Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 13, No. 2, 2024

Research on Gait Recognition Based on Deep Transfer Features

Xuedong Yu*, Linxing Peng, Yangde Ji, Ziheng Jiang

College of Intelligent Manufacturing and Elevator Technology, Huzhou Vocational and Technical College, Huzhou, 313099, China
*Corresponding author: 2022024@huvtc.edu.cn

Abstract: Gait recognition is to determine the identity information of pedestrians through the difference of their walking postures, which has received more and more attention from researchers in recent years. The current existing gait recognition methods have the situation of low recognition rate and difficult to be applied on the ground. This paper proposes a gait recognition method based on DenseNet201 deep network transfer features, aiming to improve the gait recognition rate and accelerate the rapid application of gait recognition. In this paper, a human body region segmentation method is first designed to divide the arm region and leg region from the whole pedestrian gait image. Then the pre-trained deep network model DenseNet201 after parameter fine-tuning is used to extract the depth transfer features of the whole human body region, the arm region, and the leg region in the three segmented regions, and do the sum-averaging fusion process for the depth transfer features of each segmented region. Finally, a discriminant analysis classifier is used to classify and recognize the fused depth migration features. After experiments on the CASIA-B gait database collected by the Institute of Automation of the Chinese Academy of Sciences, it is proved that the division of the arm region and the leg region has obvious effect on the improvement of the gait recognition rate, and the features extracted from the pedestrian gait images by using the deep transfer network have a good characterization performance.

Keywords: DenseNet201, Body Region Segmentation, Deep Network Transfer Features, Fusion.

1. Introduction

With the continuous forward development of economy and society, public security is getting more and more attention, how to use modern information technology and intelligent way to guarantee social security is a problem we need to solve urgently. Gait recognition [1], is one of the most promising biometric identification technologies in the field of individual identification in long-distance situations. Due to the important scientific and practical significance of gait recognition technology, it has attracted many scholars to invest in it [2-4].

Deep convolutional neural networks are widely used in the field of image classification and recognition due to their superb analytical expressiveness and feature learning ability [5]. Unprocessed images can be directly input into deep convolutional neural networks, which can efficiently reduce the pre-processing procedures for images. In recent years, more and more researchers at home and abroad have used deep convolutional neural networks to solve the practical problems faced by gait recognition [6-8].

Zhang et al [9] used CNN to extract spatial features in the gait of trained pedestrians and LSTM network to extract temporal and spatial features in the pedestrian gait sequences, optimizing the structure and parameters of the LSTM network in the gait recognition model. Gul et al [10] used gait energy maps as model inputs and captured spatial-temporal features of the gait sequences by training a three-dimensional convolutional deep neural network (3D CNN) and thus recognize the pedestrian identity. Mehmood et al [11] used a pre-trained VGG16 network model to extract features, remove redundant features and thus classify and recognize them. Das et al [12] proposed a two-stage pipeline consisting of an occlusion detection and reconstruction framework, where occlusion detection is performed by employing the VGG-16 model, followed by the use of LSTM-based network called RGait-Net to reconstruct the occlusion frames in the sequence, thus solving the occlusion problem in gait recognition. Guo et al [13] proposed a 3D skeleton based human gait recognition framework, where a self-encoder is constructed using a graph convolution based encoder and a physically based decoder, and the extracted features are fed into an RNN for classification and recognition. Pan et al [14] proposed to propose a Lower-Upper Generative Adversarial Network (LUGAN) to generate multi-view pose sequences for each single view sample to reduce cross-view variance and solve the cross-view gait recognition problem. Kumar et al [15] proposed to accomplish occlusion detection using a VGG16 network, conditional variational auto-encoder for feature encoding, and bi-directional Long Short Term Memory (Bi-LSTM) for predicting occluded frames. Finally the image reconstruction is done by decoder.

Encouraged by the above research results, this paper proposes a gait recognition method based on the transfer features of the DenseNet201 deep network as follows:

(1) Considering the variable regions during the pedestrian's walking process, the arm and leg regions in the pedestrian image are segmented to highlight the differentiating information of the pedestrian's identity.

