Cross-Domain Person Re-identification Combining Feature Concatenation and Attention

Feng Pan [1.2, a], Lin Wang [1.2, b], Yansha Zhang [1.2, c], Jie Wang [1.2, d]

[1] College of Data Science and Information Engineering Guizhou Minzu University, Guizhou, China

[2] Key Laboratory of Pattern Recognition and Intelligent System of Guizhou Province, Guizhou, China

[a] 172408874@qq.com, [b] wanglin@gzmu.edu.cn, [c] 1638292011@qq.com, [d] 939284639@qq.com

Abstract. To improve the insufficient generalization and poor cross-domain capability of the existing direct cross-dataset person re-identification methods, a cross-domain person re-identification method combining feature concatenation and attention (FCANet) is proposed. The deep features of the network are concatenated to complement the feature information and obtain discriminatively feature, and the position attention module is introduced to enhance the data feature representation capability of the cross-domain task, using the joint training network of label smooth cross-entropy loss and triplet loss, model training in the source domain, and directly deploy to the target domain for testing. To verify the performance of the proposed method, it was experimented on three public datasets of Market1501, DukeMTMC-reID and MSMT17, which mAP and Rank1can reach 51.4% and 62.7% on Market1501. The results show that the proposed method has good performance in improving the generalization of cross-domain tasks, and the recognition accuracy outperforms the domain generalization algorithms of comparison.

Keywords: cross-domain person re-identification; feature concatenation; position attention module; domain generalization.

1. Introduction

Gheissari [1] first proposed the concept of Person Re-identification (Re-ID) in 2006, which plays an important role in public safety by the process of retrieving specific pedestrians under non-overlapping surveillance devices. Cross-domain Re-ID aims to make up for the visual limitations of a single, fixed camera and identify the same person under the cross-monitoring equipment, which is an important challenge of intelligence in the field of public security, can be applied to intelligent security, arrest of fugitives and other fields, which has important research significance.

With the development of deep learning, the use of deep learning technology to solve the problems in the field of Re-ID is favored by scholars. For example, Yang [2] take ResNet50 as the backbone network and introduce CBAM module to enrich the attention map by effectively combining spatial and channel attention. Fu [3] proposed a simple and effective horizontal pyramid matching (HPM) method to extract the local characteristics of person, which enhanced the local recognition ability of person. At present, Re-ID technology based on deep learning has achieved good performance on a single data set. However, due to the large deviation of data distribution, when the model trained in a certain data set is deployed to the target environment, the performance will be sharply reduced, which seriously limits the promotion and application of person re-identification technology.

Based on the problem of poor generalization of the Re-ID method in new unknown scenarios. Some scholars have proposed unsupervised domain adaptive method, the main idea is to use the clustering algorithm [4], [5] to generate pseudo-labels of the target domain image features without label information, using the cluster generated pseudo-labels on the target domain for supervised learning, or using the generated adversarial network [6] to migrate the labeled source domain data to the target domain, and then the model is trained with the generated target domain data. The unsupervised domain adaptive approach does provide performance gains compared to the traditional Re-ID approach. However, the unsuper- vised domain adaptive approach requires further learning on

the target domain, consumes a lot of resources and time when deployed, and is susceptible to background noise, limiting the application of the model in real-world scenarios.

In view of the insufficient generalization of cross-domain Re-ID, this article regards cross-domain performance decline as a domain generalization problem, directly using the Re-ID model trained by the source domain for any unknown target domain. A cross-domain person re-identification method FCANet with feature concatenation and attention mechanism is proposed to reduce the loss of feature information, pay attention to more detailed features, and make the final person feature description more discriminative. The main work of this article are as follows:

(1) The feature concatenation strategy is proposed, through which the concatenation depth features can capture more delicate feature information when the information loss is less, and can better distinguish the same person and the similar person.

(2) There is a domain offset in cross-domain evaluation, but there is an interrelated dependency between different features of the same person in the target domain, and the position attention module is introduced to selectively aggregate the characteristics of different positions and enhance the data feature representation ability of cross-domain tasks.

(3) The metric learning adopts the union of the label smooth cross-entropy loss function and the triplet loss function, and adds the re-ranking [7] algorithm when testing.

