# Part-Based Bilinear CNN For Person Re-Identification

Li LI, Jianwu Dang, Yangping Wang, Song Wang, Zhenhai Zhang

School of Electronics and Information Engineering Lanzhou Jiaotong University, Lanzhou, Gansu Province China Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphic & Image Processing, Lanzhou China

Gansu Provincial Key Lab of System Dynamics and Reliability of Rail Transport Equipment, Lanzhou China

Lanzhou Bocai Technology Co.,Ltd

E-mail:lili-94@qq.com Tel:+86-13571806490

Abstract— Aiming at the problems of image misalignment and

the weak discriminative feature of Person Re-Identification(Re-ID), based on the fine-grained network bilinear CNN, a multipart ReID network is proposed. The branch network is used to learn the part features to reduce the influence of the misalignment problem of the datasets image on the ReID effect, and the compact bilinear pooling(CPB)s used for the fusion of each part of the branch network to generate discriminative feature. Weighted values of block feature and global feature loss are used to optimize the network .The validity of the proposed network structure is verified on the dataset CUHK03 and Market-1501. The results show that the proposed model has higher average recognition accuracy than traditional algorithms and other similar network models.

#### I. INTRODUCTION

In recent years, the state has promoted "Safe City", "Skynet Project" and "Snow Bright Project" to build multi-level video surveillance and network applications. Person Re-Identification(ReID) has been widely used in large-scale image retrieval and video analysis. Person re-identification aims to correctly match images of the same person taken from nonoverlapping cameras or one single camera across time However, it is still a challenging task, because varying illumination, viewpoints, poses and occlusions of body can make two images of different persons more similar than the same person the person re-identification includes two aspects: extraction of discriminative features from person images, and design of the distance metric for comparing these features. To handle the imaging variations caused by different viewpoints, poses and illumination, many robust descriptors have been developed, including color histograms[1,2,3],local binary patterns[4,5] etc. After obtaining pedestrian features, similarity measure learning is needed to map pedestrian features to metric spaces that are closer to the same person but farther from different people. There are many metric learning approaches in the literature, such as Mahalanobis Metric Learning (KISSME)[6] Local Fisher Discriminant Analysis (LFDA) [7], Marginal Fisher Analysis (MFA) [7], Large Mar-

gin Nearest Neighbor (LMNN) [7], , Locally Adaptive Decision Functions (LADF) [8]and attribute consistent matching [9], etc. Deep learning has also been applied to person Re-Identification, in view of its great success in various computer vision and pattern recognition tasks. Reference [10] and [11] first use in-depth learning to deal with pedestrian recognition. In order to deal with the problem of insufficient training data in Person re-identification, image pairs or triples are usually used to calculate the loss. As in [14] method, each input image is divided into three overlapping parts, which are connected through a part-specific convolution network, and the output is connected to form the final representation. Comparing the Representation of Two Images Using Cosine Similarity Measure. Ahmed [12] improve their deep learning architecture by computing cross-input neighborhood differences. Varior [13] use the long short-term memory (LSTM) to learn the dependence between local regions, aiming to fully exploit the spatial relationships. Ustinova introduce Bilinear CNN [14] to learn discriminative descriptors and histogram loss [15] to train deep network respectively.

## II. RELATED WORK

#### A. Bilinear CNN

Bilinear CNN was first proposed by [16] and achieved good results on several fine-grained recognition and face recognition datasets. The structure diagram is shown in Fig 1.It extracts features from two CNNs, which are multiplied by the outer product at each spatial location, as shown in (1). The merged vectors are regularized by sinusoidal square operation and L2, as shown in (2) (3). The next step is to classify the whole connection layer as the traditional convolutional neural network. Two separate paths correspond to components and texture detectors. Bilinear operations can simulate local pairwise feature interactions. This part explains the effectiveness of bilinear CNN. Bilinear operations can be abstracted into the following formulas:

$$Bilinear(s, \boldsymbol{I}, \boldsymbol{f}_A, \boldsymbol{f}_B) = \boldsymbol{f}_A(s, \boldsymbol{I})^T \boldsymbol{f}_B(s, \boldsymbol{I})$$
(1)





where s is the location of the region in the graph. I is person image.  $f_A$  and  $f_B$  are feature extractor.

