Journal of Computers Vol. 34 No. 1, February 2023, pp. 187-197
doi: 10.53106/199115992023023401014

Body Correlation Network for Person Re-Identifications

Bailiang Huang[1], Yan Piao[1*], Yanfeng Tang[1], Baolin Tan[2]

[1] School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130000, Jilin Province, China
ww5171351@126.com
[2] Shenzhen Yinglun Technology Co. LTD, Longgang District, Shenzhen, China
david.tan@yinglun-tech.com

Received 27 April 2022; Revised 18 July 2022; Accepted 10 August 2022

Abstract. Visual information accounts for approximately 80 to 85 percent of the information available daily in modern cities. As such, person re-identification, an instance-level image retrieval task, has become an important research topic in computer vision, machine learning, and other fields in recent years. Traditional person re-identification methods based on convolutional neural networks only extract the global feature information of people. Thus, when external factors such as changes in occlusion and illumination disturb people, the recognition performance of these methods substantially decreases. We therefore develop body correlation network (BC-Net), which takes full advantage of images of body parts and the correlations between them. Specifically, BC-Net uses body part feature information and correlation feature information as nodes and edges, respectively, and then uses graph convolutions to learn the overall topology of people. To improve use of crucial feature information, we also design a unique method of propagation between nodes and edges. We conduct extensive comparative experiments on the Market-1501 and DukeMTMC-reID datasets, and the results demonstrate that BC-Net outperforms other state-of-the-art techniques.

Keywords: computer vision, deep learning, graph convolution, person re-identification

1 Introduction

Gradual improvements in public security awareness have led to the wide deployment of cameras in significant locations in cities, such as streets, stations, airports, and hospitals, resulting in the generation of large amounts of image data. Traditional manual data-processing methods are unable to process these large amounts of data, and thus intelligent video surveillance systems have become necessary. Person re-identification (Re-ID) is a crucial part of these systems, as it uses computer vision technology to determine whether there is a pre-selected person in images or video sequences. That is, it uses a monitored image of a person to retrieve that person's image from image data collected on multiple devices [1]. Person Re-ID can be used in security, criminal investigation tracking, and other fields, and can also compensate for the visual limitations of a camera. Thus, it has high market application prospects and research value.

The challenge in the person Re-ID task is that various factors, such as shooting angle, pedestrian gait, pedestrian posture, illumination change, and object occlusion, affect the human attributes in images. Before the advent of convolutional neural networks (CNNs), studies were mainly performed using manually designed distinguishing features and similarity measures, and have resulted in many research achievements and valuable findings. However, the effectiveness of manually designed visual features largely depends on the prior knowledge of the designer, and thus the quality of these features is determined by the manual adjustment of their parameters by designers. As feature parameters have many limitations, the recognition effect achieved is usually not ideal, and the generalization ability of features is relatively weak [2]. Recently, with the development of machine learning and the continuous improvement of computing power, CNNs have been widely used in person Re-ID tasks. CNNs can obtain feature information progressively from the bottom to the top level of a large-scale dataset, extract the feature information suitable for identification or classification, and then learn this information. CNN-based methods can therefore combinejointly optimize feature expression and similarity measures, resulting in better models than traditional methods. For example, the Rank-1 of the CNN-based FPB models applied to the Market-1501 dataset reached 96.1% [3], which far exceeds the accuracy of the traditional method.

However, most CNN-based person Re-ID methods only mechanically extract and learn the global features of

* Corresponding Author

pedestrian images. These contain simple, single global feature information that cannot meet the needs of complex and changeable practical applications. Some studies have added attention modules to enhance the ability of networks to extract feature information from pedestrian images [4-6]. However, such approaches have problems: they can lose feature information and extract invalid feature information. Accordingly, inspired by the human ability to distinguish people in real life, we solve these two problems by considering the feature information of various human body parts and the correlation information between these parts, which we used to devise a body correlation network (BC-Net). In contrast to studies that have used body parts to improve networks performance for person Re-ID tasks and thus manually annotated attribute information to construct a graph structure [7-9], we use graph convolutional networks (GCNs) to efficiently construct the feature information of each body part. We also use body parts as nodes and combinations of multiple parts as edges. The GCN is used to update and learn the topological information of parts and the correlations between body parts. We train the BC-Net model on two classical datasets, and the experimental results show that its outperforms many state-of-the-art methods.

Our main contributions to the field are as follows.

- We devise a novel architecture, BC-Net, that employs body parts as nodes and multiple body parts as edges for learning the part feature information and correlation feature information of people.
- We devise a unique transmission method to enhance the transmission of information on part features and correlations between these.
- We devise a simple and effective attention aggregation mechanism to enhance the final graph-level feature representation.
- We conduct a rigorous experimental comparison using two person Re-ID datasets, which demonstrates the effectiveness of BC-Net.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces BC-Net and provides a detailed overview of the nodes and edges in the construction method, the propagation mode, and the aggregation mode. Section 4 presents the datasets and implementation details and comprehensively analyzes the experimental results obtained using the two benchmark datasets. Section 5 summarizes our findings.

