Hindawi
Mobile Information Systems
Volume 2022, Article ID 5292134, 10 pages
https://doi.org/10.1155/2022/5292134

Research Article

Application of Multi-Feature Fusion Based on Deep Learning in Pedestrian Re-Recognition Method

Ke Han, Ning Zhang, Haoyang Xie, and Qianlong Wang

School of Software, North China University of Water Resources and Electric Power, Zhengzhou 450046, China

Correspondence should be addressed to Ke Han; hanke@ncwu.edu.cn

Received 11 July 2022; Revised 24 August 2022; Accepted 5 September 2022; Published 16 September 2022

Academic Editor: Le Sun

A system known as pedestrian recognition makes use of several cameras to identify the surrounding area and quickly identify and match the target demographic. Based on pedestrian recognition, the picture model, pedestrian features, and other information, the features are developed to have a high degree of generalizability, distinctiveness, and accuracy. The application approach for pedestrian re-recognition based on deep learning for numerous features is proposed in this paper. The suggested approach successfully preserves high-level semantic information, which helps network members extract all of the pedestrian properties. As external material and semantic information were combined horizontally and vertically, environmental interference was decreased, and people's ability to create networks was enhanced. The voice channel of the speech system was introduced in order to fully utilize the global information network, and the connection between the channels was carefully addressed in order to enhance the global information network's capacity for expression. The null convolution reduced the operational continuity of the identification information. To increase the consistency of the data, the multi-level spatial convolution structure was merged with the entire image in this paper. After numerous experiments, the three groups were 89.5%, 89.5%, and 89.1%, respectively, compared to 1501, DukeMTMC-reID, CUHK03, and other medial groups, and the experimental results were 85% and 89.5%, respectively. The multimode feedback MP3 module was taken from the MP3 module in order to gain richer and denser multimode feature information. Comparing the module's initial response level (RANK1) with the various cycles yields the average accuracy for each cycle (catalog). The experiment demonstrates that the two mixed pile groups can enhance the modulus of the mixed pile group and get better results. The multi-level multi-scale pole function effectively combines the characteristics of pedestrians in various scales, and the addition of the ASP module enhanced the network context information's overall ability to be represented, aided in this chapter's research method's ability to more thoroughly analyze scene structure, and increased the precision of pedestrian re-recognition.

1. Introduction

With the rapid development of deep learning, deep learning has won great competitive conditions in the fields of pattern recognition and pedestrian re-recognition. This historic expedition provides a compact summary of the work, most of which dates back more than a thousand years. The difference between shallow learners and deep learners lies in the depth of their credit knowledge trajectory, which is a causal chain that can be learned between behavior and effect [1]. Deep learning enables multi-layer processing to express relevant research models through multi-layer abstract samples. Related research methods have been greatly improved in many fields, such as pedestrian re-recognition, visual recognition, drug detection, and genomics object detection [2]. In the research field of deep learning, the related research of pedestrian re-recognition, multi-feature method has made a great breakthrough through combination in the function field of deep learning. Many studies show that deep learning has established relevant research models in multi-feature recognition of pedestrian re-recognition. Experiments show that deep learning has data analysis and interpretation in the field of pedestrian re-recognition [3]. Deep learning is also a powerful form of machine learning. In Europe, people pay more and more attention to biological science, which enables computers to solve perceptual problems such as image and speech recognition. These deep learning technologies, such as deep artificial neural network, use multi-layer processing technology to find the model and structure of very large datasets [4]. With the latest development of digital technology, the size of datasets has become too large, and traditional data processing and machine learning technologies cannot effectively cope with it. However, analyzing complex, high-dimensional, and noise-contaminated datasets is a great challenge, and it is very important to develop new algorithms that can summarize, classify, extract important information, and transform it into understandable forms. In order to solve these problems, the deep learning (DL) model has shown excellent performance in the last ten years [5]. Recently, deep learning has achieved outstanding performance in many fields. Compared with the traditional shallow model, the deep learning method uses a deeper architecture to model the complex distribution of real datasets more powerfully. This paper proposes two fusion methods based on deep learning, which can fuse two branches of the network into a unique feature [6]. In the traditional region of interest (ROI) detection methods based on prior knowledge, the global search solution is usually used to deal with high-resolution remote-sensing images, which leads to too complex calculation. To solve this problem, this study proposes a faster and more efficient ROI detection algorithm based on multi-scale feature fusion, in which, the input image is processed along two feature channels: intensity and direction [7]. In face-to-face re-recognition, extracting image features is an important step in retrieving pedestrian images. At present, most of the methods only extract global or local features of pedestrian images. When learning image features, some inconspicuous details are easily overlooked, which is inefficient or robust for scenes with great differences. Moreover, a multi-feature fusion relative model for pedestrian re-recognition is proposed to synthesize the description state of pedestrian re-recognition according to judgment [8]. Because pedestrians are the key surveillance targets in video surveillance systems, many researches focus on cross-camera pedestrian re-recognition surveillance algorithms. At present, pedestrian re-recognition model not only faces the difficulty of network model training because of the huge number difference between different types of training samples, but also needs to reduce the influence of a large visual performance difference on model recognition accuracy [9]. Object tracking with multiple functions does not perform well in complex scenes and occlusion. The trajectory object is represented by the fusion of all features under linear weighting, and a new method is proposed to estimate the fusion coefficient according to the weight distribution and spatial concentration of all particles, thus improving the reliability of multi-feature fusion [10]. At present, the accuracy of image retrieval is a difficult problem, mainly because of the existence of feature extraction methods. In order to improve the accuracy of image retrieval, an image retrieval method based on multi-features (comprehensive analysis based on basic features) is proposed [11]. In this paper, in order to improve the robustness and performance of re-ID, we propose a re-ID method. By reordering the refined re-ID results (i.e., initial list) obtained by kernel-local Fisher discriminant analysis (kLFDA) and marginal Fisher analysis (MFA), the probability of correct targets in the initial result list can be improved and the robustness can be enhanced [12]. Pedestrian re-recognition (i.e., re-recognition) is an essential prerequisite in multi-camera video surveillance research. The premise is that pedestrian targets should be accurately re-recognized in multi-camera networks with non-overlapping visual fields before other post-event advanced utilization (i.e., tracking, behavior analysis, activity monitoring, etc.,). Driven by the latest development of deep learning technology, important re-ID problems are usually solved by deep discriminant learning or deep generative learning technology [13]. In practice, pedestrian re-recognition usually uses automatic detectors to obtain restricted images of pedestrians. However, in this process, there are two errors in the detector: excessive background and partial missing [14]. A distance measurement method based on depth function: firstly, the resolution is extracted by CN feature of the last layer, and the position and channel of each position are identified. Finally, the quality of the image is extracted from the image. Secondly, using a new improved method, convolution expansion method, and sliding frame transform function method, the qualitative solution of low channel vector is obtained. Thirdly, a distance education algorithm based on image discriminant analysis is proposed. Finally, a fusion method is studied, and the artificial cooperation task with depth resolution is completed [15].

