Human Gait Recognition Based on Frame-by-Frame Gait Energy Images and Convolutional Long Short Term Memory

Xiuhui Wang
China Jiliang University, Hangzhou 310018, China
E-mail: wangxiuhui@cjlu.edu.cn

Wei Qi Yan
Auckland University of Technology, Auckland 1010, New Zealand
E-mail: weiqi.yan@aut.ac.nz

Human gait recognition is one of the most promising biometric technologies, especially for unobtrusive video surveillance and human identification from a distance. Aiming at improving recognition rate, in this paper we study gait recognition using deep learning and propose a novel method based on convolutional Long Short-Term Memory (Conv-LSTM). Firstly, we present a variation of Gait Energy Images, i.e., frame-by-frame GEI (ff-GEI), to expand the volume of available GEI data and relax the constraints of gait cycle segmentation required by existing gait recognition methods. Secondly, we demonstrate the effectiveness of ff-GEI by analyzing the cross-covariance of one person's gait data. Then, taking use of the temporality of our human gait, we design a novel gait recognition model using Conv-LSTM. Finally, the proposed method is evaluated extensively based on the CASIA Dataset B for cross-view gait recognition, furthermore the OU-ISIR Large Population Dataset is employed to verify its generalization ability. Our experimental results show that the proposed method outperforms other algorithms based on these two datasets. The results indicate that the proposed ff-GEI model using Conv-LSTM, coupled with the new gait representation, can effectively solve the problems related to cross-view gait recognition.

Keywords: Gait classification; deep learning; Long Short-Term Memory (LSTM); frame-by-frame GEI(ff-GEI).

1. Introduction

Compared to traditional biometrics, such as face and fingerprint, etc., gait has tremendous advantages, e.g., non-offensive, remote-suitable, and easy to be collected.[1] There has been rapid research progress in gait recognition over the last decade.[2–4] However, there are still a myriad of challenges in gait recognition, e.g., cross-view, different clothing, and multiple carryings, which make gait recognition become tough. This dilemma is basically due to extract essential features from human gait. Benefited from recent development in deep learning,[5] nowadays the problems are being resolved from numerous aspects, such as convolutional neural networks (CNNs) ,[6–9] deep Bayesian networks ,[10–13] Boltzmann machine ,[14–17] encoder–decoder networks ,[18,19] ensemble methods, and others.[20–22]

In this paper, we propose a novel gait recognition method based on Long Short-Term Memory (LSTM)[23] with convolutional layers, namely, Conv-LSTM. Compared with Markov-based methods, LSTM excludes strong hypothesis, it can retrieve longer time-series sequence. To the best of our knowledge, this is the first time that this approach is proposed for gait recognition, our contributions are detailed as follows:

- *A new gait representation, i.e., frame-by-frame GEI (ff-GEI).* The new representation can expand the available dataset and relax the constraints of gait cycle segmentation required by existing gait recognition methods.
- *A novel Conv-LSTM network model for gait recog-*

nition. By utilizing the intrinsic temporality of human gait, we propose a novel neural network Conv-LSTM to achieve gait classification and recognition, which adds convolutional layers to the traditional LSTM model.

- *The proposed method greatly improves the recognition rates based on two open-accessed gait databases.* We conduct thorough comparison and analysis of the proposed method and multiple existing methods based on the CASIA Dataset B and OU-ISIR Large Population Dataset, we demonstrate that our method performs well comparing to the existing approaches.

The remaining sections of this paper are organized as follows. In Section 2, our related work on gait recognition will be reviewed. In Section 3, the new gait feature representation ff-GEI will be presented on details; then a novel gait recognition method Conv-LSTM will be presented to achieve gait classification and recognition. Our experiments are fulfilled based on the CASIA gait database and the OU-ISIR gait database so as to evaluate the proposed algorithms in Section 4. Finally, the conclusion and future work of this paper will be remarked in Section 5.

2. Related work

Followed the research trends in deep learning (DL), most of the existing approaches for gait recognition can be classified into two categories: DL-free methods and DL-based methods. The former takes use of traditional machine learning ways to extract specified gait features and achieve the pattern classification; while the latter is end-to-end oriented, utilizes deep neural networks and huge training data for gait recognition.

The traditional gait recognition methods mainly focus on one or more directions in gait recognition: well-designed gait features and related dimensionality reduction,[24–30] 3D reconstruction methods,[31–35] ingenious view-invariant gait features,[36,37] view transform models.[38–40] In the work,[24] feature presentation of Gait Energy Images (GEI) was first proposed, which was computed by using properly aligned human silhouettes in gait sequences. This gait feature presentation and its varieties[25,26] are employed in a slew of subsequent research literatures[2,3,27–30] In the work,[31–35] 3D structure of each gait was reconstructed so as to generate arbitrary 2D views by projecting the 3D model.

Generally, those methods can achieve promising gait recognition rate, but they are required to capture data using multiple calibrated cameras which are unavailable in most practical applications. The use of artificial view-invariant features for gait recognition,[36,37] which can perform very well in specific cases, but is hard to be generalized for more applications. It needs to learn projection transformation,[38,40] with which one can transform gait features from different views to one or more common views. Those methods are based on comparing the normalized gait features extracted from any two videos to calculate the similarity between them. In a word, traditional methods can alleviate the influence of various external factors (such as view variations and clothing differences), however, there lack effective feature extraction and modeling methods to solve the highly nonlinear correlation problem between gait features in complex cross-view conditions.

On the other hand, DL-based gait recognition methods[41,42] mainly focus on convolutional neural network (CNN). An extensive evaluation was proposed[41] in terms of a cross-view and cross-walking condition, with various preprocessing methods and CNN network architectures. The input and output were designed[43] for CNN-based cross-view gait recognition. In a project,[42,44] multiple special CNN networks were designed to address the problem of cross-view gait recognition. A gait recognition method[45] was developed based on a deep CNN network with 3D convolutions, which used a specified input including both grayscale images and optical flow enhanced color invariance.

The method proposed in this paper belongs to the second category. Different from the existing work, we investigate gait recognition using Long Short-Term Memory (LSTM), which is a special type of Recurrent Neural Networks (RNN) having excellent advantages in handing sequential data. Considering that human gait is time series-based and each gait is affected by its previous status, this is the main reason why we select LSTM instead of other deep learning models.

