Multimedia Tools and Applications (2021) 80:1565–1581
https://doi.org/10.1007/s11042-020-09777-7

Gait classification through CNN-based ensemble learning

Xiuhui Wang[1] · Ke Yan[1]

Received: 9 April 2020 / Revised: 26 August 2020 / Accepted: 28 August 2020 /
Published online: 8 September 2020

Abstract
Gait is a biological characteristic for video surveillance and many other applications, which can be used to identify individuals at a large distance. In this paper, a gait classification framework based on CNN Ensemble (GCF-CNN) is proposed, which includes three modules: 1) Feature extraction and preprocessing: use random sampling with replacement strategy to generate a serial of training sets from gait silhouette images; 2) Gait models training: construct and train primary CNN classifiers using different hyper-parameters, and train them a secondary classifier to combine them; 3) Gait classification: utilize the trained two-level classifier to achieve gait classification. In addition, the proposed classification framework is evaluated on the CASIA Gait Database and OU-ISIR Gait Database. And it is demonstrated by comprehensive experiments that the proposed classification framework can achieve outstanding performance in correct classification rate with respect to several state-of-the-art methods.

Keywords Ensemble learning · CNN · Gait recognition

1 Introduction

Gait is an appealing biometric feature which can be used for human recognition at a distance. Compared with other means of biometric authentication like fingerprints or faces, gait recognition can be applied without alerting or disturbing the target subjects [3, 16, 31]. On the other hand, the research on efficient and practical gait classification methods still remains a formidable challenge and an area of active research, namely, most of the existing gait recognition algorithms only work well under the best condition of image and video acquisition [3, 31].

Recently, with the successful applications of Deep Learning (DL) technologies [9] in object detection, segmentation and recognition from images and videos [22, 23], some

✉ Xiuhui Wang
wangxiuhui@cjlu.edu.cn

[1] Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University, No. 258, Xueyuan Street, Hangzhou, 310018, China

researchers have preliminarily applied deep learning to human gait recognition [19, 31]. Convolution Neural Networks (CNN), as an outstanding representative of DL technologies, has many advantages such as combining local perceptions, weights sharing and spatial down-sampling to make full use of limited sample data.

Besides, Ensemble Learning is one of the DL technologies which combines multiple primary learners through a fusion strategy to improve the overall generalization performance [28].Ensemble learning has attracted wide attentions due to its easily understandable structure and promising classification performance by combining primary learners into a stronger one. Elghazel et al. [2] proposed an ensemble method, Random Cluster Ensemble, which estimates the out-of-bag feature importance from an ensemble of partitions. Each partition is constructed using a different bootstrap sample and a random subset of the features. Tekin et al. [17] presented a systematic ensemble learning method called Hedged Bandits, which comes with both long run and short run performance guarantees. Their approach yields performance guarantees with respect to the optimal local prediction strategy, and is also able to adapt its predictions in a data-driven manner.

Inspired by the decision tree algorithm [11], in this paper, we integrate multiple heterogeneous CNN networks to achieve diverse gait feature extraction, and propose a novel gait classification framework based on CNN Ensemble (GCF-CNN). The GCF-CNN method can reduce the huge demand for training samples when using deep CNN in some degree, and thus alleviates the problem of limited data in existing open-accessed gait database and most practical applications. At the same time, the proposed framework also retains the power ability of deep CNN to extract and express diverse gait features. Our work in this paper is summarized as follows:

(1) Systematic review and discussion. We provide a comprehensive survey of existing gait classification approaches published over the past decade. Based on using or not Deep Learning (DL) technologies, we group these methods into two classes, briefly introduce the representative ones and point out the pros and cons of each class.
(2) A novel gait classification framework through CNN-based Ensemble Learning. The proposed gait classification framework consists of three folds: 1) Use bootstrap-aggregating strategy to sample the GEIs extracted from original gait silhouette images to shape a serial of training sets; 2)Train diverse CNN primary classifiers which are different in hyperparameters and training sets; 3)Construct and train a secondary classifier to ensemble CNN models.
(3) Comprehensive evaluation using two famous gait databases. We thoroughly evaluate the proposed classification framework using the CASIA Dataset A, Dataset B, and OU-ISIR LP Dataset. The experiments on the CASIA Dataset A and B are conducted for evaluating the gait classification performance under cross-view conditions, and the experiments on the OU-ISIR LP Dataset are for verifying the generalization ability with large-scale data.

The rest of this paper is organized as follows. Related work is reviewed in Section 2. Detailed description and demonstration of the proposed gait classification framework is presented in Section 3. Experimental results and discussion are proposed in Section 4. Finally, concluding remarks and future work are given in Section 5.

2 Related work

Extensive efforts have been devoted to solve gait classification under different conditions, such as cross-view, clothing variations, and with or without a loading. According to whether Deep Learning (DL) technologies are involved, these researches can be roughly classified into two major categories, i.e. DL-free methods and DL-based methods.