2. Deep Learning and Pedestrian Re-Recognition

2.1. Overview of Deep Learning. In recent years, the improvement of teaching depth has promoted the development of many research topics. The researchers have applied fast and in-depth search methods to pedestrian recognition, which is accurate and effective compared to the traditional pedestrian recognition methods. Before deep learning was widely used, researchers used machine learning to solve related problems. Data must be processed during feature extraction and selection before it can be used. It was more difficult to complete the learning task, but that was not the case. When the correlation coefficient edge of the deep learning neural network is a master, and the weight of the network is too large to be upgraded. But as scientists improve neural networks, model tuning became easier, making more and more people aware of their deep research. Deep learning especially implies adding many hidden layers in traditional artificial neural networks. Then, perform nonlinear operation on the hidden layer to solve machine learning problems such as the increase of the hidden layer and the increase of neurons connection weight, resulting in the entire neural network parameters and the complexity of asymmetry. For this problem, some scholars combine the control law and non-control law to solve this problem. The subjects were initialized using unsupervised learning of the deep learning training process to provide better initial values. Then, the control method was used to determine the appropriate learning speed and to adjust the network parameters to accelerate the convergence speed.

2.2. Pedestrian Re-Recognition. Pedestrian recognition is a technology, which can compare pedestrian recognition with the same camera or different cameras at different angles, times, and positions, and input training and extraction images that may belong to the same pedestrian from the candidate image set. Pedestrian recognition is usually an image search technology, that is, there is no pedestrian test set in the training set. Based on this feature, pedestrian recognition is considered as a classification task in training, and it is considered as a research problem in training. The complexity and accuracy of time and space are the two most commonly used indicators for performance evaluation. In most scenarios, time complexity concentrates on time complexity, which is important for algorithms that require real-time performance. Even better performance time sacrifices time. The comparison of time complexity is based on a consistent hardware infrastructure, otherwise the comparison is meaningless. Accuracy and precision are usually based on the same dataset. If different algorithms are compared, a number of samples can be obtained for accurate classification and prediction. Because of the high requirements for in-depth learning and in-depth learning hardware, the hardware composition of researchers was different in the research of pedestrian recognition algorithm. Therefore, when evaluating the recognition results, it is not the time complexity, but the accuracy of the algorithm. It was emphasized. The corresponding RANK-k represents the probability value that the first k-bit pedestrian sample images in the similarity sorting result of the pedestrian sample database contain the target pedestrian for the target pedestrian to be searched. At present, when evaluating the performance of pedestrian re-recognition algorithm, the commonly used values of k are 3, 5, 10, and 20. By the same token, the larger the RANK-k, the better the performance of the network model. Compared with RANK-1, RANK-k can better evaluate the performance of the model, which reflects the comprehensive retrieval ability of the model to find the target in the pedestrian image database. At present, the common performance indexes for measuring pedestrian recognition performance are correct speed and average accuracy. In this section, the above two performance evaluation indicators are introduced in detail. Initial matching rate and RANK 1 matching rate represent the similarity of target detection, pedestrian model library function and configuration, and the initial pedestrian model image belongs to the likelihood value of the target. The formula for calculating RANK 1 shown in (1) is the number of images aggregated in the pedestrian model base: RANK-1 first matching rate, the first matching rate represents the probability value that the first pedestrian sample image belongs to the target to be searched after comparing the feature similarity of the pedestrian sample database gallery and sorting it from small to large.

