ISBN: 978-981-18-3959-7
2022 the 12th International Workshop on Computer Science and Engineering (WCSE 2022)
doi: 10.18178/wcse.2022.06.028

Pose-Guided Multi-Granularity Feature Learning for Occluded Person Re-Identification

Jianhua Shu [1+] and Jingsheng Lei [1]

[1] College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, 201306, China

Abstract. Persons are often occluded in real-world applications of person re-identification. To alleviate the occlusion problem, this paper proposes a pose-guided multi-granularity feature learning method for occluded person re-identification. At first, the residual atrous spatial pyramid pooling module is used to expand the receptive field to extract more hierarchical pedestrian features. Next, the visible head-and-shoulder region feature and the underneath region feature from the heatmap extracted by the pedestrian estimation algorithm are calculated. Finally, the multi-granularity strategy is adopted to learn the pedestrian features of different hierarchies of information in the visible pedestrian region. Experimental results on the Occluded-DukeMTMC, Occluded-REID, and Market1501 datasets demonstrate the effectiveness of our proposed method.

Keywords: occluded person rei‑dentification, pose estimation, multi-granularity, residual atrous spatial pyramid pooling.

1. Introduction

Person re-identification is a technology to identify the presence of a specific pedestrian in an image or video using computer vision techniques, i.e., to retrieve the pedestrian from a surveillance pedestrian image across devices [1]. There are various person re-identification methods based deep learning, which can be classified into representation learning methods [2-4] and metric learning methods [5-6], by type of loss function, and global feature methods[7] and local feature[8-9] methods, by range of feature extraction. Representational learning methods treat the person re-identification task as a classification or validation task, such as classification loss [2-3] and validation loss [4]. Metric learning requires finding a metric function that learns the distance between data objects, such as contrast loss [5] and triplet loss [6]. The algorithms proposed in some papers [7] [9] have achieved good results on the holistic person re-identification dataset [10-11] but failed to perform as expected on the occluded person re-identification dataset [12-13], so there is a need to develop better algorithms for occluded person re-identification.

Person re-identification under occlusion currently faces two main challenges. First, the occluded regions may be mismatched due to the similarity of colors, textures, etc. As shown in Fig. 1, making it difficult to find a robust feature in the occlusion case. Second, pedestrian re-identification under occlusion generally compares the visible pedestrian regions in the occluded images, but the existing occlusion dataset is not labelled with visible regions, which may cause some interference to model training. Miao [13] released the Occluded-Duke dataset for visible region matching with the help of pose estimation algorithm, but the key point features of pose and global features were directly spliced together for training, narrowing the effective scope of the key point features of pose. He [14-16] proposed a series of matching methods based on the pyramid reconstruction model. Although they avoid the alignment features, reconstruction loss cannot completely portray the correspondence between the features of pedestrians. Wang [17] used a pose estimation model to extract key point features, and used the graph convolutional network to model key point features and graph matching to compute similar pedestrian images. Despite the novelty of the processing method, the graph convolutional layer and graph matching are limited by the performance of the pose

[+] Corresponding author. Tel.: + 086-18569037396
E-mail address:szqtiny@gmail.com.

estimation algorithm, and the influence of similar background features to pedestrians is overlooked, so the performance is unsatisfactory.

To alleviate the occlusion problem, we propose a pose-guided multi-granularity occluded person re-identification method. Firstly, the residual pyramid module is used to expand the receptive field to acquire more hierarchical pedestrian information, thus retaining more visible pedestrian features. Next, the visible pedestrian head-shoulders region features and the lower region features are calculated from the pedestrian pose and pedestrian image features extracted by the pose estimation algorithm. Finally, the fine-grained features of the visible pedestrian region are extracted using a multi-granularity learning strategy.

Fig. 1: Retrieval results under occlusion.

