(\cdot)$ are the activation functions, W_o and b_o are the weight and bias of the output gate respectively, O_t is the output of the output gate, and h_t is the output of current LSTM cell.

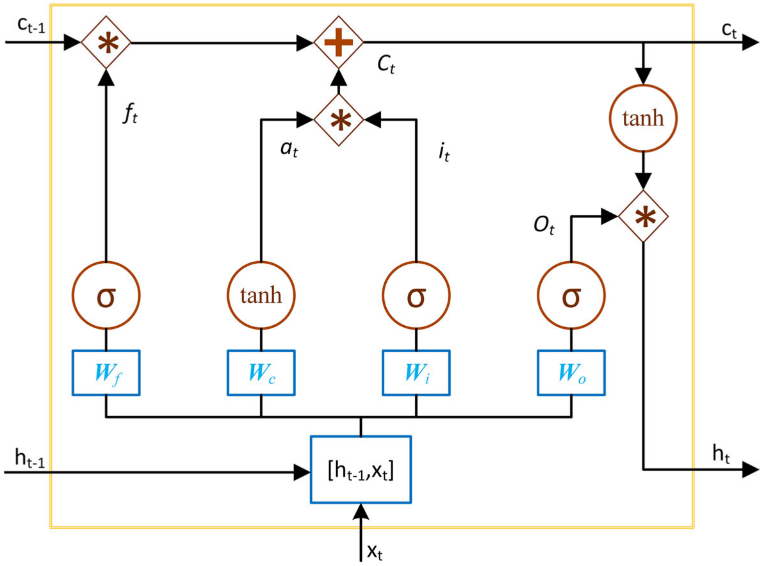

Fig. 3 The internal architecture of a LSTM cell

Next, we will compare the training process of Res-LSTM with plain LSTM. For the residual LSTM network, the forward propagation is linear, and the output of each layer is equal to its input plus the residual [11].

Given a Res-LSTM with N layers, and assuming that p_i is the weight of the p-th 1×1 convolutional layer in the shortcut connection i, then, the input of the $(i+1)$-th layer can be calculated by:

$$x_{i+1} = p_i \cdot x_i + f(x_i, W_i), \tag{8}$$

where W_i is the weight of the residual part in the i-th layer. In the forward propagation, we can calculate the input of the $(l+1)$-th layer by:

$$x_{l+1} = \left(\prod_{i=1}^{l} p_i\right) \cdot x_1 + \sum_{i=1}^{l} f(x_i, W_i), \tag{9}$$

where $l \in [1, N-1]$, W_i and p_i are the weights of residual and shortcut parts respectively.

In the back propagation, the cost function is represented by δ, and its gradient $\frac{\partial \delta}{\partial x_i}$ is decomposed into two components:

$$\frac{\partial \delta}{\partial x_i} = \frac{\partial \delta}{\partial x_{i+1}} \cdot p_i + \frac{\partial \delta}{\partial x_{i+1}} \cdot \frac{\partial f(x_i, W_i)}{\partial x_i)}, \tag{10}$$

where the first item propagates through the residual layers, and the later directly propagates along the shortcut connections.

3.3 Gait classification based on the proposed Res-LSTM

The Res-LSTM network training and gait classification are obtained through all steps in Algorithm 1.

Algorithm 1 Gait classification algorithm based on Res-LSTM.

Input: Gait silhouette images of different subjects.
Output: Average classification correct rate.
Step 1: Generate primary gait features. Construct ffGEIs from given gait silhouette images and form a training dataset G_{tr} and a testing dataset G_{te}. In our experiments, we set the span for constructing ffGEIs to 5, which means that when given N gait silhouette images, we can get up to N - (5-1) ffGEIs.
Step 2: Initialize the Res-LSTM network. The initial parameter values of Res-LSTM are randomly generated by normal distribution.
Step 3: Use the training dataset to train the Res-LSTM. This step consists of four processes: 1) By feeding ffGEIs in G_{tr} into the Res-LSTM network, we calculate the output of each layer and obtain the final predictive result; 2) According to the loss function, we can calculate the total loss C of the Res-LSTM network, and stop training if C is less than the given threshold; 3) The error term is backward propagated through each layer of the Res-LSTM network and the related gradients of each weight and bias are computed; 4) Update weight and bias of the Res-LSTM.
Step 4: Conduct gait classification using the testing dataset G_{te}. After the training process has been finished, we conduct gait classification with samples in G_{te}, and obtain the related classification results.
Step 5: Compute and return average classification correct rate.

