Pedestrian Re-recognition Based on Hybrid Network

Yuchang Si[1]

Software College, Shenyang Normal University
Shenyang 110034 China
siyuchang@163.com

Abstract. With the rapid development of related computer vision algorithms, the large-scale use of video surveillance systems has not only improved traffic safety, but also promoted the development of intelligent high-speed. However, due to the complexity of the application scene, especially in the face of complex scene occlusion factors, the noise generated by the occlusion inevitably leads to the loss of the feature information of the identified person or object, which poses a great challenge to the existing pedestrian re-recognition algorithms. Therefore, this paper proposes a novel pedestrian re-recognition based on hybrid network. Feature extraction is carried out on four cooperative branches: local branch, global branch, global contrast pool branch and associated branch, and powerful diversity pedestrian feature expression ability is obtained. The network in this paper can be applied to different backbone networks. Through experimental comparison, the proposed algorithm has certain advantages compared with the latest methods, and the ablation experimental analysis further proves the effectiveness of the proposed network structure.

Keywords: Pedestrian re-recognition, Hybrid network, Feature extraction, Backbone network.

1. Introduction

With the significant improvement of the standard of highway construction, video surveillance system has been put into use on a large scale, and it is widely distributed in the station, tunnel, bridge, interworking, high slope, service area and other places prone to safety accidents, which plays a decisive impact on the construction and operation safety of highways [1-4]. Especially in the safety of the main body of highway construction and operation, the video surveillance system covers three stages from pedestrian detection, pedestrian tracking to pedestrian re-identification. Among them, pedestrian re-recognition is a challenging problem in the field of video surveillance, and its main task is to search and identify pedestrians in the cross-camera. Over the past 10 years, researchers have proposed many related algorithms based on pattern recognition theory and machine learning. Pedestrian re-recognition technology has attracted extensive attention from academia and industry, and has become a research hotspot and difficulty in artificial intelligence, computer vision and other fields in recent years [5-7].

Pedestrian re-recognition essentially uses computer vision technology to determine whether a particular pedestrian is present in an image or video sequence. The difficulty of this technology mainly lies in two aspects: feature extraction and distance measurement (or similarity measurement). Among them, the difficulty of feature extraction is that the actual application environment of the video surveillance system is an uncontrolled environment, and the feature representation faces the influence of many use scene factors such as complex lighting conditions, local occlusion, changeable shooting Angle, changeable pedestrian posture, and changeable clothing appearance [8-10].

The partial occlusion of personnel is common in the construction and operation of expressways. For example, in the construction process, construction vehicles and equipment will block the staff. In the operation management, toll station facilities, guardrail, trees or sign pillars will also form a shield for personnel. Occlusion will lead to the reduction of the visual area of the target pedestrian in the image and the loss of feature information, and with the noise effect generated by the occlusion, the accuracy of pedestrian retrieval and recognition will be reduced. Most of the existing pedestrian re-recognition algorithms are not suitable for the special case of occlusion, so solving the local occlusion problem becomes a key factor that hinders the implementation of pedestrian re-recognition algorithm in the expressway video surveillance system [11-14].

Feature extraction is the basis of pedestrian re-recognition, and the quality of feature expression directly affects the recognition performance. No matter what feature extraction method is used, its essence is to represent the inherent feature information in the image or video in the form of vectors according to certain rules. Therefore, how to calculate the similarity between features in the feature space, that is, the distance between similar feature vectors, is crucial to the re-recognition of pedestrian identity. Traditional pedestrian re-recognition algorithms are mostly based on manual feature extraction, which are generally divided into spatial, temporal or spatio-temporal mixed features. However, represented by AlexNet [15,16] method, the emergence of deep learning methods, especially convolutional neural network (CNN) method, has been rapidly applied and developed in the field of pedestrian re-recognition by its powerful generalization ability and feature expression ability. The biggest difference with the traditional manual feature extraction method is that the feature is learned from a large data set, and the feature is extracted from the low level to the high level through a neural network similar to the human brain, so as to obtain the depth feature suitable for recognition. In reference [17], the filter pairing neural network model based on CNN was adopted to solve the impact of complex background such as character misalignment, illumination and occlusion on pedestrian re-recognition in images. Reference [18] applied spatial transformation network to pedestrian re-recognition task and achieved good results. Reference [19] systematically analyzed the pedestrian re-recognition problem and applies the multi-task loss method to improve the performance of the algorithm based on the proposed occlusion data set. References [20-22] introduced gesture information to improve the ability to extract salient features of people, and tried to solve the difficulty of pedestrian re-recognition caused by occlusion.

