Gait Recognition Based on Lightweight CNNs

Wang Hongru[1],Yuan Haoran[2]

[1]Professor,College of Information and Communication Engineering, Harbin Engineering University
(145 Nantong, Nangang District, Harbin, Heilongjiang, 150001,China)
E-mail: whrhit@hrbeu.edu.cn
[2] College of Information and Communication Engineering, Harbin Engineering University
(145 Nantong, Nangang District, Harbin, Heilongjiang, 150001,China)
E-mail: 971682945@hrbeu.edu.cn

This paper proposed a gait recognition algorithm of lightweight convolutional neural network that can reduce the computing resources required for deep convolutional networks and decrease training costs. The algorithm used simple linear operations to obtain the same number of feature maps as the output of the original convolutional layer from a part of the basic feature maps, and embedded them in the original network structure in the form of modules, reducing the amount of parameters of each volume while ensuring the recognition rate. Combined with the comparative test data, the new L-ResNet-50 (Light ResNet-50) model achieved a 50% reduction in training time and training parameters compared with the traditional ResNet-50 model, which verified the algorithm's feasibility.

Key Words *: Gait recognition, CNNs, Deep learning, L-ResNet-50*

1. INTRODUCTION

Gait feature is an essential biological feature of humans. Compared with biometric technology of transmission, such as iris recognition, fingerprint recognition and facial recognition[1)]. Gait feature-based recognition technology has advantages such as far recognition distance, low-resolution requirement and no needing target coordination[2)] . In recent years, gait recognition technology based on video sequence has been repeatedly proposed, because a video image can display the entire walking process.

The most-researched method is the gait energy map (GEI)[错误!未找到引用源。] . GEI is obtained by averaging a complete gait cycle image and combining it into an image. Multiple groups of labeled GEI were used to compose a single data set to extract the gait features of the data set and the gait features in the GEI of the subjects. The most advanced gait recognition method finds the most similar tag in the gallery according to these characteristics to determine its identity. GEI constitutes the training set and test set, as shown in

Figure 1.

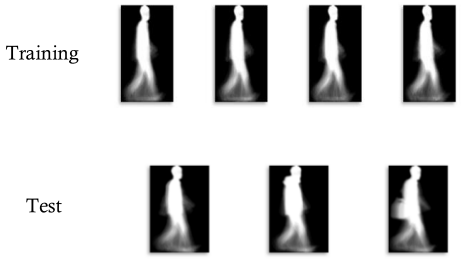

Fig1. GEI samples of training and test set

Convolutional neural networks(CNNs) have excellent performance in the field of image classification. CNNs extract advanced features in the image through multi-layer convolution operation and continuously updates parameters in the kernel through a back-propagation(BP) algorithm to complete the learning of these advanced features[3]. Yan et al.[5]proposed to send an energy map into CNN to extract advanced gait features, and introduced multi-task learning model to achieve better performance than a traditional gait recognition algorithm. Tan[6] et al. proposed to put GEI into the training matching model using a two-channel convolutional neural network, to match gait features to identify human identity, containing a robustness to cross-views gait recognition. Li[7]et al. proposed a recognition method based on deep learning VGG-19 network. After periodic detection, gait sequences were directly sent into VGG network for feature extraction. Finally, a combined Bayes model was used for gait recognition.

The methods proposed by the above researchers have achieved ideal recognition effects on specific data sets, and the comparison of experimental results shows that the number of layers in the deepening network can improve the recognition accuracy under the occlusion condition and the cross-view condition. Deeper networks acquires more advanced features, which is conducive to the classification learning of the model. But a deep network also means a larger parameters scale, making network training complicated. Sometimes it even needs to rely on high-performance computation resources, which is not conducive to the development and popularization of this technology.

In this paper, a lightweight deep convolutional neural network method is proposed, which obtains the characteristic information equivalent to the original network and reduces the number of parameters in network training, so that the model can be easily deployed on embedded devices. Different from the traditional convolution network, the conventional convolution process is divided into two parts: The first part is the same as the normal convolution operation, but only generates a certain number of feature maps; The second part, uses basic feature maps generated in the previous part to obtain the remaining feature graph through a simple linear operation[8]. Finally, adding the feature images of the two parts gets the number of feature images equal to the number of feature images output by the original convolution layer, and simultaneously , realizing this operation in each convolution layer to finally reduce the number of the whole model parameters finally. The experimental results prove that the improved deep network can achieve the same recognition rate as the original network, while the training parameters are reduced by a factor of two, so that the loss value of the network converges faster and reduces the training time, making it easier to deploy on mobile devices

2. RELATED WORK

(1) Input data

In this part, the Gait Energy Image GEI mentioned above is used as the learning objective of CNN. The synthetic GEI image first takes the silhouette of the walking target in the video and extracts the target from the background. Then a series of morphological processing is done to the extracted target to generate the ideal binarization image. Gait period is usually used to represent the gait energy, reflecting the static and dynamic information of human walking. In GEI, the occurrence frequency of each pixel is calculated and the corresponding brightness is set to reflect the gait information of walking.

$$GEI(x,y) = \frac{1}{N}\sum_{t=1}^{N} I_t(x,y) \quad (1a)$$

Where, N represents all the images within a less frequent period, I (x, y) is a gait image, and t represents the number of frames.