$$y = \sin(x)\sqrt{|x|}$$
 (2)  $z = \frac{y}{\|y\|_{z}}$  (1)

Since the overall architecture is directed acyclic graph, the parameters can be trained by the gradient of back propagation classification loss. The bilinear form simplifies the gradient of the convergent layer. If the outputs of the two networks are A and B the sizes are L \* M and L \* N, then the size of fusion feature  $x = A^T B$  is M \* N. dl/dx is the gradient of loss function to x. Then according to the chain rule of gradi-

ent, the gradients of A and B are  $\frac{dl}{dA} = B(\frac{dl}{dx})^T$ 

and  $\frac{dl}{dB} = A(\frac{dl}{dx})$ , respectively. The gradients of the classi-

fication and normalization layers are easy to calculate, and the gradients of the layers below the assembly layer can be calculated using chain rules. The calculation of the gradient is shown in Fig 2.

#### B. Compact Bilinear Pooling

Bilinear features usually have high dimensions, which will cost a lot of computation and storage, but also affect the speed of operation.Ref [16] proposes two compact bilinear pooling methods to reduce feature dimensions.

The bilinear descriptors obtained by comparing each image descriptor can be regarded as second-order polynomial kernels, which are executed differently in linear classifiers. In the method of [17], the polynomial kernel approximates to two low-dimensional methods: Random Maclaurin (RM) [18] and Tensor Sketch (TS) [19]. If the output dimension is not very low, TS is better than RM, so TS is used as an approximation method in this paper. The flow chart of the algorithm is as follows:

input:  $x \in R^c$  represents the extracted features

output:  $f(x) \in \mathbb{R}^d$  represents the mapped feature and satisfies  $d \ll c^2$ .

1. Generating Random Uniform Distributions  $h_k \in N^c$  and  $s_k \in N^c$ 

where  $s_k(i) \in \{+1, -1\}$   $h_k(i) \in \{1, 2, ..., d\}$ , k=1,2. Sketch function is defined as

2. Sketch function is defined as  

$$Cx(x,h,s) = \{(Cx)_{1},...,(Cx)_{d}\}$$
where  $(Cx)_{j} = \sum_{i:h(j)=j} s(j)x_{j}$ 

$$f(x) = FFT^{-1} (FFT (Cx(x,h_{1},s_{1}))^{\circ}$$

$$FFT (Cx(x,h_{2},s_{2})))$$

where  $\circ$  denotes the multiplication of the corresponding elements.



Fig 2 Computing gradients in the bilinear CNN model.

### III. THE PROPOSED METHOD

Person Re-Identification and fine-grained classification have many similarities, for example, different people may look very similar in appearance, which requires a more de-





tailed part to distinguish whether the same person or not. At present, there are some ideas on fine-grained classification for Person Re-Identification. Reference [20] divides a pedestrian picture into three parts, extracts features by three convolutional neural networks, and fuses features of each part by bilinear pooling. Reference [21] adds two branch networks on the basis of bilinear CNN, and takes the characteristics of branches and backbone networks as the general features.

Reference [20] divides pedestrian images into three parts directly. Because pedestrians in the widely used public data set are detected by DPM algorithm, there will be some problems of pedestrian misalignment. Directly dividing pedestrian images may cause some parts without pedestrian information, resulting in the final fusion feature containing too much background information.

Reference [21] uses branch network, but only splices the features of branch network and backbone network. There is no fusion processing for the characteristics of each branch network.

The difference between proposed method and the above method is that the image features are extracted first, and then the existing features are divided into two parts by using branch network, so as to minimize the impact of pedestrian misalignment on the results. The features of each part learnt by the two branch networks are fused by using compact bilinear pooling to generate discriminant part features. Then the features of multiple branch networks and backbone networks are classified separately, and multiple losses are used to guide the training of the network. In this way, the network can learn the characteristics of pedestrian pictures in different locations, so that the integrated pedestrian features can contain more information about pedestrians, and have more discriminant.