2 Related Work

Person Re-ID aims to find a target person from numerous images, and some advanced mainstream algorithms have been applied for person Re-ID and achieved unexpected accuracy gains [10-12].

Traditional global extraction methods are critical in learning the overall feature information of people, which enables people's images to be retrieved from various sources. However, these methods may ignore crucial distinguishable structures and regional feature information. The part-level feature information extraction method can solve this problem well [13-14]. Especially by applying its carefully designed attention mechanism, thereby improving the ability of networks to extract part-level feature information [15-20]. Chen et al. [21] devised an attention pyramid method to capture part-level feature information. Li et al. [22] designed two mechanisms to enhance the learning of inseparable and diverse properties of body parts. In some studies [23-26], authors have used part-level feature information, such as the human body, posture, and clothing, to standardize the description of people, thus capturing more detailed identifiable feature information and improving the effectiveness and robustness of a network. Yang et al. [27] enhanced the effectiveness and robustness of a network by distributing posture information to various body parts. Huang et al. [28] used the clothing characteristics of people and standardized the description of people by embedding the clothing state consciousness of each part; this approach can enhance distinguishable feature information. Li et al. [29] used the process of attribute detection to generate the corresponding part detector and used attribute information to refine the description of the parts. They combined the refined local features and the overall features to generate a final feature representation. Rao et al. [30] used the principle that the relative height of body parts in the skeleton is fixed to capture the various relationships between adjacent body part nodes and the cooperative relationships between body parts at various levels, which enabled more detailed identifiable feature information to be captured than by some other methods.

A graph is a data format that can be used to represent various networks, including social networks, communication networks, and relational networks. The nodes in a graph represent individuals in a network, and the edges represent connections between these individuals. In person Re-ID tasks, graph structure data are required to represent the clothing and attribute changes of characters, the relationship between various body parts, and the skeleton-structure relationship. Therefore, the emergence of graph CNNs provides a new method for solving person Re-ID tasks [7-8, 31]. Typically, researcher [9, 32-34] have used key points or attribute information of the human body to construct the graph structure and support learning of the topological structure of the feature graph

during information transmission and updating. Nguyen et al. [9] combined the probability of character attribute labels appearing in the dataset with global body feature information, collated the resulting data in a graph, and used GCNs to learn the topological structure of character attributes. Chen et al. [34] used the feature information of the key nodes of the human body to form GCNs and enhanced representation learning from the feature graph during information transfer and update. Recently, GCNs have also achieved good performance in some unsupervised person Re-ID tasks [35-36]. Shen et al. [7] established a graph describing the pairwise relationship between a probe and gallery images. They propagated and updated the graph in an end-to-end manner and predicted pairwise relationship features. This can achieve accurate similarity estimation and provide more accurate information for relationship fusion than other methods. Similarly, Bai et al. [37] fused multi-domain feature information by graph convolution to minimize the distance between domains.

In the current study, inspired by part-level feature information and graph structure, we derive a human body correlation network. This network uses the part-level feature information of each body part in a targeted manner and uses a graph structure to learn the correlation information between body parts. The final global representation is highly distinguishable in the image-based person Re-ID task.

3 Materials and Methods

The full-scale and attention mechanism can effectively improve a network's ability to extract the feature information of body parts. However, because of the lack of correlation between body parts, the feature information of body parts is lost in the presence of a chaotic background or a dislocated structure, which degrades re-identification performance. Accordingly, to fully invoke the feature information of each body part and extract the correlation between body parts, we design a BC-Net learning framework, as detailed below.

3.1 Construction of Nodes and Edges

We aim to capture the feature information and the correlation of four body parts by constructing a set of a fixed graph, expressed as $G=(S^*, L^*)$. The graph contains four nodes $S^*=\{S_1^*, S_2^*, S_3^*, S_4^*\}$ and eleven edges $L^*=\{L_1^*, L_2^*,\ldots, L_{10}^1, L_{11}^1\}$. Each edge contains two to four different nodes. The construction process is shown in Fig. 1.