(2) Using the DenseNet201 deep transfer learning network pre-trained on ImageNet as a feature extractor can efficiently and quickly complete the extraction of identity features.

(3) The global identity information and local identity information are fused to fully reflect the individual differences in pedestrian identities, helping to further improve the gait recognition rate.

2. The Proposed Algorithm

2.1. The Segmentation of gait image regions

The human body engages in a continuous and cyclical movement during walking. This is why individuals can recognize one another from a distance based on their walking postures; different people exhibit distinct walking styles. Throughout the walking process, various parts of the body work in harmony to facilitate forward movement, with varying degrees of motion among each body part. According to an analysis of human physiology, the following observations can be made: (1) During walking, the leg region exhibits the greatest amplitude of movement, while the head shows minimal motion, characterized by only a slight bobbing. (2) The amplitude of swinging motions also varies; some individuals have a pronounced forearm swing, while others exhibit a more significant backward arm swing. Additionally, the range of motion in the front and back leg movements differs among individuals. In light of these observations, it is useful to categorize the human body into left and right regions, as well as upper and lower regions, as illustrated in the left half of Figure 1.

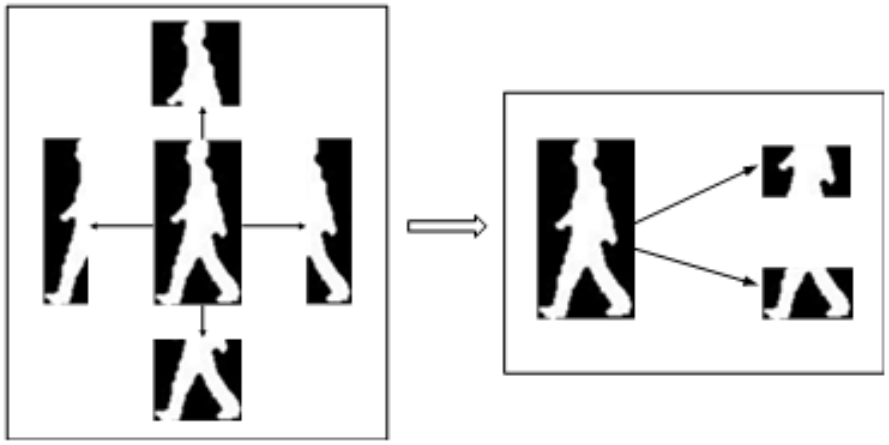


Figure 1. Human body segmentation

As the experiments continue, it is found that the human walking image differences are mainly reflected in the arm region and the leg region, the first human body region division failed to focus on the consideration of the change of the region of great variability, so the final experiments used the second human body region division, the human body silhouette contour map for the arm region and the leg region to do the slice, as shown in the right half of Figure 1.

As shown in Figure 2 below, in order to facilitate the smooth running of the cut-off experiment, a uniform cut-off criterion was delineated to take the height of the human side profile image as the benchmark, and the part of the cut-off area to as the arm swing part and the part of the cut-off area to as the leg swing part, as shown in Equation (1).

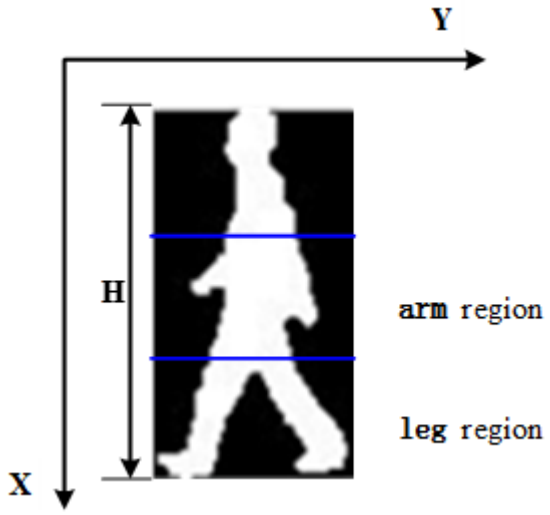


Figure 2. The Segmentation of gait image regions

$$W_{arm} = S\left(\frac{1}{3}H:\frac{2}{3}H\right) \qquad W_{leg} = S\left(\frac{2}{3}H:H\right) \quad (1)$$

Where S represents the entire body contour region, W_{arm} represents the arm region and W_{leg} represents the leg region.