In addition, traditional methods for gait representation cannot preserve spatiotemporal relationship in each gait sequence completely, the relevant contour synthesis of the original data results in sig-

nificant reduction in the available gait datasets. This will definitely affect the training process of deep neural network and lead to overfitting. Thus, we present a new representation of gait features, i.e., ff-GEI, which has the superior advantages to original silhouettes GEIs. The ff-GEI can express the available gait data and relax the constraints of gait cycle segmentation required by the existing gait recognition methods.

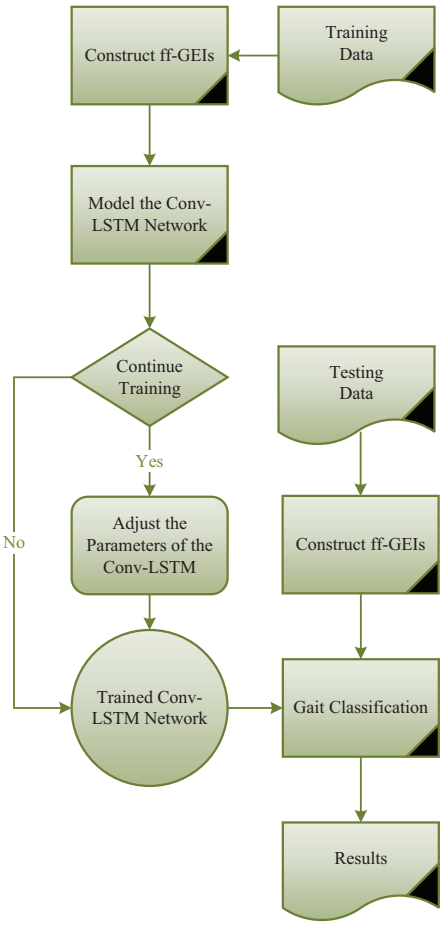

Figure 1. The diagram of Conv-LSTM-based gait recognition.

3. Our method

Gait feature representation and classifier design are the crucial steps in gait recognition. In this section, we describe our proposed approach, which uses ff-GEI as the input and takes full use of deep learning technology to provide effective gait classification. The ff-GEI is a variation of GEI aiming at expanding available and trainable gait data. Besides, considering the intrinsic temporality of human gait and the capability of LSTM for sequential data processing, we design a Conv-LSTM-based gait classifier and train it using ff-GEIs. The proposed solution mainly includes three steps: 1) training and test ff-GEIs, 2) modeling and training the Conv-LSTM network, and 3) classifying the test data using the trained Conv-LSTM. The diagram of this proposed method is shown in Fig. 1.

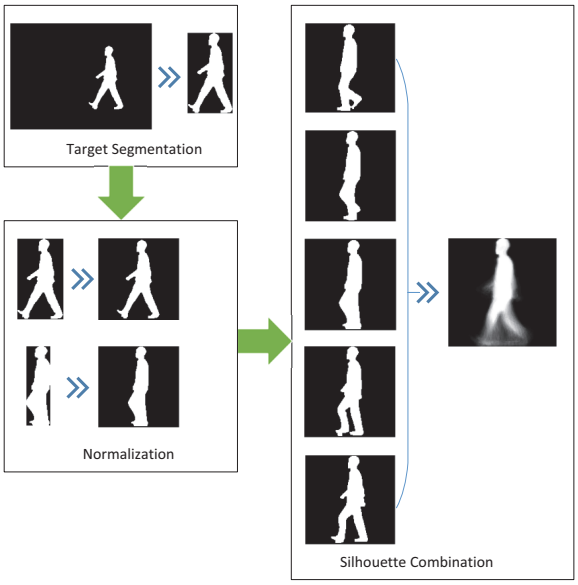

Figure 2. The diagram of Conv-LSTM-based gait recognition.

3.1. *The ff-GEI construction*

Currently, most public gait databases, such as CASIA and OU-ISIR, provide completed silhouette sequences that have been extracted from original human walking videos. In this paper, our focus is on how to preprocess the videos and feed the relevent silhouette sequences into our gait classifier effectively. As shown in Fig. 2, the procedure of ff-GEI construction includes three steps:

(a) *Segmentation.* We use filtering techniques as well as erosion and dilation operations in morphological operations to handle the original silhouette sequence. The silhouette region of a visual object is segmented from each preprocessed image and converted into a binary image.

(b) *Normalization.* The normalization includes two operations: size normalization and center alignment. The first operation aims to proportionally resize all the silhouette images into a standard

scale. In the resizing process, instead of altering the size of silhouette regions, we modify the size of image templates. The alignment operation aligns each silhouette region at the center of these templates.

(c) *Combination.* The normalized templates are combined with the silhouette, finally we get the ff-GEI gait representation.

Given the segmented silhouette sequence $I_t(u, v)$ at time t, ff-GEI is defined as

$$F_t(u, v) = \frac{1}{2m+1} \sum_{t=i-m}^{i+m} I_t(u, v), \tag{1}$$

where m is the frame number of a half of gait period from which we get the cross-relationship information, t is the frame number in current sequence, i is the middle point of current period, and (u, v) is the 2D image coordinate.

As ff-GEI has the attributes of basic GEI, ff-GEI is less sensitive to silhouette noises than the original representation of silhouette images.[24] Compared with the basic GEI, ff-GEI only adds historical information to each frame in the original silhouette sequence and does not significantly reduce the number of gait training samples; thus, it is more suitable for training deep neural networks which usually need massive data to ensure the accuracy of training results.

3.2. *Conv-LSTM modeling and training*

Our Conv-LSTM network consists of three convolutional layers, three pooling layers, one fully connected (FC) layer, three LSTM layers, and one softmax layer as shown in Fig. 3. Firstly, along with the walking direction, ff-GEIs are fed into three 2-tuple sets (convolutional layer, pooling layer). A max pooling is used for the pooling layers, which create an invariance to small shifts and distortions. Secondly, the local feature maps extracted from convolutional layers are reassembled into an entire graph through a fully connected layer. ReLU activation functions are employed in all the convolutional layers and the fully connected layer to reduce computational complexity and alleviate gradient vanishing problem. Then, the refined features are converted into vector sequences which are applied as the input of three LSTM layers to be further optimized. Finally, a softmax layer is employed to generate final classification results. We use a softmax function and cross entropy to predict the gait. In addition, for eliminating the joint adaptability of neuron nodes and enhancing the generalization ability of the final model, both the fully connected layers and LSTM layers adopt a dropout technique.[46]

In the training phase, we use backward propagation[47] (BP) to compute the gradient of weight and bias of neuron as well as update the related weights and biases by using stochastic gradient descent[48] (SGD). The training process consists of two iterative steps: forward pass for training and backward pass for reducing the cost.