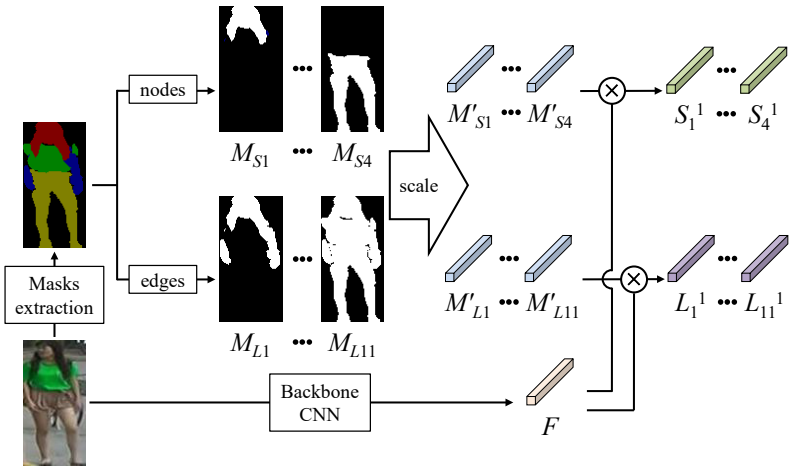

Fig. 1. Construction of nodes and edges

We use a backbone CNN to obtain the feature information tensor F of people, and use the mapping of body parts on F as the features of nodes in the graph. We use the superior performing Self-Correction for Human Parsing model [38] to preprocess images, during which we combine relevant parts to accurately segment the human body into four parts (head, arm, body, and leg). This affords a set of binary masks of the same size as the input image that are denoted as $M_{Si}=\{M_{S1}, M_{S2}, M_{S3}, M_{S4}\}$. To facilitate the operation, we scale M_{Si} to the same size as the feature map F and obtain $M'_{Si}=\{M'_{S1}, M'_{S2}, M'_{S3}, M'_{S4}\}$ after the $L1$ normalization operation. The feature $S_i^1=M'_{Si}F=\{S_1^1, S_2^1, S_3^1, S_4^1\}$ of the four primary nodes can be obtained from this body part mask M'_{Si} and the feature map F.

We define eleven fixed edges to establish the spatial correlation of four nodes. Each edge contains two to four different body part nodes, and the connection relationship between an edge and each node is shown in Table 1. We take the total mapping of each body part node on F as the edge feature. For example, the edge L_{10} is connected to nodes S_2, S_3, and S_4, and the corresponding body parts are arms, body, and legs. We add the masks of the three body parts to obtain $M_{L10}=M_{S2}+M_{S3}+M_{S4}$. The mask M_{L10} is scaled to the same size as the feature map F, and M'_{L10} is obtained after the $L1$ normalization operation. Finally, the edge L_{10} is characterized by $L_{10}=M'_{L10}F$. Therefore, from the body part mask M'_{Lj} and the feature map F, we can obtain the features of eleven primary edges $L_j^1=M'_{Lj}F=\{L_1^1, L_2^1, L_3^1, L_4^1, L_5^1, L_6^1, L_7^1, L_8^1, L_9^1, L_{10}^1, L_{11}^1\}$.

Table 1. Connection relationship between edges and nodes

Nodes	Edges										
	L_1^k	L_2^k	L_3^k	L_4^k	L_5^k	L_6^k	L_7^k	L_8^k	L_9^k	L_{10}^k	L_{11}^k
S_1^k	●	●	●				●	●	●		●
S_2^k	●			●	●		●	●		●	●
S_3^k		●		●		●	●		●	●	●
S_4^k			●		●	●		●	●	●	●

3.2 Propagation of Nodes and Edges

When designing the propagation scheme, we require that the various body parts of the human image have distinct correlations. For example, when a person wears clothes with clear patterns, the body node should be able to provide more distinguishable feature information than other nodes. Therefore, in an edge containing the body node, this node has a higher correlation with the edge than the other nodes. Similarly, each edge containing the body node has a different correlation with the body node. Therefore, we design a unique propagation mechanism to strengthen the mutual transmission of important feature information and the correlation between nodes and edges. The scheme is shown in Fig. 2.

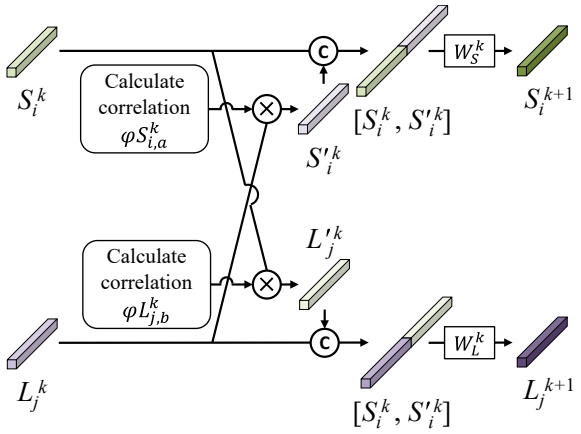

Fig. 2. Schematic diagram of the propagation scheme for nodes and edges

We define the correlation between edges and a node by computing the cosine distance between the node S_i and the edges containing the node S_i. In the k-th layer of the graph neural network, the correlation between the node S_i^k and the a-th edge containing the node is calculated as follows:

$$eS_{i,a}^k = \cos\left(S_i^k, Q(i)_a^k\right) . \tag{1}$$

where cos indicates the cosine distance calculation and $Q(i)_a^k$ represents the set of edges associated with the node S_i^k. For example, for the node S_1^k, $Q(1)_a^k=\{L_1^k, L_2^k, L_3^k, L_7^k, L_8^k, L_9^k, L_{11}^k\}$ can be determined based on Table 1. We use the softmax function to sequentially normalize the correlations of the seven edges containing the node S_i^k.