2.2. The DenseNet201 Deep Transfer Learning Network

The more complex the deep convolutional neural network structure, the better the classification and recognition effect will be, but the deepening of the network layers may bring the gradient disappearance, network degradation and other problems. This problem can be well solved by using DenseNet201 transfer learning network, in which each layer is connected to all previous layers in the dense connection module of the network structure to realize feature reuse. The structure of human gait is more complex, and the gait image during the movement process contains rich feature information, using DenseNet201 to combine shallow features and deep features can extract gait features with stronger characterization ability.

As shown in Figure 3, the network of DenseNet201 is mainly composed of four parts, namely: Dense Block, Transition layer, Global Average Pooling and Softmax. The network has two main features: one is that it establishes a dense connection between all the previous layers and the later layers, and the features can be reused; the other is that each layer of the network structure is narrower, which reduces the learning redundancy. This network structure also has several outstanding advantages: it mitigates gradient vanishing, maximizes feature reuse, makes the network parameters better fitted, and greatly reduces the number of parameters.

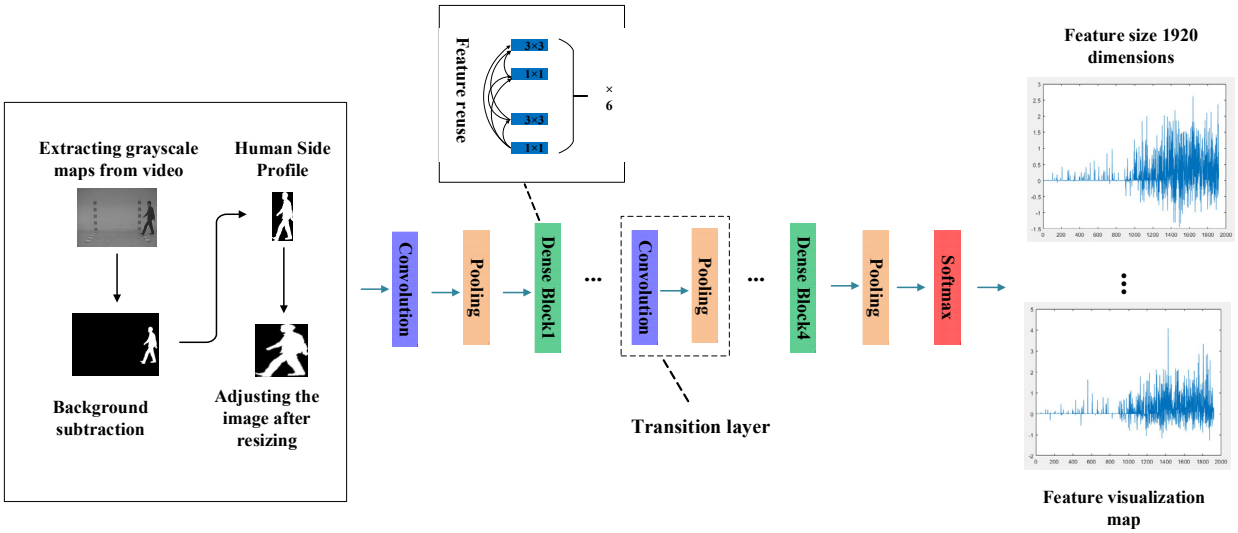


Figure 3. Detailed structure of the DenseNet201 transfer network

(1) Dense Block

In a traditional convolutional neural network, if you have L layers, then there will be L connections; while in the dense connectivity layer there will be L(L+1)/2 connections in each layer, because the input of each layer is the output of all the previous layers, and the network will connect the output of all the previous layers, to achieve the reuse of features, to improve the efficiency of the whole network. The structure of the Dense Block, the dense connectivity layer, is that the activation function is in the front and the convolutional layer is in the back. The output of the first layer in the dense connectivity module is:

$$x_l = H_l([x_0, x_1 \ldots, x_{l-1}]) \tag{2}$$

Where $H(\cdot)$ represents a combination of a series of complex functions, which include operations such as activation function, convolution, and so on. In this paper the activation function Relu is used in the densely connected layer to increase the nonlinearity of the network, in the densely connected first layer the output after the activation function is given in equation (3). The Relu activation function is given in equation (4).

$$x_l = f(x_l) \tag{3}$$

$$f(x) = max(0, x) \tag{4}$$

Then after the activation function, the convolution operation is started on the human gait image, in terms of the convolution layer settings, in order to extract more discriminative features on the gait image, 1×1 and 3×3 convolution kernels are used alternatively and the step size is set to 2. The number of convolution layers set in different Dense Blocks varies, and the parameter settings for the dense block layers are detailed in Table 1.

Table 1. Parameter settings of dense block layers

Name	Number of kernel	Size of kernel	Step
Dense Block 1	6	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}$	2
Dense Block 2	12	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}$	2
Dense Block 3	48	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}$	2
Dense Block 4	32	$\begin{bmatrix}1\times1\\3\times3\end{bmatrix}$	2

(2) Transition layer

The densely connected approach of DenseNet201 causes the gait feature maps to expand, and in order to make the feature maps of each layer of the same size, the structure of the transition layer Transition is used in the network to connect between two neighboring Dense Blocks, which reduces the size of the feature maps. The Transition layer consists of a 1 × 1 convolutional layer and a 2 × 2 average pooling layer. The parameters of each transition layer are set as shown in Table 2 below.

Table 2. Parameter settings of the transition layer

Name	Number of kernel	Size of kernel	Step
Convolution	32	1×1	2
Average Pooling	\	2×2	2

(3) Global Average Pooling

Global Average Pooling is placed after the densely connected layer and the transition layer, which reduces the feature dimensions and regularizes the entire network

structurally, and it prevents the layer from overfitting, integrates the global feature information, and results in improved performance of the network model.

(4) Softmax

The main role of the Softmax layer is to map all the outputs of multiple neurons into the range (0,1) and the sum of these values is 1. The label value corresponding to the maximum value of the output is taken to be the desired multiclassification recognition result and the output of the Softmax layer is given below in equation (5).

$$\sigma(z) = \frac{e_K^z}{\sum_{k=1}^{K} e_k^z} \tag{5}$$

Where z is the output value of the average pooling layer and $k = 1,2,3\cdots,K$ is the number of classifications, which corresponds to 124 classes in the gait database in this paper.