However, due to the inherent defects of CNN, such as information loss caused by pooling layer, convergence caused by gradient descent, etc., after generative adversarial network (GAN) was proposed [23-25], the research on pedestrian re-recognition based on GAN, represented by domestic scholars, increased exponentially. The research based on GAN or its improved model is to learn the essential characteristics of real data through GAN, simulate the real distribution of data, and achieve better recognition performance through recognition and generated antagonism. However, due to the defects of GANs, such as training instability and mode collapse, how to effectively solve the influence of shielding and other factors on pedestrian re-recognition is a hot research issue at present.

In recent years, Transformer-based algorithms have been gradually introduced into the field of computer vision from natural language processing, and its effectiveness in the field of image recognition has been proved by some scholars [26]. Reference [27] proposed TransReID model to solve the global information loss problem in CNN model. The model based on Transformer needed the support of large data sets, and the training time was relatively long. Layer-to-layer focus on modeling global features, lack of relationship and interaction with local features. Local features were crucial for pedestrian re-recognition, especially when faced with the influence of scene factors such as occlusion. In reference [28], sliding windows were designed to simulate local relationships, and the width and height of feature maps were constantly reduced. Swin-Transformer [29,30] is proposed to effectively take into account the global and local feature relationships and improve training efficiency.

Therefore, this paper proposes a pedestrian re-recognition network based on multi-branch cooperation. Feature extraction is carried out on four cooperative branches: local branch, global branch, global contrast pool branch and associated branch, and powerful diversity pedestrian feature expression ability is obtained. The network in this paper can be applied to different backbone networks. In this paper, ResNet and lightweight network OSNet [31] are used as backbone networks to build new networks and achieve start-of-the art performance on multiple pedestrian re-recognition data sets.

2. Proposed Model

This paper proposes a pedestrian re-recognition network (BC-Net) based on multi-branch collaboration. OSNet is used as the backbone to build BC-Net. The input image size is $H \times W \times C$, where the height is $H = 3$, the width is $W = 3$, and the number of channels is $C = 3$. The shared network of BC-OSNet adopts the first five layers of OSNet, including three convolutional layers and two transition layers. Then four cooperative branches are used for feature extraction, including local branches, global branches, global contrast pooling branches and associated branches. The four branches are used in order to learn rich but differentiated features.

Similar to the network structure of BC-OSNet, BC-Net is constructed by using ResNet as the backbone network, abbreviated as BC-ResNet. In the BC-ResNet network architecture, the first four layers of ResNet (Resnet-101-ibn-a used in the experiment) are used as a shared network, and conv_1, conv2_x, conv3_x, conv4_x, conv5_x are used as part of a collaborative branch to learn multiple features.

2.1. Multi-branch Network Architecture

The first branch is the local branch. In this branch, the feature map is divided into four horizontal grids, and the local features of $1 \times 1 \times C$ are obtained using average pooling. Notably, the concatenation of four local features into a column vector produces a single ID prediction loss, and each local feature is learned by the ID prediction loss of the PCB [32,33]. The final cascade features are:

$$f = [f_1^T, f_2^T, \cdots, f_4^T]^T. \tag{1}$$

Where f_1, f_2, f_3 and f_4 represent the four column vectors of the horizontal partition feature graph.

Marked data set $(x_i, y_i), i = 1, 2, \cdots, N_s$. ID predicted loss is:

$$L = -\frac{1}{N_s}\sum_{i=1}^{N_s}\log_2(\frac{exp((W^{y_i})^T f^i + b_{y_i})}{\sum_j exp((W^j)^T f^j + b_j)}). \tag{2}$$

Where W^{y_i} and W^j are columns y_i and j of the weight matrix W. N_s is the number of data sets. f^i, f^j is characterized by columns i and j. b_{y_i} and b_j are offset. Local branches provide more effective and differentiated information than PCB.

The second branch is the global branch. For BC-OSNet, the difference with local branching is that GeM pooling is performed directly after convolution fourth layer and convolution fifth layer. Note that the GeM initialization parameter is set $p_k = 3.0$ to obtain a 512-dimensional vector:

$$GeM(f_k = [f_0, f_1, \cdots, f_n]) = [\frac{1}{n}\sum_{i=1}^{n}(f_i)^{p_k}]^{\frac{1}{p_k}}. \tag{3}$$

Where the GeM operator f_k is a single feature graph. When $p_k \to \infty$, GeM corresponds to maximum pooling, and when $p_k \to 1$, GeM corresponds to average pooling.