$$\varphi S_{i,a}^{k} = \frac{\exp(eS_{i,a}^{k})}{\sum_{a} \exp(eS_{i,a}^{k})} \ . \tag{2}$$

We aggregate the edges containing the node S_i^k into S'^k_i by weighted summation, depending on the correlation between the node S_i^k and edges.

$$S_i'^k = \sum_a \varphi S_{i,a}^k Q(i)_a^k \ . \tag{3}$$

We define the correlation of nodes to an edge by calculating the cosine distance between the edge L_j and nodes contained in the edge L_j. In the k-th layer of the graph neural network, the correlation between the edge L_j^k and the b-th node it contains is calculated as follows:

$$eL_{j,b}^k = \cos\left(L_j^k, P(j)_b^k\right) \ . \tag{4}$$

$P(j)_b^k$ represents the set of nodes connected by the edge L_j^k. For example, $P(1)_b^k = \{S_1^k, S_2^k\}$ for the edge L_1^k can be determined based on Table 1. We use the softmax function to sequentially normalize the correlations of all nodes in the edge L_j^k.

$$\varphi L_{j,b}^k = \frac{\exp(eL_{j,b}^k)}{\sum_b \exp(eL_{j,b}^k)} \ . \tag{5}$$

We aggregate the nodes containing the edge L_j^k into L'^k_j by weighted summation, depending on the correlation between the edge L_j^k and the connected nodes.

$$L_j'^k = \sum_b \varphi L_{j,b}^k P(j)_b^k \ . \tag{6}$$

Splicing S_i^k and S'^k_i, L_j^k and L'^k_j. After being updated by the fully connected layer, the node feature S_i^{k+1} and the edge feature L_j^{k+1} are obtained.

$$S_i^{k+1} = \text{LeakyReLU}(W_S^k [S_i^k, S_i'^k]) \ . \tag{7}$$

$$L_j^{k+1} = \text{LeakyReLU}(W_L^k [L_j^k, L_j'^k]) \ . \tag{8}$$

W_S^k is the trainable weight matrix of nodes in layer k, and W_L^k is the trainable weight matrix of edges in layer k. Node features $S_i^K = \{S_1^K, S_2^K, S_3^K, S_4^K\}$ and edge features $L_j^K = \{L_1^K, L_2^K, L_3^K, L_4^K, L_5^K, L_6^K, L_7^K, L_8^K, L_9^K, L_{10}^K, L_{11}^K\}$ are obtained after the feature information is updated and propagated for K-1 rounds.

3.3 Aggregate Operation

Finally, we also aggregate the updated four nodes and eleven edges. The node features and edge features are concatenated as a graph-level feature representation for each person image. The nodes and edges of various body parts in person images should have different degrees of importance after being affected by occlusion, illumination, color and other factors [33, 39]. Therefore, we consider it necessary to design an attention mechanism to address this problem. We first calculate the node aggregation weight αS_i and the edge aggregation weight αL_j. Then, we complete the final aggregation operation by performing a weighted summation of nodes and edges according to their respective aggregation weights.

$$\alpha S_i = \frac{\exp(S_i^K)}{\sum_i \exp(S_i^K)} \quad . \tag{9}$$

$$\alpha L_j = \frac{\exp(L_j^K)}{\sum_j \exp(L_j^K)} \quad . \tag{10}$$

$$S = \sum_{i=1}^{4} (\alpha S_i) S_i^K \quad . \tag{11}$$

$$L = \sum_{j=1}^{11} (\alpha L_j) L_j^K \quad . \tag{12}$$

4 Experiments

We conducted extensive experiments to verify the effectiveness of the BC-Net.

4.1 Experimental Setup

Dataset. We conducted experimental validations on two widely used person Re-ID datasets, Market-1501 [40] and DukeMTMC [41]. Market-1501 is a large-scale person Re-ID dataset collected on the campus of Tsinghua University, China. Its training set consists of 12,936 images from 751 identities, and its test set consists of 19,732 images from 750 identities. There is no intersection between these two sets, and the dataset also contains distracting images and low-resolution images. DukeMTMC-reID was collected at Duke University, USA, and is very close to a real-world situation because pedestrians in the images may be dressed in similar clothes or occluded. Its training set consists of 16,522 images from 702 identities, its test set consists of 17,661 images from 702 identities, and its query set consists of 2,228 images collected by each camera from 702 identities in the test set.