Suppose there are N ff-GEI samples corresponding to M individuals in the training dataset, our Conv-LSTM network consists of P layers in total, then the loss function is defined as

$$L = -\frac{1}{N} \cdot \sum_{n=1}^{N} \left[\hat{y}_n log y_n + (1 - \hat{y}_n) log(1 - y_n) \right], \tag{2}$$

where y_n is the predictive output of the Conv-LSTM network, $\hat{y}_n$ is the corresponding label. In the first step, we calculate the output y^p $(p = 2, 3, \cdots, P)$ of each layer by using

$$y^p = f(z^p) = f(w^p y^{p-1} + b^p), \tag{3}$$

where w^p is weight of the p-th layer, b^p is bias of the p-th layer, and f is the general activation function in the p-th layer.

When the output of a Conv-LSTM network does not coincide with the expected value, a backpropagation operation is conducted. Then, we compute the final error term δ^P between the output result and the expected value, feed δ^P back into the network and obtain the error term with respect to each layer

$$\delta_i^P = \frac{\partial L}{\partial y_i^P} f'(z_i^P) \tag{4}$$

and

$$\delta_i^p = \sum_j (w_{j,i}^{p+1} \delta_j^{p+1} f'(z_i^p)), \tag{5}$$

where δ_i^p is error term of the i-th neuron in the p-th layer, z_i^p is weighted input of the i-th neuron in the p-th layer, y_i^P is output of the i-th neuron in the P-th layer, $w_{j,i}^{p+1}$ is weight on the connection from the i-th neuron in the p-th layer to the j-th neuron in the $(p+1)$-th layer.

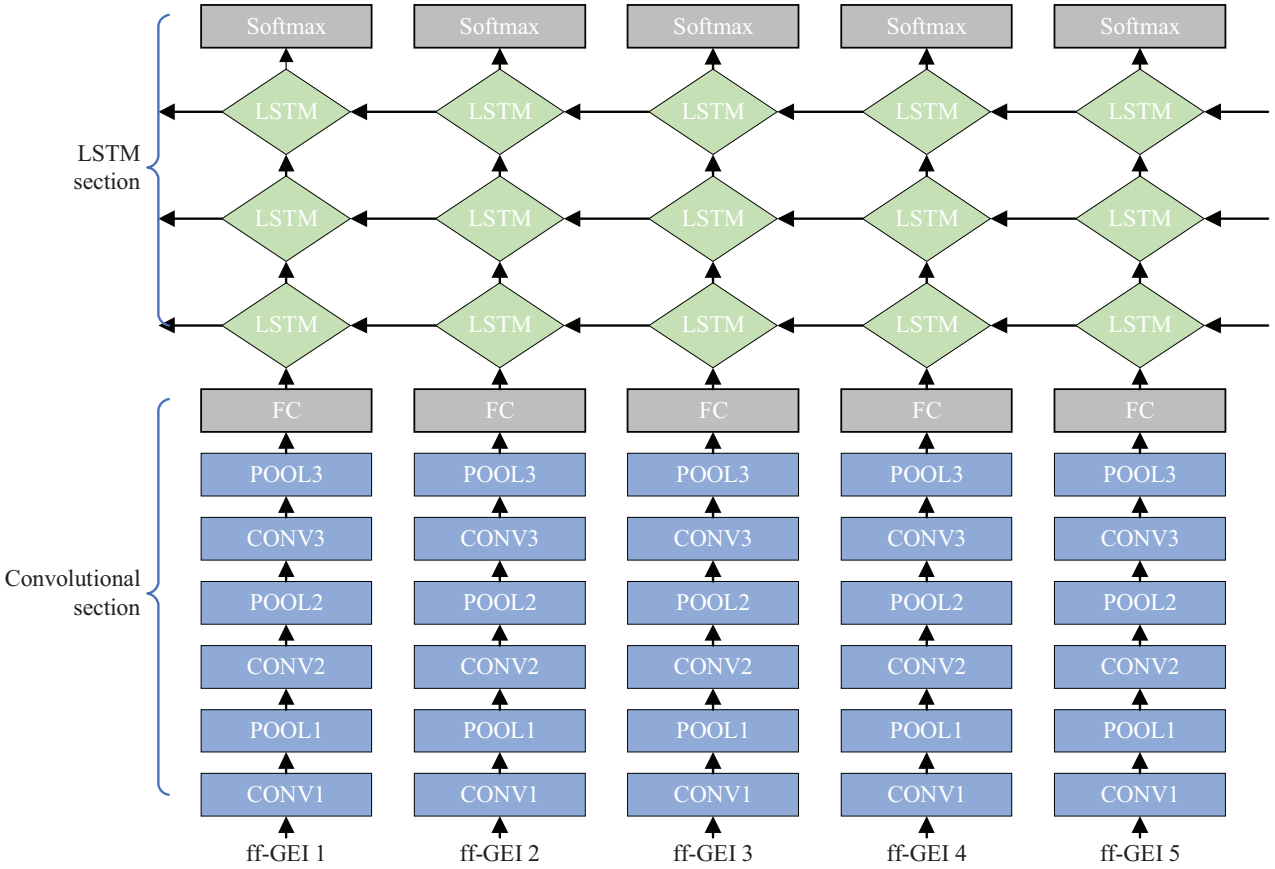

Figure 3. The architecture of the proposed gait recognition network based on Conv-LSTM.

Finally, we obtain gradients of the loss function with regard to each weight and bias

$$\frac{\partial L}{\partial w_{i,j}^{p}} = y_j^{p-1}\delta_i^p \tag{6}$$

and

$$\frac{\partial L}{\partial b_i^p} = \delta_i^p. \tag{7}$$

According to the LSTM and convolution shown in Fig. 3, we use various ways to calculate error term δ and partial derivatives of the loss function with respect to each weight and biase.