In the proposed gait recognition scheme, the residual LSTM network is formed by introducing shortcut connections, and the subsequent layers can directly combine the output of the shallow layers which are not adjacent. Compared with the classical LSTM network, the shortcut branch is equivalent to a simple identity mapping implementation, which does not produce additional parameters and thus reduce the computational complexity.

4 Experiments

In this section, we presented two comparative experiments for showing the performance of the proposed framework. The first experiment was used to evaluate the correct recognition rate under cross-view condition. The second one was to evaluation the generalization ability of the proposed gait recognition method using large-scale dataset. Besides, we used Python 3.6 programming language and scikit-learn 0.22 Toolkit to implement all the experiment on Dell precision T7820 with two 5220R CPU and 256G memory.

Several existing gait recognition methods were selected in our experiments to demonstrate the performance of the proposed approach. In addition, in order to facilitate the quantitative analysis of experimental results, we used cumulative match characteristic (CMC) as the evaluation criterion in our experiments, which is a well-known measurement to evaluate the classification capabilities of a recognition system.

4.1 Gait databases

The first gait dataset used in our experiments is the CASIA Dataset B [37], which is kindly provided by the Institute of Automation Chinese Academy of Sciences and public available. CASIA Dataset B is a large cross-view gait dataset, which is created in January 2005. There are totally 124 persons, and the gait data was captured from 11 views, as shown in Fig. 4. Three conditions namely view angle, clothing and carrying condition changes, are separately considered. There were 93 males and 31 females, 123 Asians and 1 European among all subjects. Most subjects were young people and they aged between 20 and 30.

Another dataset we used is the OU-ISIR Large Population Dataset [14], which consists of 62,528 subjects (with age ranging from 2 to 95 years), as shown in Fig. 5. The camera was set at a distance of approximately 8 meters from the straight walking course and a height of

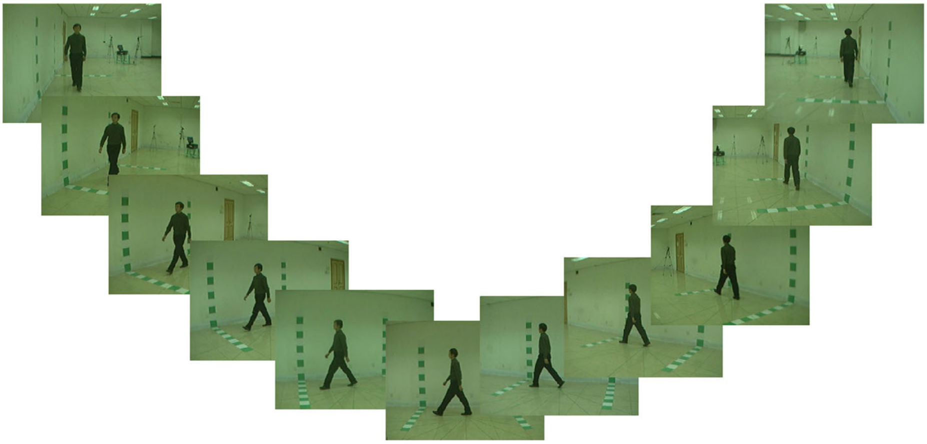

Fig. 4 Examples from CASIA Dataset B

Fig. 5 Examples from OU-ISIR LP dataset

approximately 5 meters. The image resolution and frame rate were 1280 × 980 pixels and 25 fps, respectively. Each subject was asked to walk straight three times at his/her preferred speed. The first sequence with or without COs (if he/she did not have COs) is called the A1 sequence, and the second and third sequences without COs are called A2 and A3 sequences, respectively.

