Gait Recognition based on Improved LeNet Network

Xiaoqiang Shao[1] ,Xinchao Nie[1] ,Xuan Zhao[1] ,Runyang Zheng[1] ,Defeng Guo[1]
1. College of Electrical and Control Engineering, Xi'an University of Science and Technology
Xi'an, China
1005507659@qq.com, 15621396713@163.com, 1124774458@qq.com, 1173080774@qq.com,2670314531@qq.com

***Abstract*—Today, biometrics play an important role in personal identification systems, but there are also many problems. For example, face recognition is susceptible to lighting, makeup, age and distance, and fingerprints are easily forged. Gait recognition is to identify the identity through the walking posture of the person. The gait features are characterized by long-distance, non-contact and non-disguise, which can overcome the defects of face and fingerprint features. However, gait recognition is susceptible to covariates such as clothing, backpacks, and perspectives. In order to reduce the influence of covariates on gait recognition, this paper conducts an in-depth study on gait recognition algorithms. The main contents are as follows: for the residual network model, not only can the network convergence speed be accelerated, but also the recognition rate can be improved. Therefore, the idea of the residual network is combined into the LeNet network model to improve the LeNet network model. Verify the performance of the improved LeNet network model, the improved LeNet network model improved the recognition rate of normal, backpack and coat walking by 0.3%, 0.5%, and 0.8% respectively. Therefore, the improved LeNet network model makes the network converge faster and the recognition rate is improved to some extent.**

Keywords—Biometrics; Gait recognition; Gait energy map; LeNet

I. INTRODUCTION

As a behavioral feature, gait is a technique for identifying individuals based on the differences in gaits of different individuals. Since the traditional method is to extract features manually, and the deep learning method is to automatically extract features, the deep learning method is more complete and more convenient to operate than the traditional methods. The gait recognition methods based on deep learning can be roughly divided into two categories: a method based on Convolutional Neural Networks (CNN) and a method based on Three-Dimension Convolutional Neural Networks (3DCNN).

Munif Alotaibi et al proposed a specialized deep convolutional neural network (SDCNN) for gait recognition, which is less sensitive to occlusion and other covariates [1]. In order to highlight the importance of different body parts, Huimin Wu et al. proposed the Feedback Weight Convolutional Neural Network (FBW-CNN) method [2], which first extracted from the input sequence through the pre-trained CNN model. A set of depth feature vectors is then trained on the kernel function of the fully connected layer as the guiding weight of each receptive field of the input sequence, and finally the input layer is updated by the weighted receptive field and the CNN model is retrained using the new image. He Zhengyi proposed a gait recognition and simulation method based on integrated CNN model and deep information network for difficult identification and prediction of multiple gait sequences [3]. Li Ying proposed a gait recognition algorithm based on the perspective of three-dimensional convolutional neural network for the problem of cross-view in gait recognition [4]. The algorithm first extracts local and global features through three-dimensional convolutional neural network and migration learning methods, then combines local and global features by serial fusion and selects features using Principal Component Analysis (PCA). Feature classification is performed by a method of support vector machine. At present, although researchers at home and abroad have made more in-depth research on the theory and methods of gait recognition technology, and the recognition rate has improved, there are still some problems to be solved in this field [5]. The main problems are as follows:

(1) Acquisition of gait contour map: extracting the foreground image from the gait video sequence and performing preprocessing such as binarization and normalization to obtain the gait contour map, but the background image is complex and constantly changing. This will affect the quality of the gait profile, resulting in a lower recognition rate. Therefore, obtaining a complete and accurate gait contour image from a complex and constantly changing background image is the key to gait recognition;

(2) Clothes: When a pedestrian wears a long or loose dress, this will change the size of the pedestrian and block the part of the leg, thus affecting the identification result of the identity;

(3) Carrying objects: different carriers will block the legs to varying degrees, thus affecting the integrity of the gait profile;

(4) Viewing angle: Since the video capture device is generally fixed, the captured gait images are not completely the same perspective;

(5) Speed: People's walking state is different (normal walking, jogging and running, etc.), and the speed is different. For the same person, the difference in human gait in different walking situations increases the difficulty of gait recognition [6].