## A. Description Of The Network Structure

The structure of the network is summarized in Fig 3. It consists of bilinear convolution neural network and two parts of the network. The backbone Network extracts the overall characteristics of pedestrians, and the branch network extracts more fine-grained features of each part of the body. The two network feature vectors are fused by compact bilinear pooling to generate more discriminant overall feature and block feature. The convolution layer is used to reduce the dimension of





all the above features, and then the soft Max layer is used to classify them. In the first convolution layer, we pass RGB image of size 60 \* 160 \* 3 through 32 learned filters of size 7 \* 7. The second convolution layer consists of 64 filters with a size of 5\*5. Behind them are max pool layers that halve feature sizes. The third convolution layer is also composed of 64 filters of 3 \* 3 size. The steps of the three convolutions are 2 \*2, 1 \*1 and 1\*1, respectively. The activation function used after each convolution layer is ReLU. Finally, 64 feature vectors with a depth of 20 \* 8 are obtained from each branch. The two features are fused in the depth direction by using compact bilinear pooling to obtain 20 \*8 \*512 feature vectors. Then the adaptive global average pooling is used to downsample the vectors to get the features of 1 \* 1 \* 512. Finally, the convolution of 1 \* 1 and the softmax layer are used to classify the vectors.

#### B. Description Of Part Net

Persson re-identification is usually processed partially or neighbourly to preserve more spatial information. Inspired by this, this paper uses an additional part of the network to represent the learning part and uses some features to classify to promote the network to learn more fine features. The feature mapping of the first maximum pooling layer is used as input. Part of the network architecture is shown in Fig 4. The first



Fig 5 The influence of the number of partition on the result

largest pool layer has a feature graph size of 40 \* 15, which is cut into k (k = 5) equal non-overlapping parts. Each part is passed to a subnet. Each subnet consists of two convolution layers. In order to achieve more fine functionality, some features of the two branch networks are fused using compact bilinear pooling. After dimensionality reduction using adaptive global average pooling and 1 \*1 convolution, classification operations are performed using soft Max layer. The channels of the convolution layer are 64. The filter size is 3 \*3 and the step size is 1 \*1. And padding is set to default. After each convolution layer, ReLU is used as the activation function.

The feature segmentation of the branch network on the specific pedestrian is shown in Fig 7. The feature extracted from the branch network is divided horizontally into five parts, which are represented by different colors, corresponding to the slicing operation in the process of Fig 4. Then all the features of a branch are obtained by convolution, ReLU and other operations. After fusing the same color features of the two branches, the par features for classification are obtained.

#### IV. EXPERIMENT Datasets And Evaluation Protocols

А. In this paper, experiments are carried out on the mainstream Person Re-Identification data sets CUHK03 [22] and Market-1501 [27]. CUHK03 is a Person Re-Identification data set published by the Chinese University of Hong Kong. The data sets are divided into two subsets: "labeled" and "detected". In the "detected" part, the boundary of all pedestrian images is generated by DPM; in the "labeled" part, In the labeled section, the boundary of the pedestrian image is all labelled manually. This experiment presents the results of single-shot query. In this paper, an image is randomly selected for each person from the test set. This paper uses the latest data set containing 1467 identities and randomly divides it into training set (1267) and verification set (100) test set (100). Results Average values were obtained in five random groups. For each split, the results are averaged on 100 sets of random query libraries.



Fig 6 The influence of the proportion of backbone loss to total loss on the results



Fig 7 Visualization of Branch Network Extraction Features

Market-1501 was collected on the campus of Tsinghua University. It collected 1501 pedestrian pictures from six cameras. Each pedestrian is captured by at least two cameras, and there may be multiple images under one camera. There are 751 people in the training set, including 12,936 images, with an average of 17.2 training data per person; 750 people in the test set, including 19,732 images, with an average of 26.3 test data per person.