4.2 Experiment on CASIA Dataset B

This comparative experiment was conducted on CASIA Dataset B and all the gait samples were randomly grouped into two sets, training set and test set according to the ratio of 80% and 20%. Three deep learning-based methods are involved: basic LSTM, Conv-LSTM [29], and LB-CNN [32]. Furthermore, to comply with related literatures, ffGEIs were used as inputs of all these methods. The experimental results are shown as in Fig. 6, in which the horizontal axis is the rank and the vertical axis is the recognition rate respectively.

Figure 6 shows that the proposed Res-LSTM method performs better than other methods in terms of correct recognition rate. Firstly, compared with the methods that don't

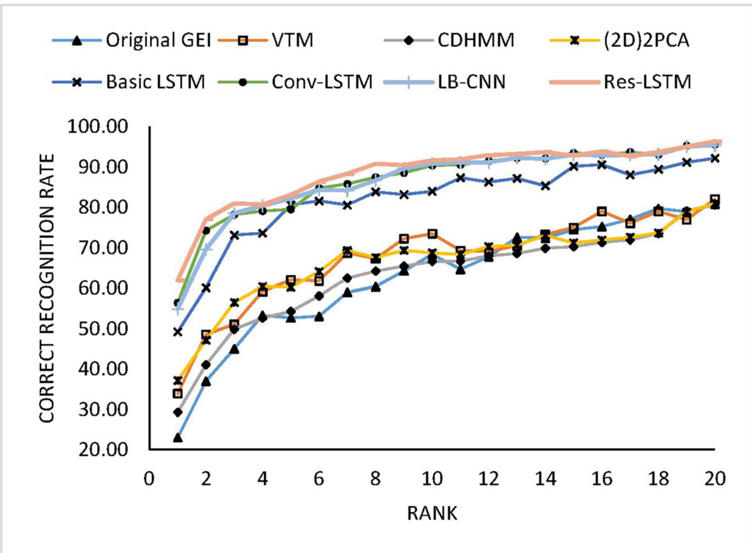

Fig. 6 CMC curves of several methods on CASIA Dataset B

use deep learning technologies, our method has great advantages in correct recognition rate. For example, our correct recognition rate at Rank 1 is almost 63%, while the correct recognition rates of most DL-free methods are under 38%. Secondly, our method performs better than basic LSTM in terms of correct recognition rate, when the rank value is less than 6. The main reason is that the basic LSTM only preserves information of the past because the only inputs it has seen are from the past; while Res-LSTM will operate the inputs through two paths, the residual path and the shortcut path.

4.3 Experiment on OU-ISIR large population dataset

In this section, the generalization capability of the Res-LSTM-based method was assessed on the OU-ISIR Large Population Dataset. Three deep learning-based methods are involved: basic LSTM, Conv-LSTM [29], and input/output [22]. All the subjects are randomly grouped into training set and test set by using the ratio of 80% and 20%. In each run, we train the Res-LSTM network using the training set and keep another for testing. Figure 7 is the CMC curves of eight different approaches.

Figure 7 shows that the proposed method greatly outperforms than those DL-free methods. The main reason is that Res-LSTM can better explode the intrinsic differences between different human gaits. Furthermore, Fig. 7 also shows that the proposed Res-LSTM-based approach greatly improves the scores from the basic LSTM. This owns to the fact that, when forward operations are conducted, the Res-LSTM network can preserve information from both the residual path and the shortcut path.