978-1-7281-4390-3/20/$31.00 ©2020 IEEE

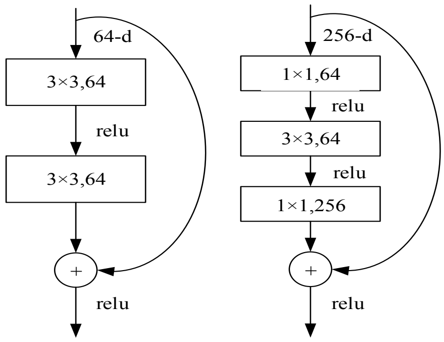
In view of the above problems, this paper mainly solves the problem that the recognition rate is not ideal due to covariates (clothes and backpacks), and proposes a gait recognition algorithm based on improved LeNet network, so as to improve the recognition rate to some extent.

II. NETWORK COMMONLY USED CONVOLUTIONAL NEURAL NETWORK MODEL

A. Network model introduction

Commonly used convolutional neural networks include LeNet, AlexNet [7], VGGNet, GoogLeNet, ResNet, and SeNet. Next, a brief introduction to these common network models.

In 1994, Yan LeCun, one of the three giants of deep learning, proposed the LeNet network model, which is used for the recognition and classification of handwritten characters with an accuracy rate of 98%. LeNet is a 5-layer network structure: three convolutional layers, two pooling layers and two fully connected layers. The convolution kernels of the three convolutional layers are all 5×5, and the activation function uses the Sigmoid function. In 2012, Alex Krizhevsky presented the AlexNet network model and won the ImageNet competition. For the AlexNet network model, the pooling layer, LRN (Local Response Normalization, LRN) and dropout layer are ignored. It consists of 5 convolution layers and 3 fully connected layers. Subsequent deepening of the VGGNet network model, using a small 3 × 3 convolution kernel, the same convolution module stacked, making the network more non-linear expression capabilities. The GoogLeNet network model consists of 22 layers and uses convolution kernels of size 1 × 1, 3 × 3 and 5 × 5 to make the network have multi-scale representation capabilities. As the depth of the convolutional neural network continues to deepen, degradation problems arise. The degradation problem is not over-fitting, but as the number of network layers increases, but the accuracy on the training set appears saturated or even decreases. In 2015, a deep convolutional neural network model, the residual network, was proposed. The residual network mainly solves the degradation problem and is easier to optimize, so that the network not only speeds up convergence, but also improves the recognition rate. The residual network achieved the first place in Image detection, ImageNet positioning, COCO detection and COCO segmentation tasks [8]. The residual network (ResNet) is composed of several residual modules stacked, as shown in Fig 1. Fig. 1(a) is a "basic" residual consisting of two convolutional layers with a convolution kernel size of 3 × 3. The difference unit, Fig. 1(b) is a "bottleneck" residual unit composed of three convolutional layers, where in the first convolutional layer and the third convolutional layer are respectively reduced in dimension and dimension. The depth residual network models of 18, 34, 50, 101 and 152 layers are given in [9].



(a) Basic Architectures (b) Bottleneck Architectures

Fig.1. Two residual units

III. LENET NETWORK MODEL

The whole process of gait recognition can be divided into three stages: gait detection, gait feature extraction and classification recognition. The gait recognition algorithm based on the improved LeNet network model is mainly divided into network training and testing processes.

A. Gait data set partition

The gait data set CASIA-B contains a sequence of gait contour images of 124 individuals, each of whom has a sequence of gait contour images of three different walking states at 11 viewing angles. Each person has 10 gait image sequences, 6 of which are normal walking (nm), 2 are backpack walking (bg), and 2 are gait sequences in coat walking (cl) state. For the experiments in this paper, the first set of training samples and test samples (nm-nm), the four gait sequences of each person's normal walking state (nm-01, nm-02, nm-03 and nm-04), a total of 5,456 sheets were used to train the model, two gait sequences (nm-05 and nm-06), a total of 2,728 sheets were used to test the network model; the second group of training samples and test samples (bg-bg), half of each person's backpack walking status (bg-01), a total of 1364 gait energy as a training sample, the other half (bg-02) a total of 1364 pieces as a test sample; the third group of training samples and test samples (cl-cl) Each person wore half of the coat status (cl-01) with a total of 1364 sheets as training samples and the other half (cl-02) with a total of 1364 sheets as test samples.