2. Method

This section introduces our proposed method, including the residual atrous spatial pyramid pooling module (Residual ASPP module) for expanding the receptive field, the visible pedestrian feature extraction module for extracting visible pedestrian head-shoulders region features and underneath region features, and the multi-granularity feature learning module for learning multi-granularity local features. The three modules are jointly trained in an end-to-end way. A structure diagram of the proposed method is shown in Fig. 2.

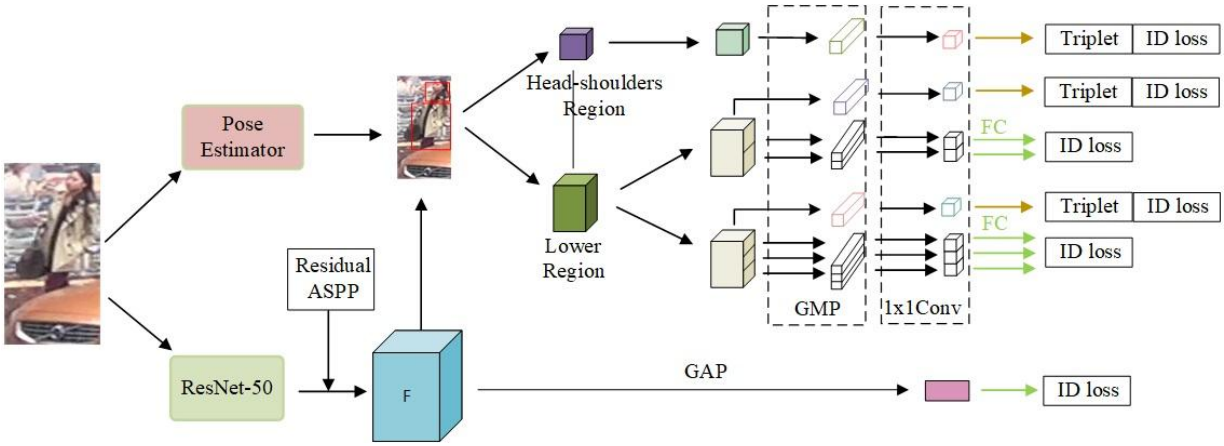

Fig. 2: Network structure.

2.1. Residual ASPP Module

The person re-identification images of occluded scenes generally contain complex information such as pedestrians, backgrounds, and occlusions, etc. In order to extract as much multilevel information as possible about the pedestrian features from the original images before calculating the visible pedestrian features, we introduce the residual atrous spatial pyramid pooling (Residual ASPP) module proposed in literature [18] after ResNet50 [19] to expand the receptive field. The Residual ASPP module uses dilated convolutions [20]

and the residual module to expand the receptive field. The diagram of the Residual ASPP module is shown in Fig. 3, where *Conv* denotes the 1 × 1 convolution. We set dilated rates at 1, 2, 3, 6, 12, and 24, and the Residual ASPP module to cascade twice.

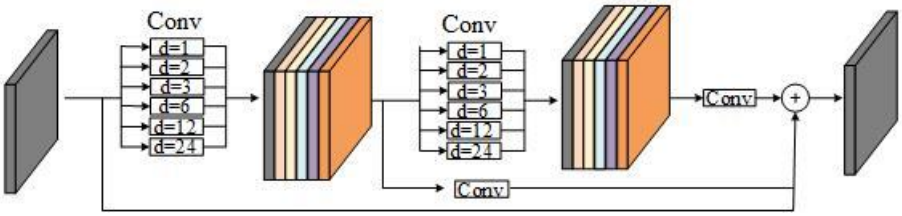

Fig. 3: Schematic diagram of the Residual ASPP module.