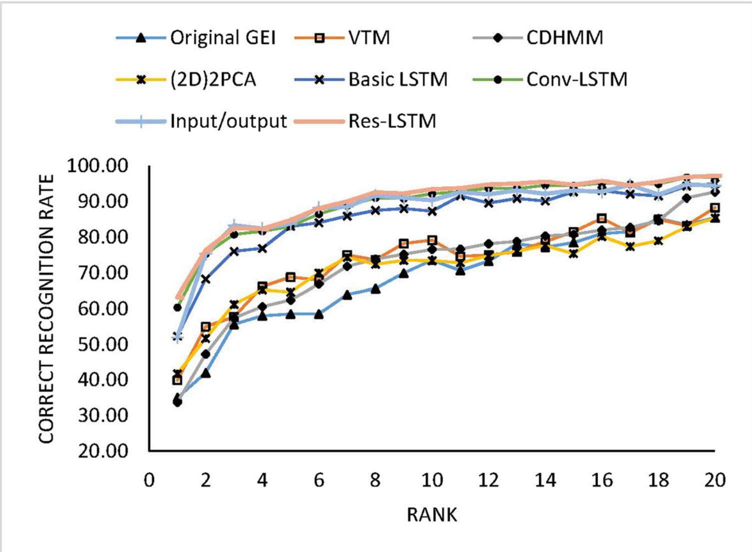

Fig. 7 CMC curves of several methods on OU-ISIR large population dataset

5 Conclusion, discussion and future work

In this work, a new gait recognition model, named Res-LSTM, was presented to take full use of the inherent feature expression ability of LSTM and temporal characteristics of human walking. Our contributions are that: (1) propose a new gait recognition method based on Res-LSTM to extract high-discriminative gait features and achieve gait classification, and (2) greatly advance the correct recognition scores on the CASIA Dataset B and OU-ISIR LP Dataset.

The limitation of our method is that, like most existing methods, only laboratory data is used as input of the proposed network, and only a few fixed sampling angles are considered in the experimental results. How to solve the problem of gait recognition in real environment with arbitrary angle changes of view is one of our future works.

Acknowledgements This research was funded by the National Natural Science Foundation of China, grant number No. 61602431 and the Natural Science Foundation of Zhejiang Province, grant number No.Y20F020113.

References

1. Abdulsattar F, Carter J (2016) Performance analysis of gait recognition with large perspective distortion. In: IEEE International Conference on Identity, Security and Behaviour Analysis. Sendai, Japan, pp 1–6
2. Battistone F, Petrosino A (2019) Tglstm: A time based graph deep learning approach to gait recognition. Pattern Recogn Lett 126:132–138
3. Boulgouris NV, Huang X (2013) Gait recognition using hmms and dual discriminative observations for sub-dynamics analysis. IEEE Trans Image Process 22(9):3636–3647
4. Deng M, Wang C, Cheng F, Zeng W (2017) Fusion of spatial-temporal and kinematic features for gait recognition with deterministic learning. Pattern Recogn 67:186–200
5. Ercolano G, Rossi S (2021) Combining cnn and lstm for activity of daily living recognition with a 3d matrix skeleton representation. Intel Serv Robot 14:1–11
6. Feng Y, Li Y, Luo J (2016) Learning effective gait features using lstm. In: 23rd International conference on pattern recognition. Cancun, pp 325–330
7. Gao Z, Nie W, Liu A, Zhang H (2016) Evaluation of local spatial-temporal features for cross-view action recognition. Neurocomput. 173(P1):110–117
8. Gao Z, Xuan H, Zhang H, Wan S, Choo KR (2019) Adaptive fusion and category-level dictionary learning model for multiview human action recognition. IEEE Internet of Things Journal 6(6):9280–9293
9. Han J, BirBhanu (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322
10. Han J, Pauwels E, Zeeuw P (2012) Employing a rgb-d sensor for real-time tracking of humans across multiple re-entries in a smart environment. IEEE Trans Consum Electron 58:255–263
11. He K, Zhang X, Ren S (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition, pp 770–778. Las Vegas, NV
12. Hong C, Yu J, Wan J, Tao D, Wang M (2015) Multimodal deep autoencoder for human pose recovery. IEEE Trans Image Process 24(12):5659–5670
13. Hong C, Yu J, Zhang J, Jin X, Lee K (2019) Multimodal face-pose estimation with multitask manifold deep learning. IEEE Transactions on Industrial Informatics 15(7):3952–3961
14. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inf Forensics Secur 7(5):1511–1521
15. Kai C, Qiang H (2016) Training deep bidirectional lstm acoustic model for lvcsr by a context-sensitive-chunk bptt approach. IEEE-ACM Transactions on Audio Speech and Language Processing 24(7):1185–1193
16. Le Cun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(5):436–445
17. Muramatsu D, Shiraishi A, Makihara Y, Uddin MZ, Yagi Y (2015) Gait-based person recognition using arbitrary view transformation model. IEEE Trans Image Process 24(1):140–154