2.3. Gait Recognition Based on DenseNet201 Deep Network Transfer Features

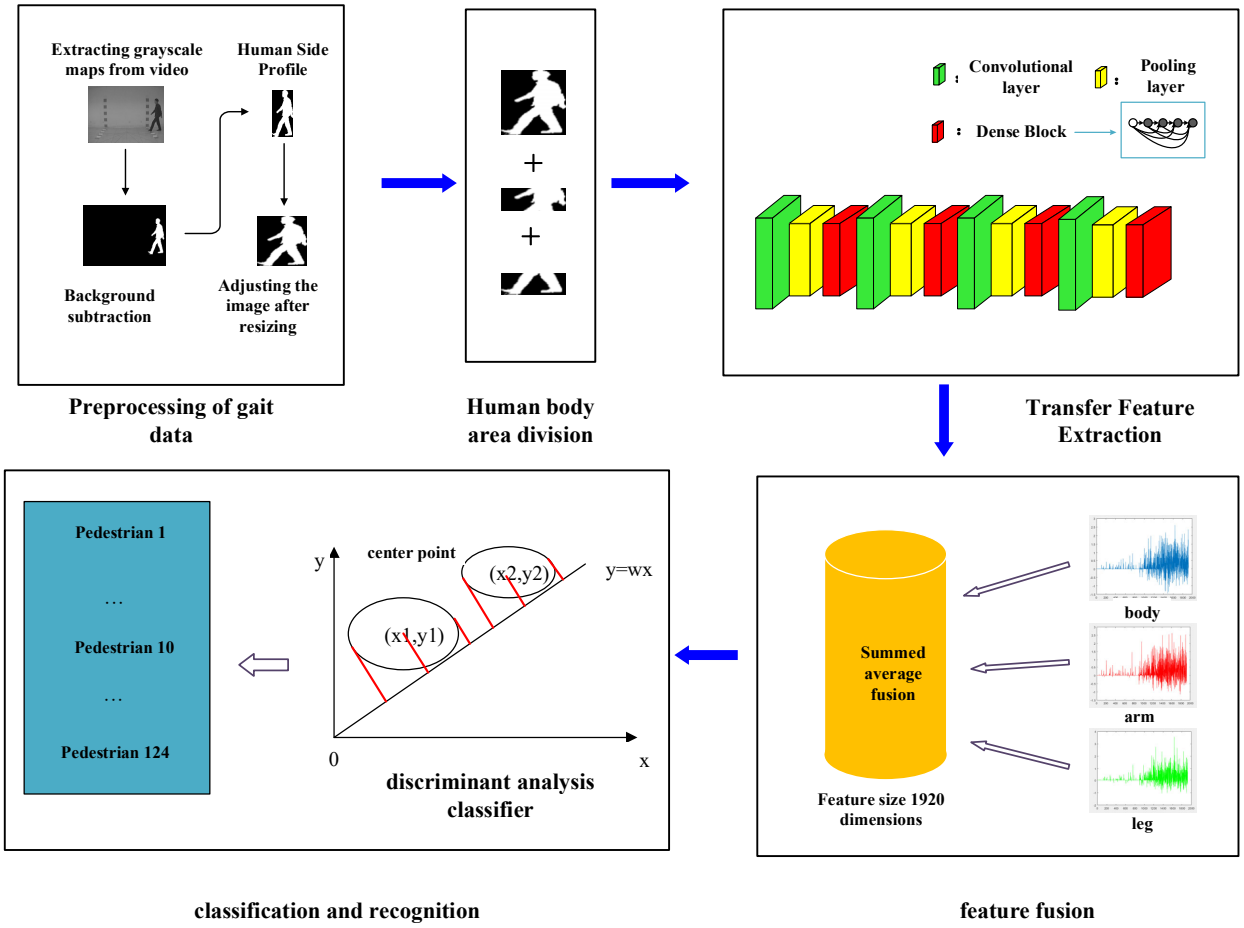


Figure 4. Gait recognition method based on DenseNet201 transfer features

The whole human body area contains all the features of pedestrian identity, and the transfer features extracted for the whole human body area are more comprehensive, which have covered all the features and can fully demonstrate the pedestrian identity. However, due to the large area extracted, it instead contains many redundant features in it, which brings inconvenience to the later classification and recognition. The arm swing area and leg swing area are periodic movements, and their feature changes are more obvious, therefore, in this paper, the depth transfer features are extracted for both the arm swing area and the leg swing area. The feature variation of the arm part and leg part is strong, and the extracted pedestrian transfer features of the whole body are very comprehensive, so this paper takes the transfer features of the whole human body, the transfer features of the arm swing part, and the transfer features of the leg swing part to be summed and averaged fusion, and the three complement each other to improve the gait recognition rate.

Pedestrians in the process of walking, itself is the arms and legs cooperate with each other to show the salient features, so in our classification and identification process should be combined together, so that the features obtained by the classifier is more robust. The features extracted from the whole body region of the pedestrian are more comprehensive, and the fusion of the features extracted from the whole body region, the arm region and the leg region can completely show the identity features of the pedestrian and make the pedestrian identity information more discriminative. As shown in Figure 4, the whole body part, arm swing part, leg change part of the three regions were input to the DenseNet201 network to extract features, after fine-tuning the transfer learning network to extract the 1920-dimensional features of the avg_pool layer, at this time to obtain the three 1920-dimensional features, which are summed and averaged fusion of the three features, the fused features are input to the discriminant analysis classifier for gait recognition.