The DL-free gait classification methods mainly focus on gait feature processing, such as new gait feature representations [5, 21, 25, 26], 3D gait reconstruction [1, 12, 20, 34] or view transform models (VTMs) for gait features [7, 8, 13]. As a novel feature presentation, Gait Energy Images (GEI) was first proposed in [5], which was computed by averaging properly aligned human silhouettes in gait sequences. GEI and its varieties are widely employed in many subsequent research literature. Tao et al. [21] develop a general tensor discriminant analysis (GTDA) as a preprocessing step for LDA, and successfully apply it in human gait recognition. In addition, a Gabor wavelets-based gait recognition algorithm was proposed in [26], which employs the two-dimensional principal component analysis ($(2D)^2$PCA) method for reducing feature dimension. Ariyanto et al. [1] reconstructed the 3D structure of each gait to generate arbitrary 2D views by projecting the 3D model. Tang et al. [20] propose gait partial similarity matching that assumes a 3D object shares common view surfaces in significantly different views, in which 3D parametric body models are morphed by pose and shape deformation from a template model using 2D gait silhouette as observation. Normally, these 3D-reconstruction-based methods can obtain higher classification scores, but they generally require multiple calibrated cameras which are unavailable in most gait databases and practical scenarios. In [8], to address the problems in cross-view gait recognition, a motion co-clustering is carried out to partition the most related parts of gaits from different views into the same group, and inside each group, and then a linear correlation between gait information across views is further maximized through canonical correlation analysis. Most researches on VTMs need to learn projection transformations [7, 13], with which one can transform gait features from different views to one or more common views. These approaches compare the normalized gait features extracted from any two videos to calculate the corresponding similarity. In short, traditional DL-free methods for gait classification can reduce the influence of various covariant factors, like view changes and different clothing condition, with or without a bag. However, there are still little effective feature extraction and modeling methods to solve the highly nonlinear correlation between gait features in complex walking environments.

On the other hand, DL-based gait classification methods combine gait feature processing and classifier designing together by some DL technologies. Currently, DL-based gait recognition methods mainly focused on convolution neural network (CNN) and recurrent neural network (RNN) [19, 29, 31]. CNNs are a specialized kind of neural network for processing data that has a known grid-like topology [4], which use convolution in place of general matrix multiplication in at least one of their layers. Convolution layers in a CNN have the advantages of local receptive fields and shared weights. Each neuron in a convolution layer will be connected to a small region, which is also its called local receptive field, of its input neurons. Pooling is an operation, which almost all convolutional networks employ. A pooling function replaces the output of the net at a certain location with a summary statistic of the nearby outputs. Pooling layers are used immediately after each convolutional layer, which make the gait representations smaller and more manageable.

An extensive study was conducted in [31] with respect to a cross-view and cross-walking condition, with various preprocessing methods and CNN network architectures.

Besides, this research presented a CNN-based method for gait recognition and three CNN-based Network, namely, Local@Bottom, Mid-level@Top and Global@Top. Takemura et al. [19] presented 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. In [6, 18], several specific CNN network models were proposed to solve the problem of multi-view gait recognition. Furthermore, Wolf et al. [30] discussed a gait recognition approach based on a deep convolution neural network with 3D convolutions, which used a specified input including both gray-scale images and optical flow.

The proposed approach in this paper belongs to the DL-based methods. However, different from existing researches, we present a gait classification framework based on Ensemble Learning(EL) technologies. Instead of using the CNN standalone, we use several diverse CNN in our classification framework, and train them using bootstrap-aggregating strategy. In addition, by using the output of each CNN as training sets, we construct and train a secondary classifier to combine the primary CNN models.

3 Methodology

The proposed gait classification framework is schematically shown in Fig. 1. Firstly, from each gait silhouette sequence, several GEIs are constructed based on gait cycles which will be collected as a original gait sample set G. Secondly, we reconstrct G by bagging-like strategy, i.e. random sampling with replacement, to obtain a serial of training sets $\{g_i \| i = 1, 2, \cdots, n\}$ with a silght difference, where N is the number of primary learners. Then, a

