LFN: Based on the Convolutional Neural Network of Gait Recognition Method

Xiaopeng Zhu ¹, Li Jun Yun ¹, ², *, Feiyan Cheng ¹ and Chunjie Zhang ¹

¹ School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
² Yunnan Key Laboratory of Opto-electronic Information Technology, Kunming 650500, China

554618609@qq.com, *, b Corresponding author Email: byunlijun@ynnu.edu.cn

Abstract. In this paper, a novel method based on convolutional neural network (CNN) to address gait recognition was proposed. Gait is a unique biologic feature that the feature almost cannot be altered. Existing gait recognition virtually based on a traditional method such as Gait Energy Image (GEI). GEI has various gait silhouette sequence images and put this image together. Meanwhile, those images must in the same gait period, which lacks of flexibility. To address this issue, a different gait recognition model, called LFN, based on a convolution neural network (CNN) was proposed. The model is composed of three convolution layers in parallel. It can be trained in an end-to-end manner and each sample gait silhouette in the uniform gait period does not to be required. Our working base on the OU-ISIR large population gait dataset. Being based on the result of the experiments, our network can learn the significant features of each sample gait silhouette sequence effectively, at the same time, our model can achieve high and stable accuracy in three experiments.

Keywords: Gait recognition, Convolution neural network, LFN, Gait dataset.

1. Introduction

Gait is similar to iris, facial, fingerprint, an important feature of biologic, meanwhile is one of complex behavioral trait, was paid attention to academic research that everyone has their own gait habit. Likewise, the profile of the private individuals may be altered but subtle differences in muscle strength, tendon and bone length, bone density, coordination, and so on, everyone has their gait [1], so that gait recognition has been applied to various scenarios which are utilized in the investigation into crime that is a most important field. So that gait can be used in many scenes in identity authentication [2].

However, compare to iris recognition, facial recognition, fingerprint recognition, the gait recognition is hardly researched. Such as, in iris recognition research Mahesh Kumar Morampudi, Sowmya Veldandi, Munaga V. N. K. Prasad, U. S. N. Raju, they propose a novel Privacy Preserving (PP) multi-instance iris remote authentication system to accord with attacks at the malicious server and over the transmission channel [3]. In this study, the security and accuracy of iris recognition was improved. In facial detection, Manoj Prabhakaran Kumar, Manoj Kumar Rajagopal, the normalized minimal feature vectors using semi-supervised Twin Support Vector Machine (TWSVM) learning was
selected [4]. Also in the area of facial recognition, Oualid Laiadi, Abdelmalik Ouamane, Abdelhamid Benakcha, Abdelmalik Taleb-Ahmed, Abdenour Hadid, they use the metric learning (SILD-WCCN/LR) approach to kinship verification combined with deep and shallow features (multi-view features) to address the kinship verification [5]. Yonghong Liu, Baicun Zhou, Congying Han, Tiande Guo, Jin Qin, designed a novel method based on deep learning for aligned fingerprints matching [6]. These fields have a lot of study but in gait recognition the machine learning is rare proposed.

The gait data are also very easy to collect that it does not depend on distance, speed, clothing and does not have to touch each other. On the other hand, iris recognition requires the target to be within 30 centimeters, face recognition need the target to keep within 5 meters, and the target of gait recognition only stay within 50 meters under ideal conditions. Compare with the traditional Gait Energy Image (GEI) method [7], based on the convolution neural network has some advantages for instance convolution neural network (CNN) doesn't require gait silhouette in the same gait period and we don't need to generate a lot of GEI.

Nowadays, a few different ways about gait recognition are proposed by researchers. For example, markerless gait utilize camera or depth sensor to capture the body images, at the same time to directly calculate the joint positions from the body images [8]. After markless the gait capture system reduce the lots of times, so that could be used in widespread applications.