The LSTM part consists of a softmax layer and several LSTM layers. The activation function in the softmax layer is typically defined as

$$y_i^s = f(z_i^s) = \frac{e^{z_i^s}}{\sum_{j=1}^{M} e^{z_j^s}}, \tag{8}$$

where z^s is weighted input of the softmax layer, y^s is output of the softmax layer, z_i^s and y_i^s are the i-th neuron of z^s and y^s, respectively. Applying the chain rule, we obtain error term δ^s with respect to the softmax layer from Eq.(4) and Eq.(8)

$$\delta_i^s = \mathrm{y}_i - \hat{y}_i. \tag{9}$$

Then, based on the Eq.(6) and Eq.(7), we notice partial derivatives of the loss function with respect to the weights and biases in the softmax layer are

$$\frac{\partial L}{\partial w_{i,j}^{s}} = y_j^{s-1}(\mathrm{y}_i - \hat{y}_i) \tag{10}$$

and

$$\frac{\partial L}{\partial b_i^s} = \mathrm{y}_i - \hat{y}_i. \tag{11}$$

Next, we will discuss how to reduce backward propagate error term δ through the LSTM layers of the Conv-LSTM network and calculate the corresponding gradients of each weight and bias. In a LSTM layer as shown in Fig. 4, the error term is propagated in two directions: one is along with temporal direction, i.e., from the present to the preceding one; the other is from the present to the previous.

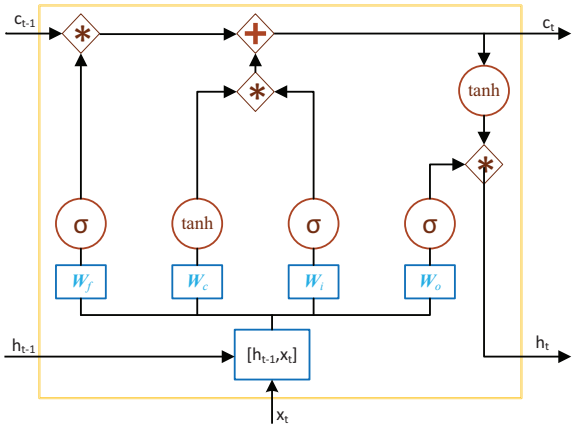

Figure 4. The diagram for one LSTM layer in Conv-LSTM.

In each LSTM layer, there are four 2-tuple sets of weight and bias, i.e., (W_f, b_f), (W_i, b_i), (W_o, b_o), (W_c, b_c) which correspond to forget gate, input gate, output gate, and candidate cell layer. At time t, the related error terms are defined as

$$\delta_f^t = \frac{\partial L}{\partial \mathbf{y}_f^t} f'(W_f[h_{t-1}, \mathrm{x}_t] + b_f) = \frac{\partial L}{\partial \mathbf{y}_f^t} f'(\mathbf{z}_f^t), \quad (12)$$

$$\delta_i^t = \frac{\partial L}{\partial \mathbf{y}_i^t} f'(W_i[h_{t-1}, \mathrm{x}_t] + b_i) = \frac{\partial L}{\partial \mathbf{y}_i^t} f'(\mathbf{z}_i^t), \quad (13)$$

$$\delta_o^t = \frac{\partial L}{\partial \mathbf{y}_o^t} f'(W_o[h_{t-1}, \mathrm{x}_t] + b_o) = \frac{\partial L}{\partial \mathbf{y}_o^t} f'(\mathbf{z}_o^t), \quad (14)$$

and

$$\delta_c^t = \frac{\partial L}{\partial \mathbf{y}_c^t} f'(W_c[h_{t-1}, \mathrm{x}_t] + b_c) = \frac{\partial L}{\partial \mathbf{y}_c^t} f'(\mathbf{z}_c^t), \quad (15)$$

where δ_f^t is error term of the forget gate, $\partial \mathbf{y}_f^t$ is output of the forget gate, $\mathbf{z}_f^t$ is weighted input of the current LSTM layer, the meanings of δ_i^t, δ_o^t, δ_c^t, $\mathbf{y}_i^t$, $\mathbf{y}_o^t$, $\mathbf{y}_c^t$, $\mathbf{z}_i^t$, $\mathbf{z}_o^t$ and $\mathbf{z}_c^t$ are similar to these notations, h_{t-1} is output of current LSTM layer at time t-1, x_t is input of current LSTM layer at time t. According to the roles, all the weight matrices W_f, W_i, W_o and W_c, can be further split as W_{fh}, W_{fx}, W_{ih}, W_{ix}, W_{oh}, W_{ox}, W_{ch}, and W_{cx}.

The convolutional part consists of a fully connected layer, several convolutional layers, and pooling layers. The forward pass and backward pass of the fully connected layer are similar with that of the softmax layer in the LSTM part, except for the activation function is ReLU function: $f(x) = \max(0, x)$, $x \in (-\infty, +\infty)$. Hence, we will focus on how to train the convolutional layers and pooling layers. A max pooling is used for pooling layers, which can create an invariance to small shifts and distortions. In addition, ReLU activation functions are used in all the convolutional layers to reduce computational complexity and alleviate gradient vanishing problem.

For each convolutional layer, local connection mode is adopted, the loss transmission depends on the convolution kernel. In the process of backpropagation, we firstly find which nodes are connected to the convolutional layer. Suppose that y_i^p is the i-th feature map of the p-th convolutional layer, then

$$y_i^p = f(\sum_j \left(y_j^{p-1} * k_{i,j}^p + b_i\right)), \quad (16)$$

where $f(\cdot)$ is the activation function of the p-th layer, y_j^{p-1} is the output of the $(p-1)$-th layer, $k_{i,j}^p$ is the convolution kernel used by the connection from the i-th feature map in the p-th layer to the j-th feature map in the $(p-1)$-th layer, and b_i is the corresponding bias. Given the error term δ^{p+1} of the next pooling layer, the error term δ_i^p with respect to the i-th output feature map of the current convolutional layer can be expressed as

$$\delta_i^p = U(\delta_i^{p+1}) \circ f'(z_i^p), \quad (17)$$

where $U(\cdot)$ is the upsampling function which restores δ^p to the size before pooling operation, then the values of each pooling area are input to the places where we get the maximum elements, z_i^p is the corresponding weighted input.

Within each pooling layer, there are no parameter that needs to be trained, so we only resolve the problem of backward pass of the error term. We set error term of the previous convolutional layer as δ^{p+1}, then the error term δ^p of the current pooling layer is defined as

$$\delta^p = \delta^{p+1} * R(W^{p+1}) \circ f'(z^p), \quad (18)$$

where $R(\cdot)$ is a rotation function which rotates the convolution kernel for 180°.