2. Related work

Person re-identification as an instance-level recognition task, in 2019, Zhou et al. designed a novel deep CNN network in Re-ID, called omni-scale network (OSNet) [8], OSNet extracts the multi-scale features of person, which can not only capture different spatial scales but also encapsulate a synergistic combination of multiple scales, using a unified aggregation gate to dynamically fuse multi-scale features. The amount of parameters in the OSNet network is only 2.2M, which is much less than the ResNet50 standard in Re-ID.

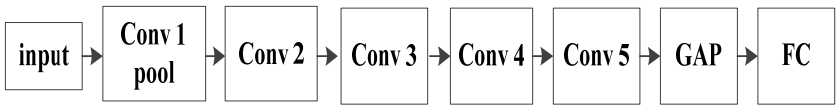

Figure 1. The OSNet network architecture

Neural architecture search (NAS) [9] aims to automate the network architecture, and the early NAS methods were usually based on reinforcement learning or evolutionary algorithms. Literature [8] aims to solve the domain gap between different Re-ID data sets in pedestrian reconstruction and cross-domain assessment. Auto-searched instance normalization (AIN) is introduced to improve the generaliz- ation performance for cross-domain recognition.

3. Cross-Domain Person Re-identification Combining Feature Concatenation and Attention

3.1 Feature Concatenation Module

There are domain gaps in cross-domain re-ID. The feature description effect of a single feature map is not ideal, and the existing method uses the fusion of global and local features as the feature description sub-characters of person, and there is a problem that the local feature are excessively divided into images to make the feature lose information. Inspired by the literature, the OSNet-AIN is selected as the backbone network for feature extraction, in order to make the proposed feature more robust, the image features of the last two convolutional layers in the extracted network are concatented together, so that the feature information is complementary and the information loss is reduced. The full connection layer in the network is replaced with the BN layer to obtain the final feature

description, thus improving the training speed of the network and improve the performance of cross-domain person re-identification tasks. Feature concatenation module is shown in Figure 2.

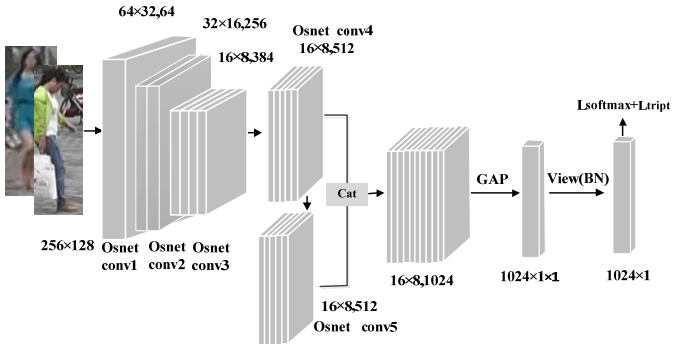

Figure 2. Feature concatenation module

3.2 Position Attention Module

Cross-domain person re-identification is affected by the factors of multi-complex scenes, and the inter-domain gap between domains makes the cross-domain recognition accuracy low. Although there is a domain gap between cross-domain person, there are interrelated dependencies between different characteristics of the same person. The Position Attention Module[10] (PAM) is introduced to obtain the dependence of any two positions in the feature map, and similar features are considered to be related to each other. The idea of the PAM module is to selectively aggregate features at each position by feature-weighted sums of all positions. A schematic diagram of the structure of the PAM module is shown in Figure 3.

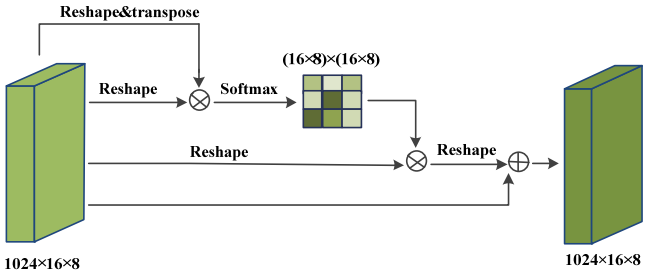

Figure 3. The Position Attention Module

The PAM module is introduced to establish a rich context relationship on the depth features after concatenation, which then enhance the representation ability of features. After concatenation, the depth feature A changes the feature dimension multiplication to obtain the correlation degree matrix between any two points, and after softmax normalization, it can be obtained that the attention map S ((16×8) × (16×8)) of each position to other positions can be obtained, where the more similar the feature, the greater the response value, the darker the color. The response values of the attention map S are then weighted to the feature D, and the points at each position fuse similar features in the global space through the attention map S. Finally, the weighted feature map and the depth feature map are added to the corresponding elements to obtain the feature map F with more characterization ability.