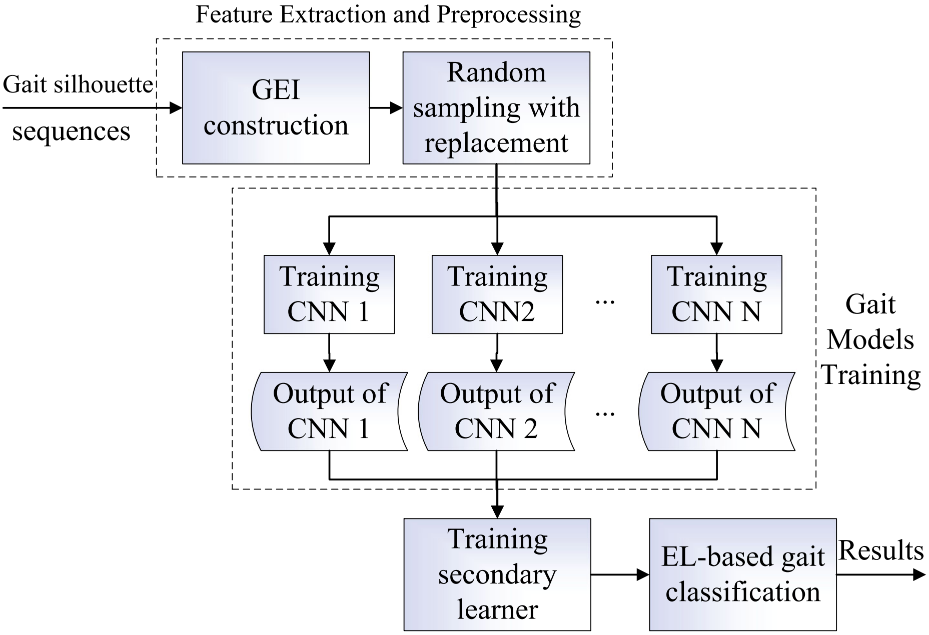


Fig. 1 Work flowchart of the proposed gait classification framework

set of CNNs are trained using the extracted training sets $\{g_i \| i = 1, 2, \cdots, n\}$, and the corresponding outputs are used as inputs to train a secondary learner. Additionally, the primary CNN learners and their training process are diverse in hyper-parameters, such as number of nodes, batch size and learning rate.

3.1 Feature extraction and preprocessing

The gait energy image (GEI) [5] is an effective and efficient gait representation for individual recognition, which considering both the spatial and temporal characteristics of human gaits. Compared with other gait representation, GEIs can reflect major shapes of gait silhouettes and their changes in a given cycle, and are relatively insensitive to local distortion of contour edge and small-scale changes. As shown in Fig. 1, after the gait silhouette sequences are input, the follow-up procedure, namely, feature extraction and preprocessing, consists of two stages: 1) GEI construction, and 2) random sampling with replacement.

In the GEI construction stage, a silhouette preprocessing procedure is firstly carrried out on the inputs, which mainly includes two operations: scale normalization and central alignment. The first operation is used to adjust the canvas of each silhouette image into the same width and height, while keeps the silhouette region unchanged. The second operation aims to assure that the same human parts is roughly combined to the same point in the constructed GEI. There are some basic assumptions about these operations: 1) the movement of each human parts in a walking cycle is relatively stable, especially, limbs and legs swinging together with the same rhythm, 2) during a cycle of walking, there was no significant change in direction, and 3) the radial distortion of the camera used to collect gait data is small enough. Under these assumptions, we can construct a GEI by

$$I(x, y) = \frac{1}{N} \sum_{i=1}^{N} S_i(x, y) \tag{1}$$

where S is a preprocessed silhouette sequence in one gait cycle, $S_i(x, y)$ is the i^{th} image in the given gait cycle, N is the number of images in S, and (x,y) is the 2D image coordinate. Fig. 2 shows some samples of 11 different views from CASIA Gait Dataset B.

During the procedure of random sampling with replacement, we reconstruct the GEI-based training set to produce a serial of diverse training sets. Random sampling with replacement is a truthful mechanism that utilizes sampling operations in order to achieve

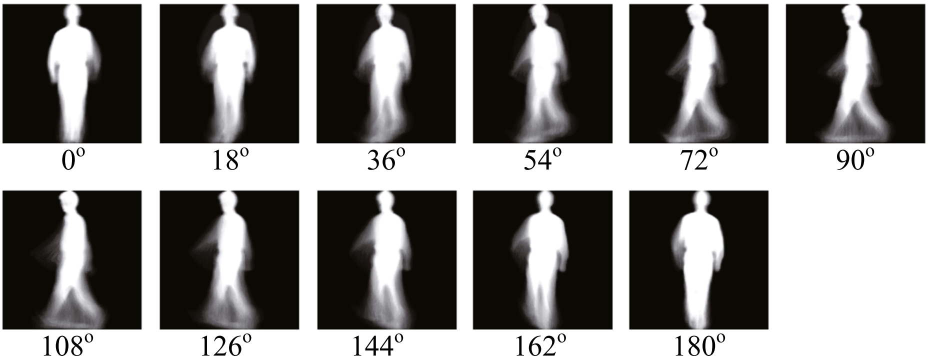


Fig. 2 GEI examples of different views from the CASIA Dataset B

approximately-optimal gain in prior-independent mechanisms [15]. This method reduces the variance of its primary learners by introducing randomness in the process of constructing the model. Given an original GEI set G of size M, the probability that a sample g in G is selected at a time is $\frac{1}{M}$. Then the probability that g is not selected in M times of sampling is

$$\lim_{M\to+\infty}\left(1-\frac{1}{M}\right)^M=\frac{1}{e}\approx 0.3679, \tag{2}$$

where e is the base of natural logarithm. Thus, when the original GEI set $\{G\}$ is large enough, we can get a series of new training sets $\{G_i|i=1,2,\cdots,N\}$ with a difference of 36.79%. In other words, each set G_i is expected to have 63.21% of unique GEIs from G.

3.2 Gait models modeling and training

In this section, we will discuss how to train the primary CNN classifiers and the corresponding secondary classifier. To be convenient, similar CNN architectures are used for both the primary and secondary classifiers. In each primary CNN classifier, several convolution-ReLU-pooling triples (CRP-T), and one fixed fully-connected layer are used, as shown in Fig. 3a. There are five primary CNN models, as shown in Table 1,where CRP-Ts means convolution-ReLU-pooling triples and FMs means feature maps. Among the five primary CNN classifiers, we set different padding and kernel parameters to make the input and output size of each classifier consistent. In addition, we add a ReLU layer and a SOFTMAX layer to produce temporary classification information that will be used to calculate the iteration error in modeling training process of each primary classifier.