3. Experimental Setup and Analysis of Results

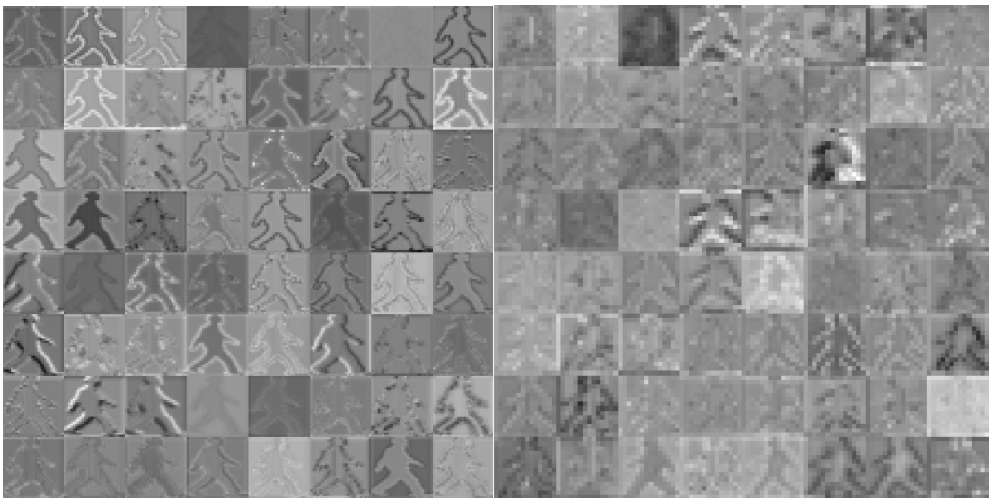
3.1. Experimental setup

In this paper, the CASIA-B gait database collected by the Institute of Automation of the Chinese Academy of Sciences was used for the experimental study of gait. This database is a data collection of 124 volunteers, while using 11 cameras to collect data from 11 angles for each volunteer, the shooting viewpoints are 0° to 180°, and one viewpoint is taken every 18° (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180°), there are normal walking, backpack, and coat-wearing conditions. In the normal walking condition, there are six repetitions of nm01-nm06 to collect data; in the backpack condition, there are two repetitions of bg01-bg02 to collect data; and in the coat wearing condition, there are two repetitions of cl01-cl02 to collect data. All data need to go through gait data preprocessing operation, deep transfer feature extraction, and classification and recognition steps. Experiments are conducted on the gait data in three conditions in turn, with nm01-nm04 as the training set and nm05-nm06 as the test set for the normal walking case; bg01 as the training set and bg02 as the test set for the backpack walking case; and cl01 as the training set and cl02 as the test set for the overcoat walking case. The specifics of the experimental setup are shown in Table 3:

Table 3. Experimental setup for the CASIA-B dataset

parameters	training set	test set
dataset	nm01-nm04	nm05-nm06
	bg01	bg02
	cl01	cl02

3.2. Feature visualization and analysis



(a) Shallow visualization (b) Deep visualization

Figure 5. Visualization of gait features extracted by DenseNet201

In order to more intuitively show the effect of the transfer learning network on the extraction of pedestrian gait images, the effect of the DenseNet201 transfer learning network on the extraction process of gait images is demonstrated. For deep feature extraction, the network parameters before the conv5_block3_concat layer are first frozen, and then the pedestrian gait image is input into the network to fine-tune the network parameters after the conv5_block3_concat layer, and the fine-tuned transfer network is used to extract deep features from the pedestrian gait data. The features extracted using the deep network are deeper layer by layer, in order to be able to clearly see the changes of image features in the transfer learning network, this paper will DenseNet201 network extracted features to do shallow visualization and deep visualization processing, as shown in Figure 5, it can clearly see that the shallow features extracted by the transfer network are mainly extracted from the pedestrians' body shape contour and other basic features, which is in favor of image details; while with the deepening of the layers, the deep network focuses more on abstract features and the proposed features are more discriminative.