$$\text{rank} - 1 = \frac{\sum_{i=1}^{m} S_i}{m}. \quad (1)$$

The corresponding classification results show that the first pedestrian sample image of the rough classification results of the pedestrian sample library is the probability value of the target pedestrian. When evaluating the performance of existing pedestrian recognition algorithms, the typical values are k 3, 5, 10, and 20. A higher taxonomy also indicates good performance of the sample network. Compared with the 1-classifier, the k-classifier can better evaluate the performance of the model, reflecting the comprehensive search ability of the model in the pedestrian image library. Due to multiple factors, such as the effect of previous identification hardware on the pedestrian average simulation accuracy, the sample images used for network model training are usually collected by two cameras with fewer training samples. All pedestrians can be as accurate as detecting Viper data in the warehouse and pedestrian model databases. For these simple datasets, RANK-k curves and CMC can be used to comprehensively evaluate the performance of the network model. With the rapid development of computer science, the application of CNN in pedestrian recognition technology has overcome many obstacles, and the composition of pedestrian information set is becoming more and more complex. The performance of the network model is comprehensively evaluated by using only the above two indicators. The proposed scheme can simultaneously estimate the recall rate and accuracy of the network model, and reflect the average accuracy of various prediction results at the end of the network, which is more suitable for the current complex datasets, and can evaluate the network model comprehensively and objectively. According to the pedestrian sample images and models displayed in the pedestrian, the obtained prediction results can be divided into four types: cut, false positive, false negative, and true negative. Where, the accuracy is the ratio of the real number (tp) in all the prediction accuracy results in the network model (tp + fp). The memory rate is the ratio of the actual score to all the correct scores. (2) and (3) introduce the calculation formula as follows:

$$AP = \int_0^1 p(r)\mathrm{d}r, \quad (2)$$

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q}. \quad (3)$$

3. Convolution Neural Networks in Deep Learning

3.1. Structure of the Convolution Neural Networks. Currently, the convolution neural network has been widely used as a deep network structure model for deep learning. These features are more prominent in the field of computer vision through local observation, weight distribution, and segmentation sampling and other means. The input images can be opened directly without, which is important for many topics in the field of computer vision. The basic structure CNN consists of input layer, convolution layer, pole layer, complete interface layer, and output layer. According to the configuration of the network, as shown in Figure 1, generally, the convolution layer and slurry layer are fed for many times, and the convolution layer and slurry layer are arranged alternately, that is, the convolution layer and bonded to the slurry layer, and the slurry layer is tracked by the convolution layer.

3.2. Levels and Functions of Deep Learning

3.2.1. Convolution Layer, Pooling Layer and Full Connection Layer. The convolution layer in deep learning is the core structure in the deep learning structure, in which, the convolution layer is mainly a deeper analysis of the neural network to obtain higher level characteristics of the analyzed objects. For example, using face images to train the convolution neural network, the first convolution layer learns the line segments and curves of the whole face; the second convolution layer learns the contour of the facial features; the third convolution layer learns the contour of the face; and so on, as the number of layers increases, you can learn the complete characteristics of the whole face. When performing the upper-layer data analysis of the convolution layer, the convolution calculation formula is

$$y_{mn} = f\left(\sum_{j=0}^{Q-1}\sum_{i=0}^{P-1} x_{m+i,n+j} w_{ij} + b\right) 0 \le m \le M, 0 \le n \le N. \quad (4)$$

In formula (4): $x_{m+i,n+j}$ is the pixel value of two-dimensional input data at point $(w+i, n+j)$, w_{ij} is the value of convolution kernel with the size of $P\times Q$ at (i, j), b is the offset value, which usually represents the intercept between spatial coordinates and the origin, and can be understood as a supplement to the linear transformation of input, M and N are the dimension size of input image, respectively, f represents the activation function, and y is the output after convolution operation. After the input image is summed by convolution operation and offset, a two-dimensional feature map is obtained, that is, convolution layer C_x. At the same time, because the action area of convolution kernel on the original image is local, the convolution neural network greatly reduces the number of parameters of weight w compared with the traditional neural network which uses full connection in each layer.

$w_{ij}\,(i, j)b$ in (4): is the value of the convolution kernel of size $P\ Q$, the offset value, usually represents the intercept between the spatial coordinate and the origin, can be understood as a supplement after the linear transformation of the input. Since the operational range of the original image resolution verification is local, the number of parameters for resolving the weights of the neural network is greatly reduced than the traditional completely associated neural network. For example, when the dimension of the input data is 1000, the number of nodes in the hidden layer is 1000. If fully connected, the 1010 parameter is used; if the convolution neural network is used for 100 different convolution kernels, only 104 parameters are required, greatly reducing the number of parameters. The activity of the font convolution layer is a static feature of the image. The image feature extraction section is also applied to the rest of the image. Therefore, the index of statistical operation is necessary, that is, the post-operation is the main purpose of feature extraction. By retaining useful features as much as possible, we reduce the data volume computation, reduce the convolution image resolution, improve the translation sensitivity, and accelerate the network learning speed. There are usually two ways to complete the maximum font and the average font. The maximum pole configuration requires a cross section of different pixels, and the resulting maximum value is the pole configuration value; the font average is to find the average of each component, and then, use the average result after the font operation. In the neural network convolution structure, the entire connection is made after multiple convolution layers and one or more connection layers. The entire connection layer means that every cell in the layer is connected to all the properties of the preceding layer. The connection layer integration of the convolution layer and the well-class separation method is completed.