2.3. Visible Pedestrian Feature Extraction Module

Intuitively, the recognition of visible pedestrians involves two regions: the head-shoulders region and the underneath region. If the head-shoulders region is clearly visible, the pedestrian can be recognized easily; else if the head-shoulders region is blurred, the identification would otherwise have to be assisted by the clothes and bags. Based on this assumption, we divided the visible pedestrians into the head-shoulders region and the underneath region. Using the pose estimation algorithm, we obtained heatmaps of the key points of the pedestrians and concatenated all the key points in the head-shoulder region to obtain a joint heatmaps of the head-shoulders region $Heatmap_{hs}$, and correspondingly a joint heatmap of the key points in the lower region $Heatmap_{lower}$. The key point heatmap also contains the confidence scores of the detected key points. If the confidence scores are lower than the threshold, the detected key points are considered to be occluded and ignored without concatenating the regions. The visible pedestrian heatmaps and pedestrian features of the two regions are element-wise products for extracting the visible pedestrian head-shoulders region features and the lower region features.

$$\begin{cases} F_{hs} = F \otimes Heatmap_{hs} \\ F_{lower} = F \otimes Heatmap_{lower} \end{cases} \tag{1}$$

2.4. Multi-Granularity Feature Learning Module

The head-shoulders region features and the lower region features produced by the visible pedestrian feature extraction module are still relatively coarse. In the case where the whole body is basically visible, the lower region features are those of most of the body regions, and the granularity will appear too large compared to the key point features to perceive the local body features. We resorted to the multi-granularity strategy in literature [4] to narrow the lower region to use it as a classification task to learn local features. As the granularity of the expression region changed, the model was trained to learn features with different granularity information from the lower region features. Distinguished from literature [4], in which the multi-granularity strategies were directly used for the holistic person re-identification task, we selectively used it for the visible pedestrian region in the occluded person re-identification task.

As shown in Fig. 2, the head-shoulder region features and the lower region features are divided into three branches. At the top is the head-shoulders region feature branch, which uses Global Max Pooling (GMP) for the head-shoulders region features to acquire the texture information from the head-shoulders region, and then a 1 × 1 convolution and ReLU to downsize the features from 2048 dimensions to 256 dimensions. This branch mainly learns the feature information of the head-shoulder region. The second and third branches deal with the lower region features, using the same network architecture as in the head-shoulders region feature branch. The difference is that the second and third branches chunk the lower region feature map horizontally using different chunking granularity, and then each branch goes through the same network as the head-and-shoulder region feature branch to independently learn the local feature representation.

In the training phase of the network, the person re-identification task is considered as an image multiclassification problem. In order to learn distinctive features, the global features of the head-shoulder region features and the lower region features are trained using triplet loss for metric learning, while the other features are trained using the softmax loss for classification learning.

In the testing phase, all feature maps are downsized from 2048 to 256 dimensions and then concatenated to extract the final local features. We concatenated pedestrian global features and final local features to learn robust features with differentiation. Then we used the cosine distance between image features to represent the similarity.

3. Experiments

3.1. Datasets and Settings

Our method is evaluated on four datasets, including two occluded and two holistic re-identification datasets.

One occluded dataset is Occluded-Duke, which includes 15,618 training images from 702 people, 17,661 gallery images from 1110 people, and 2,210 images in the query set from 519 people; the other occluded dataset is Occluded-REID, which contains 2000 images from 200 people, with 5 full-body images and 5 occluded images from each person's identities. Market-1501 consists of 32,668 images from 1,501 person identities, most of which are holistic. DukeMTMC-reID contains 16,522 training images from 702 people, 2,228 images in the query set from 702 people, and 17,661 gallery images from 702 people, most of which are also holistic.

In performance evaluation, we utilized the standard metrics as in most person re-identification literatures, including the cumulative matching cure (CMC) and the mean Average Precision (mAP). All the experiments were performed in single query setting.

3.2. Implementation Details

We utilized ResNet50 as the base CNN backbone by removing its global average pooling layer and fully connected layer. The pedestrian feature F was $2048 \times 16 \times 8$ in size. The pose estimation algorithm adopted HRNet [21] and performed BN processing on the features after the triplet loss before calculating the softmax loss.