4. Experiments

In this section, based on the CASIA Dataset B[49] and OU-ISIR Large Population Dataset,[50] we designed and implemented two experiments for evaluating the proposed method. Experiment I is used to evaluate the correct recognition rate with view variations based on CASIA Dataset B (tagged as CASIA B). The CASIA B contains the videos of 124 walking

individuals including 11 views, distributed from 0 to 180° under three kinds of walking conditions, i.e., normal conditions, wearing coats, and carrying bags. Experiment II is to verify its generalization ability with a large scale of samples, which takes use of the OU-ISIR Large Population Dataset (tagged as OU-ISIR LP). The OU-ISIR LP comprises of two subsets: A and B. Dataset A is a set of two sequences per subject while Dataset B is a set of one sequence per subject. In addition, each of the main subsets is further divided into 5 subsets based on the observation angles 55°, 65°, 75°, and 85° including all four angles. The dataset consists of over 4,000 individuals walking on the ground surrounded by 2 cameras at 30 fps with 640×480 pixels.

4.1. *Benchmarks and evaluation criterion*

To evaluate the proposed gait recognition method in a reliable and equal way, we selected three state-of-the-art methods[30,41,43] in both comparative experiments. We also implemented the VTM-based method[38] and the method of original GEI algorithm[24] as a baseline. Besides, our previous gait recognition method based on (2D) 2PCA is added to compare with the proposed method.

- *Input and output architectures*:[43] In this method, both gait verification and identification are considered. The Siamese network with a pair of inputs and contrastive loss is employed for gait verification, a triplet network with a triplet of inputs and triplet ranking loss is used for gait identification.
- *LB network model*:[41] The LB (Local @ Bottom) network was proposed in 2017 and presented by using three network architectures, namely, LB, MT (Mid-Level @ Top), and GT (Global @ Top). The MT network is similar to the LB network, except that two extra nonlinear projections are applied before comparing the differences between the GEI pairs. There is not significant difference between the LB and MT networks in our tests. The GT network suffers from severe overfitting and has less satisfactory performance.
- *(2D)2PCA*:[30] In the method, a Gabor wavelet-based gait recognition algorithm is proposed, which employs two-dimensional principal component analysis ((2D)2PCA) to reduce the dimension of feature space. In addition, the multi-class support vector machine (SVM) is adopted to classify different gaits.
- *View transformation model*[38] (VTM): VTM is a typical method to solve the problem of cross-view gait recognition, we implement it as a baseline and compare the performance with other approaches. The view transformation model is obtained with a training set of individuals from multiple views. In gait recognition, the model transforms a gallery of features into the same view as the input vector.
- *The original GEI method*:[24] This method uses the original GEI as the feature vectors and the minimum Euclidean distance classifier for individual recognition. Therefore, we adapted this method as the baseline of traditional approaches for gait recognition.
- *SimpleRNN and SimpleCNN methods*: According to the structures of traditional CNN network and RNN network,[51] we implemented the simpleRNN and SimpleCNN methods using the same configurations as proposed in Conv-LSTM approach.

To evaluate the performance of the proposed method, we use Receiver Operating Characteristic (ROC) [52] as the evaluation criterion in our experiments. The ROC is a well-accepted metric to evaluate the performance of a recognition system. ROC analysis enables us to select optimal models and to discard suboptimal ones independently from the class distribution. In addition, a ROC curve can be generated by plotting the cumulative distribution function of the Genuine Accept Rate (GAR) along y axis as well as the cumulative distribution function of the False Match Rate (FPR) along x axis.

Assume that there are M individuals (tagged as $p_1, p_2, \cdots, p_M$) in one gait dataset, the current test sample is tagged as g_x, then TPR and FPR are defined as

$$TPR = \frac{TP}{TP + FN} \tag{19}$$

and

$$FPR = \frac{FP}{FP + TN}, \tag{20}$$

where TP(True Positive) means hit, TN (True Negative) refers to orrect rejection, FP (False Positive) stands for false alarm, FN(False Negative) indicates missing.

If we take the sampe g_x into consideration, TP means g_x really belongs to the n-th person,

$n \in [1, M]$; TN (True Negative) refers to that g_x is not identified as the n-th person, it is as same as the ground truth; FP (False Positive) stands for that g_x is identified as the n-th person by comoputer but it does not match with the ground truth; FN(False Negative) indicates that g_x is not identified.

4.2. *Experiment I. Comparisons on several methods by using CASIA Dataset B*

This experiment was conducted based on CASIA Dataset B. All the gait samples including 11 views and three different conditions are mixed together for training, thus, 110(10×11) videos are available for each individual. We used leave-one-out cross-validation strategy to construct training set and test set. The mixed gait data are split into 100 subsets. In each test, we randomly selected 99 subsets for training and the remaining one for test. We repeated the test for 1000 times and obtained the average correct recognition rates and standard deviations.

By implementing seven most recent or baseline methods based on the CASIA Dataset B, the average correct recognition rates (CRRs) and standard deviations are presented in Table 1, the ROC curves are shown in Fig. 5.

Table 1 shows that our Conv-LSTM-based method performs the best compared to others in terms of average correct recognition rate. On the one hand, compared with traditional methods, our method has a very significant improvement. For example, our average correct recognition rate is 95.9%, while the best DL-free method only has the rate 88.5%. On the other hand, our method performs excellent than many existing DL-based methods. Furthermore, not only the proposed method has an average correct recognition rate comparable to that of the input/output[43] method, but also the standard deviations of the proposed method show this method is robust.

In addition, Fig. 5 shows the ROC curves of eight methods in our experiment, where the dotted line is the pure opportunity line (POL) with zero recognition ability. According to the characteristics of ROC curves,[52] the farther a ROC curve is from the POL line, the better the classifier is performed in the gait recognition. From Fig. 5, we see intuitively that the classification capability of the proposed method is superior to that of the other seven methods, especially those DL-free ones.