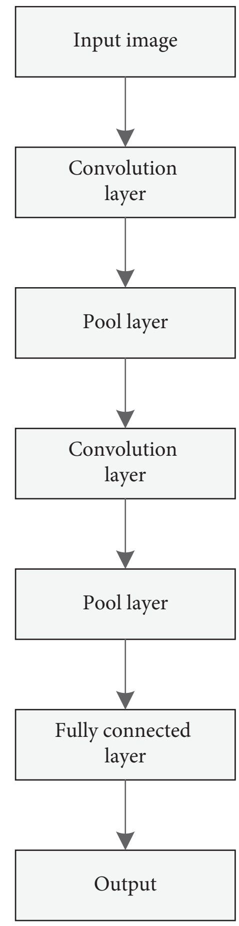

Figure 1: Convolution neural network structure.

3.2.2. Activation Function. In the data output of deep learning, the correlation analysis is conducted by the activation function, which directly affects the performance of

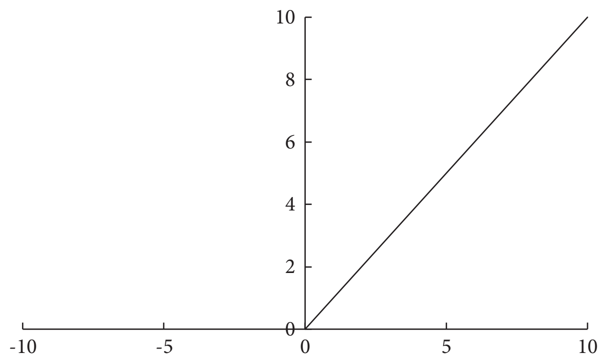

Figure 2: The ReLU function.

the convolution neural network. Activation function is very important for neural networks. At present, the selected activation function is mostly based on previous tuning experience or verified through experiments. However, the average person often does not have enough empirical knowledge to independently choose the activation function, in this case, a typical activation function is selected through experimental verification. The sigmoid function is in a mathematical form such as

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{5}$$

The input values of the activation function are mostly between 0 and 1. The sigmoid function is widely used very early, since the sigmoid function value is 0 at input infinitesimal size and 1 at input infinity, the sigmoid function is easy to derive. If the gradient is 0 when the parameters in the convolution neural network model are not updated and the model performance is poor; (2) the output value of the sigmoid function is never equal to 0. The activation function is also one of the functions of nonlinear saturation in mathematical models, and its mathematical form is the formula as follows:

$$f(x) = \frac{e^{x} + e^{-x}}{e^{x} + e^{-x}}. \tag{6}$$

The difference between the tannin function and the sigmoid function is that the tannin function compresses the input values −1 and 1. The mean value of the tan function results is zero. In this case, the convergence rate is much faster than the sigmoid function. The performance improves compared to sigmoid, but the tanh function still does not solve the problem of gradient dispersion. The ReLU function is a nonlinear unsaturated activation function in its mathematical form:

$$f(x) = \max(0, x). \tag{7}$$

When the input value of the calculated activation function is negative, the output value is zero. When the input value is positive, the output value of the activation function is equal to the input value. The linear study of the activation function is not saturated, which avoids the slope dispersion and converges faster after passing down. The mathematical representation of the activation function is easily computed by simply a threshold as shown in Figure 2. The output is 0 when the input is negative, so the activation function has excellent features.

3.3. Training of Convolution Neural Network

3.3.1. Deep Learning Propagation Algorithm. The forward neural network structure of deep learning is divided into deep learning input and deep learning output. The progress of forward learning is to pass the input values to the output through the neural network. The advance after deep learning is a process of solving the correlation coefficient by using the neural network structure. In the forward propagation phase, the input is often one or more images, where the image can be viewed as a matrix. In the forward propagation process, the weight vector in each layer is first initialized, usually in order to extract more depth features, the weight vector in each layer is different. Then, the input value is multiplied by the weight, in the next correlation step, the input value is transmitted to the output layer to obtain the final relevant parameters, as visible :

$$Y = F_n\left(\ldots\left(F_2\left(F_1\left(X \cdot W^1\right)W^2\right)\ldots\right)W^N\right). \tag{8}$$

Back-propagation algorithm is the core algorithm of neural networks. It is a back-propagation algorithm with neural network theory support. The basic idea of the algorithm is: the input data is input by the input layer, the hidden layer during the propagation process (the hidden layer of convolution neural network is multiple convolution layer and pool layer), the output data is output layer, and the output value is output by the target layer. During the backward propagation process, the error between the target values is minimized. The basic mathematical idea of the underlying propagation algorithm is the application of the link capability rule in computation. The derivative formula is

$$\frac{\partial z}{\partial x} = \frac{\widehat{\partial\partial z}}{\partial y} \cdot \frac{\partial y}{\partial x}. \tag{9}$$