In this paper, Cumulative Match Characteristic curve(CMC)and Mean Average Precision (mAP) are used to evaluate the performance of pedestrian recognition methods on data sets. CMC regards Person Re-Identification as a sort problem. The probability of successful first matching in the database is expressed by rank1, that is, the probability that the first image in the sort table is the correct result. The average value is obtained through experiments



Fig 8 The result on Market-1501

many times. mAP is the average of AP, which is the evaluation criterion when Person Re-Identification is regarded as a problem of image retrieval. The formulas of AP and mAP are as follows.

$$AP = \frac{\sum_{k}^{n} (P(k) \times B(k))}{N}$$
(4)

where k is the serial number of the search image, P (k) is the proportion of the image related to the retrieved image. When the retrieved image is the correlation image of the first k image, the value of B (k) is 1, otherwise it is 0.N is the number of revelant images.

mAP=
$$\frac{\sum_{q=1}^{Q} AP(q)}{Q}$$
(5)

where Q is the number of queries

B. Implementation Details

In order to verify the effectiveness of the proposed method, two experimental groups with slight changes in network model are designed. One is to use the backbone network instead of the branch network structure. As shown in Fig 3, the branch network and all the layers behind it are removed directly. The second is to combine the features of branch networks directly with those of backbone networks, instead of using a compact bilinear pool to fuse and train with multiple classifiers. The network layer and pooling layer of the two experimental groups are consistent with those used in this paper. All images were adjusted to 160\*160 and batch size to 256. All images in each epoch were reset in random order according to the label. The data of this experiment did not do any data enhancement (i.e., mirror, flip, etc.). The initial learning rate is  $1*10^{-4}$ , which is reduced to  $1*10^{-5}$  after 10,000 iterations. The whole training lasts 20,000 iterations. This paper uses Adam to optimize the training. In this paper, the branch network divides the feature into five parts. The loss function uses cross-entropy loss. The total loss of the whole network is weighted by the weight of 0.6 and 0.4 respectively.

## C. Parameter Analysis

In this experiment, the number of pedestrian feature segmentation K and the weight loss of backbone network lambda are tested through the verification set, and good re-

sults are selected to run on the test set. In order to control the number of variables, the loss weight of backbone network is 0.5 when analyzing the effect of the change of K on the results. When the number of partitions is 1, it is equivalent to removing the branch network and merging the backbone network with a compact bilinear pool. Different number of segmentation will fine-tune the middle layer network to ensure the matching of feature dimensions. Then, the optimal k-value is used to analyze the influence of the weight change of the backbone feature loss on the results. The results are illustrated in Fig. 5 and Fig. 6 respectively. It can be seen that the network performance is best when the number of partitions is 5. This is because when the number of segmentation is too small, the features learned contain a lot of background information. When the number of segmentation is too large, it pays too much attention to the local feature, which divides some discriminant features into several local features, and loses the discriminability, which leads to the performance degradation. When the number of partitions is 5, the weights of the sum of backbone loss and partial loss are 0.4 and 0.6 respectively, the network performance is the best. This shows that the feature extracted by the branch is more discriminant than the feature extracted directly, so the weight is slightly larger. In this paper, only between weights is considered. The results were compared when the interval was 0.1, the lower limit was 0.2and the upper limit was 0.8.

D. Experimental Results And Analysis TABLE I RESULTS ON CULIVO2 LABELED

RESULTS ON CUHRO3 LABELED				
Method	Rank1(%)	Rank5(%)	Rank10(%)	
LOMO+XQDA[1]	52.20	85.80	92.14	
Ahmed[12]	54.74	86.30	93.88	
Ensembles[23]	62.1	89.1	94.3	
FusedModels[24]	72.43		95.51	
End-to-end	65.65	91.28	96.29	
CAN[25]				
MR-B-CNN[20]	65.04	91.60	95.84	
Ustinova[15]	65.77	92.85	97.62	
MCBC[21]	71.62	94.86	98.12	
Backbone and	69.51	92.36	96.14	
part concatenated				
Backbone CPB	62.42	87.92	94.59	
Ours	73.58	95.89	98.10	