Table 1. The esults of comparative experiments on CASIA Dataset B

Methods	Type	Average CRR(%)	Standard Deviation
$(2D)2PCA$[30]	DL-free	87.2	0.45
VTM[38]	DL-free	88.5	0.41
$OriginalGEI$[24]	DL-free	81.0	0.40
$Input/output$[43]	DL-based	94.3	0.43
LB[41]	DL-based	92.8	0.46
$SimpleRNN$	DL-based	91.5	0.48
$SimpleCNN$	DL-based	90.1	0.41
Conv-LSTM	DL-based	**95.9**	**0.31**

There are two reasons why our method achieves this performance:

- We use the gait feature representation ff-GEIs as the inputs of our network model. The ff-GEIs expand volume of the available dataset which is very important in training a complex deep learning model, meanwhile reduces the errors due to using gait cycle segmentation required by those gait recognition methods.
- The deep learning model Conv-LSTM is used to achieve gait recognition. By adding convolutional layers to the traditional LSTM model, Conv-LSTM extracts the essential characteristics of human gait by using the convolutional layers, at the same time it makes full use of RNNs to tackle the sequence of gait data.

Table 2. Results of comparative experiments on OU-ISIR LP

Methods	Type	Average CRR(%)	Standard Deviation
$(2D)2PCA$[30]	DL-free	88.9	0.39
VTM[38]	DL-free	85.2	0.37
$OriginalGEI$[24]	DL-free	84.3	0.29
$Input/output$[43]	DL-based	98.3	0.34
LB[41]	DL-based	94.6	0.37
$SimpleRNN$	DL-based	93.1	0.42
$SimpleCNN$	DL-based	91.9	0.38
$Conv-LSTM$	DL-based	99.1	0.23

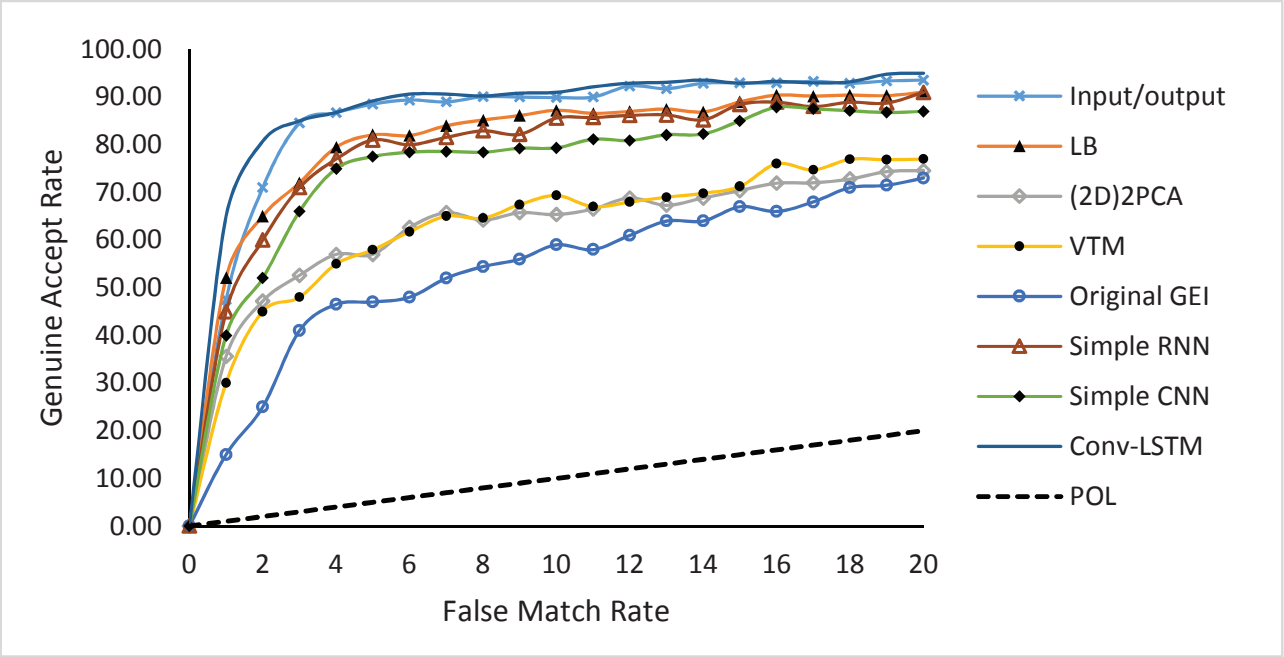

Figure 5. The ROC curves of the methods based on CASIA Dataset B.

4.3. *Experiment II. Comparisons of several methods by using OU-ISIR Large Population Dataset*

To evaluate the generalization ability of our method, we designed this experiment based on OU-ISIR dataset which consists of approximately 4000 subjects. In this experiment, we still used the leave-one-out cross-validation strategy to construct training set and test set. Instead of using the subset A and subset B of OU-ISIR LP as training set and test set separately, we mixed them together and evenly split the results into 100 sets. In each run, we trained the relevant gait classifiers using 99 parts of the sample sets for training and the rest one for test. In total, we repeated the loop for 1000 times and obtained the corresponding average correct recognition rate and the standard deviation.

The average correct recognition rates and standard deviations are listed in Table 2, the ROC curves are shown in Fig. 6. The results reveal that the proposed method performs excellent compared to others.

According to Table 2, we see that DL-based methods outperform the DL-free methods in terms of average correct recognition rates. This is the reason why deep learning network models can better explode the essential human gaits. Furthermore, though the proposed Conv-LSTM-based method and the input/output[43] method are comparable to the results of average correct recognition rate, but from their standard deviations, we see that the proposed method is more robust. Furthermore, Fig. 6 shows that, compared with the other seven methods, the ROC curve of the proposed Conv-LSTM-based method is farther from the pure opportunity line (POL), so its classification results are acceptable.

5. Conclusion and future work

The most challenging problems in gait recognition are feature representation and classifier design. In this paper, we present a novel gait classifier which takes full use of deep learning technology. In order to train this DL-based classifier, we design a new gait representation ff-GEI to meet the needs of DL-based classifier for the large-scale training data. Finally, in order to test the proposed gait recognition model, we have designed and implemented a gait classification algorithm. To the best of our knowledge, this is the first time gait recognition is developed based on the Conv-LSTM algorithm.