Implementation Rules. We used ResNet-50 [42] as the backbone and pre-trained it on the ImageNet dataset [43]. The size of probe images was uniformly adjusted to 256×128 for training and testing. Data expansion included random erasures and random flips. Each character identity was set to a unique classification. In the graph propagation, we set K to 3. We adopted the cross-entropy loss function and the triplet loss function to jointly supervise BC-Net for end-to-end training and used the Adam optimizer to train it with 350 epochs. The learning rate started at 0.0004 and declined by 0.1 at 75, 150, 225, and 300 epochs. We concatenated the node aggregation feature S and the edge aggregation feature L as the image corresponding features extracted from the model. BC-Net extracts the features of a given query image and compares them with the features of other images in the candidate set. Thus, we used the Euclidean distance as the similarity measure to sequentially sort the images. If an image has high similarity with the identity of the query image, it is ranked near the top of the list; otherwise, it is ranked near the bottom of the list. If most of the matching images at the top of the list are the correct identity, this shows that the re-identification performance of the network is high. We used two metrics to compare the performance of the network with that of other state-of-the-art methods: cumulative matching characteristics (CMC) and mean average precision (mAP).

4.2 Model Component Analysis

We separately examined the contributions of the main components of the network by experiments on the Market-1501 dataset, using mAP and Rank-1 as the evaluation criteria. The results are shown in Table 2.

Table 2. Ablation study on main components

	Settings	Market1501	
		mAP	Rank-1
1	BC-Net	**89.4**	**95.4**
2	BC-Net (3 nodes)	83.1	92.9
3	BC-Net (5 nodes)	89.5	96.1
4	BC-Net (no propagation)	80.4	89.2
5	BC-Net (no relevance)	87.5	93.0
6	BC-Net (sum)	88.0	94.6
7	BC-Net (OSNet)	88.2	95.3

We changed the number of body part nodes to three (settings 2) and five (settings 3). The mAP of settings 2, comprising three body part nodes (head, body, and limbs), decreased by 6.3% compared to that of settings 1. The decrease in the number of nodes and edges caused a reduction in the ability of the network to extract correlation feature information for body parts. When the number of body part nodes was five (head, chest, abdomen, arms, and legs), the edges increased to 26 and the parameters of the graph operation part increased by nearly three times. However, the mAP of settings 3 only increased by 0.1%. Considering the computational cost and performance improvement, we recommend four body part nodes (settings 1) are used.

To verify the role of information transfer between nodes and edges, we removed the propagation between nodes and edges (settings 4) to make them independent. The mAP of settings 4 was significantly (9%) less than that of settings 1. Thus, the network could not extract rich feature information from nodes and edges because of the lack of information transmission, resulting in a decline in network-identification performance.

Next, we eliminated the correlation calculation between nodes and edges (settings 5) and calculated S'^{k}_{i} and L'^{k}_{j} separately in a cumulative manner. The mAP for settings 5 decreased by 4.2%. Correlation calculation can connect the feature information between nodes and edges and transfer it between nodes and edges according to its degree of importance, thus enabling critical feature information to be more highly utilized. Therefore, correlation calculation between nodes and edges improves the recognition performance of the network.

We changed the weighted summation in the aggregation operation to accumulation (settings 6), and the mAP of settings 6 reduced by 1.9% compared with that of settings 1. The accumulation did not make the critical feature information in nodes and edges stand out, indicating that our aggregation operation with the attention mechanism was effective.

We replaced the backbone network with OSNet (settings 7) [5], a lightweight network with small model size. The mAP of OSNet increased from 84.9% to 88.2%, demonstrating the versatility and advancement of our network architecture.

4.3 Comparisons with State-of-the-Art Approaches

Table 3 shows an experimental comparison results obtained with our method and state-of-the-art models on the Market1501 and DukeMTMC-reID datasets. We used mAP, Rank-1, Rank-5, and Rank-10 as evaluation indicators for person Re-ID. The state-of-the-art models used for comparison were LightMBN [44], FPB [3], CDNet [45], L3DS [46], and PAT [22]. We also used GCN-based models, such as GPS [9], SGGNN [7], PAAN [8], and the APDR [29] model with body parts, to enhance re-identifications accuracy. In the experiment, we observed that different GPUs had different recognition performances with trained models because of their varying computing capabilities. We therefore trained all models from scratch on the same GPU according to their requirements (retaining the initial pre-training), to objectively and fairly compare their recognition performance.

Compared with GCN-based models (GPS, SGGNN, and PAAN), our method achieved a more than 4.8% mAP performance improvement on the two datasets. In the graph propagation, we consider the correlation between nodes and edges, and thus the transmission of information between nodes and edges is introduced to ensure full utilization of the critical feature information. Unlike the GPS model, we did not use manually marked attribute information but introduced the correlation information between body parts into the network. This makes the feature information extracted from the network more comprehensive and detailed than it is without such an approach. Unlike the APDR model using body parts, which applies detectors or horizontal image segmentation methods to segment body parts, we used masks to accurately segment body parts. In this way, we significantly reduced

the error and difficulty of extracting body parts from the network, thus enabling our method to improve the mAP by 12.7% (compared with APDR model) on the DukeMTMC dataset. Compared with the FPB model, the mAP score of BC-NET on the Market-1501 dataset increased by only 0.3% but the mAP score of BC-NET on the DukeMTMC dataset increased by 4.7% (from 72.4% to 77.1%). This is because BC-NET exploits the correlation between body parts to enhance the feature extraction of people in the network. Compared with state-of-the-art techniques, our method achieved the best recognition performance, thus demonstrating the benefit of using body parts to construct graphs in the person Re-ID task.