Compare to fingerprint recognition, iris recognition and face recognition, gait recognition requires video data collection of the identified object in advance. And the gait is based on a video sequence, which captures continuous changes in the process of walking and finds its unique characteristics, so this method is relatively safe. With the data collection method, specialized equipment is not needed that only need to collect the gait of the target and process these gait images, for example, cropping, binarization, etc. Gait recognition has already and the process applied to the investigation into a lot of cases to identify who is suspecting. This method has been achieved a very well precision rate.

However, in gait recognition tasks, the accuracy is impacted by many factors, such as the traveling speed, the angle of camera, clothing, and impedimenta, etc. So a great method must solve these problems.

2. Related Work
In this section, briefly description will be reviewed from the work of others and us, others have suggested and the general structure of our model.

2.1. Others works.
In the early work, the RGB images were captured from the camer that be used in gait recognition tasks [9]. Later, the depth camera was invented that the main view gait outlines was used in gait recognition [10][11]. Thus, the field of gait recognition has been greatly developed.

In recent years, many gait recognition methods have been proposed. For example, Maryam Babaee and Linwei Li, etc. [12], proposed a deep learning model to transform incomplete GEI to the corresponding complete GEI obtained from a full gait cycle. Francesco Battistone and Alfredo Petrosino, etc. [13] adopts a robust deep learning model based on graphics. This method called Time based Graph Long Short-Term Memory (TGLSTM) network that is able to dynamically learn graphs when they may change during the time [9]. Another appearance-based method as gait entropy image (GEnI) [14], the novel gait representation was proposed. The method based on computing entropy, a GEnI encodes in a single image the randomness of pixel values in the silhouette images over a complete gait cycle. It thus captures mostly motion information and is robust to covariate condition changes that affect appearance. And Jose Portillo-Portillo, Roberto Leyva, Victor Sanchez, etc. they proposed a view-invariant gait recognition algorithm, which builds a unique view invariant model taking advantage of the dimensionality reduction provided by the Direct Linear Discriminant Analysis (DLDA)[15], this method used GEI and DLDA, which can solve the errors caused by the problem of Angle of view.
On the other hand, Rubén Delgado-Escaño, Francisco M. Castro, Julián Ramos Cózar, etc. [16] uses a non-traditional approaches. They are approaching need input the original data, and the model to have the ability to automatically learn the best representations without any constraint introduced by the preprocessed features [11]. They adopt OU-ISIR dataset that is a series of intertial sensors to get original data of gait.

For the tensor analysis to identify human gaits is a significant method. However, tensor analysis method is susceptible when the angle has been changed. Xianye Ben, Peng Zhang, ZhihuiLai, etc. [12] offered a new method that is a general tensor representation framework for cross-view gait [17]. And the gait recognition precision will be impacted by many factors, such as clothing, carrying and view angle to address this issue, Ziyuan Zhang, Luan Tran, Xi Yin, etc. proposed an AutoEncoder framework to explicitly disentangle pose and appearance features from RGB imagery and the LSTM-based integration of pose features over time produces the gait feature [18]. Zifeng Wu et al. [19] proposed a gait recognition method network based on deep convolutional neural network (CNN). Using a small group of tagged videos of people walking from multiple perspectives, the deep network can be trained to identify people.

2.2. Our model.
In this area of research, many researchers has adopted the method of Gait Energy Image (GEI) [7] to solve the gait recognition task. The Gait Energy Image (GEI) was proposed by Ju Han, Bir Bhanu, to address the problem of the lack of training templates, they take advantage of combing statistical gait features from real and synthetic templates [7].

In our model, the binarization human silhouette is used as model input [20], so the silhouette no more processing is required. The method has been used in computer vision domains like content recommendation [21] and image captioning [22] put the features together. Our model is called, LFN, it has three convolutional layers in parallel, and one shared fully connected block. The systemic test and contrast experiment result showed that LFN has high accuracy rate with others models.

3. GaitSet and Model
In this section, the structure of model and gaits will be explained in detail. The overall model struct is shown in Figure 1.