TABLE II

RESULTS	ON CUHK	03 DETECTI	ED	
Method	Rank1(%)	Rank5(%)	Rank10(%)	
LOMO+XQDA[1]	46.00	82.50	88.55	
Ahmed[12]	44.96	76.40	83.47	
Fused Models[24]	72.04		96.00	
Varior[26]	68.1	88.1	94.6	
End-to-endGAN[25]	63.05	82.94	88.17	
MR-B-CNN[20]	60.22	88.10	94.22	
MCBC[21]	64.69	90.28	95.2	
Backbone and part con-	63.03	89.10	94.13	
catenated				
Backbone CPB	57.47	87.48	92.48	
Ours	66.81	91.32	95.87	
Table III				

RESULTS	ON CUHK03 MAI	RKET-1501

Method	Rank1(%)	Rank5(%)	Rank10(%)	mAP
End-to-end CAN[25]	48.24	-	-	24.43
DiscrNullSpace[28]	61.02	-	-	35.68
siamLstm[13]	61.60	-	-	35.31

Varior[26]	65.88	-	-	39.55
MCBC[21]	66.36	85.01	90.17	41.17
Backbone and part concatenated	72.42	89.21	93.45	57.32
Backbone CPB	70.64	87.43	93.10	56.95
Ours	74.37	90.13	94.73	59.16

The results of MCBC and the proposed network are shown in Fig 8, where (a) and (c) are the results of the proposed network and (b) and (d) are the results of MCBC. The green box represents the correct re-identification, the yellow box represents misaligned people, and the red box represents the wrong re-identification. It can be seen that the identification ability of (a) and (c) is significantly better than that of (b) and (d).

Tables I and II are the experimental results of the proposed method on CUHK03 dataset. In this paper, MCBC with similar network structure is used as the experimental benchmark for comparison. In labeled subset, the Rank 1 values of MCBC and Experiment 1 are 71.62% and 69.51%, respectively. This is because the backbone network of MCBC uses compact bilinear pooling to fuse, while Experiment 1 uses feature vector splicing directly, so the effect is worse. The Rank1 values of Experiment 1 and this method are 69.51% and 73.58% respectively, which shows that the result of using compact bilinear pooling is 4.07% higher than that of using direct splicing in the proposed network structure. The rank 1 values of Experiment 2 and this method are 62.42% and 73.58% respectively, which shows that after using branch network, each branch learns the characteristics of different parts of pedestrians separately and effectively improves the recognition rate. The trend of results on detected subset is basically consistent with that of analysis and labeled.

Table III is the result of experiments on Market-1501 dataset using the proposed method. In this data set, two control experimental groups like CUHK03 are also set up. With the increase of branch CPB, the accuracy of CPB is improved by 3.73% compared with that of only using backbone network. It can be seen clearly that after using branch network for block processing, the performance of re-recognition is improved to reduce the impact of misalignment on matching results. The accuracy of branch and trunk is improved by 1.95% when CPB is directly connected with trunk and block features. This shows that under the framework of bilinear CNN, the fusion effect of block feature is better than that of stitching. The experimental results are superior to the traditional algorithm and the same type of shallow convolution neural network algorithm.

#### V. CONCLUSION

Based on bilinear convolution neural network, a pedestrian recognition network based on body and block features is proposed in this paper. Firstly, on the basis of multi-branch network, dimensionality reduction and classification of each block's features are carried out, and different classifiers are used to learn the features to reduce the impact of pedestrian image misalignment on the recognition rate. Secondly, in view of the problem that the dimension of partial feature fusion of bilinear branching networks is too high, the compact bilinear pooling method is adopted to obtain more discriminant feature vectors, but the dimension is much smaller than the original bilinear pooling feature vectors, which can reduce the dimension and ensure the running speed of the network. The validity of the proposed network is verified on CUHK03 and Market-1501 datasets. The experimental results show that the recognition accuracy of the proposed network is higher than that of the traditional algorithm and the same type of network.