For example, in the first convolution layer (CONV1), as the kernel is 5*5, a neuron will correspond to 25 pixels of the input GEI image. Moreover, the same weights and bias for each neurons in one convolution layer, which means that all the neurons in the same layer will detect exactly the same type of gait features at different parts of an input GEI. To put it in more formal terms, convolution layers are intuitively designed to insensitive to translation changes of images. Based on experience, instead of using a sigmoid or tanh activation function, we employ a rectified linear unit, i.e. ReLU, which is defined as

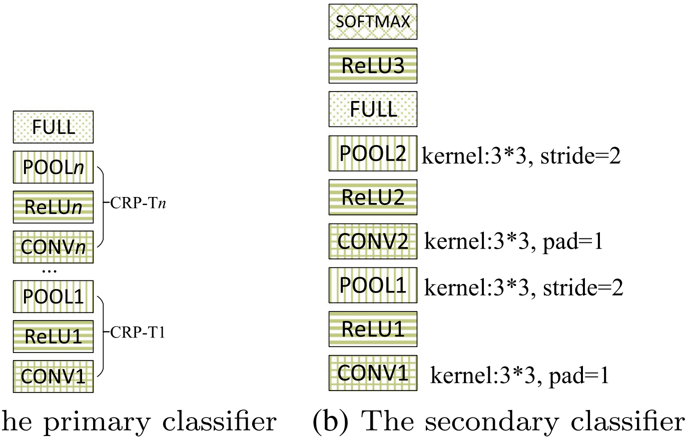
$$f(x)=max(0,x). \tag{3}$$

In the proposed network, we use a max-pooling strategy with 3 by 3 pooling windows and a stride length of 2. An extra fully-connected layer will be used to integrate global information from across the entire input GEI. The output of the fully-connected layer in each primary classifier will be used as inputs of the secondary classifier.

On the other hand, we construct the secondary classifier as shown in Fig. 3b. Similar with the primary classifiers, there are still several convolution-ReLU-pooling triples and a fully-connected layer. In the last, a SOFTMAX layer is adopted to calculate the final classification results. But different from the primary CNN classifiers, we adopt fixed number

Table 1 Configuration of five primary CNN models

	CRP-Ts	FMs	Kernel size	Pad size
CNN-1	1	64	3∗3	1
CNN-2	1	64	5∗5	2
CNN-3	2	64|128	5∗5 | 3∗3	2|1
CNN-4	2	64|128	11∗11 | 5∗5	4|2
CNN-5	3	64| 128 |256	11∗11 | 5 ∗ 5 | 3∗3	4| 2 |1



(a) The primary classifier (b) The secondary classifier

Fig. 3 Architectures of primary and secondary classifiers

of layers in the secondary classifier, and invariable number of feature maps in each convolution layer. It should be noted that the second classifier is used to ensemble the primary classifiers. Different from existing methods, the primary classifiers in our method focus on different gait feature extraction, and their outputs are vectors, not simple classification results. Therefore, at the fusion stage, instead of using average or voting method, we use a secondary classifier to combine the results of all primary classification.

In the training phase, we use backward propagation [33] to compute the gradients of each weight and bias of each neuron as well as update the related weights and biases by using stochastic gradient descent [10]. The training process consists of two iterative steps: forward pass of the training data and backward pass of the loss. Suppose there are N GEI samples corresponding to M individuals in the training dataset, then the loss function can be defined as:

$$L = -\frac{1}{N} \cdot \sum_{n=1}^{N} \hat{y}_n \log y_n, \tag{4}$$

where y_n is predictive output of our CNN models, $\hat{y}_n$ is the corresponding label vector. In the first step, we calculate output y^p $(p = 2, 3, \cdots, P)$ of each layer and the final error vector δ^P.

$$z^p = w^p y^{p-1} + b^p \tag{5}$$

and

$$y^p = f(z^p) \tag{6}$$

where w^p is weight of the p-th layer, b^p is bias of the p-th layer, and f is activation function of the p-th layer. When output of the current CNN network is not consistent with our expectation, the back-propagation is completed. We compute final error δ^P between the actual result and the expected value, feed δ^P back into the network and obtain the error vector with respect to each layer

$$\delta_i^P = \frac{\partial L}{\partial y_i^P} f'(z_i^P) \tag{7}$$

and

$$\delta_i^p = \sum_j (w_{j,i}^{p+1} \delta_j^{p+1} f'(z_i^p)), \tag{8}$$

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. Finally, we can obtain gradient of the loss function with regard to each weight and bias

$$\frac{\partial L}{\partial w_{i,j}^p} = y_j^{p-1}\delta_i^p \tag{9}$$

and

$$\frac{\partial L}{\partial b_i^p} = \delta_i^p. \tag{10}$$

3.3 Algorithm of gait classification based on ensemble learning

After finishing the training of gait models, we can further describe our framework for gait classification in detail, as Algorithm 1.

Algorithm 1 Gait classification algorithm based on ensemble learning.

Input: Labelled gait silhouette set $S = \{(s_{i,j}, b_i) | i = 1, 2, \cdots, N;\ j = 1, 2, \cdots, M\}$. N is the number of individuals in a given gait database, M is the number of silhouette sequences for each individual in S.

Output: The correct classification rates R.