3.3 Loss Function

Train the network using a joint combination of a label smooth cross-entropy loss function and a triplet loss function, with expression such as equation (1). The triplet loss function with balance weight is added to make the image distance of the same pedestrian closer, and the distance of different pedestrians is as far as possible, which improves the recognition accuracy of the model.

$$L = L_{soft\max} + \alpha L_{tript}. \tag{1}$$

Where $L_{soft\max}$ representing the cross-entropy loss function with label smooth, L_{tript} representing the triplet loss function, α is 0.5.

$$L_{soft\max} = -\sum_{i=1}^{n} P(x_i)\log(Q(x_i)). \tag{2}$$

$$Q_i = \frac{\exp(z_i)}{\sum_{j=1}^{K}\exp(z_j)}. \tag{3}$$

$$p = (1-\varepsilon)y + \frac{\varepsilon}{K}. \tag{4}$$

Where $Q(x_i)$ is the confidence probability of each category, $P(x_i)$ is the probability after label smoothing, K is the number of label categories, ε is the number of arbitrary small, when the label y is true, $y=1$.

4. Experiments

4.1 Datasets and Settings

In order to verify the effectiveness of the proposed method FCANet, the method was tested on four publicly available and commonly used datasets of Market 1501, DukeMTMC-reID and MSMT17, and the results were compared with other commonly used person re-identification algorithms. The details of the four datasets are shown in Table 1. The experiment is based on the Pytorch framework, using the Torchreid library to build a network, and the operating system is Ubuntu 18.04.5 LTS (x86_64).

(1) The GPU is Tesla P100-PCIE-16GB, the memory is 16GB, and the version of Nvidia is 460.32.03.

(2) The CPU is Intel(R) Xeon(R) CPU @ 2.20GHz.

Table 1. Datasets Details Introduction

datasets	cameras	IDs (T-Q-G)	Images (T-Q-G)
Market1501	6	751-750-751	12936-3368-15913
Duke	8	702-702-1110	16522-2228-17661
MSMT17	15	1041-3060-3060	30248-11659-82161

T: Train. Q: Query. G: Gallery

4.2 Implementation Details

The pre-trained osnet_ain_x1_0 network fine-tuned on the ImageNet datasets is used as the base network for training, the FCANet network proposed are used to extract the features, the input image size is 256×128, the data enhancement method is random flip, color jitter, and the margain of the triplet loss function is set to 0.3. The number of epochs is 150, during the first 10 epochs, the ImageNet pre-trained base network is frozen, the gradient is updated using the amsgrad optimizer, the initial learning rate is 0.0015, the learning rate is decayed by 0.1 per 60 epochs, the weight decay is 5e-4, and batch match-size is set to 64.

The performance of person re-identification model was evaluated using the mean Average Precision (mAP) and Cumulative Matching Characteristics (CMC) curves. CMC refers to the probability of successful matching of the previous K images, and this paper uses the probability of successful matching of the top 1 image, which are recorded as R-1.

4.3 Performance Contrast

To evaluate the effectiveness of the proposed cross-domain Re-ID method, the proposed method is compared with the popular algorithm for Re-ID in recent years on four publicly available Re-ID datasets. There are two groups of popular algorithms for comparison, firstly, an unsupervised domain adaptive method that needs to be trained on target domain data, including: 1) Domain invariant feature representation:TJ-AIDL[11](CVPR'18), PAUL[12] (CVPR'19); 2) Style migration-based approaches: SPGAN[6](CVPR'18), CSGLP[13]; 3) Methods for forming pseudo-labels based on clustering: SAL[14], NSSA[15]. Secondly, domain generaliz- ation methods that only need to be trained on the source domain and deployed directly in the target domain, including: ECNbaseline[16](CVPR'19), QAConv[17] (ECCV'20), OSNet-AIN[8](TPAMI'21), SBS[18] (ICCV'21), PABO-QAConv[19] and other algorithms.