The gradient optimizer adopted Adam, a learning decay strategy with an initial learning rate of 0.005 and a decay of 10 times in 40, 90, and 140 rounds until the end of training, with 240 training rounds. 64 training batches were used, and 16 pedestrians were selected each time, with 4 images for each pedestrian.

3.3. Performance Evaluation

In this section, the proposed method is compared with the current state-of-the-art methods on the four datasets Occluded-Duke, Occluded-ReID, Market-1501, and DukeMTMC-reID in Table 1 and Table 2.

Table 1: Results comparisons on Occluded-Duke and Occluded-ReID

Methods	Occluded-Duke		Occluded-ReID	
	Rank-1	mAP	Rank-1	mAP
Ad-Occluded[22]	44.5	32.2	41.3	38.9
PCB[3]	42.6	33.7	-	-
DSR[14]	40.8	30.4	59.3	53.2
SFR[15]	42.3	32	-	-
PGFA[13]	51.4	37.3	-	-
IVP[23]	56.3	43.5		
PVPM[24]	-	-	66.8	59.5
PDVM[25]	53.0	38.1	-	-
HOReID[17]	55.1	43.8	80.3	70.2
PAFM[26]	55.1	42.3	76.4	68.0
ABSA[27]	55.4	43.8	82.5	71.8
HG[28]	61.4	50.5	82.3	71.7
Ours	61.5	50.4	80.7	70.6

Table 2: Results comparisons on Market-1501 and DukeMTMC-reID

Methods	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank1	mAP
PCB[3]	92.3	77.4	81.8	66.1
AlignedReID[7]	91.8	79.3	-	-
MGN[4]	95.7	86.9	88.7	78.4
DSR[16]	58.8	67.2	50.7	70.0
SFR[17]	63.9	74.8	56.9	78.5
VPM[29]	93.0	80.8	83.6	72.6
PGFA[15]	91.2	76.8	82.6	65.5
HOReID[19]	94.2	84.9	86.9	75.6
HG[28]	95.6	86.1	87.1	77.5
Ours	94.3	86.4	87.2	77.9

On the Occluded-Duke dataset, the proposed method achieved 61.5% Rank-1 and 50.4% mAP. The results of the proposed method and two types of methods on Occluded-Duke and Occluded-ReID are shown in Table 1. The first type of methods applied on the holistic person re-identification dataset are Ad-Occluded [22], and PCB [3], among which only PCB achieved a superior performance. The second type is dedicated to occlusion, including DSR [14], SFR [15], PGFA [13], HOReID [17], and HG[28] et al, which had a significant performance improvement over common person re-identification methods on the occluded datasets, except for the earlier DSR and SFR, which had a more general performance. PGFA, IVP, PVPM, PDVM, HOReID, and the proposed method are all pose-guided methods, among which PGFA is the earliest proposed method in terms of time, IVP has made some changes in PGFA, HOReID uses the graph matching principle to train and match for the refined pedestrian key points, PVPM introduces pseudo-label training based on graph matching, and PDVM uses a self-attentive mechanism to focus on non-pedestrian regions. Overall, the performance of each of them has been improved differently. HG method uses a student-teacher framework with denoising autoencoder to perform the occlusion task, achieving the same effect as our method. In our proposed method, we used the residual pyramid module to expand the receptive field to extract more hierarchical pedestrian features, and pedestrian key points to detect the visible pedestrian region and conducted multi-granularity analysis to achieve the best results among several methods.

On Occluded-ReID, our proposed method achieved 80.7% Rank-1 and 71.1% mAP. Fewer methods were tested on this dataset, and the first and second types of methods had inferior performance. Methods that are specialized in masking continued to perform at a high level, and our proposed method were about 10% higher in both Rank-1 and mAP compared to the PVPM method, slightly better than HOReID.