3.3.2. Deep Learning Back-Propagation-Related Algorithm. In the back-propagation algorithm of deep learning, the proposed image training sample can be taken as the output value for deep learning function solution, in which, the weight is required. The training of the activation function can get the image parameters, and the parameter expressions such as formula as follows:

$$f(W, b, x) = y = \text{sigmoid}\left(\sum_i x_i w_i + b\right) = \text{sigmoid}(z). \tag{10}$$

The loss function of deep learning is selected to the mean variance and the cross function in the sample training, such as equations as follows:

$$J(W,b,x,y)=\frac{1}{2}\frac{1}{n}y-h_{w,b}(x)^2, \tag{11}$$

$$J(\hat{y},y)=\frac{1}{n}\sum_{i=1}^{n}-[y_i\log\hat{y}_i+(1-y_i)\log(1-\hat{y}_i)]. \tag{12}$$

In deep learning neural network, the combination of loss function to paranoid function can determine the expressions as follows:

$$\nabla_W J(W,b,x,y)=\frac{\partial}{\partial W}J=\frac{\partial J}{\partial z}\frac{\partial z}{\partial W}=-(y-a)x^T, \tag{13}$$

$$\nabla_b J(W,b,x,y)=\frac{\partial}{\partial b}J=\frac{\partial J}{\partial z}\frac{\partial z}{\partial b}=-(y-a). \tag{14}$$

The update rule of deep learning is shown in (15), in which, the learning rate is used to control the speed of gradient descent. The above is the process of updating parameters by gradient descent method.

$$\begin{aligned}W&:=W+\alpha\nabla_W J(W,b,x,y),\\ b&:=b+\alpha\nabla_b J(W,b,x,y).\end{aligned} \tag{15}$$

3.4. Optimization of the Convolution Neural Network

3.4.1. Stochastic Gradient Descent Method. The iterative inputs from the stochastic gradient and gradient iteration methods are compared and compressed to correct the stochastic gradient iteration sample in the iterative method. The gradient component extracted from the sample quickly obtains the optimal solution of the model when the sample size is large enough. Periodic division of each sample using random descent requires a long time and significantly improved computational complexity. Therefore, this section introduces several methods to avoid overload, including stochastic gradient methods and time algorithms.

3.4.2. Momentum Algorithm. Momentum algorithms (momentum) of deep learning convolution neural networks are usually used together with stochastic gradient descent and can accelerate the convergence of stochastic gradient descent, especially at high curvature, small but consistent gradients, updating rules such as formulas as follows:

$$v:=\alpha v-\varepsilon\nabla_\theta\left(\frac{1}{m}\sum_{i=1}^{m}L\left(f\left(x^{(i)};\theta\right),y^{(i)}\right)\right), \tag{16}$$

$$\theta:=\theta+v. \tag{17}$$

3.5. Loss Function. To effectively train the network model proposed in this chapter, the network model is trained using a combination of triple-loss function and cross-loss function. Use the cross-table function of the classifier. And the classifier predicts the identity of the pedestrian in the loss function of deep learning. Limit the severe loss function to triplicate to increase the capacity of the 3D cumulative module. This chapter randomly selects different ID samples, each ID contains different video sequences, and each video sequence contains one boundary frame. The specific loss function is shown in formula as follows:

$$L_{\text{triplet}}=\frac{1}{P}\frac{1}{K}\sum_{i=1}^{P}\sum_{a=1}^{K}\left[m+\overbrace{\max_{p=1\cdots K}D\left(f_a^i,f_p^i\right)}^{\text{hardest positive}}-\underbrace{\min_{\substack{j=1\cdots P\\ n=1\cdots K\\ i\neq j}}D\left(f_a^i,f_n^j\right)}_{\text{hardest negative}}\right]_+. \tag{18}$$

Cross-entropy loss function is often used in the loss function algorithm of deep learning, including the correlation derivation function of the cross-entropy loss function, such as formulas as follows:

$$L_{soft\max}=-\sum_{i=1}^{P}\sum_{a=1}^{K}\log\frac{e^{W_{y_{aj}}^{T}x_{a,i}}}{\sum_{k=1}^{C}e^{W_k^T x_{a,i}}}, \tag{19}$$

$$L_{\text{total}}=L_{\text{soft}\max}+L_{\text{triplet}}. \tag{20}$$

4. Experimental Results and Analysis

4.1. Pedestrian Re-Recognition Model Based on Multi-Scale Feature Fusion. Here, we propose a pedestrian for studying feature fusion on a basis. First, the stripe model module is introduced in the residual block, and the residual block convolution module is introduced after the stripe model to capture the content semantics in different directions. When extracting content data, the block does not have valid correlation results and blocks. Then, the RESNET50 layers are combined and the content data at different scales are compiled to facilitate the design of the network structure.

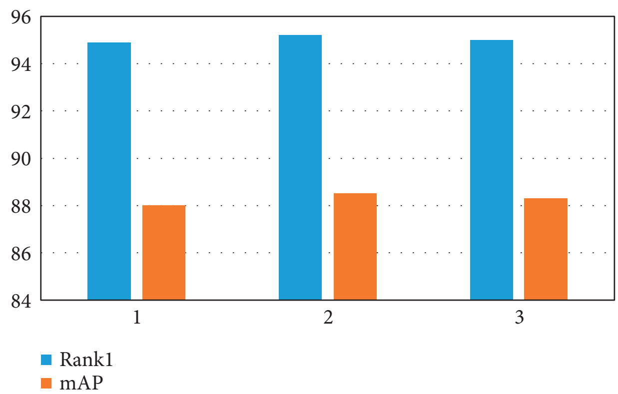

Figure 3: Performance comparison of different cycle times.