18. Muramatsu D, Makihara Y, Yagi Y (2016) View transformation model incorporating quality measures for cross-view gait recognition. IEEE Trans Cybern 46(7):1602–1615
19. Perrett T, Damen D (2019) Ddlstm: Dual-domain lstm for cross-dataset action recognition, pp 7844–7853
20. Sarkar S, Phillips PJ, Liu Z, Vega IR, Grother P, Bowyer KW (2005) The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27(2):162–177
21. Shiraga K, Makihara Y, Muramatsu D, Echigo T, Y. Yagi (2016) VGeinet: View-invariant gait recognition using a convolutional neural network. In: International conference on biometrics. Halmstad, pp 1–8
22. Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2018) On input/output architectures for convolutional neural network-based cross-view gait recognition. IEEE Trans Circuits Syst Video Technol 28(1):1–13
23. Tang J, Luo J, Tjahjadi T, Guo F (2017) Robust arbitrary-view gait recognition based on 3d partial similarity matching. IEEE Trans Image Process 26(1):7–22
24. Thapar D, Nigam A, Aggarwal D, Agarwal P (2018) Vgr-net: A view invariant gait recognition network. In: IEEE 4th International conference on identity, security, and behavior analysis. Singapore, pp 1–8
25. Wang X, Yan WQ (2020) Cross-view gait recognition through ensemble learning. Neural Comput and Applic 32(11):7275–7287
26. Wang X, Wang J, Yan K (2018) Gait recognition based on gabor wavelets and $(2D)^2 PCA$. Multimed Tools Appl 77(10):12545–12561
27. Wang X, Feng S, Yan WQ (2019) Human gait recognition based on self-adaptive hidden markov model. IEEE Transactions on Computational Biology and Bioinformatics, pp 1–12
28. Wang X, Zhang J, Yan WQ (2019) Gait recognition using multichannel convolution neural networks. Neural Computing and Applications
29. Wang X, Yan WQ (2020) Human gait recognition based on frame-by-frame gait energy images and convolutional long short term memory. Int J Neural Sys 30(1):1950027
30. Watanabe Y, Kimura M (2020) Gait identification and authentication using lstm based on 3-axis accelerations of smartphone. Procedia Computer Science 176:3873–3880
31. Wolf T, Babaee M, Rigoll G (2016) Multi-view gait recognition using 3d convolutional neural networks. In: IEEE International conference on image processing. Phoenix, AZ, pp 4165–4169
32. Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A comprehensive study on cross-view gait based human identification with deep cnns. IEEE Trans Pattern Anal Mach Intell 39(2):209–226
33. Yanghao L, Naiyan W, Jianping S, Xiaodi H, Jiaying L (2018) Adaptive batch normalization for practical domain adaptation. Pattern Recognition
34. Yu S, Chen H, Reyes EBG, Poh N (2017) Gaitgan: Invariant gait feature extraction using generative adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, pp 532–539
35. Yu J, Tan M, Zhang H, Tao D, Rui Y (2019) Hierarchical deep click feature prediction for fine-grained image recognition. IEEE Trans Pattern Anal Mach Intell, pp 1–1
36. Yu S, Chen H, Wang Q, Shen L, Huang Y (2017) Invariant feature extraction for gait recognition using only one uniform model. Neurocomputing 239:81–93
37. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: International conference on pattern recognition, pp 441–444. Hong Kong, China