The performance indicator pairs of different algorithms are shown in Table 2, where A→B represent datasets A as the source domain (training set), datasets B as the target domain (test set), and M and D represent datasets Market1501 and DukeMTMC-reID, respectively.

Table 2. Performance comparison with popular algorithms

Method	M→D		D→M	
	mAP	R-1	mAP	R-1
TJ-AIDL[11]	23	44.3	26.5	58.2
2 SPGAN[6]	22.3	41.1	22.8	51.5
3 ECNbaseline[16]	14.8	28.9	17.7	43.1
4 PAUL[[12]	35.7	56.1	36.8	66.7
5 CSGLP [13]	27.1	47.8	31.5	61.2
6 NSSA [15]	45.5	65.5	**47.9**	**76.2**
SAL [14]	48.5	67.6	38.7	65.3
QAConv[17]	28.7	48.8	27.2	58.6
QAConv+re	47.8	56.9	45.8	65.7
OSNet-AIN	30.5	52.4	30.6	61
SBS[18]	26.4	45	24.5	53.1
PABO-QAConv [19]	38.6	60.4	44.1	74.9
PBPM [20]	32.1	52.3	27.6	58.2
FCANet	**36.6**	**57.1**	**31.4**	**61.4**
+reranking	**51.4**	**62.7**	**46.4**	**65.1**

Table 2 compares the FCANet with the popular cross-domain Re-ID algorithm. Obviously, the mAP and Rank1 evaluation indicators of the FCANet network are higher than the listed domain generalization methods[16],[17],[8],[18]; and achieves better performance on target datasets than some unsupervised domain adaptive methods[11],[6],[12],[13],[20]. In the evaluation of datasets M→D, the mAP and Rank1 evaluation indicators of FCANet network reached 51.4% and 62.7%, respectively, in datasets D→M, the mAP and Rank1 reached 46.4% and 65.1%, respectively. Through the comparison of results, it can be seen that after the introduction of various strategies, the mAP and Rank1 on the two public datasets exceed the compared domain generalization methods, indicating that the proposed method has better recognition effect than other domain generalization methods.

The performance comparison results with popular algorithms verify the effectiveness of the proposed method. To further verify the generalization of the method, the network models trained on the Market1501 and DukeMTMC-reID datasets are deployed to the MSMT17 datasets that is more realistically adapted to the scenario. Popular algorithms for comparison are PTGAN[21](CVPR'18) based on style migration, method for forming pseudo-labels based on clustering: ECN(CVPR'19)[16],OSNet-AIN[8] based on domain generalization. The results are shown in Table 3. MS represent the datasets MSMT17.

Table 3. Performance comparison of different algorithms on the MSMT17 datasets

Method	M→MS		D→MS	
	mAP	R-1	mAP	R-1
PTGAN[21]	2.9	10.2	3.3	11.8
ECN [16]	8.5	25.3	10.2	30.2
OSNet-AIN[8]	8.2	23.5	10.2	30.3
FCANet	**10.7**	**28.6**	**10.4**	**30.9**

As can be seen from Table 3, the FCANet network is well generalized on large datasets MSMT17. When the Market1501 and DukeMTMC-reID datasets is the source domain, the mAP and Rank1 of FCANet is better than the current popular algorithm, and reached 10.7%, 28.6 (10.4%, 30.9). Through the comparison of experimental results, it can be seen that after the introduction of various strategies, the improvement effect on the public datasets is obvious, which verifies the generalization of the proposed method.

4.4 Ablation Experiments

In order to verify the contribution of feature concatenation operation and position attention module to OSNet-AIN network, the network was trained on the datasets Market 1501, DukeMTMC-reID and MSMT17, the cross-domain datasets were evaluated during the test, and the experimental results are shown in Table 4. OSNet-AIN is the backbone network of the experiment, M, D, MS represent the datasets Market 1501, DukeMTMC-reID and MSMT17, and subsequent experiments are in the form of M, D, and MS.