The third branch is the Global Contrast Pooling (GCP) branch [34,35]. GCP obtains local contrast features based on average pooling and maximum pooling. For BC-OSNet, after convolutional layer 5 outputs the feature map, it is divided into six horizontal grids, and 256-dimensional feature vectors are obtained by GCP. To better understand the internal structure of GCP, f_{avg} and f_{max} are used to represent the feature maps obtained by average pooling and maximum pooling, respectively.

Average pooling is operated on each local feature ($f_{avg} = \sum_{i=1}^{n} AP(f_i)$), while maximum pooling is performed on the output feature map of convolution layer 5. Contrast features represent information about the differences between them.

$$f_{cont} = \frac{1}{n-1}(f_{avg} - f_{max}). \tag{4}$$

In order to reduce the dimension, the Bottleneck layer is used to handle f_{cont} and f_{max} with the channel number C, and the corresponding dimensionality reduction features are represented by f'_{cont} and f'_{max} with the channel number c. The global comparison features are:

$$q_0 = f'_{max} + B(C(f'_{max}, f'_{cont})). \tag{5}$$

Where C represents the concatenation of f'_{cont} and f'_{max} to form a column vector with the number of channels $2c$. B represents the bottleneck layer operation, through which the channel dimension is reduced from $2c$ to c.

The fourth branch is a one-to-many associated branch. Local features contain information about the various regions, but do not reflect the connections between them, and local association branches associate image regions with the corresponding remaining parts. Similar to GCP, six horizontal grid features are computed through the local relation module, which are $f_1, f_2, \cdots, f_6$.

First, it is obtained by average pooling:

$$r_i = 0.2\sum_{j\neq i} f_j. \tag{6}$$

The bottleneck layer then processes f_i and r_i, reducing the number of channels from C to c, generating f'_i and r'_i. Using association networks, it calculates local relational features:

$$q_i = f'_i + B(C(f'_i, r'_i)), i = 1, 2, \cdots, 6. \tag{7}$$

q_i is a 256-dimensional vector.

2.2. Classifier

For a single pedestrian sample in a given feature space, suppose the number of intra-class similarity and inter-class similarity is K and L, expressed as $(p_i)(i = 1, 2, \cdots, K)$ and $(n_j)(j = 1, 2, \cdots, L)$, respectively. In pedestrian re-recognition task, the aim of training is to minimize the intra-class distance and maximize the inter-class distance. In other words, the expected intra-class similarity p is as close to 0 as possible, and the inter-class similarity n is infinitely close to 1. In order to make the model more efficiently complete the training, based on the Loss functions

Triplet Loss and Circle Loss, a more suitable loss function, namely STR-Loss, is designed for the training of the network structure.

$$STR - Loss = \frac{1}{N} max0, \log[\sum_{i=1}^{K}\sum_{j=1}^{L} exp(\gamma(n'(n_j - m) - p'(p_i - 1 + m)))]. \quad (8)$$

Where γ is a scale factor. m is similarity separation deviation. N is the current batch number.

In order to improve the training process, a weight factor n' and p' are added to n and p, respectively.

$$n' = |n + m|. \quad (9)$$

$$p' = | - (p - 1) + m|. \quad (10)$$

Where n' and p' are non-negative weight parameters, which are used to control the gradient of n and p. When n and p deviate greatly from the optimal values, there is a large gradient. When the deviation is small, the gradient is smaller. This design allows each similarity score to select the optimization weight according to its optimization state, which makes the model training process more targeted and saves the training cost.

3. Experiment and Analysis

3.1. Data Set

The experiment uses two publicly available data sets for pedestrian re-recognition, Market-1501 [36] and DukeMTMC-reID9 [37], and a data set specifically customized for occlusion scenarios, namely, Occluded-Duke7. The Market1501 data set consists of 32668 images taken by six cameras, including 1501 people. The DukeMTMC-reID data set consists of 36411 images taken from eight cameras, including 1404 people. Occluded-Duke data set is manually filtered from DukeMTMC-reID data set, including 15618 images with 702 figures. Its test set consists of 17661 gallery images and 2210 query images containing 1110 people. Compared with the first two public data sets, Occluded-Duke is more challenging to verify the algorithm because it contains a large number of occluded figure images.