Step 1: *Create a serial of training sets.* Based GEI set S' from the input gait silhouette set S to characterize human walking properties, as described in (1), a group of slightly different GEI set S'' are extracted from S' using random sampling with replacement strategies.

Step 2: *Train the primary CNN classifiers.* By adding a temporary SOFTMAX layer to each primary CNN classifiers, we train them using S'' based on the BP algorithm [33]. These gait models are similar in structure, but different in number of layers, number of feature maps and size of convolution kernels. In addition, training sets and hyper-parameters, such as learning rate and batch size are also different.

Step 3: *Construct new training set S_{tr}.* Based on the trained primary classifiers, we construct a new trainning set S_{tr} and the corresponding labels from the final fully-connected layer of each primary classifier.

Step 4: *Train the secondary CNN classifier.* Using the new straining set S_{tr}, we train the secondary classifier similar with backward propagation [33] by computing the gradients of parameters and updating the related parameters by stochastic gradient descent [10].

Step 5: *Calculate the correct clsssification rates.* With the trained secondary classifier, the classification proceeds based on the testing set S_{te}, and we can obtain the related correct classification rate R.

4 Experimental results

In this section, three widely-used gait databases,1) CASIA Dataset A, 2) CASIA Dataset B, and 3) OU-ISIR LP Dataset are used to evaluate the performance of the proposed classification framework. Five existing methods and a simple CNN method are added to the comparison experiments,

In addition, for the convenience of quantitative comparison and analysis, we used Cumulative Match Characteristics (CMCs) as the evaluation criterion in our experiments, which is

Fig. 4 Some Samples from the CASIA Dataset A

a well-accepted measurement to judge the classification capabilities of recognition systems. Furthermore, CMC enables us to select possibly optimal models and discard sub-optimal ones independently by sorting the scores of candidates.

4.1 Experiments on CASIA Dataset A

In this section, experiments are carried out on CASIA Dataset A to evaluate the performance of proposed gait classification framework. This Dataset was created on Dec. 10, 2001, including 20 persons, as shown in Fig. 4. Each person has 12 image sequences, 4 sequences for each of the three directions, i.e. parallel, 45 degrees and 90 degrees to the image plane. The length of each sequence is not identical for the variation of the walker's speed, but it must range from 37 to 127. The Dataset A includes 19139 images.

In training of the secondary classifier, three sequences of each directions in Dataset A are selected as the training set and the other sequences are used as the testing datasets. Besides, we implemented four existing approaches [5, 14, 26, 27], a simple CNN-based method and the proposed GCF-CNN method. The corresponding CMC curves are shown in Fig. 5. From the results shown in Fig. 5, the proposed GCF-CNN method outperforms the other five methods, especially compared with DL-free methods. In addition, the correct recognition rates and standard deviation of the proposed method are compared with those of existing methods in Table 2, which show that the proposed GCF-CNN method performs better in terms of correct recognition rate and stability.

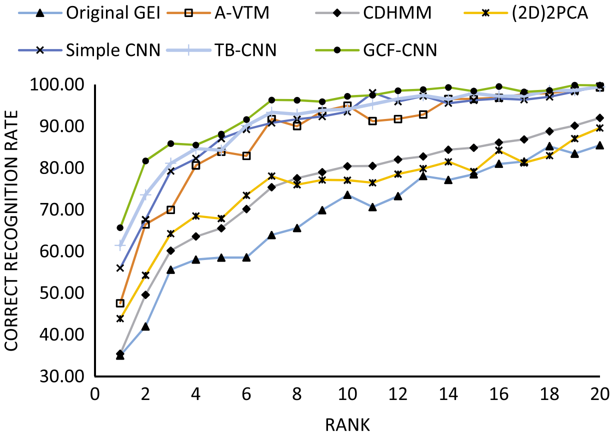


Fig. 5 The CMC curves of different approaches in Experiments on CASIA Dataset A

Table 2 Comparison with existing methods on CASIA Dataset A in terms of correct recognition rates (CRR) and standard deviations (SD)

METHOD	Rank-1(%)	Rank-2(%)	Rank-3(%)	Rank-4(%)	Rank-5(%)	Rank-10(%)	Rank-15(%)	Rank-20(%)	AVERAGE CRR(%)	SD
Original GEI	35.00	42.00	55.59	58.00	58.47	73.54	78.48	85.44	85.44	4.28
A-VTM	47.55	66.47	69.96	80.60	83.92	94.98	96.60	99.31	99.31	3.69
CDHMM	35.39	49.58	60.18	63.56	65.52	80.40	84.83	92.01	92.01	5.30
(2D)2PCA	43.83	54.23	64.25	68.50	67.85	77.08	79.13	89.61	89.61	4.12
Simple CNN	55.99	67.60	79.25	82.27	87.08	93.49	96.25	99.68	99.68	3.51
TB-CNN	61.44	73.56	81.13	84.65	84.24	94.14	97.96	99.49	99.49	2.07
GCF-CNN	65.64	81.68	85.83	85.54	88.13	97.15	98.42	99.83	99.83	1.44