Table 4. Performance comparison of different algorithms on the MSMT17 datasets

Method	M→D		D→M		M→MS		D→MS	
	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1
OSNet-AIN	30.5	52.4	30.6	61.0	8.2	23.5	10.2	30.3
+Cat	33.9	54.8	29.9	59.8	9.1	25.0	10.1	30.5
+Cat+PAM	**35.3**	**55.7**	**30.9**	**60.7**	**9.3**	**25.5**	**10.9**	**30.9**

From Table 4, it can be seen that concatenation operations and PAM strategies contribute to the network. When only contribution, in the evaluation of datasets M→D, the values of mAP and Rank1 reached 33.9% and 54.8%, respectively, and after the two strategies were carried out at the same time, the evaluation effect of each cross-domain datasets was improved, among which the evaluation of datasets D→MS was increased by 0.8% and 0.4% compared with mAP and Rank1, which are contribution only. The results show that the recognition effect of the two strategies is better, and the effectiveness of the method is verified.

Most of the existing re-ID methods use multi-loss training strategies, such as the cross-entropy loss function and the triplet loss function at the same time. To do this, the triplet loss function is added to the training strategy of the network, which adds a balance weight α to measure the triplet loss, α takes 0.5. The identification effect is shown in the Table 5.

Table 5. Comparison of different loss performance

Method	M→D		D→M		M→MS	
	mAP	R-1	mAP	R-1	mAP	R-1
FCANet α=0	35.3	55.7	30.9	60.7	9.3	25.5
FCANet α=0.5	**36.6**	**57.1**	**31.4**	**61.4**	**10.7**	**28.6**

As can be seen from Table 6, there is a difference between the effect of using both cross-entropy loss and triple-action loss and using only cross-entropy loss. The identification accuracy using two losses is better than using only one loss. On the Duke datasets, the mAP and Rank1 improved slightly,

0.5% and 0.7%, respectively, due to the fact that the scenes between the two datasets were similar, both taken on university campuses.

5. Conclusion

For the cross-domain Re-ID task, most existing methods have insufficient generalization and poor cross-domain capability. FCANet, a cross-domain Re-ID method of feature concatenation and fusion attention mechanism, enables the network to pay attention to more local details by using the concatenation feature to improve the pedestrian feature discrimination. At the same time, the position attention module is introduced to correctly look at the dependency relationship of any two positions in the feature map, reduce the loss of feature information, and make the final person feature descriptor more discriminative. Experiments on three public and commonly used datasets of Re-ID verify the effectiveness of concatenation feature strategies and attention modules, and the recognition accuracy outperforms the state-of-the-art algorithms of comparison. However, when conducting cross-domain assessment of real-world scenarios, the recognition accuracy rate needs to be improved. Therefore, the follow-up work will continue to study cross-domain Re-ID, and design a network model with stronger generalization, so as to improve the accuracy of Re-ID in real-world scenarios.

References

[1] Gheissari Niloofar, Thomas B. Sebastian, and Richard Hartley. "Person reidentification using spatiotemporal appearance," 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). Vol. 2. IEEE, 2006, pp.1528-1535, doi:10. 1109/CVPR.2006.223.

[2] Yang, Wenjie, Huang Houjing, Zhang Zhang, Chen Xiaotang, Huang Kaiqi and Zhang Shu. "Towards rich feature discovery with class activation maps augmentation for person re-identification," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, pp.1389-1398, doi:10.1109/CVPR.2019. 00148.

[3] Fu Yang, Wei Yunchao, Zhou Yuqian, et al. "Horizontal pyramid matching for person re-identification," Proceedings of the AAAI conference on artificial intelligence. Vol. 33. No. 01. 2019, pp.8295-8302, doi:10.1609/aaai.v33i01.33018295.

[4] Ge, Yixiao, Dapeng Chen, and Hongsheng Li. "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification," arXiv preprint arXiv:2001.01526 (2020). doi:10.48550/arXiv.2001.01526.