Table 3. Comparison results of models trained on the same equipment

Methods	Market1501				DukeMTMC			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
LightMBN	88.4	94.1	96.5	97.9	72.6	86.7	91.9	94.6
FPB	89.1	95.3	95.9	98.2	72.4	84.7	89.3	93.5
CDNet	84.1	93.6	95.7	96.2	68.5	81.2	83.4	86.1
L3DS	85.3	94.0	95.8	97.6	68.0	81.4	84.2	86.8
PAT	87.9	93.9	95.2	96.1	70.2	83.1	86.7	89.6
GPS	84.6	92.8	94.0	95.8	70.9	84.0	87.9	90.9
SGGNN	81.4	90.5	91.9	93.3	63.2	77.3	79.9	82.0
PAAN	76.9	89.2	90.9	92.0	62.7	79.0	80.5	81.1
APDR	80.2	88.0	89.5	92.6	64.4	78.6	80.1	84.8
BC-Net	**89.4**	**95.4**	**97.6**	**99.3**	**77.1**	**89.0**	**93.8**	**95.4**

To demonstrate the effectiveness of BC-Net more intuitively, we show the activation images of two sets of query images in Fig. 4. It can be seen that for the same input image, BC-Net effectively extracted the distinguishing feature information of people. BC-Net was able to extract the primary feature information of the hands (pink handbag) and legs (white shoes) from the first input image, despite its being blurred image quality. However, the baseline method without BC-Net incorrectly extracted information from the background (the white hat of others in the background) of the first input image. This shows that because BC-NET uses masks to segment the human body, it can extract the feature information of people without being affected by background information and other noises. BC-Net also effectively extracted the feature information of multiple parts (hair, shoulders, abdomen, handbag, shoes) from the second and third input images. This shows that because BC-Net strengthens the mutual transmission ability of critical feature information and correlations between nodes (body parts) and edges (combination of body parts), it extracts more comprehensive feature information than other methods.

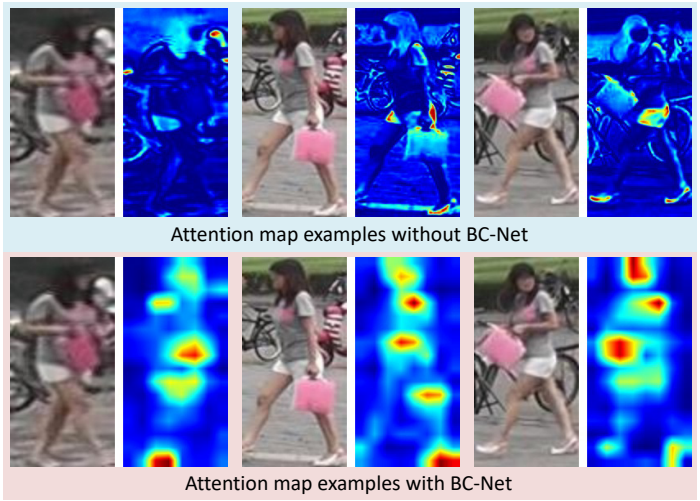

Fig. 3. Attention map of query image examples

We show examples of three sets of random retrievals from the Market-1501 dataset in Fig. 3. In the first row of Rank-4 and Rank-5 and the second row of Rank-5, BC-Net achieved a good re-identification effect even when the retrieval image is blurred. This shows that BC-Net can extract body part feature information and correlation feature information with strong distinguishability, thus improving the network's retrieval ability. Moreover, because of BC-Net's use of correlation feature information between body parts, it achieved correct retrieval results in the case of a partial occlusion, such as Rank-4 in the second row and Rank-5 in the third row. However, in the third row of Rank-4, it made an incorrect retrieval, the arm part of the detection image is completely occluded and thus its features and correlations could not be correctly captured. Overall, our experimental results show that BC-Net successfully constructs body part feature information and correlation feature information, and its overall performance is very competitive with that of state-of-the-methods.