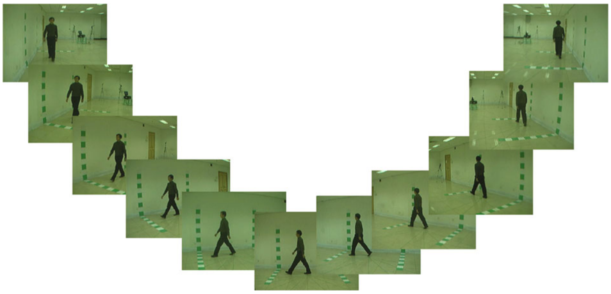


Fig. 6 Examples from the CASIA Dataset B

4.2 Experiments on CASIA Dataset B

This section reports experimental results on CASIA Dataset B to examine the cross-view classification performance of the proposed EL-based framework. CASIA Dataset B is a large multi-view gait database [32], which is created in January 2005. There are 124 subjects, and the gait data was captured from 11 views, as shown in Fig. 6. Three different 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. Dataset C was collected by an infrared camera in Aug. 2005. It contains 153 subjects and takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag.

For all the involved methods, the training set and testing set are constructed with the same division, i.e. 50% for training and 50% for testing. The CMC curves of six methods are shown in Fig. 7. The results show that our GCF-CNN method performs better than others in terms of correct classification rate. On the one hand, compared with DL-free gait recognition methods, our method has a very significant advantage in correct classification rate. On the other hand, our method performs better than the simple CNN-based method in terms of correct classification rate. In addition, Fig. 7 also illustrates the fact that, compared with the results in Experiment 1, the performance of all the six methods have slightly decreased. The reason is that, in Dataset B, there is much variation factors, such as wearing a coat, with or without a bag. Furthermore, the correct recognition rates and standard deviation of the proposed method are compared with those of existing methods in Table 3, which show that the proposed GCF-CNN method can improve the correct recognition rate and has strong stability.

4.3 Experiments on OU-ISIR LP dataset

In this section, we further reports experimental results on OU-ISIR LP Dataset to verify the generalization ability of the proposed approaches. The OU-ISIR LP Dataset [24] consists of 4016 subjects (with age ranging from 1 to 94 years), as shown in Fig. 8.The camera was set at a distance of approximately 8 m from the straight walking course and a height of approximately 5 m. The image resolution and frame rate were 640 by 480 pixels and 30 fps, respectively. Each subject was asked to walk straight three times at his/her preferred speed. Each dataset comprises two main subsets, A and B. A is a set of two sequences (gallery and probe sequences) per subject. B is a set of one sequences per subject. In addition, each of the

Table 3 Comparison with existing methods on CASIA Dataset B in terms of correct recognition rates (CRR) and standard deviations (SD)

METHOD	Rank-1(%)	Rank-2(%)	Rank-3(%)	Rank-4(%)	Rank-5(%)	Rank-10(%)	Rank-15(%)	Rank-20(%)	AVERAGE CRR(%)	SD
Original GEI	33.25	39.90	52.81	55.10	55.55	69.86	74.55	81.17	57.77	5.78
A-VTM	45.65	63.81	67.16	77.38	80.57	91.18	92.74	95.34	76.73	4.15
CDHMM	32.91	46.11	55.97	59.11	60.93	74.77	78.89	85.57	61.78	5.58
(2D)2PCA	40.76	50.43	59.75	63.70	63.10	71.68	73.59	83.34	63.29	5.49
Simple CNN	52.63	63.54	76.54	77.33	81.86	87.88	91.30	94.00	78.13	3.88
TB-CNN	58.13	73.47	79.80	78.31	82.54	89.35	91.51	94.80	80.99	2.83
GCF-CNN	62.36	77.59	81.54	81.26	83.72	92.29	94.50	95.00	83.53	1.89

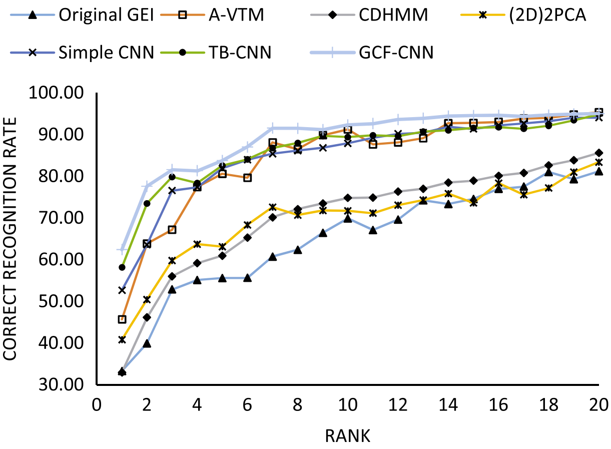


Fig. 7 The CMC curves of different approaches in Experiments on CASIA Dataset B

main subsets is further devided into 5 subsets based on the observation angles, 55 degrees, 65 degrees, 75 degrees, 85 degrees, and including all four angles.

In these experiments, sequence A and B are selected as the training set and testing set respectively. The CMC curves of six methods are shown in Fig. 9. From Fig. 9, the correct classification rate of the proposed approach is higher than the other methods for rank number less than 15. In addition, compared with the results in Experiment 2, the overall correct classification rates increase because of less interference factors in OU-ISIR LP Dataset, such as clothing and carrying condition changes, and smaller variations in view angle. Besides, the correct recognition rates and standard deviation of the proposed method are compared with those of existing methods in Table 4, which show that the proposed GCF-CNN method has good performance in correct recognition rate and robustness.