Finally, the multi-scale characteristics of the network structure are further maintained through the connection of the debris life pool module, so as to reduce the impact of the environment on people and improve the performance of pedestrians.

4.1.1. SPM Module. Using horizontal and vertical bar image operation in SPM module to obtain context information in different spatial dimensions is helpful for the network to improve scene analysis. It is a multi-scale method of deep learning to avoid the complexity of information, so as to improve the accuracy and precision of multi-scale research in pedestrian. First, the deep learning function introduces both horizontal and vertical layers. A one-dimensional convolution treatment yields both the horizontal and vertical feature tensors. Then, the extension function is used to copy the original image size tag card, and the obtained results are multiplied according to the input function, connecting the output tensor with the input tensor, to help the network establish the performance of the external contact scene.

4.1.2. MPM Module. For different characteristics of sensory officialdom, semantic information cannot be completely retained, easy to be recognized by pedestrians. This chapter extracts feature information at different scales to improve the ability to solve pedestrian characteristics. Based on the selected model, write a remote global data model, focusing on the network short message format. The pile body model adopts 2×2, 3×3, and 4×4 pile cores, which shows the calculation results of the first two piles and the first three piles, respectively, and is combined with the superposition results. Given the low increment of multifunctional information in neural networks, all semantic information can be extracted during network transmission to improve the analytical power of the network. Therefore, before the remaining RESNET50 network structure, the MP3 module is introduced. Finally, the mixed pool module is adjusted according to the characteristics of the pedestrian identity authentication. In the input function, the MP3 module compiles different scale functions. To obtain rich, denser feature information, the feedback MP3 module is extracted from the MP3 module to obtain new features. The average accuracy for each cycle is obtained by comparing the module's initial response level (RANK1) to the different cycles (catalog). The practice shows that both the groups of mixed piles can improve the modulus of the mixed pile group and obtain better performance. Some of the experimental results are shown in Figure 3.

4.1.3. Residual ASPP Module. To maintain the effective properties of the network, the RESNET50 network of ASPP is created combined with empty convolution, thus achieving efficiency by extending the separation of multiple properties and further improving the independence of the system. The structure of this paper is improved. This function uses only a valid ASP module. Moreover, the different voids have different effects on the module extraction. The tunneling velocity causes large variation in feature scale extraction and is absent in the intermediate features of the information network. Furthermore, the transformation, experimental verification, and optimal selection are performed according to the task characteristics of pedestrians in this chapter. Some of the experimental results are shown in Figure 4.

4.2. Network Ablation Analysis. The algorithm in this chapter to eliminate SPM, MPM, and remaining ASPP modules. To verify the module validity, we use this method to present an abstract dataset for each module in TAG-1501. The AGW's most studied quasi-network is used in deep learning multi-scale pedestrian, and the SPM plate is added to this quasi-network, and then, the algorithm research in this chapter is integrated. The ablation analysis is shown in Table 1. (1) Adopt AGW baseline as baseline network. (2) Add SPM modules only in baseline. (3) Add MPM modules to baseline only. (4) Add only the residential ASPP module in baseline (5) Adding MPM module and residual ASPP module to (2) is the final algorithm of this chapter.

4.3. Comparative Experimental Analysis. The algorithmic examples in this chapter are compared to the recent functionality of the general datasets. Figures 5 and 6 show the comparison between methods and other methods in different media types. Results are presented, and the method of this step can achieve better results. The algorithm is mainly due to the different structure of horizontal and vertical Wells and different multi-target data acquisition content, and then, combines emptiness to maintain the network state and effectively extract various features. As shown in Figure 5, the multi-scale fusion method proposed here is RANK1, at 5% and 17%, when compared to the fusion method of the market-1501 dataset. Compared to image-based offset methods, RANK1 and catalog increased by 17% and 32%, respectively. The analysis of the experimental results on the dataset is shown in Figure 5:

As shown in Figure 6, the relevant experimental methods on the dataset of deep learning pedestrian, where the experimental methods in this chapter are about 5 percentage points higher compared to the other experimental methods.

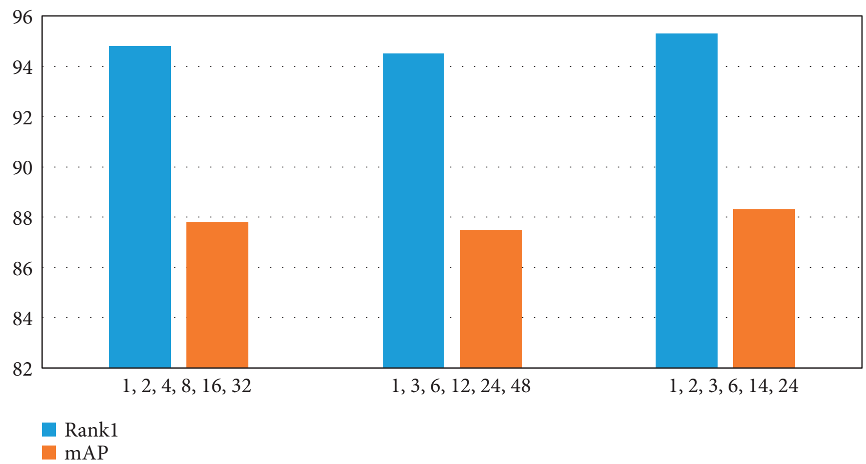

Figure 4: Comparison of void rate effects on performance.