B. LeNet network model parameters

Common convolutional neural network models include LeNet, AlexNet, VGGNet, GoogLeNet, and ResNet. Since the gait database capacity is not very large, this paper first selects the LeNet model to train the model and test mode in the gait database. The LeNet model consists of a three-layer convolution kernel with a 5 × 5 convolution layer, two layers of pooled layers, and two layers of fully connected layers. For the gait dataset of this paper, the number and step size of the convolution kernel in the LeNet model are adjusted, as shown in Table I.

TABLE I. LENET NETWORK MODEL PARAMETERS

Layer name	Output size	Filter size
Conv1	64×64	$[5 \times 5, 1, 4, \ \mathit{stride} = 1]$
Max-pool	32×32	$[2 \times 2, \mathit{stride} = 2]$
Conv2	32×32	$[5 \times 5, 4, 8, \mathit{stride} = 1]$
Max-pool	16×16	$[2 \times 2, \mathit{stride} = 2]$
Conv3	16×16	$[5 \times 5, 8, 8, \mathit{stride} = 1]$
Fc1	Number of neurons: 1000	
Fc2	Number of neurons: 124	

For Table.I, Conv1, Conv2, and Conv3 represent the first, second, and third convolutional layers respectively; Max-pool represents the largest pooled layer; and Fc1 and Fc2 represent the first and second fully connected layers, respectively. The fi rst fully connected layer has 1000 neurons, and the second full y connected layer has 124 neurons; [5×5,1,4, stride=1] indicat es that the convolution kernel size is 5×5, and the input channe l is 1, the output channel is 4, and the step size is 1.

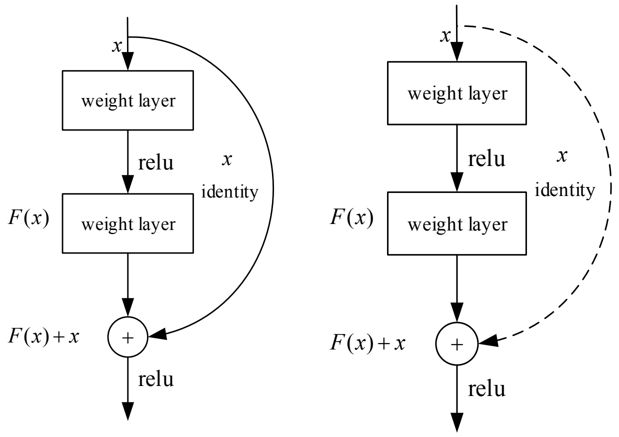
IV. RESEARCH ON GAIT RECOGNITION ALGORITHM BASED ON IMPROVED LENET NETWORK MODEL

In 2015, a deep convolutional neural network model, the residual network, was proposed. The residual network mainly solves the degradation problem and is easier to optimize, which speeds up network convergence. Literature [10] and [11] have applied the residual network to the field of face recognition and achieved good results. This section mainly optimizes the LeNet network model and introduces the ideas proposed in the residual network into the LeNet network, thus accelerating the convergence speed of the model and improving the recognition rate.

A Residual network thinking

In theory, the performance of the network model can be improved by increasing the depth and width of the network. Networks with a large number of layers generally perform better than networks with fewer layers. For example, a network A with a large number of layers and a network B with a small number of layers, then the performance of the network A is at least the same as the performance of the network B. All the parameters of network B are migrated to the front layer of network A. However, the latter layer of network A mainly makes an equivalent mapping, so that the same effect of network B can be achieved. For the VGGNet network model, the number of network layers is increased based on the AlexNet network model to improve network performance.

If the layer behind the deep network is an identity map, the network model degenerates into a shallow network, so the main task is to learn the identity mapping function. There are two ways to learn an identity map using some of the network layers behind the network model: direct fitting H(x)=x and residual function F(x)=H(x)-x. The direct fitting method is more difficult, which may be the reason why the deep network is difficult to train. The H(x)=F(x)+x function is based on the neural network through the shortcut connection (the so-called "quick connection" is skipping one or multi-layer connection) way to achieve. As shown in Figure 2, for the residual unit, if the input x matches the dimension of the output feature map of the second layer convolutional layer, the shortcut connection is a solid line, that is, the Fig. 2(a); otherwise, the dimension does not match, then the shortcut Connection is a dotted line, that is, Fig. 2(b).