4.4 Comprehensive analysis

In this section , we will comprehensively discuss the average correct recognition rate (ACRR), standard deviations (SD) and time complexity of the proposed GCF-CNN models. This experiment is conducted on CASIA Dataset B, and the results are shown in Table 5, where NPC refers to the number of primary classifiers in one model, ACRR-1 is the average correct recognition rate with Rank 1, and AT means the average time to complete one time of gait recognition. The training set and testing set are constructed by randomly selecting 50% for training and the rest 50% for testing. The hardware environment of this experiment includes: Intel Xeon Silver 4214, 2.2G CPU; 128GB (64GB * 2) RDIMM Memory; 2 * RTX 2080Ti GPU.

Fig. 8 Samples from the OU-ISIR LP Dataset

Table 4 Comparison with existing methods on OU-ISIR-LP in terms of correct recognition rates (CRR) and standard deviations (SD)

METHOD	Rank-1(%)	Rank-2(%)	Rank-3(%)	Rank-4(%)	Rank-5(%)	Rank-10(%)	Rank-15(%)	Rank-20(%)	AVERAGE CRR(%)	SD
Original GEI	33.95	40.74	53.92	56.26	56.72	71.33	76.12	82.88	58.99	5.61
A-VTM	46.60	65.14	68.56	78.99	82.24	93.08	94.67	97.32	78.33	3.54
CDHMM	34.15	47.84	58.08	61.34	63.23	77.58	81.86	88.79	64.11	5.51
(2D)2PCA	42.51	52.60	62.33	66.44	65.81	74.76	76.75	86.92	66.02	4.93
Simple CNN	54.31	65.57	78.98	79.80	84.47	90.68	96.28	97.00	80.89	3.03
TB-CNN	59.05	75.53	81.48	81.08	82.02	90.46	96.75	97.61	83.26	2.56
GCF-CNN	64.33	80.04	84.11	83.83	86.36	95.21	96.45	98.00	86.04	1.19

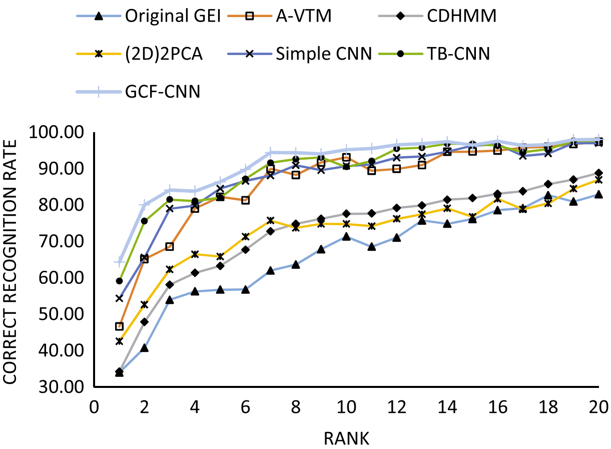


Fig. 9 The CMC curves of different approaches in Experiments on OU-ISIR LP Dataset

From Table 5, we can see that with the increasing of NPC, ACCR-1 and AT increased, while SD decreased. This means that when we increase the number of primary classifiers, the average recognition rate and algorithm stability are improved to a certain extent, but the time consumed is also increasing. This is because, with the increase of the number of primary classifiers, the expression ability and feature extraction ability of the GCF-CNN model are improved, but, at the same time, the complexity of the model increases, resulting in the complexity of the corresponding recognition algorithm.

On the other hand, we also can see from Table 5 that with the increase of NPC, the increasing speed of ACCR-1 gradually slowed down, but AT increased steadily. This shows that when the number of primary classifiers reaches a certain threshold, to further increase the number of primary classifiers to improve the average correct recognition rate and algorithm robustness is not a good choice. Therefore, when we finally implement the GCF-CNN model, we adopt a relatively good compromise scheme, that is, NPC value is set to 5.

5 Conclusion and future work

In this paper, a gait classification framework based on CNN Ensemble (GCF-CNN) is proposed. We first utilize a Bagging-like strategy to preprocess the traditional GEIs to create slightly different training sets. Then, diverse CNN primary learners are trained separately with different hper-parameters and training sets. Finally, after obtaining the output of each CNN, we use them as inputs to train a secondary learner for combining the primary learners.

Table 5 Experiment results of GCF-CNNs with different number of primary classifiers

NPC	1	2	3	4	5	6	7	8	9	10
ACRR-1(%)	57.23	62.78	64.25	65.19	65.64	65.70	65.81	65.88	65.90	65.93
SD	2.31	2.01	1.80	1.59	1.39	1.31	1.25	1.24	1.20	1.17
AT(ms)	7.52	9.13	10.66	12.15	13.57	14.98	16.27	17.84	19.11	20.59

To the best of our knowledge, this is the first time we work for gait recognition using Ensemble learning technologies. The proposed framework is evaluated on the CASIA Dataset B and OU-ISIR LP Dataset and performs better under different conditions with respect to several existing approaches.

The limitation of our method is that the framework only uses homogeneous CNN classifiers. How to integrate other types of classifiers, such as support vector machine (SVM) and decision tree (DT), is a valuable research direction. In addition, our comparative experiments only use experimental datasets. How to further evaluate the proposed method in practical application is our future work.