Table 1: Analysis of network ablation.

	SPM	MPM	Residual ASPP	RANK1	mAP
Baseline				95.1	87.8
This chapter algorithm	√			95.2	88.2
This chapter algorithm		√		95.0	88.3
This chapter algorithm			√	95.5	88.2
This chapter algorithm	√	√	√	95.9	88.5

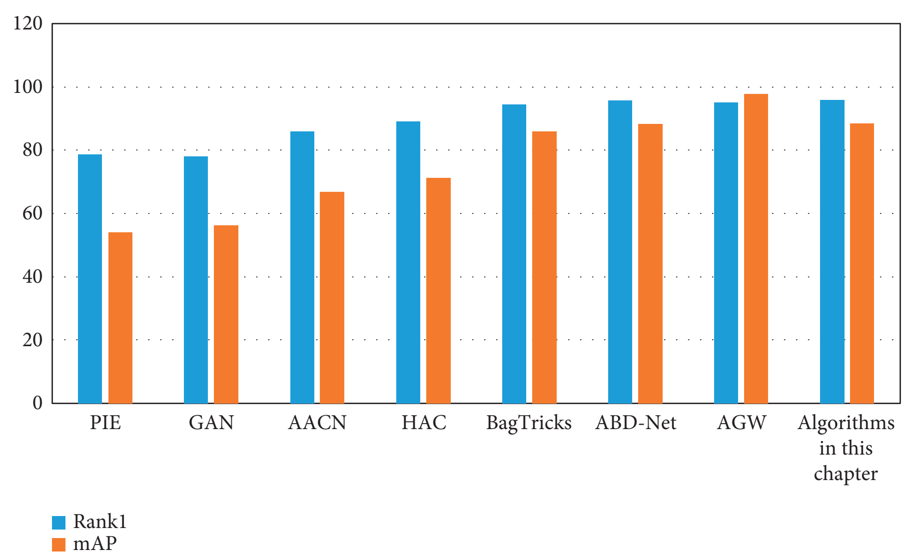

Figure 5: Comparison of experimental results on the market-1501 dataset.

The experimental method of the datasets is as shown in Figure 6:

This chapter presents a new model for pedestrian detection. Through multi-level multi-scale pole function, effectively combined the characteristics of pedestrians at different scales in the network, add ASP module, further improve the overall representation ability of network context information, help the research method of this chapter to better analyze the scene structure, and improve the research accuracy of pedestrian. In addition, the proposed algorithm is tested on generic datasets of Market-1501, CUHK03, and all metrics are greatly improved. Experiments show that the model proposed in this chapter is feasible and effective. However, the approach presented in this chapter does not effectively utilize global information. The next step is to improve the accuracy of the pedestrian recognition model from the perspective of the overall feature relationship. As shown in Figure 7, the methods in this chapter are compared to work related to CUHK03 in datasets not using the structure of this chapter. The AGW used in this paper is the basic model, which is based on the improved dredging method. The experimental results

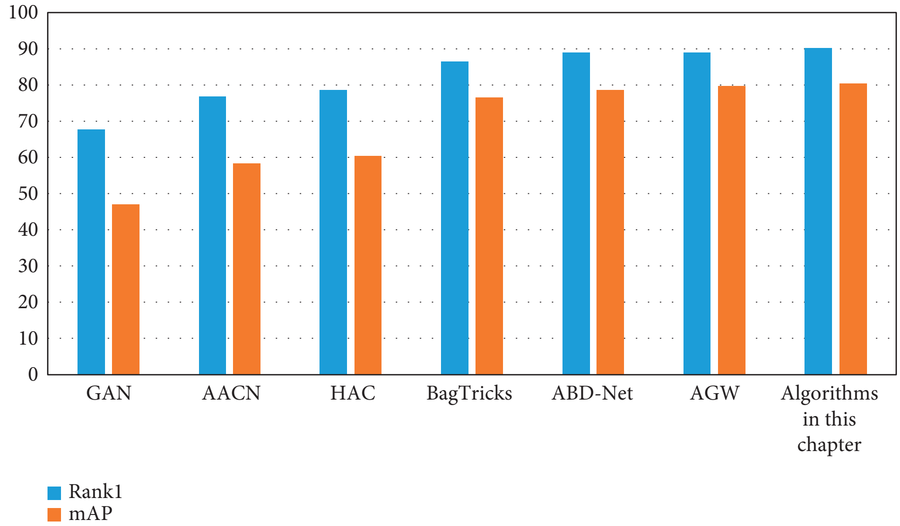

Figure 6: Comparison of experimental results on the DukeMTMC-reID dataset.