From the experiments, the occluded person re-identification algorithm was found to significantly outperform the metrics of the holistic person re-identification algorithms on the occluded datasets, and the occlusion situation is a factor worth consideration for a robust person re-identification algorithm.

Methods designed for occluded cases can usually achieve better performance in the corresponding scenarios, but they may have average performance on datasets mainly consisting of full-body visible cases. By comparing them with holistic person re-identification methods on the two typical holistic person re-identification datasets, Market-1501 and DukeMTMC-reID, the our proposed method exhibits the competitive performance, achieving 94.3% Rank-1 and 86.4% mAP on the Market-1501 dataset, while 87.2% Rank-1 and 77.9% mAP on the DukeMTMC-reID dataset. The results are shown in Table 2.

Person re-identification methods generally have poorer metrics than holistic person re-identification methods when applied for obscured scenes, and a small number of methods have a similar performance to holistic person re-identification methods. For example, DSR, SFR, VPM, etc. designed for early occluded person re-identification methods have poorer performance. Ordinary person re-identification dataset contains a large number of full-body visible pedestrian images, requiring the model to be more sensitive to local feature information; unfortunately, the model of early designed occluded person re-identification methods had poor feature extraction ability which fell short of the requirement. HOReID, on the other hand, uses the graph matching method to find an alternative way to achieve similar results to MGN. HG uses the student-

teacher framework with denoising autoencoder to address re-identification problem. The proposed method resorts to the multi-granularity design of MGN. Although many optimization measures are included, it is still affected by the performance of the pose estimation algorithm itself, which is slightly inferior to the MGN method, but slightly superior to the HOReID method, and the same effect has been achieved as the HG method.

The main difference between the holistic person re-identification dataset and the occluded person re-identification dataset is the number of full-body visible pedestrians. In the holistic person re-identification dataset the pedestrian images with most visible body regions and a single background occupy a larger proportion. If there is no distinctive feature, then simply capturing as many relevant features as possible from the pedestrian images becomes the key to distinguishing the pedestrians, even including the background features. Holistic person re-identification methods can take full advantage of the above settings to design networks that can extract as many fine-grained features as possible to achieve good performance without considering the contamination of the occluded background. In contrast, if this type of method is applied in the occlusion scenario, the performance will be significantly degraded due to the large number of occlusion cases that carry many contaminated pedestrian features. Furthermore, the occlusion method tends to miss some detailed information when dealing with the occlusion situation, leading to insufficient feature extraction ability. The proposed method has combined pose estimation methods and multi-granularity size modules to handle both occlusion scenarios and rich feature extraction capabilities, thus achieving competitive performance on holistic person re-identification datasets.

3.4. Discussion

In this section, our proposed Residual ASPP module, visible pedestrian feature extraction module, and multi-granularity feature learning module are evaluated. The following experiments were all performed on the Occluded-DukeMTMC dataset.

Table 3: Ablation experiment of modules

Methods	Rank-1	mAP
Baseline	53.7	44.5
Baseline+ASPP	54.3	44.9
Baseline+VRM	55.3	45.6
Baseline+ASPP+VRM	56.4	46.5
Baseline+ASPP+VRM+MGN	61.5	50.4

First, each of these modules has been validated. The experimental results are shown in Table 3. Baseline denotes the network of independent key point features and global features computed by the pose estimation algorithm and the ResNet50 algorithm. The initial Rank-1 and mAP of this Baseline network are 53.7% and 44.5%, respectively. After adding the residual pyramid pooling module (Residual ASPP, ASPP), richer pedestrian features had less impact on the individual pose key point features, and the overall model performance was only marginally improved. When the key points were changed to two types of region features (Visible Region Module, VRM) training, Rank-1 and mAP achieved a small improvement by 1.6% and 1.1%, respectively, with more images of pedestrians and backgrounds that were more similar in the occlusion dataset. When ASPP and VRM were combined into the model, the Rank-1 and mAP of the model became 56.4% and 46.5%, respectively, a more significant improvement than the single VRM, and the large receptive field of ASPP has brought about significant performance improvement. When the Multiple Granularity Network (MGN) module was added finally, Rank-1 improved by 5.1% and mAP improved by 3.9%. The performance improvement of the MGN module was rather obvious, and a significant portion of the slightly obscured or fully visible pedestrian images in the obscured dataset was retrieved more often than ever, which, combined with the previous module optimizations, has resulted in the best performance of 61.5% Rank-1 and 50.4% mAP.