The limitation of this proposed method is that only experimental gait datasets with large scale samples are adapted. How to resolve the gait recognition problems in real applications with small amount of

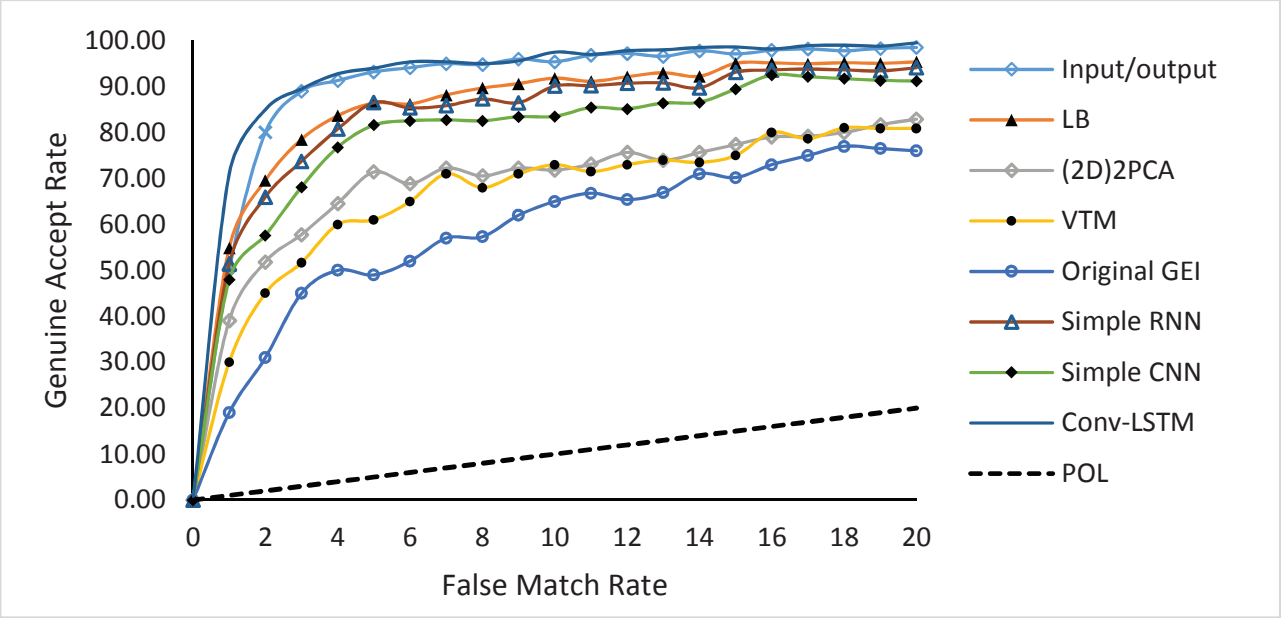

Figure 6. The ROC curves for gait recognition methods based on OU-ISIR LP.

training data or irregular distribution of samples is our future research mission.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NSFC) (Project No. 61303146 and Project No. 61602431).

Bibliography

1. S. Sarkar, P. Phillips and Z. Liu, The humanid gait challenge problem: Data sets, performance, and analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(02) (2005) 162–177.
2. X. Huang and N. Boulgouris, Gait recognition with shifted energy image and structural feature extraction, *IEEE Transactions on Image Processing* **21**(04) (2012) 2256–2268.
3. N. Boulgouris and X. Huang, Gait recognition using HMMs and dual discriminative observations for sub-dynamics analysis, *IEEE Transactions on Image Processing* **22**(09) (2013) 3636–3647.
4. H. Aggarwal and D. Vishwakarma, Covariate conscious approach for gait recognition based upon Zernike moment invariants, *IEEE Transactions on Cognitive and Developmental Systems* **PP**(99) (2017) p. 1.
5. Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature* **521**(5) (2015) 436–445.
6. P. Wang and X. Bai, Regional parallel structure based CNN for thermal infrared face identification, *Integrated Computer-Aided Engineering* **25**(3) (2018) 247–260.
7. M. Molina-Cabello, R. Luque-Baena, E. López-Rubio and K. Thurnhofer-Hemsi, Vehicle type detection by ensembles of convolutional neural networks operating on super-resolved images, *Integrated Computer-Aided Engineering* **25**(4) (2018) 321–333.
8. S. Li, X. Zhao and G. Zhou, Automatic pixel-level multiple damage types detection of concrete structure using fully convolutional networks, *Computer-Aided Civil and Infrastructure Engineering* **34**(7) (2019) 616–634.
9. R. Wu, A. Singla, M. Jahanshahi, E. Bertino, B. Ko and D. Verma, Pruning deep convolutional neural networks for efficient edge computing in condition assessment of civil infrastructures, *Computer-Aided Civil and Infrastructure Engineering* **34**(9) (2019) 774–789.
10. V. Schetinin, L. Jakaite and W. Krzanowski, Bayesian learning of models for estimating uncertainty in alert systems: Application to aircraft collision avoidance, *Integrated Computer-Aided Engineering* **25**(3) (2018) 229–245.
11. X. Luo, H. Li, X. Yang, Y. Yu and D. Cao, Capturing and understanding workers' activities in far-field surveillance videos with deep action recognition and bayesian nonparametric learning, *Computer-Aided Civil and Infrastructure Engineering* **34**(4) (2019) 333–351.
12. X. Liang, Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization, *Computer-Aided Civil and Infrastructure Engineering* **34**(5) (2019) 415–