3.2. Experimental Setting and Evaluation Criteria

The experiment is based on the deep learning framework Pytorch, using GPU to train on the TeslaV100 deep learning server. In order to verify the classification performance of the model, the model is compared with several algorithms discussed more recently. The experiment uses quintuple cross-validation, a re-sampling process used to evaluate independent data models on a limited data sample. Prior to the implementation of the five-fold cross-validation scheme, the data sets are divided into training data (90% of the total) and test data (10% of the total). The experiment adopts the end-to-end supervised learning method, the parameters are randomly initialized at the beginning, and optimized by cross entropy loss and Adam optimizer, the learning rate is 0.001, and the training period is 250.

In terms of evaluation criteria, Rank-1 recognition rate and mean Average Precision (mAP) in the cumulative matching characteristic curve are used for evaluation. Rank-1 recognition rate represents the ratio of the number of correctly judged label tests in the first recognition to the total number of test samples after matching according to similarity. The average accuracy mean represents the mean of the average accuracy of the model across all categories in the data set.

3.3. Comparison Experiments

In order to demonstrate the effectiveness of the proposed method, four newer occlution-oriented pedestrian re-recognition algorithms, including Pose Transfer [38], TCSDO [39], PGFA [40], HOReID model [41], and Swin Transformer, are compared on three data sets. Table 1 shows the experimental comparison results on the three data sets.

Compared with four convolutional neural network-based methods (Pose Transfer, TCSDO, PGFA and HOReID), The un-optimized baseline network Swin Transformer has achieved relatively good results on the three data sets, which shows that Swin Transformer has excellent potential in pedestrian re-recognition tasks, and how to further optimize it is worth studying. Swin Transformer can learn global information at the initial stage of network training, and its use of information is better than that of convolutional neural networks. In terms of the utilization

Table 1. Evaluation results of different methods on three datasets/%

Data set	Occluded Duke	Occluded Duke	Market-1501	Market-1501	DukeMTMC-reID	DukeMTMC-reID
Method	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
Pose Transfer	48.6	35.3	87.7	69.0	68.7	48.2
TCSDO	50.2	36.3	89.4	73.3	74.3	60.7
PGFA	51.5	37.4	91.3	76.9	82.7	65.6
HOReID	55.5	43.9	94.3	85.0	87.0	75.7
Swin Transformer	54.9	42.8	92.9	79.8	85.8	75.3
Proposed	58.4	47.3	95.6	88.3	88.2	80.4

of feature information, the HOReID model that comprehensively utilizes high-level information is better than the PGFA model that only extracts key point information. Through the targeted extraction and processing of pedestrian image information, the HOReID model achieves higher scores.

The method proposed in this paper further optimizes Swin Transformer according to the characteristics of pedestrian re-recognition tasks, enabling the model to achieve better performance in all three data sets. The optimization method adopted in this paper is similar to the Pose Transfer method, both of which belong to increasing the data amount by generating occluded samples, but the method in this paper has huge advantages. On the one hand, it is affected by the difference of network structure; on the other hand, the Pose Transfer method uses GAN to generate images of people with different poses, which is not targeted to optimize the occlusion problem, and the backward network performance leads to the great difference between the pose diversity and the pose of people in the actual scene.

3.4. Ablation Experiment

In order to further verify the effectiveness of the proposed method, the Occluded sample generation module and STR loss function were ablated on the Occluded-Duke dataset. In this experiment, the performance of the VIT (Vision Transformer) network [42], which is based on Transformer, is also tested, and the model training time is added to the evaluation criteria, as shown in Table 2.

Table 2. Results of ablation experiments on Occluded-Duke dataset

Method	Training time/min	Rank-1/%	mAP/%
VIT	129	54.7	42.9
Swin Transformer	36	54.9	42.8
Remove the sample generation module	15	55.2	43.0
Cross entropy loss function	53	57.7	46.4
Proposed	32	58.4	47.3

As can be seen from the results in Table 2, compared with VIT, the method proposed in this paper greatly saves training time while maintaining performance. Although the occlusion sample generation module is added to the network structure, the training time is relatively extended, but the Rank-1 and mAP index are improved on the whole. Meanwhile, with the assistance of the loss function STR-Loss, this paper proposes that the model is improved from the training time to the evaluation index.