Fig. 1 The structured flowchart of the LFN
The model of convolution layer's detail data include kernel, channel and image resolution will be given in Table 1. There are two keys pointed out specially in the model that the first, except for the pooling layer of third parallel layer in our model, all pooling layers was used pooling kernel size of 2. In the pooling layer of third parallel layer, the size of 5 was chosen. The decision was made to extract
a wider range of gait characteristics. Later, the Rectified Linear Unit (ReLU), that is a common activation function in artificial neural network, was selected. To use the ReLU that the problem of gradient explosion and gradient disappearance can be avoided.

Table 1. The model of convolution layer's detail data

| layers   | kernel | channel | resolution |
|----------|--------|---------|------------|
| input    |        | 3       | 128*128    |
| Conv0    | 3      | 32      | 126*126    |
| Conv1_1  | 3      | 64      | 61*61      |
| Conv1_2  | 3      | 128     | 28*28      |
| Conv2_1  | 3      | 64      | 61*61      |
| Conv2_2  | 3      | 128     | 28*28      |
| Conv3    | 3      | 256     | 12*12      |
| Conv4    | 3      | 256     | 61*61      |

3.1. GaitSet.
The OU-ISIR Gait Database of Large Population Dataset is used in this paper [23], that the OU-ISIR Gait Database is meant to aid research efforts in the general area of developing, testing and evaluating algorithms for gait-based human identification. The Institute of Scientific and Industrial Research (ISIR), Osaka University (OU) has copyright in the collection of gait video and associated data and serves as a distributor of the OU-ISIR Gait Database.

And The data have been collected by since March 2009 through outreach activity events in Japan. The approved informed consent was obtained from all the subjects in this dataset. The data set consists of persons walking on the ground surrounded by the 2 cameras at 30 fps, 640 by 480 pixels. The datasets are basically distributed in a form of silhouette sequences registered and size-normalized to 88 by 128 pixels size [24].

A dataset is defined, it includes N people with an identity tag, \( y_i \in \{1, 2, 3, \ldots, N\} \) in this dataset every subject has about one hundred and fifty silhouettes, therefore, each subject of n silhouettes can be classified as a collection, \( X_i = \{X_j^i \mid j = 1, 2, \ldots, n\} \) [1] By the way, 20% of them were randomly divided into test datasets, and the remaining 80% were used as training datasets.

3.2. Model.
Recently, the gait recognition task nearly utilizes the GEI method to solve this issue, but generating a GEI silhouette has more trouble. To address this task, a model based on convolutional neural network (CNN) is proposed, it called LFN, in this paper that the model has three convolutional layers in parallel, and one shared fully connected block. When the data input to the model, at first the data will be processed by a convolutional layer, soon afterward, after activation layer and pooling layer treatments, the data will be copied in three as shown in Figure 1.

In this addition, The gait recognition task is solved by the following three steps:

\[
F_i = F_3 \left( F_{1_2} \left( F_{1_1} \left( F_0 \left( X_i \right) \right) \right) \right) + F_{2_2} \left( F_{2_1} \left( F_0 \left( X_i \right) \right) \right) + F_4 \left( F_0 \left( X_i \right) \right) \tag{1}
\]

Where \( F_0 \) is the first convolutional layer of LFN, the first convolutional layer aims to extract the gait silhouette features preliminarily.

The F1_1, F1_2, F2_1, and F2_2 are convolutional, activation and maxpooling layers of the first two layers in the model of parallel structure. And the maxpooling of kernel size and stride all are 2. Those convolutional layers are designed to learn more and more minutiae, and when using a convolutional layer to accomplish this task, the model does not produce a stable effect.