430.
13. Y. Huang, J. Beck and H. Li, Multitask sparse Bayesian learning with applications in structural health monitoring, *Computer-Aided Civil and Infrastructure Engineering* **34**(9) (2019) 732–754.
14. M. Rafiei, W. Khushefati, R. Demirboga and H. Adeli, Supervised deep restricted Boltzmann machine for estimation of concrete compressive strength, *ACI Materials Journal* **114**(2) (2017) 237–244.
15. M. Rafiei and H. Adeli, A novel machine learning based algorithm to detect damage in highrise building structures, *The Structural Design of Tall and Special Buildings* **26**(18) (2017) 1–11.
16. M. Rafiei and H. Adeli, A novel unsupervised deep learning model for global and local health condition assessment of structures, *Engineering Structures* **156**(1) (2018) 598–607.
17. M. Rafiei and H. Adeli, Novel machine learning model for construction cost estimation taking into account economic variables and indices, *Journal of Construction Engineering and Management* **144**(12) (2018) 1–9.
18. S. Bang, S. Park, H. Kim and H. Kim, Encoder–decoder network for pixel-level road crack detection in black-box images, *Computer-Aided Civil and Infrastructure Engineering* **34**(8) (2019) 713–727.
19. K. Maeda, T. Ogawa, M. Haseyama and S. Takahashi, Convolutional sparse coding-based deep random vector functional link network for distress classification of road structures, *Computer-Aided Civil and Infrastructure Engineering* **34**(8) (2019) 654–676.
20. J. Torres, A. Galicia, A. Troncoso and F. Martinez-Alvarez, A scalable approach based on deep learning for big data time series forecasting, *Integrated Computer-Aided Engineering* **25**(4) (2018) 335–348.
21. J. Torres, A. Galicia, A. Troncoso and F. Martinez-Alvarez, Large scattered data interpolation with radial basis functions and space subdivision, *Integrated Computer-Aided Engineering* **25**(1) (2018) 49–62.
22. H. Hashemi and K. Abdelghany, End-to-end deep learning methodology for real-time traffic network management, *Computer-Aided Civil and Infrastructure Engineering* **33**(10) (2018) 849–863.
23. S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation* **9**(8) (1997) 1735–1780.
24. J. Han and B. Bhanu, Individual recognition using gait energy image, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(02) (2006) 316 – 323.
25. D. Tao, X. Li, X. Wu and S. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(10) (2007) 1700 – 1715.
26. X. Yang, Y. Zhou, T. Zhang, G. Shu and J. Yang, Gait recognition based on dynamic region analysis, *Signal Processing* **88**(9) (2008) 2350 – 2356.
27. P. Theekhanont, W. Kurutach and S. Miguet, Gait recognition using GEI and pattern trace transform, *Information Technology in Medicine and Education*, (Hokodate, Japan, 2012), pp. 936–940.
28. T. Connie, M. Goh and A. Teoh, A grassmannian approach to address view change problem in gait recognition, *IEEE Transactions on Cybernetics* **47**(06) (2017) 1395 – 1408.
29. Y. Guan, C. Li and F. Roli, On reducing the effect of covariate factors in gait recognition: A classifier ensemble method, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**(07) (2015) 1521 – 1529.
30. X. Wang, J. Wang and K. Yan, Gait recognition based on Gabor wavelets and (2D)2PCA, *Multimedia Tools and Applications* **77**(10) (2018) 12545 – 12561.
31. G. Zhao, G. Liu, H. Li and M. Pietikainen, 3D gait recognition using multiple cameras, *International Conference on Automatic Face and Gesture Recognition*, (Southampton, UK, 2006).
32. G. Ariyanto and M. Nixon, Model-based 3D gait biometrics, *International Conference on Biometrics*, (Washington, DC, USA, 2011).
33. F. Abdulsattar and J. Carter, Performance analysis of gait recognition with large perspective distortion, *IEEE International Conference on Identity, Security and Behavior Analysis*, (Sendai, Japan, 2016).
34. J. Luo, J. Tang and T. Tjahjadi, Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis, *Pattern Recognition* **60** (2016) 361 – 377.
35. J. Tang, J. Luo and T. Tjahjadi, Robust arbitrary-view gait recognition based on 3D partial similarity matching, *IEEE Transactions on Image Processing* **26**(1) (2017) 7–23.
36. M. Goffredo, I. Bouchrika, J. Carter and M. Nixon, Self-calibrating view-invariant gait biometrics, *IEEE Trans. Systems, Man, and Cybernetics, Part B* **40**(4) (2010) 997 – 1008.
37. W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma and H. Li, A new view invariant feature for cross-view gait recognition, *IEEE Transactions on Information Forensics and Security* **8**(10) (2013) 1642—-1653.
38. Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo and Y. Yagi, Gait recognition using a view transformation model in the frequency domain, *IEEE ECCV*, (Graz, Austria, 2006), pp. 151–163.
39. W. Kusakunniran, Q. Wu, H. Li and J. Zhang, Multiple views gait recognition using view transformation model based on optimized gait energy image, *IEEE ICCV*, (Kyoto, Japan, 2009), pp. 1058–1064.
40. W. Kusakunniran, Q. Wu, J. Zhang, H. Li and L. Wang, Recognizing gaits across views through correlated motion co-clustering, *IEEE Transactions on Image Processing* **23**(2) (2014) 696–709.

41. Z. Wu, Y. Huang, L. Wang, X. Wang and T. Tan, A comprehensive study on cross-view gait based human identification with deep CNNs, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(02) (2017) 209 – 226.
42. N. Jia, V. Sanchez and C. Li, Learning optimized representations for view-invariant gait recognition, *International Joint Conference on Biometrics*, (Denver, USA, 2017), pp. 774–780.
43. N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo and Y. Yagi, On input/output architectures for convolutional neural network-based cross-view gait recognition, *IEEE Transactions on Circuits and Systems for Video Technology* **28**(1) (2018).
44. K. Shiraga, Y. Makihara and D. Muramatsu, GEINet: View-invariant gait recognition using a convolutional neural network, *International Conference on Biometrics*, (Halmstad, Sweden, 2016).
45. T. Wolf, M. Babaee and G. Rigoll, Multi-view gait recognition using 3D convolutional neural networks, *IEEE International Conference on Image Processing*, (Phoenix, USA, 2016), pp. 4165–4169.
46. B. Ko, H. Kim, K. Oh and H. Choi, Controlled dropout: A different approach to using dropout on deep neural network, *IEEE International Conference on Big Data and Smart Computing (BigComp)*, (Jeju, South Korea, 2017), pp. 358–362.
47. R. Zhang, Z. Xu, G. Huang and D. Wang, Global convergence of online BP training with dynamic learning rate, *IEEE Transactions on Neural Networks and Learning Systems* **23**(2) (2012) 330–341.
48. X. Li, Preconditioned stochastic gradient descent, *IEEE Transactions on Neural Networks and Learning Systems* **29**(5) (2018) 1454 – 1466.
49. S. Yu, D. Tan and T. Tan, A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, *International Conference on Pattern Recognition*, (Hong Kong, China, 2006), pp. 441–444.
50. H. Iwama, M. Okumura, Y. Makihara and Y. Yagi, The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition, *IEEE Transactions on Information Forensics and Security* **7**(5) (2012) 1511–1521.
51. I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning* (MIT Press, 2016). *http://www.deeplearningbook.org* .
52. *R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha and A. W. Senior, The relation between the ROC curve and the CMC,* IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, (Buffalo, USA, 2005), pp. 15–20.*