4. Conclusion

In this paper, we propose a pedestrian re-recognition network (BC-Net) based on multi-branch collaboration, which has a four-branch architecture of local branch, global branch, local association pooling and contrast branch to obtain more diverse and higher resolution features. Experiments using OSNet and ResNet as BC-Net backbone networks confirm the effectiveness of the four-branch framework. In addition, BCNet has applied microstructures and techniques, including GeM. ablation experiments to show that four-branch collaboration can learn complementary features. BC-Net obtained Start-of-the-art results on three pedestrian re-recognition databases. Based on the fusion innovation of multiple pedestrian re-recognition networks, BCNet proves the excellence of multi-branch networks through experiments. The disadvantage is that the number of network parameters and training time have increased significantly. At present, the effective improvement of pedestrian re-recognition performance through multi-branch cooperation lacks theoretical support, which needs further proof and analysis.

5. Conflict of Interest

The authors declare that there are no conflict of interests, we do not have any possible conflicts of interest.

Acknowledgments. None.

References

1. Patrikar D R, Parate M R. Anomaly detection using edge computing in video surveillance system[J]. International Journal of Multimedia Information Retrieval, 2022, 11(2): 85-110.
2. Rajavel R, Ravichandran S K, Harimoorthy K, et al. IoT-based smart healthcare video surveillance system using edge computing[J]. Journal of ambient intelligence and humanized computing, 2022, 13(6): 3195-3207.
3. Yin S, Li H, Laghari A A, et al. An Anomaly Detection Model Based On Deep Auto-Encoder and Capsule Graph Convolution via Sparrow Search Algorithm in 6G Internet-of-Everything[J]. IEEE Internet of Things Journal, 2024.
4. Kong M, Guo Y, Alkhazragi O, et al. Real-time optical-wireless video surveillance system for high visual-fidelity underwater monitoring[J]. IEEE Photonics Journal, 2022, 14(2): 1-9.
5. Zhao Y, Zhou H, Cheng H, et al. Cross-modal pedestrian re-recognition based on attention mechanism[J]. The Visual Computer, 2023: 1-14.
6. Han K, Zhang N, Xie H, et al. Application of Multi-Feature Fusion Based on Deep Learning in Pedestrian Re-Recognition Method[J]. Mobile Information Systems, 2022, 2022.
7. Yin S. Object Detection Based on Deep Learning: A Brief Review[J]. IJLAI Transactions on Science and Engineering, 2023, 1(02): 1-6.
8. Zhao Y, Sun G. Multi Strategy fusion pedestrian re-recognition method for community[C]//Proceedings of the 8th International Conference on Computing and Artificial Intelligence. 2022: 407-411.
9. Zhang Y, Cao J, Li X, et al. Pedestrian Re-recognition Based on Memory Network and Graph Structure[C]//International Forum on Digital TV and Wireless Multimedia Communications. Singapore: Springer Nature Singapore, 2022: 3-15.
10. Fan Y, Li H, Sun B. Cycle GAN-MF: A Cycle-consistent Generative Adversarial Network Based on Multifeature Fusion for Pedestrian Re-recognition: Cycle GAN-MF[J]. IJLAI Transactions on Science and Engineering, 2024, 2(1): 1-9.
11. Wu D, Huang H, Zhao Q, et al. Overview of deep learning based pedestrian attribute recognition and re-identification[J]. Heliyon, 2022, 8(12).
12. Wu Z. Pedestrian re-recognition method based on dual attention and GRU[C]//Journal of Physics: Conference Series. IOP Publishing, 2022, 2303(1): 012071.
13. Shan L, Xiong Q, Li K, et al. The Improved Pedestrian Target Detection and Re-recognition Algorithm[C]//2023 International Conference on Cyber-Physical Social Intelligence (ICCSI). IEEE, 2023: 586-591.
14. Jiang Y, Yin S. Heterogenous-view occluded expression data recognition based on cycle-consistent adversarial network and K-SVD dictionary learning under intelligent cooperative robot environment[J]. Computer Science and Information Systems, 2023, 20(4): 1869-1883.
15. Chen H C, Widodo A M, Wisnujati A, et al. AlexNet convolutional neural network for disease detection and classification of tomato leaf[J]. Electronics, 2022, 11(6): 951.
16. Ullah A, Elahi H, Sun Z, et al. Comparative analysis of AlexNet, ResNet18 and SqueezeNet with diverse modification and arduous implementation[J]. Arabian journal for science and engineering, 2022, 47(2): 2397-2417.
17. Huang Y, Xie G, Yang K, et al. Domain Adaptive Pedestrian Re-recognition for Oilfield Operations Security[C]//Proceedings of the 2023 7th International Conference on Computer Science and Artificial Intelligence. 2023: 277-281.
18. Shan Y, Yu G, Gao Y. Pedestrian Re-identification Based on Gait Analysis[J]. IEEE Access, 2023.
19. Jianheng L. Improved Method for Pedestrian Recognition Based on Generative Adversarial Networks[J]. Journal of Artificial Intelligence Practice, 2023, 6(2): 23-30.
20. Hong F, Lu C, Tao W, et al. OMNet: Object-Perception Multi-Branch Network for Pedestrian Re-Identification[J]. Big Data Research, 2022, 27: 100302.
21. Xu Z, Chen J, Chen Z, et al. A video target re-recognition method based on adaptive attention enhancement and multi-scale feature fusion[J]. IEEE Access, 2023.
22. Liu H, Zheng T, Sun F, et al. ER-DeepSORT: Pedestrian Multiobject Tracking with Enhanced Reidentification[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2023, 18(3): 427-435.
23. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
24. Gui J, Sun Z, Wen Y, et al. A review on generative adversarial networks: Algorithms, theory, and applications[J]. IEEE transactions on knowledge and data engineering, 2021, 35(4): 3313-3332.
25. Yin S, Wang L, Wang Q, et al. M2F2-RCNN: Multi-functional faster RCNN based on multi-scale feature fusion for region search in remote sensing images[J]. Computer Science and Information Systems, 2023 (00): 54-54.
26. Zhang Q, Xu Y, Zhang J, et al. Vitaev2: Vision transformer advanced by exploring inductive bias for image recognition and beyond[J]. International Journal of Computer Vision, 2023, 131(5): 1141-1162.
27. Meng L, Li H, Chen B C, et al. Adavit: Adaptive vision transformers for efficient image recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12309-12318.