And the function of \( F_3 \) is merging the feature of the first two layers and to extract more and more detail feature.
The $F_i$ is convolutional, activation and maxpooling layer, beyond the first two layers the maxpooling layer of $F_i$ the maxpooling kernel size and the stride of 5 is designed and this is designed to include biggish detail features. Subsequently, the first two layers and $F_i$ are added so that more detail can be obtained at the same time. After passing through the activation and maxpooling layers of kernel size and the stride of 2, those features will be passed to the first fully connected layer.

And after designing the activation layer, in order to prevent the LFN form overfitting in the process of feature learning. In our early experiments, after lots of iterations the result appeared overfitting. In order resolve this problem, a dropout object is set after the second fully connected layer. In the dropout layer have a half of the features are randomly discarded. According to the experiment, we proved that the test data of loss doesn't converge ahead of time, our approach has a certain effect.

3.3. Loss function.
In the LFN, Negative log-likelihood loss (NLLLoss)[25] is used as a loss function. And the NLLLoss formulated as:

$$L(Y, P(Y | X)) = -\log P(Y | X) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log(p_{ij})$$

(2)

The NLLLoss of input is a log probability and a target label so it does not compute log probability for us. And in this function, the Y means output, the X is input, the L is loss function. N is a number of samples; M means number of class. Judge the result of output was used $y_{ij}$, and the $p_{ij}$ means the model to predict the probability that the input instance $x_i$ belongs to category j. For the reasons given above, in the last layer of the model, the log-softmax function is selected. The formulated as:

$$LogSoftmax(x_i) = \log \left( \frac{\exp(x_i)}{\sum_j \exp(x_j)} \right)$$

(3)

But it was verified by our later experiments that if you choose the crossentropyloss function and the softmax function you can get the same effect.

4. Experiments
The experiment mainly consists of two parts. The first part, Comparing LFN with three convolution layers in parallel. Testing the LFN with one convolutional layer, the first two convolutional layers and the three convolutional layers in turn. The second part, LFN was compared to the most advanced gait recognition methods. And an Adam is used as an optimizer[26].

4.1. Self contrast.
In the experimental setup part of this paper, we compared the single-layer LFN model, the double-layer LFN model and the complete LFN model in the same data set and under the same super-parameter setting conditions, and analyzed the recall rate, accuracy, F1 value, etc.
Fig. 2 The accuracy of single-layer LFN

As shown in Figure 2, this figure is the accuracy of the training set and test set of the single-layer LFN model after 1000 rounds of training. It can be seen from the figure that after about 900 rounds of iterative training, the accuracy of the single-layer LFN model converged below 97.5% with a large fluctuation.

Figure 3 below shows the performance of the LFN model with double-layer structure after 1000 rounds of training.

Fig. 3 The accuracy of double-deck LFN

It can be seen from figure 3 above that, under the same super-parameter setting condition, the average accuracy of the LFN model with two-layer structure on the training set is improved to a
certain degree, reaching about 97.5%, after 800 rounds of iterative training, compared with that of the LFN model with single-layer structure.

Figure 4 shows the accuracy of the LFN model with complete structure. Similarly, the number of training sessions and the setting of super-parameters were consistent with the previous two experiments.

![Accuracy of complete LFN](image)

**Fig. 4** The accuracy of complete LFN

It can be seen from figure 4 above that after 900 rounds of training, the accuracy of LFN model in the test set can converge to more than 97.5%, which is further improved compared with the single-layer LFN model and the double-layer LFN model. And judging from the rising trend of accuracy, there is the possibility of further improvement.

The figure 5 below shows the comparison of the accuracy of each structure in the testing set in the three experiments.

![Comparison of the accuracy of three structures](image)

**Fig. 5** Comparison of the accuracy of three structures
As can be seen from figure 5, the green broken line is the LFN model with complete structure, the orange is the accuracy curve of the double-layer structure, and the blue is the accuracy curve of the single-layer structure. As can be seen from the figure above, the green curve is at the top of the three curves, the orange curve is at the middle, and the blue curve is at the bottom, which is also consistent with our conclusion. Under the same data set and the same super-parameter setting, the complete LFN model has the best recognition effect on human gait.