(2) Lightweight model

The lightweight design of the model mainly focuses on four aspects: reducing the amount of calculation, the training parameters, the actual running time and simplifying the underlying implementation. Researchers have successively proposed depthwise separable convolution, group convolution, new activation function and other methods[9]. The actual running time of some existing structures is analyzed, and some architecture design principles are put forward, according to which the architecture is re-designed.

Model pruning connections is to compress the model by cutting off unnecessary links between neurons[10] to reduce parameters. The Model quantization processes the representative weights and activation functions [11], which is conducive to the compression of the model and the improvement of the training speed. Knowledge distillation uses the large model to teach the smaller ones[12], improving the performance of the model. These methods need to be based on excellent pre-training models and do not change the basic framework of CNN.

In 2017, Mobilenet_v1[13] proposed effective network architecture and divided the convolutions into two parts using depthwise convolutions to build a lighting, low-latency model that can easily fit mobile design requirements or embedded visual applications.

3. MODEL CONSTRUCTION

(1) Generate more feature maps through linear operations

Although a series of lightweight network architectures such as MobileNet_V1 compress the improved training efficiency of the model, there isn't a clear reduction in training parameters. Through the visualization operation of CNN's feature maps, researchers found a large number of similar feature images[14] in the feature map generated by a convolutional layer. If there is a mathematical correlation between these similar feature images, simple operations can transform them from some basic feature images. Figure 2.is the output of a gait energy graph after the convolution operation. Pair-to-pair-to-similar situations in this group of feature maps. Compared with the complex convolution process, these simple operations do not need to take up unnecessary computing resources. In the back propagation algorithm, only the convolution kernel parameters that generate the basic feature map need to be updated, which reduces the scale of training parameters.

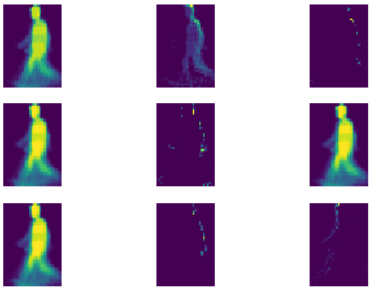

Fig.2 Output of partial feature graph of ResNet-50

According to the above theory, the new lightweight convolution operation is designed, and the convolution process is also divided into two parts: The first part is the traditional convolution operation, but only generates half of the original basic feature map; The second part is to use these basic feature map to get the rest of the feature map through linear transformation. The lightweight module divides a traditional convolution process into these two parts, as shown in Figure 3.

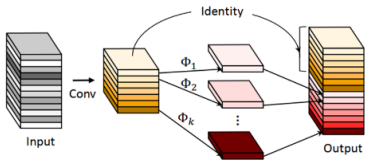

Fig.3 Specific operation of the lightweight module

Where ϕ stands for linear operation, there are

many options. DepthWise is the most effective for the re-convolution of the basic feature map after testing.

(2) Lightweight RESNET-50 network

ResNet is a deep network model proposed in 2015 [15], which builds a deep network by stacking residual blocks to solve the problem that the optimization effect becomes worse as the network deepens. The residual block can be understood as a seed network. The lightweight module mentioned above is used to replace the residual block in the original ResNet-50, to reduce the training parameters of the original network and reduce the calculation cost to achieve the lightweight of the ResNet-50 model. For example, in the original network, the first layer of convolution uses a 5×5 convolution kernel, which needs to generate 16 feature maps. Now, only 8 basic feature graphs are required to be generated, and the remaining 8 feature maps are generated by linear operation of the basic feature maps. Theoretically, the training parameters needed for this layer are reduced to half of the original ones.

Figure 4. shows a composition diagram of the 'lightweight residual module'.

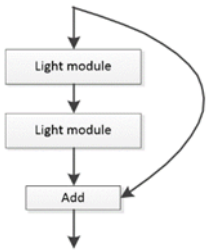

Fig.4 Lightweight residual module

4. EXPERIMENTS

In this section, the lightweight modified L-RESNET-50 network will be used to complete the gait recognition task, and the specific experimental data in the process of model training will be collected and compared with the original network in feature map similarity, recognition rate under various conditions, number of training parameters and training time.

(1) Data preparation

The experiment is based on the CASIA-B dataset of the Chinese Academy of Sciences, which consists of 124 gait targets and 10 walking sequences under each target: Six groups were normal (NM), two groups were wearing overcoats (CL) and two groups were wearing backpacks (BG).Each case includes 11 different gait sequences (0°, 18°, 36°,..., 180 °). The GEI synthesized by plus or minus 0°-18° is too similar to be of research significance among the eleven angles. Therefore, four GEIs synthesized by normal walking gait sequences were selected as the training set from each perspective after excluding the four angles. The remaining GEIs constituted the test set according to different viewing angles and shielding conditions.