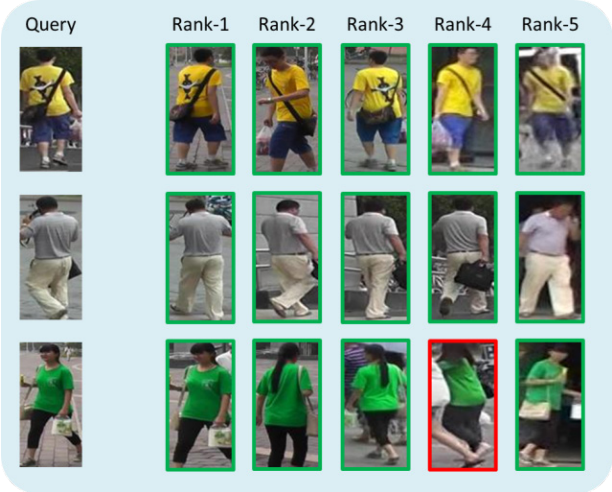

Fig. 4. Examples of three sets of random retrievals on the Market-1501 dataset

5 Conclusion

In this paper, we introduce BC-Net, a model that captures correlation information between body parts to solve the problem of image-based person Re-ID. The well-designed graph structure of the model means that it explicitly utilizes the feature and correlation information of human body parts, and thus demonstrates better accuracy than other approaches in target searching in person Re-ID tasks. We develop a unique propagation mechanism to enhance the mutual transmission of critical feature information and correlations between nodes and edges during graph operation. We also develop an attention mechanism to enhance the final graph-level feature representation for aggregation operations of nodes and edges. We demonstrate the state-of-the-art performance of BC-Net by conducting comparative experiments on two benchmark datasets. In future work, we will introduce an attribute information generator for human body parts to ensure that the human attribute information does not depend on manual annotation, which will enable the overall framework to extract more detailed human information.

References

[1] W. Chen, Y. Lu, H. Ma, Q. Chen, X. Wu, P. Wu, Self-attention mechanism in person re-identification models, Multimedia Tools and Applications 81.4(2022) 4649-4667.

[2] Y.-S. Li, Research on person re-identification method based on monitoring, [dissertation] Chengdu: University of Electronic Science and Technology of China, 2020.

[3] S. Zhang, Z. Yin, X. Wu, K. Wang, Q. Zhou, B. Kang, FPB: Feature Pyramid Branch for Person Re-Identification, <https://arxiv.org/abs/2108.01901>, 2021 (accessed 24.03.22).

[4] H. Cai, Z. Wang, J. Cheng, Multi-scale body-part mask guided attention for person re-identification, in: Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[5] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Omni-scale feature learning for person re-identification, in: Proc. of the IEEE/CVF International Conference on Computer Vision, 2019.

[6] Y. Rao, G. Chen, J. Lu, J. Zhou, Counterfactual attention learning for fine-grained visual categorization and re-identification, in: Proc. of the IEEE/CVF International Conference on Computer Vision, 2021.