#### ACKNOWLEDGMENT

This research was supported by the National Nature Foundation of China (Grant No. 41761082, 61162016, 61562057), the Gansu Science and Technology Project (Grant No. 18JR3RA104, 1504FKCA038), the Science and Technology Project of Gansu Education Department (Grant No. 2017D-08) and Lanzhou Talent Innovation and Entrepreneurship Project(2015-RC-86). all support is gratefully acknowledged.

#### REFERENCE

- Liao S, Hu Y, Zhu X, et al. Person re-identification by local maximal occurrence representation and metric learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 2197-2206.
- [2] Zhao R, Ouyang W, Wang X. Person re-identification by salience matching[C]//Proceedings of the IEEE International Conference on Computer Vision. 2013: 2528-2535.
- [3] Zhao R, Ouyang W, Wang X. Unsupervised salience learning for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3586-3593.
- [4] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on pattern analysis and machine intelligence, 2002, 24(7): 971-987.
- [5] Xiong F, Gou M, Camps O, et al. Person re-identification using kernel-based metric learning methods[C]//European conference on computer vision. Springer, Cham, 2014: 1-16.
- [6] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 2288-2295.
- [7] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10(Feb): 207-244.
- [8] Li Z, Chang S, Liang F, et al. Learning locally-adaptive decision functions for person verification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3610-3617.
- [9] Khamis S, Kuo C H, Singh V K, et al. Joint learning for attribute-consistent person re-identification[C]//European Conference on Computer Vision. Springer, Cham, 2014: 134-146. [10]Yi D, Lei Z, Liao S, et al. Deep metric learning for person re-identification[C]//Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014: 34-39.
- [10] Yi D, Lei Z, Liao S, et al. Deep metric learning for person reidentification[C]//2014 22nd International Conference on Pattern Recognition. IEEE, 2014: 34-39.
- [11] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 152-159.
- [12] Ahmed E, Jones M, Marks T K. An improved deep learning

architecture for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3908-3916.

- [13] Varior R R, Shuai B, Lu J, et al. A siamese long short-term memory architecture for human re-identification[C]//European Conference on Computer Vision. Springer, Cham, 2016: 135-153.
- [14] Ustinova E, Ganin Y, Lempitsky V. Multi-region bilinear convolutional neural networks for person reidentification[C]//Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on. IEEE, 2017: 1-6.
- [15] Ustinova E, Lempitsky V. Learning deep embeddings with histogram loss[C]//Advances in Neural Information Processing Systems. 2016: 4170-4178.
- [16] Lin T Y, RoyChowdhury A, Maji S. Bilinear cnn models for fine-grained visual recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1449-1457.
- [17] Gao Y, Beijbom O, Zhang N, et al. Compact bilinear pooling[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 317-326.
- [18] Kar P, Karnick H. Random feature maps for dot product kernels[C]//Artificial Intelligence and Statistics. 2012: 583-591.
- [19] Pham N, Pagh R. Fast and scalable polynomial kernels via explicit feature maps[C]//Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013: 239-247.
- [20] Ustinova E, Ganin Y, Lempitsky V. Multi-region bilinear convolutional neural networks for person reidentific • ation[C]//2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2017: 1-6.
- [21] Liu J, Yang Z, Zhang T, et al. Multi-part compact bilinear CNN for person re-identification[C]//Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017: 2309-2313.
- [22] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 152-159.
- [23] Paisitkriangkrai S, Shen C, Van Den Hengel A. Learning to rank in person re-identification with metric ensembles[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1846-1855.
- [24] Subramaniam A, Chatterjee M, Mittal A. Deep neural networks with inexact matching for person reidentification[C]//Advances in Neural Information Processing Systems. 2016: 2667-2675.
- [25] Liu H, Feng J, Qi M, et al. End-to-end comparative attention networks for person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3492-3506.
- [26] Varior R R, Haloi M, Wang G. Gated siamese convolutional neural network architecture for human reidentification[C]//European Conference on Computer Vision. Springer, Cham, 2016: 791-808.
- [27] Zheng L, Shen L, Tian L, et al. Scalable person reidentification: A benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1116-1124.
- [28] Zhang L, Xiang T, Gong S. Learning a discriminative null space for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1239-1248.