(2) Experiments and Results

a) Similarity Comparison of Feature Maps

After replacing the original residual module with the lightweight module, some feature maps comparison experiments were first carried out to calculate the MSE (Mean Squared Error) value of the features obtained by the linear operation and the original feature maps on the pixel value, to judge whether the original feature image can be replaced. MSE data of the first three layers are selected as shown in Table 1.

Table 1 MSE values of the top three convolution layers

	Conv1	Conv2	Conv3
MSE (10^{-3})	3.3	11.0	23.9

As shown in Table 1, all MSE values are very small, indicating a strong correlation between feature mappings under the two operations. These original feature mappings can be obtained by a linear transformation of basic feature mappings.

b) Comparison with the Original Network

The same data set was used to carry out image classification comparison experiments with the ResNet-50 and the lightweight network. In the experiment, it was found that the recognition rates of

the new lightweight network and the original network were the same under the condition of shading and from various angles, as shown in Table 2 and Table 3.

Table 2 Gait recognition rate at different views, test the ability of cross-view recognition of the model

view	ResNet-50 (acc)	L-Resnet-50 (acc)
36	98%	98.1%
54	96.5%	97%
72	94.2%	93.1%
90	92.2%	91.5%
108	93.7%	93%
126	95.5%	95.3%
144	97%	96.7%

Table 3 Gait recognition rate under different walking conditions

Conditions	ResNet-50 (acc)	L-Resnet-50 (acc)
BG	83.2%	82.3%
CL	77.3%	76%

However, the training parameters are reduced by 50% compared with the original network, and the floating-point computation is also reduced by nearly half, as shown in Table 4.

Table 4. Network parameters and FLOPS

Model	Weights (M)	Flops (B)
ResNet-50	25.6	4.1
L-ResNet-50	13.0	2.2

In the case of the same hyperparameters, the training time required by the model is shorter, while in the convergence speed of the loss curve, the new lightweight network is faster than the traditional ResNet-50 network. Fig. 5 shows the loss function curve of the lightweight network trained and tested at a 90° perspective. Convergence can be achieved after only 100 batch processes.

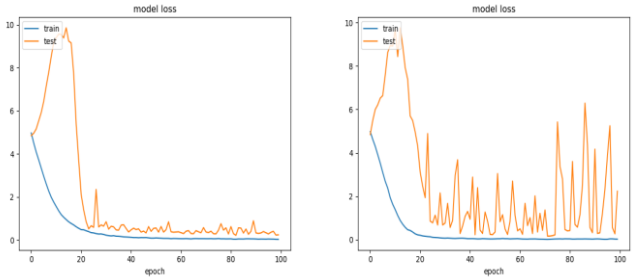

Fig.5. Loss function under 36°

5. CONCLUSION

Designing a lightweight module modified a deep convolutional neural network (RESNET-50) is lightweight modified reducing the operation parameters of a large network by nearly 50% and accelerating the training speed of the network. The convolutional neural network can effectively learn more advanced features from GEI, and guarantee the accuracy of gait recognition under different perspectives and shielding conditions, laying a foundation for the future production of gait recognition technology based on deep learning. With further study, more effective recognition methods will predictably result in more significant practicability across numerous situations.

REFERENCES

1) Phillips P J. Human identiflcation technical challenges. New York, USA: In: Proceedings of the 2002 International Conference on Image Processing.,2002.
2) Boulgouris, N. V., Hatzinakos, D., & Plataniotis, K. N. (2005). Gait recognition: a challenging signal processing technology for biometric identification. IEEE signal processing magazine, 22(6), 78-90.
3) Liao, R., Yu, S., An, W., & Huang, Y. (2020). A model-based gait recognition method with body pose and human prior knowledge. Pattern Recognition, 98, 107069.
4) Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. in NIPS, 2012.
5) Yan C, Zhang B L, Coenen F. Multi-attributes gait identiflcation by convolutional neural networks. In: Proceedings of the 8th International Congress on Image and Signal Processing. Shenyang, China: IEEE, 2015. 642¡647
6) Tan T N, Wang L, Huang Y Z, Wu Z F. A Gait Recognition Method Based on Depth Learning, CN Patent 201410587758, June 2017
7) Li C, Min X, Sun S Q, Lin W Q, Tang Z C. DeepGait: a learning deep convolutional representation for viewinvariant gait recognition using joint Bayesian. Applied Sciences, 2017, 7(3): 210
8) Kai Han,Yunhe Wang.GhostNet: More Features from Cheap Operations[R].New York, USA:CVPR2020,13 Mar 2020.
9) Andrew G. Howard, Menglong Zhu.MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J].cs.CV,2017,1704(04861v1): 17 Apr 2017
10) Song Han, Huizi Mao, and William J Dally. Deep compression: Compressing deep neural networks with pruning,trained quantization and huffman coding. In ICLR, 2016.
11) Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, and Ali Farhadi. Xnor-net: Imagenet classification using binary convolutional neural networks. In ECCV, pages 525–542.Springer, 2016.
12) Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.
13) Andrew G Howard, Menglong Zhu, Bo Chen, DmitryKalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.
14) Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In CVPR, pages 4510–4520,2018.
15) Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016.