[7] Y. Shen, H. Li, S. Yi, D. Chen, X. Wang, Person re-identification with deep similarity-guided graph neural network, in: Proc. of the European conference on computer vision (ECCV), 2018.
[8] Y. Zhang, X. Gu, J. Tang, K. Cheng, S. Tan, Part-based attribute-aware network for person re-identification, IEEE Access 7(2019) 53585-53595.
[9] B.-X. Nguyen, B.-D. Nguyen, T. Do, E. Tjiputra, Q.-D. Tran, A. Nguyen, Graph-based person signature for person re-identifications, in: Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[10] X. Chen, X. Liu, W. Liu, X.-P. Zhang, Y. Zhang, T. Mei, Explainable Person Re-Identification with Attribute-guided Metric Distillation, in: Proc. of the IEEE/CVF International Conference on Computer Vision, 2021.
[11] J. Liu, Z.-J. Zha, W. Wu, K. Zheng, Q. Sun, Spatial-temporal correlation and topology learning for person re-identification in videos, in: Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[12] Y. Bai, J. Jiao, W. Ce, J. Liu, Y. Lou, X. Feng, L.-Y. Duan, Person30k: A dual-meta generalization network for person re-identification, in: Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[13] J. Yang, X. Shen, X. Tian, H. Li, J. Huang, X.-S. Hua, Local convolutional neural networks for person re-identification, in: Proc. of the 26th ACM international conference on Multimedia, 2018.
[14] Z. Zhang, C. Lan, W. Zeng, Z. Chen, Densely semantically aligned person re-identification, in: Proc. of the IEEE/CVF conference on computer vision and pattern recognition, 2019.
[15] K. Han, J. Guo, C. Zhang, M. Zhu, Attribute-aware attention model for fine-grained representation learning, in: Proc. of the 26th ACM international conference on Multimedia, 2018.
[16] J. Guo, Y. Yuan, L. Huang, C. Zhang, J.-G. Yao, K. Han, Beyond human parts: Dual part-aligned representations for person re-identification, in: Proc. of the IEEE/CVF International Conference on Computer Vision, 2019.
[17] B. Chen, W. Deng, J. Hu, Mixed high-order attention network for person re-identification, in: Proc. of the IEEE/CVF international conference on computer vision, 2019.
[18] Z. Zhang, C. Lan, W. Zeng, X. Jin, Z. Chen, Relation-aware global attention for person re-identification, in: Proc. of the ieee/cvf conference on computer vision and pattern recognition, 2020.
[19] X. Chen, C. Fu, Y. Zhao, F. Zheng, J. Song, R. Ji, Y. Yang, Salience-guided cascaded suppression network for person re-identification, in: Proc. of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
[20] Z. Zhang, C. Lan, W. Zeng, Z. Chen, Multi-granularity reference-aided attentive feature aggregation for video-based person re-identification, in: Proc. of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
[21] G. Chen, T. Gu, J. Lu, J.-A. Bao, J. Zhou, Person re-identification via attention pyramid, IEEE Transactions on Image Processing 30(2021) 7663-7676.
[22] Y. Li, J. He, T. Zhang, X. Liu, Y. Zhang, F. Wu, Diverse part discovery: Occluded person re-identification with part-aware transformer, in: Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[23] M.-M. Kalayeh, E. Basaran, M. Gökmen, M.-E. Kamasak, M. Shah, Human semantic parsing for person re-identification, in: Proc. of the IEEE conference on computer vision and pattern recognition, 2018.
[24] J. Miao, Y. Wu, P. Liu, Y. Ding, Y. Yang, Pose-guided feature alignment for occluded person re-identification, in: Proc. of the IEEE/CVF international conference on computer vision, 2019.
[25] G.-A. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, J. Sun, High-order information matters: Learning relation and topology for occluded person re-identification, in: Proc. of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
[26] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, D. Katabi, Learning longterm representations for person re-identification using radio signals, in: Proc. of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
[27] J. Yang, J. Zhang, F. Yu, X. Jiang, M. Zhang, X. Sun, W.-S. Zheng, Learning to know where to see: a visibility-aware approach for occluded person re-identification, in: Proc. of the IEEE/CVF International Conference on Computer Vision, 2021.
[28] Y. Huang, Q. Wu, J. Xu, Y. Zhong, Z. Zhang, Clothing status awareness for long-term person re-identification, in: Proc. of the IEEE/CVF International Conference on Computer Vision, 2021.
[29] S. Li, H. Yu, R. Hu, Attributes-aided part detection and refinement for person re-identification, Pattern Recognition 97(2020) 107016.
[30] H. Rao, S. Xu, X. Hu, J. Cheng, B. Hu, Multi-Level Graph Encoding with Structural-Collaborative Relation Learning for Skeleton-Based Person Re-Identification, arXiv preprint arXiv:2106.03069, 2021.
[31] D. Chen, D. Xu, H. Li, N. Sebe, X. Wang, Group consistent similarity learning via deep CRF for person re-identification, in: Proc. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.
[32] J. Yang, W.-S. Zheng, Q. Yang, Y.-C. Chen, Q. Tian, Spatial-temporal graph convolutional network for video-based person re-identification, in: Proc. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
[33] Y. Yan, J. Qin, J. Chen, L. Liu, F. Zhu, Y. Tai, L. Shao, Learning multi-granular hypergraphs for video-based person re-identification, in: Proc. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
[34] D. Chen, A. Doering, S. Zhang, J. Yang, J. Gall, B. Schiele, Keypoint message passing for video-based person re-identification, in: Proc. Proceedings of the AAAI Conference on Artificial Intelligence, 2022.
[35] X. Liu, S. Zhang, Graph consistency based mean-teaching for unsupervised domain adaptive person re-identification, arXiv preprint arXiv:2105.04776, 2021.

[36] H. Ji, L. Wang, S. Zhou, W. Tang, N. Zheng, G. Hua, Meta pairwise relationship distillation for unsupervised person re-identification, in: Proc. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.
[37] Z. Bai, Z. Wang, J. Wang, D. Hu, E. Ding, Unsupervised multi-source domain adaptation for person re-identification, in: Proc. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[38] P. Li, Y. Xu, Y. Wei, Y. Yang, Self-correction for human parsing, IEEE Transactions on Pattern Analysis and Machine Intelligence 44(6)(2020) 3260-3271.
[39] X. Lu, W. Wang, C. Ma, J. Shen, L. Shao, F. Porikli, See more, know more: Unsupervised video object segmentation with co-attention siamese networks, in: Proc. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019.
[40] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: Proc. Proceedings of the IEEE international conference on computer vision, 2015.
[41] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: Proc. European conference on computer vision, 2016.
[42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
[43] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proc. 2009 IEEE conference on computer vision and pattern recognition, 2009.
[44] F. Herzog, X. Ji, T. Teepe, S. Hörmann, J. Gilg, G. Rigoll, Lightweight multi-branch network for person re-identification, in: Proc. 2021 IEEE International Conference on Image Processing (ICIP), 2021.
[45] H. Li, G. Wu, W.-S. Zheng, Combined depth space based architecture search for person re-identification, in: Proc. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[46] J. Chen, X. Jiang, F. Wang, J. Zhang, F. Zheng, X. Sun, W.-S. Zheng, Learning 3d shape feature for texture-insensitive person re-identification, in: Proc. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.