The figure 6 below is the result of testing the models of the three structures after we re-selected the verification set of the 1,000 human gait samples. The recall rate, accuracy and F1 values of our computer in each structure were obtained.

Fig. 6 The index comparison of the three structures

Where, subscript A represents the recall rate, accuracy and F1 value of the LFN model with single-layer structure; similarly, subscript B represents the LFN model with double-layer structure; subscript C represents the LFN model with complete structure. In the figure, the abscissa is each sample, and the ordinate is the corresponding sample number. In the three metrics in three kinds of structure, but has a complete structure LFN model volatility relative to minimum, and the volatility of almost the same as the sample after the validation set data analysis we found that volatility is that we choose to achieve validation set is the cause of image quality is appeared really exist, so lead to the distortion in some image appeared on the classification of certain deviation.
4.2. Compare with other model.
Comparing the accuracy of the model with other advanced models. Compared data is shown in Table 2. After comparing with other models, it can be found that LFN has made great progress in accuracy.

Table 2. Compare with the LFN and other model of accuracy

|                  | 500 times | 700 times | 800 times | 900 times | max   |
|------------------|-----------|-----------|-----------|-----------|-------|
| Single layer of struct | 97.43%    | 97.6%     | 97.36%    | 97.39%    | 97.85%|
| Dual layer of struct   | 97.01%    | 97.31%    | 97.5%     | 97.44%    | 97.87%|
| Completely LFN        | 97.28%    | 97.47%    | 97.77%    | 97.73%    | 98.04%|
| CNN-3D(Wu et al. 2017)[19] | ——        | ——        | ——        | ——        | 92.1% |
| CNN-Ensemble(Wu et al. 2017)[19] | ——        | ——        | ——        | ——        | 94.1% |

From the above table of experimental data, it can be seen that in the self-comparison of LFN, the performance of the model with complete structure is indeed due to the single-layer structure and the double-layer structure model. Because of the complete model has deeper network structure and more parameters, which can obtain richer detailed features. And the LFN also has good performance compared to other models.

5. Conclusion
This paper designs a new gait recognition model based on deep learning, called LFN. The LFN has the three pipes of convolutional neural networks that the struct is able to extract the different size features in the gait silhouette. The struct is capable of improving the model accuracy in a large part.

By comparing the results of multiple experiments, LFN has the best performance compared with single-layer structure and double-layer structure when using three-layer structure, meanwhile, the overall accuracy and stability of the model are the best. And LFN also has some advantages over other gait recognition models.

Acknowledgments
This study was financially supported by Key Project of Applied Basic Research Program of Yunnan Province (2018FA033).

References
[1] Imed Bouchrika, John N. Carter, Mark S. Nixon. Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras, Multimedia Tools and Applications, pp.1380-7501 (2014)
[2] Hanqing Chao, Yiwei He, Junping Zhang, etc. GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition, The Thirty-Third AAAI Conference on Artificial Intelligence(AAAI), Vol.33 (2019)
[3] Morampudi, M.K., Veldandi, S., Prasad, M.V.N.K. et al. Multi-instance iris remote authentication using private multi-class perceptron on malicious cloud server. Applied Intelligence, (2020)
[4] Kumar, M.P., Rajagopal, M.K. Detecting facial emotions using normalized minimal feature vectors and semi-supervised twin support vector machines classifier. Applied Intelligence, Vol.49, pp.4150-4174 (2019)
[5] Laiadi, O., Ouamane, A., Benakcha, A. et al. Learning multi-view deep and shallow features through new discriminative subspace for bi-subject and tri-subject kinship verification. Applied Intelligence, Vol.49, pp.3894-3908 (2019)
[6] Liu, Y., Zhou, B., Han, C. et al. A novel method based on deep learning for aligned fingerprints
[7] Ju Han, Bir Bhanu. Individual recognition using gait energy image, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.28, pp.316-322 (2006).