28. Wang Y, Huang R, Song S, et al. Not all images are worth 16x16 words: Dynamic transformers for efficient image recognition[J]. Advances in neural information processing systems, 2021, 34: 11960-11973.
29. Zhao J, Yan K, Zhao Y, et al. Transformer-based dual relation graph for multi-label image recognition[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 163-172.
30. Liu X, Wang L, Han X. Transformer with peak suppression and knowledge guidance for fine-grained image recognition[J]. Neurocomputing, 2022, 492: 137-149.
31. Pan K, Zhao Y, Wang T, et al. MSNet: a lightweight multi-scale deep learning network for pedestrian re-identification[J]. Signal, Image and Video Processing, 2023, 17(6): 3091-3098.
32. Tsai C Y, Chiu C C, Chen J S. A case-based reasoning system for PCB defect prediction[J]. Expert Systems with Applications, 2005, 28(4): 813-822.
33. Sun Y, Zheng L, Li Y, et al. Learning part-based convolutional features for person re-identification[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(3): 902-917.
34. Teng L, Qiao Y, Shafiq M, et al. FLPK-BiSeNet: Federated learning based on priori knowledge and bilateral segmentation network for image edge extraction[J]. IEEE Transactions on Network and Service Management, 2023.
35. Yin S, Wang L, Shafiq M, et al. G2Grad-CAMRL: an object detection and interpretation model based on gradient-weighted class activation mapping and reinforcement learning in remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023.
36. Zheng L, Shen L, Tian L, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1116-1124.
37. Ma L, Sun Q, Georgoulis S, et al. Disentangled person image generation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 99-108.
38. Liu J, Ni B, Yan Y, et al. Pose transferrable person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4099-4108.
39. Zhuo J, Lai J, Chen P. A novel teacher-student learning framework for occluded person re-identification[J]. arxiv preprint arxiv:1907.03253, 2019.
40. Miao J, Wu Y, Liu P, et al. Pose-guided feature alignment for occluded person re-identification[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 542-551.
41. Wang G, Yang S, Liu H, et al. High-order information matters: Learning relation and topology for occluded person re-identification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 6449-6458.
42. Mao X, Qi G, Chen Y, et al. Towards robust vision transformer[C]//Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. 2022: 12042-12051.

Biography

Yuchang Si is with the Software College, Shenyang Normal University. His research direction is image processing, computer application and AI.