Table 4: Performance comparison between different partitioning strategies

Methods	Rank-1	mAP
Head-shoulders	56.4	46.5
Upper-Lower	55.5	45.7

Two parts	58.9	48.5
Three parts	60.0	49.2
Four parts	60.2	49.1
Ours	61.5	50.4

Second, the choice of visible pedestrian region segmentation and multi-granularity strategies has been analyzed, with the experimental results shown in Table 4. For the division method of visible pedestrian regions, one is divided into the head-shoulder region and lower region (head-shoulders) chosen in the proposed method, and the other is roughly divided into the upper region and lower region (upper-lower). The experiment used the above Baseline method and added the optimization strategy of Residual ASPP, but it did not add any multi-granularity strategy after using the method. The head-shoulders region and lower region division methods chosen in our proposed method have achieved relatively good performance. There are more occluded images in this dataset, and the division of the head-shoulders region and other body regions in the occluded images is more consistent with the unique person identification, which will be more sensitive to local discriminating features. As for the multi-granularity division hierarchy strategy, we have directly adopted the division in literature [4], and the other three methods divide granularities into two layers (two parts), three layers (three parts) and four layers (four parts) for the lower region respectively. The performance of these three granularity classification methods varies depending on the degree of image occlusion. It has also been found from the experimental data that the performance of the method dividing multiple granularities into three layers and four layers are good, while the finer granularity classification is of only a little significance to the occluded pedestrian images. The method of extracting fine-grained features from two distinct dimensions used in our proposed method has achieved the best performance without considering the influence of the degree of image occlusion.

Fig. 4: Partial special image query results.

To demonstrate the improvement of our model, we conducted some retrieval experiments on baseline and our proposed method for occluded pedestrian images. Fig. 4 shows some visual comparison results of retrieving the top five. The first image of a pedestrian occluded by a backpack was retrieved correctly by both methods without reflecting a large difference. The second image is a half-body occluded pedestrian, who is occluded by a car and a backpack. The baseline method can only rely on pedestrian key point feature retrieval, and only the second one was retrieved correctly in the results of Baseline retrieval, while all the other images detected similar and incorrect local retrieval results for key point features. The proposed method has a vast receptive field for acquiring integral pedestrian information of features from large pedestrian region features, and all but the fourth image were basically correctly retrieved, probably because it

has used more color information of cloths. In the third image, a pedestrian is obscured by another pedestrian. The baseline method tends to fuse the key point features of the two pedestrians, while our method tends to fuse the regional features of the two pedestrians. Compared to the baseline method, our method considers the pedestrian in black as the query pedestrian. This is a correctly labelled image. Combined with the local feature mining of the multi-granularity feature learning module, we can even retrieve a fifth image that does not contain the pedestrian in white.

4. Conclusion

In this paper, we have proposed a novel method to solve the problem of occluded person re-identification. First, the Residual ASPP module has been proposed to extracted more hierarchical pedestrian features by enlarging receptive field. Next, the pedestrian features have been divided into head-shoulders region features and lower region features by the pedestrian estimation algorithm, and the above two features have been applied to learn the local features by multi-granularity feature learning strategy. Compared to other occluded person re-identification methods, our method has demonstrated a competitive performance in occluded and holistic tasks.

5. Acknowledgements

This work was supported by the National Natural and Science Foundation of China (Grant No. 61972357).