[8] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, et al. Real-time human pose recognition in parts from single depth images, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1297-1304 (2011).

[9] Yanxiang Wang, Bowen Du, Yiran Shen, Kai Wu, Guangrong Zhao, Jianguo Sun, Hongkai Wen. EV-Gait: Event-Based Robust Gait Recognition Using Dynamic Vision Sensors, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.6358-6367 (2019).

[10] P. Chattopadhyay, A. Roy, S. Sural, and J. Mukhopadhyay. Pose depth volume extraction from rgb-d streams for frontal gait recognition, Journal of Visual Communication and Image Representation, Vol.1, pp.53-63 (2014).

[11] Nunes J.F., Moreira P.M., Tavares J.M.R.S., GRIDDS - A Gait Recognition Image and Depth Dataset, Computational Vision and Biomechanics, Vol.34.

[12] Maryam Babaee, Linwei Lia, Gerhard Rigolla. Person Identification from Partial Gait Cycle Using Fully Convolutional Neural Network, Neurocomputing, Vol.338, pp.116-125 (2019).

[13] Francesco Battistone, Alfredo Petrosino. TGLSTM: A time based graph deep learning approach to gait recognition, Pattern Recognition Letters, pp.1-7 (2018).

[14] K. Bashir, T. Xiang, and S. Gong. Gait Recognition Using Gait Entropy Image, International Conference on Imaging for Crime Detection and Prevention (ICDP), (2010).

[15] Portillo-Portillo, J., Leyva, R., Sanchez, V. et al. A view-invariant gait recognition algorithm based on a joint-direct linear discriminant analysis. Applied Intelligence, Vol.48, pp.1200-1217 (2018).

[16] Rubén Delgado-Escaño , Francisco M. Castro, Julián Ramos Cózar, et al. An end-to-end multi-task and fusion CNN for inertial-based gait recognition, IEEE Access, Vol.7, pp.1897-1908 (2018).

[17] Xianye Ben, Peng Zhang, ZhihuiLai, etc. A general tensor representation framework for cross-view gait recognition, Pattern Recognition, Vol.90, pp.87-98 (2019).

[18] Ziyuan Zhang, Luan Tran, Xi Yin, etc. Gait Recognition via Disentangled Representation Learning, The IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.4710-4719 (2019).

[19] Wu, Z.; Huang, Y.; Wang, L.; Wang, X.; and Tan, T. A comprehensive study on crossview gait based human identification with deep CNNs, IEEE TPAMI, Vol.39, No. 2, pp.209-226 (2017).

[20] Charles, R. Q.; Su, H.; Kaichun, M.; and Guibas, L. J. Pointnet: Deep learning on point sets for 3D classification and segmentation, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.77-85 (2017).

[21] Hamilton, W.; Ying, Z.; and Leskovec, J. Inductive representation learning on large graphs, In Neural Information Processing System (NIPS), pp.1024-1034 (2017).

[22] Krause, J.; Johnson, J.; Krishna, R.; and Fei-Fei, L. A hierarchical approach for generating descriptive image paragraphs, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3337-3345 (2017).

[23] Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, IEEE Trans. on Information Forensics and Security, Vol.7, No. 5, pp.1511-1521 (2012).

[24] Takemura, N.; Makihara, Y.; Muramatsu, D.; Echigo, T.; and Yagi, Y. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition, IPSJ Transactions on Computer Vision and Applications (CVA), Vol.10, No. 4, pp.1-14 (2018).

[25] Donglai Zhu, Hengshuai Yao, Bei Jiang, Peng Yu. Negative Log Likelihood Ratio Loss for
Deep Neural Network Classification. arXiv:1804.10690 (2018)

[26] Kingma, D. P., and Ba, J. Adam: A method for stochastic optimization, In the 3rd International Conference for Learning Representations (ICLR), (2015)