3.3. Analysis of experimental results

According to the gait recognition experiments mentioned in 2.3, we can know that in the same condition, this chapter needs to do four sets of experiments: gait recognition based on the transfer features of the whole human body region, gait recognition based on the transfer features of the arm region, gait recognition based on the transfer features of the leg region, and gait recognition based on the fusion of the transfer features of multi-region, and gait experiments are conducted in three conditions of normal walking, backpacking, and wearing a coat, which are the three conditions of normal walking, backpacking, and wearing a coat in the CASIA-B gait database collected by the Institute of Automation of the Chinese Academy of Sciences. The gait experiments are conducted for three conditions of normal walking, backpacking and wearing a coat in the database.

Table 4. Gait recognition rate (%) in normal walking condition

Feature	Viewpoint										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Body	94.8	92.1	91.6	86.1	87.5	87.5	87.1	87.7	89.9	93.3	95.3
Arms	85.4	73.6	66.3	58.1	57.5	56.3	55.7	57.7	62.7	72.5	80.5
Legs	62.6	54.8	50.5	48.9	57.0	63.0	58.6	53.7	47.6	47.5	49.0
Fusion	96.2	93.2	92.7	93.5	94.9	95.2	93.1	91.0	92.6	94.8	96.0

As shown in Table 4, under the normal walking condition, we did gait recognition experiments on the whole human body, arm swing region, leg swing region, and three-region summation and average fusion. From 0° to 180° of the 11 viewpoints can be seen, the whole body region for migrating feature extraction and classification of the recognition of the best results, which is also reflected from the side of the whole body region of the characteristics of the full-body region more fully reflect the identity of the traveler, the arm swing region, although the identification of the identity of the traveler alone to identify the recognition of the lower rate, but through the whole human body region, the arm swing region, the leg swing region of the three regions of the fusion of the characteristics of the body region, the arm swing region, the leg swing region, to be able to Although the recognition rate is low, through the fusion of the whole human body region, arm swing region and leg swing region features, it can reflect the comprehensive pedestrian feature information and the regional pedestrian feature information with great variability, which complement each other and reflect the identity of the traveler in a complete and full way.

Table 5. Gait recognition rate (%) in backpacking condition

Feature	Viewpoint										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Body	92.6	88.9	84.2	78.1	83.0	85.7	84.9	84.0	87.4	90.2	92.1
Arms	80.3	68.7	59.4	52.1	55.1	61.4	60.0	59.0	62.3	69.9	77.5
Legs	50.7	44.0	33.5	34.0	41.8	47.9	44.5	39.4	35.5	36.0	40.4
Fusion	92.9	88.8	85.3	82.8	87.3	90.7	89.3	87.2	89.3	91.3	93.1

As shown in Table 5, in the backpack condition, overall, the gait recognition rate decreases due to the influence of the backpack, but the decrease is small. Similar to the normal walking condition, with the assistance of the arm swing region and the leg swing region, the recognition rate of the features fused by the three regions is relatively high; it can also be seen that the arm swing region and the arm swing region do not have as high a recognition rate as the features proposed by the whole-body region due to the one-sidedness of the extracted features.

Table 6. Gait recognition rate (%) in coat-wearing condition

Feature	Viewpoint										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Body	80.0	75.6	71.7	67.7	69.9	72.4	71.4	72.0	70.3	74.9	79.1
Arms	77.5	61.6	48.1	42.2	37.4	42.8	41.7	40.8	46.2	59.7	73.4
Legs	45.6	37.6	30.0	32.0	40.0	47.8	41.0	35.8	31.1	30.7	35.3
Fusion	82.2	73.9	68.6	68.7	72.1	78.7	76.3	74.6	71.5	75.5	81.6