Acknowledgements This work is supported by the National Natural Science Foundation of China (No. 61303146 and 61602431) and the Natural Science Foundation of Zhejiang Province (No.Y20F020113).

Compliance with Ethical Standards

Conflict of interests We declare that we have not financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled "Gait Classification Through CNN-based Ensemble Learning".

References

1. Ariyanto G, Nixon M (2011) Model-based 3D gait biometrics. In: International Conference on Biometrics. Washington, DC USA
2. Aussem A, Elghazel H (2015) Unsupervised feature selection with ensemble learning. Mach Learn 98:157–180
3. Connor ARP (2018) Biometric recognition by gait: A survey of modalities and features. Comput Vis Image Underst 167(01):1–27
4. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press
5. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(02):316–323
6. Jia N, Sanchez V, Li C (2017) Learning optimized representations for view-invariant gait recognition. In: International Joint Conference on Biometrics, pp 774–780, Denver, USA
7. Kusakunniran W, Wu Q, Li H, Zhang J (2009) Multiple views gait recognition using view transformation model based on optimized gait energy image. In: IEEE ICCV, pp 1058–1064, Kyoto, Japan
8. Kusakunniran W, Wu Q, Zhang J, Li H, Wang L (2014) Recognizing gaits across views through correlated motion co-clustering. IEEE Trans Image Process 23(2):696–709
9. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(5):436–445
10. Li X (2018) Preconditioned stochastic gradient descent. IEEE Transactions on Neural Networks and Learning Systems 29(5):1454–1466
11. Li J, Ma S, Le T, Liu L, Liu J (2017) Causal decision trees. IEEE Trans Knowl Data Eng 29(2):257–271
12. Luo J, Tang J, Tjahjadi T (2016) Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis. Pattern Recogn 60:361–377
13. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: IEEE ECCV, pp 151–163, Graz, Austria
14. 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
15. Sahu A, Runger G, Apley D (2011) Image denoising with a multi-phase kernel principal component approach and an ensemble version. In: 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pages 1–7
16. Sarkar S, Phillips P, Liu Z (2005) The humanid gait challenge problem: Data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27(02):162–177
17. Schaar MVD Tekin C, Yoon J (2015) Adaptive ensemble learning with confidence bounds. IEEE Trans Signal Process 99:1–10

18. Shiraga K, Makihara Y, Muramatsu D (2016) GEINet: View-invariant gait recognition using a convolutional neural network. In: International Conference on Biometrics. Halmstad, Sweden
19. 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 Transactions on Circuits and Systems for Video Technology, 28(1)
20. Tang J, Luo J, Tjahjadi T (2017) Robust arbitrary-view gait recognition based on 3D partial similarity matching. IEEE Trans Image Process 26(1):7–23
21. Tao D, Li X, Wu X, Maybank S (2007) General tensor discriminant analysis and Gabor features for gait recognition. EEE Trans Pattern Anal Mach Intell 29(10):1700–1715
22. Tong M, Li M, He B, Ma L, Zhao M (2020) DKD–DAD: a novel framework with discriminative kinematic descriptor and deep attention-pooled descriptor for action recognition. Neural Computing and Applications volume 32:5285–5302
23. Tong M, Zhao M, Chen Y, Houyi W (2019) D3-LND: A two-stream framework with discriminant deep descriptor, linear cmdt and nonlinear kcmdt descriptors for action recognition. Neurocomputing 325:90–100
24. Uddin MZ, Ngo TT, Makihara Y, Takemura N, Li X, Muramatsu D, Yagi Y (2018) The ou-isir large population gait database with real-life carried object and its performance evaluation. IPSJ Trans on Computer Vis Appl 10(1):1–8
25. Wang X, Feng S, Yan WQ (2019) Human gait recognition based on self-adaptive hidden markov model. IEEE Transactions on Computational Biology and Bioinformatics 1(1):1–10
26. Wang X, Wang J, Yan K (2018) Gait recognition based on gabor wavelets and $(2D)^2PCA$. Multimedia Tools and Applications 77(10):12545–12561
27. Wang X, Yan K (2016) Human gait recognition using continuous density hidden Markov models. Pattern Recognit Artif Intell 29(8):709–717
28. Wang X, Yan WQ (2020) Cross-view gait recognition through ensemble learning. Neural Comput and Applic 32:7275–7287
29. Wang X, Yan WQ (2020) Human gait recognition based on frame-by-frame gait energy images and convolutional long short term memory. International Journal of Neural Systems, 30(1)
30. Wolf T, Babaee M, Rigoll G (2016) Multi-view gait recognition using 3D convolutional neural networks. In: IEEE International Conference on Image Processing, pp 4165–4169, Phoenix, USA
31. 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(02):209–226
32. 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
33. Zhang R, Xu Z, Huang G, Wang D (2012) Global convergence of online BP training with dynamic learning rate. IEEE Transactions on Neural Networks and Learning Systems 23(2):330–341
34. Zhao G, Liu G, Li H, Pietikainen M (2006) 3D gait recognition using multiple cameras. In: International Conference on Automatic Face and Gesture Recognition. Southampton, UK

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.