[5] Zhai Yunpeng, Lu Shijian, Ye Qixiang, et al. "AD-Cluster: Augmented Discriminative Clustering for Domain Adaptive Person Re-Identification," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2020, pp.9021-9030, doi:10.1109/CVPR42600.2020.00904.

[6] Deng Weijian, Zheng Liang, Ye Qixiang, Kang Guoliang, Yang Yi, and Jiao Jianbin."Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 994-1003, doi:10.48550/arXiv. 1711.07027.

[7] Zhong Zhun, Zheng Liang, Cao Donglin, and Li Shaozi. "Re-ranking Person Re-identification with k-reciprocal Encoding," IEEE Computer Society IEEE Computer Society, 2017, pp. 1318-1327, doi:10.48550/arXiv.1701.08398.

[8] Zhou Kaiyang, Yang Yongxin, Cavallaro Andrea, and Xiang Tao. "Omni-scale feature learning for person re-identification," Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019, pp.3702 - 3712, doi:10.1109/ICCV.2019.00380.

[9] Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning," arXiv preprint arXiv:1611.01578 (2016). doi:10.48550/arXiv.1611.01578.

[10] Fu Jun, Liu Jing, Tian Haijie, et al. "Dual attention network for scene segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, pp.3146-3154, doi:10.1109/CVPR.2019.00326.

[11] Wang Jingya, Zhu Xiatian, Gong Shaogang, and Li Wei. "Transferable joint attribute-identity deep learning for unsupervised person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE (2018), pp.2275 − 2284, doi:10.1109/CVPR.2018.00242.

[12] Yang Qize, Yu Hongxing, Wu Ancong, and ZhengWeishi. "Patch-Based Discriminative Feature Learning for Unsupervised Person Re-Identification," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3633-3642, doi:10.1109/CVPR.2019.00375.

[13] Ren Chuanxian, Liang Bohua, Ge Pengfei, Zhai Yiming, Lei Zhen. "Domain adaptive person re-identification via camera style generation and label propagation." IEEE Transactions on Information Forensics and Security 15 (2019), pp.1290-1302, doi:10.1109/TIFS.2019. 2939750.

[14] Jiang Kongzhu, Zhang Tianzhu, Zhang Yongdong, Wu Feng, Rui Yong. "Self-supervised agent learning for unsupervised cross-domain person re-identification." IEEE Transactions on Image Processing 29 (2020), p.8549-8560, doi:10.1109/TIP.2020.3016869.

[15] Zhao, Yiru, and Hongtao Lu. "Neighbor similarity and soft-label adaptation for unsupervised cross-dataset person re-identification." Neurocomputing 388 (2020), pp. 246-254, doi:10.1016/j.neucom. 2019.12.115

[16] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, Yi Yang. "Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-identification," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 598-607, doi:10.1109/CVPR.2019.00069.

[17] Liao Shengcai, and Ling Shao. "Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting," European Conference on Computer Vision. Springer, Cham, 2020, pp.456-474, doi:10.1007/978-3-030-58621-8_27.

[18] Chen Xiaodong, Liu Xinche, Liu Wu, Zhang Yongdong, Mei Tao. "Explainable Person Re-Identification with Attribute-guided Metric Distillation," Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021, pp.11813-11822, doi:10.1109/ICCV 48922.2021.01160.

[19] LI Yunlong, Cheng Deqiang, LI Jiahan, Huang Ji, Zhang Jianying, Ma Haohui. "Cross-domain Person Re-identification based on Progressive Attention and Block Occlusion," Journal of Beijing University of Aeronautics and Astronautics. 2022, pp.1-13, doi:10.13700/j.bh.1001-5965.2022.0025.

[20] Yang Ping, Wu Xiaohong, He Xiaohai, Chen Honggang, Liu Qiang. "Cross-Domain Person Re-identification Method Based on Point-by-Point Feature Matching," Pattern Recognition and Artificial Intelligence. 2022,35(06), pp.516-525.doi:10.16451/j.cnki.issn1003-6059. 202206004.

[21] Wei Longhui, Zhang Shiliang, Gao Wen, Tian Qi. "Person transfer gan to bridge domain gap for person re-identification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 79-88, doi:10.1109/CVPR.2018.00016.