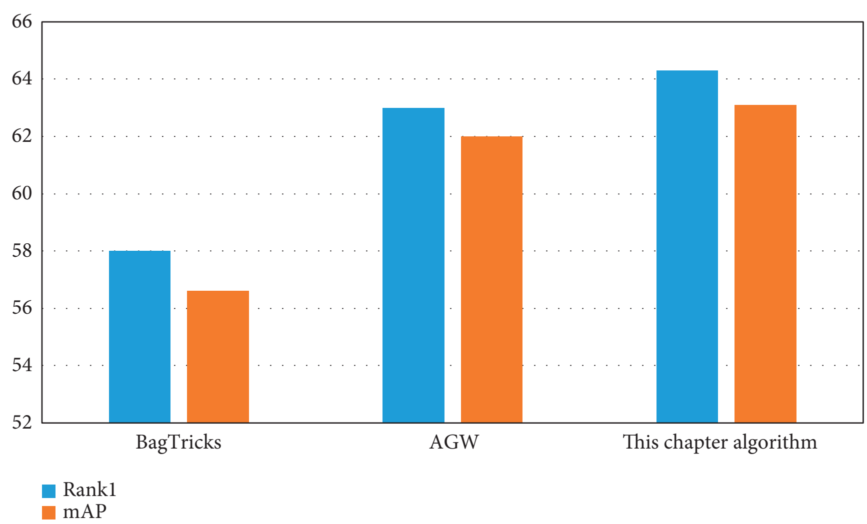

Figure 7: Comparison of test results on CUHK03 dataset.

show that the present method improves in various indexes compared with the newly proposed work. The comparison of the test results on the CUHK03 dataset is shown in Figure 7:

5. Conclusion

With the rapid development of the times, the research on the features of pedestrian re-recognition is getting deeper and deeper, and now the research method of deep learning is introduced into the research of multi-feature pedestrian re-recognition. With the development of pedestrian re-recognition research, this technology has been greatly improved. However, due to the complex and changeable application environment, pedestrian recognition is vulnerable to various disturbances. How to solve these problems is the key to improve the accuracy of pedestrian recognition, and it is also the main problem that many scholars pay attention to. Pedestrian motion detection method based on deep learning has a higher accuracy than traditional methods because of its strong automatic extraction ability and broad application prospects. The main step of this method is to extract and analyze the input pedestrian image and hierarchical evaluation of pedestrian images. This paper also uses the processing mechanism of human visual system to deeply study the feature extraction and transfer mechanism in the learning method, which improves the accuracy of pedestrian recognition task.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

This work was sponsored in part by Research Project of Cross-Mirror Tracking Optimization Algorithm Based on Convolutional Neural Network. Project number: 20B520014. This work was sponsored in part by 2022 Henan Province Science and Technology Research Project for the clothing industry of high-precision three-digit human body and clothing reconstruction and its application in virtual fitting Project No.: 222102210124 (3) This work was sponsored in part by The 2021 Key Project of Higher Education Teaching Reform Research and Practice in Henan Province is based on the characteristics of Kunpeng Ecology. Project Number: 2021SJGLX167.

References

[1] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] X. Hao, G. Zhang, and S. Ma, "Deep learning," *International Journal of Semantic Computing*, vol. 10, no. 03, pp. 417–439, 2016.

[4] N. Rusk, "Deep learning," *Nature Methods*, vol. 13, no. 1, p. 35, 2016.

[5] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 747–750, 2020.

[6] J. Li, B. Zhang, and D. Zhang, "Information fusion based on deep learning," *Information Fusion*, vol. 5, no. 7, pp. 197–256, 2022.

[7] L. Zhang, K. Yang, and H. Li, "Regions of interest detection in panchromatic remote sensing images based on multiscale feature fusion," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4704–4716, 2015.

[8] S. Wu and L. Gao, "Multi-level joint feature learning for person Re-identification," *Algorithms*, vol. 13, no. 5, p. 111, 2020.

[9] W. Qu, Z. Xu, and B. Luo, "Pedestrian Re-identification monitoring system based on deep convolutional neural network," *IEEE Access*, vol. 1, no. 99, 2020.

[10] L. I. Yuanzheng, L. U. Zhaoyang, and L. I. Jing, "Robust video object tracking algorithm based on multi-feature fusion," *Journal of Xidian University*, vol. 39, no. 4, pp. 1–6, 2012.

[11] Y. Zhang, L. I. Yunfeng, and J. Sun, "Image retrieval based on multi-feature fusion," *Journal of Computer Applications*, vol. 35, no. 2, pp. 495–498, 2015.

[12] S. Geng, M. Yu, Y. Liu, Y. Yu, and J. Bai, "Re-ranking pedestrian re-identification with multiple Metrics," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 11631–11653, 2018.

[13] W. Huang, M. Luo, and P. Zhang, "Full-scaled deep metric learning for pedestrian re-identification," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 1–31, 2021.

[14] Z. Zheng, L. Zheng, and Y. Yang, "Pedestrian alignment network for large-scale person Re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3037–3045, 2019.

[15] X. Deng, K. Liao, Y. Zheng, G. Lin, and H. Lei, "A deep multi-feature distance metric learning method for pedestrian re-identification," *Multimedia Tools and Applications*, vol. 80, no. 15, pp. 23113–23131, 2021.