As shown in Table 6, in the condition of coat-wearing, the presence of the coat makes the recognition rate decrease float more, at the same time, there is a situation that the recognition rate of the fusion features of the two viewpoints of 18° and 36° is lower than that of the whole human body in the whole body, and the reason for this is because in the case when the recognition rate of the arm swinging region and the leg swinging region is low, it may play a negative role in the recognition of the whole human body, but from the data it can be It can be seen that even though the recognition rate of the fused features has decreased compared to the recognition rate of the whole human body, the decrease is small. Moreover, according to the previous multi-group experiments, it can be seen that the average addition and fusion help to improve the recognition rate, so the case of the two viewpoints at 18° and 36° does not affect the overall conduct of the experiment.

4. Conclusions

In this paper, a gait recognition method based on deep network transfer features is proposed. Firstly, a human body region division method is proposed, which focuses on the dynamic region of pedestrian gait, the arm region and the leg region; secondly, the DenseNet201 transfer learning network, which has been pre-trained by the ImageNet dataset, is introduced to extract more discriminative human identity characterization features; and lastly, the CASIA-B dataset collected by the Institute of Automation of the Chinese Academy of Sciences (IAS) is used for the Validation, the experimental results show that the algorithm proposed in this study has a wide range of applications in human gait recognition. In the next step, we will develop a real-time human gait recognition system and apply the algorithm in practical applications. Of course, there are still some constraints in human gait recognition, such as mutual occlusion of human bodies and perspective transformation. These constraints should be solved in the subsequent research.

Acknowledgements

The authors gratefully acknowledge the financial support from Huzhou Vocational and Technical College School-level Planning Subject(2024YB13) fund.

References

[1] Connor P, Ross A. Biometric recognition by gait: a survey of modalities and features[J]. Computer Vision and Image Understanding, 2018, 167(FEB.): 1-27.

[2] Hu M, Wang Y, Zhang Z, et al. Incremental learning for video-based gait recognition with LBP flow[J]. IEEE transactions on cybernetics, 2012, 43(1): 77-89.

[3] Zeng W, Wang C, Yang F. Silhouette-based gait recognition via deterministic learning[J]. Pattern recognition, 2014, 47(11): 3568-3584.

[4] Chao H, He Y, Zhang J, et al. Gaitset: regarding gait as a set for cross-view gait recognition[C]. Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 8126-8133.

[5] Aptoula E, Ozdemir M C, Yanikoglu B. Deep learning with attribute profiles for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1970-1974.

[6] Li C, Min X, Sun S, et al. DeepGait: a learning deep convolutional representation for view-invariant gait recognition using joint Bayesian[J]. Applied Sciences, 2017, 210(7): 1-15.

[7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409.1556, 2014.

[8] Yu S, Chen H, Wang Q, et al. Invariant feature extraction for gait recognition using only one uniform model[J]. Neurocomputing, 2017, 239: 81-93.

[9] Yujie Z, Lecai C, Wu Z, et al. Research on gait recognition algorithm based on deep learning[C]//2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI). IEEE, 2021: 405-409.

[10] Gul S, Malik M I, Khan G M, et al. Multi-view gait recognition system using spatio-temporal features and deep learning[J]. Expert Systems with Applications, 2021, 179: 115057.

[11] Mehmood A, Tariq U, Jeong C W, et al. Human gait recognition: A deep learning and best feature selection framework[J]. Comput. Mater. Cont, 2022, 70: 343-360.

[12] Das D, Agarwal A, Chattopadhyay P. Gait recognition from occluded sequences in surveillance sites[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 703-719.

[13] Guo H, Ji Q. Physics-augmented autoencoder for 3d skeleton-based gait recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 19627-19638.

[14] Pan H, Chen Y, Xu T, et al. Toward complete-view and high-level pose-based gait recognition[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 2104-2118.

[15] Kumar S S, Singh B, Chattopadhyay P, et al. BGaitR-Net: An effective neural model for occlusion reconstruction in gait sequences by exploiting the key pose information[J]. Expert Systems with Applications, 2024, 246: 123181.