(a) Residual unit of "solid line" connection

(b) Residual unit connected by "dotted line"

Fig.2. Residual unit

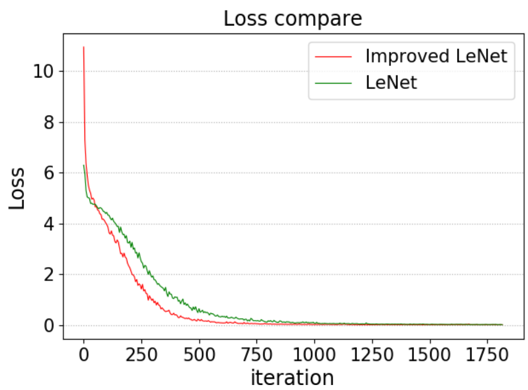
V. EXPERIMENTAL RESULTS AND ANALYSIS

The next experiments were based on a gait dataset for a comparative experiment of the LeNet network model and the improved LeNet network model.

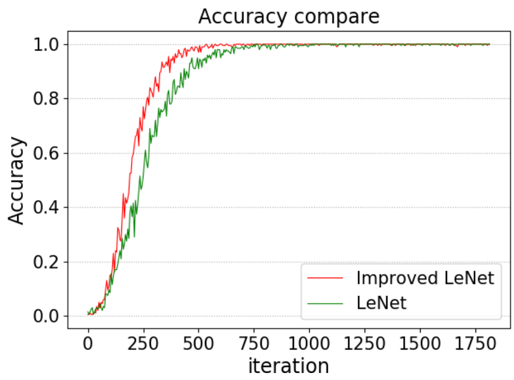
First, experiments were performed on the first set of samples (nm-nm), the second set of samples (bg-bg), and the third set of samples (cl-cl), and then analyzed whether the improved LeNet network model improved the network model's Performance.

A. Experiment under normal walking conditions

The first set of samples (nm-nm) were tested using LeNet and the improved LeNet network model. As shown in Fig.3, a comparison of loss and recognition rates for different network models is used on the first set of samples.



(a) Loss comparison



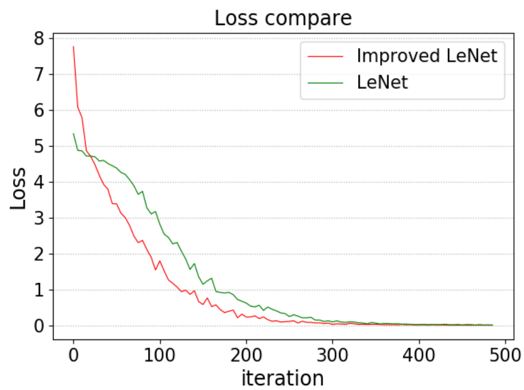
(b) Comparison of recognition rates

Fig.3. Performance comparison of the first set of samples on different network models

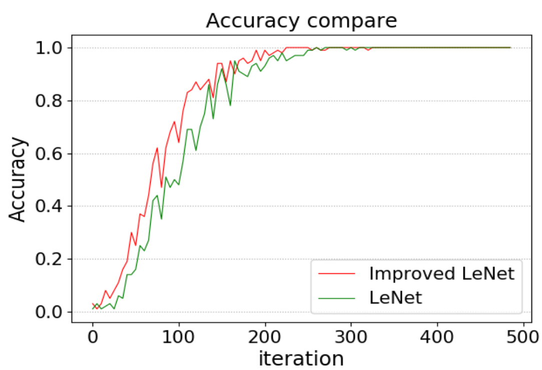
As can be seen from Fig.3, the improved LeNet network model on the first set of samples not only converges faster than the LeNet network model, but also improves the recognition rate of test samples by 0.3%.

B. Experiment with backpack walking

The second set of samples (bg-bg) were tested in the LeNet network model and the improved LeNet network model. As shown in Fig.4, a comparison of loss and recognition rates for different network models was used on the second set of samples.



(a) Loss comparison



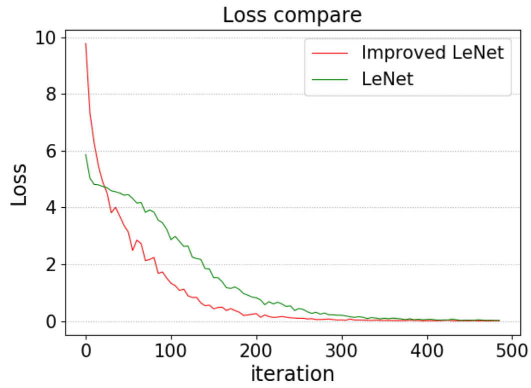
(b) Comparison of recognition rates

Fig.4. Performance comparison of the second set of samples on different network models

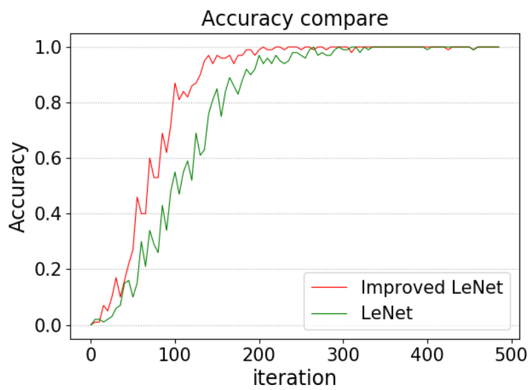
As can be seen from Fig.4, the improved LeNet network model on the second set of samples allows the model to converge faster and the recognition rate of the test samples is increased by 0.5%.

C. Experiment in a coat walking state

The third set of samples (cl-cl) were tested in the LeNet network model and the improved LeNet network model. As shown in Fig.5, a comparison of loss and recognition rates for different network models was used on the third set of samples.



(a) Loss comparison



(b) Comparison of recognition rates

Fig.5. Performance comparison of the third set of samples on different network models

According to Fig 5, the improved LeNet network model is better than the LeNet network model. The improved LeNet makes the network model converge faster and the recognition rate of test samples is also increased by 0.8%.

From the above experimental data, the improved LeNet network model is better than the LeNet network model on the nm-nm, bg-bg and cl-cl samples. In particular, the improved LeNet model makes the network converge faster, and the recognition rates on nm-nm, bg-bg, and cl-cl samples are increased by 0.3%, 0.5%, and 0.8%, respectively. On the samples of normal, backpack and coat walking, the improved LeNet is better than the LeNet network model, regardless of the convergence speed and recognition rate of the model. Therefore, the improved LeNet network model is used as the optimal network model in this paper.

VI. CONCLUSIONS

The gait recognition is susceptible to the influence of viewing angle, speed, backpack, jacket and other covariates, resulting in low recognition rate. This paper proposes a gait recognition algorithm based on improved LeNet network. The optimal network model of this paper was selected by adjusting the model parameters and a large number of contrast experiments. For the improved LeNet network model, a large number of experiments were carried out in the gait database CASIA-B. Experiments show that the improved LeNet network model improves the recognition rate to some extent. The recognition rates of the improved LeNet network model under normal (nm-nm), backpack (bg-bg), and cl-cl walking samples were 98.3%, 89.2%, and 95.8%, respectively. The model not only makes the network converge faster, but also improves the recognition rate and achieves the expected requirements.

REFERENCES

[1] Munif Alotaibi, Ausif Mahmood, Improved gait recognition based on specialized deep convolutional neural network[J], Computer Vision and Image Understanding, 2017, pp. 103-110.

[2] Huimin Wu, Jian Weng, Xin Chen, Wei Lu, Feedback weight convolutional neural network for gait recognition[J]. Journal of Visual Communication and Image Representation, 2018, pp. 424-432

[3] He Zhengyi,Zeng Xianhua,Guo Jiang.A Gait Recognition and Simulation Method Based on Integrated Convolutional Neural Network and Deep Letter Network[J].Journal of Shandong University, Engineering edition, 2018, 48(03): 88-95.D.

[4] Li Ying. Research on gait recognition method based on three-dimensional convolutional neural network[D]. Beijing Jiaotong University, 2018.

[5] He Yizhen, Zhang Junping. Deep Learning of Gait Recognition: A Review[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(05): 442-452.

[6] Yu Tao. Research on human gait recognition algorithm based on deep learning [D]. University of Electronic Science and Technology of China, 2018.

[7] L. Xiao, Q. Yan and S. Deng, Scene classification with improved AlexNet model, 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 2017: 1-6.

[8] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.

[9] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.

[10] CAO Chuan, ZHANG Hongying. Face Recognition Algorithm Based on Improved Residual Network[J]. Sensors and Microsystems, 2018, 37(08): 127-129+133.

[11] Xu Peichao, Chen Wenbai, Chen Xiangfeng, Han Hu. Application of Multitask and Resnet Network in Face Multi-attribute Recognition[J]. SmallComputer Systems, 2018, 39(12): 2720-2724.