6. References

[1] Luo Hao, Jiang Wei, Fan Xing, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032−2049.

[2] Zheng L, Zhang H H, Sun S Y, et al. Person re-identification in the wild[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, Hawaii, USA. New York: IEEE Press, 2017:3346-3355.

[3] Lin Y T, Zheng L, Zheng Z D, et al. Improving Person Re-identification by Attribute and Identity Learning[J]. Pattern Recognition, 2019, 95:151-161.

[4] Zheng Z D, Zheng L, Yang Y. A discriminatively learned CNN embedding for person re-identification.[J]. ACM Trans actions on Multimedia Computing, Communications, and Applications, 2018, Volume 14, Issue 1, Article No.: 13, pp:1-20.

[5] Wang Y C, Chen Z Z, Wu F, Wang G. Person re-identification with cascaded pairwise convolutions[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, New York: IEEE Press, 2018:1470-1478.

[6] Chen W H, Chen X T, Zhang J G, et al. Beyond triplet loss: a deep quadruplet network for person re-identification[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, Hawaii, USA. New York: IEEE Press, 2017:1320-1329.

[7] Zhang X, Luo H, Fan x, et al. AlignedReid:Surpassing human-level performance in person re-identification[EB/OL].(2018-1-31)(2022-12-1). https://arxiv.org/abs/1711.08184.

[8] Sun Y F, Zheng L, Yang Y, et al.Beyond Part Models: Person Retrieval with Refined Part Pooling(and A Strong Convolutional Baseline)[M]. Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11208: 501-518.

[9] Wang G S, Yuan Y F, Chen X, et al. Learning Discriminative Features with Multiple Granularities for Person Re-Identification[C]// Proceedings of the 26th ACM international conference on Multimedia, Seoul, Korea. New York: ACM, 2018: 274-282.

[10] Zheng L, Shen L Y, Tian L, et al. Scalable Person Re-identification: A Benchmark[J]. 2015 IEEE International Conference on Computer Vision(ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.

[11] Ristani E, Solera F, Zou R, et al. Performance meansures and a data set for multi-target, multi-camera tracking[M].

Hua G, Jégou H. Computer vision-ECCV 2016 Workshops. Cham: Springer, 2016, 9914: 17-35.

[12] Zhuo J X,Chen Z Y, Lai J H,et al. Occluded Person Re-identification[C]// 2018 IEEE International Conference on Multimedia and Expo (ICME), July 23-27, 2018, San Diego, CA, USA. New York: IEEE Press, 2018:1-6.

[13] Miao J X, Wu Y, Liu P, et al. Pose-Guided Feature Alignment for Occluded Person Re-Identification[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul,Korea. New York: IEEE Press, 2019:542-551.

[14] He L X, Liang J, Li H Q, et al. Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, New York: IEEE Press, 2018:7073-7082.

[15] He L X, Sun Z N, Zhu Y H, et al. Recognizing Partial Biometric Patterns[EB/OL]. (2018-10-17)[2021-12-1]. https://arxiv.org/abs/1810.07399.

[16] He L X, Wang Y G, Liu W, et al. Foreground-aware Pyramid Reconstruction for Alignment-free Occluded Person Re-identification[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-Nov 2, 2019, Seoul, Korea. New York: IEEE Press, 2019:8449-8458.

[17] Wang G A, Yang S, Liu H Y, et al. High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020:6448-6457.

[18] WANG Y Q，WU T H，YANG J G，et al. DeOccNet：learning to see through foreground occlusions in light fields[C]// Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision. Piscataway：IEEE，2020:118-127.

[19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016:770–778.

[20] ZHANG X H，ZOU Y X，SHI W. Dilated convolution neural network with LeakyReLU for environmental sound classification[C]// Proceedings of the 22nd International Conference on Digital Signal Processing. Piscataway：IEEE，2017:1-5.

[21] Wang J D, Sun K, Cheng T H, et al. Deep High-Resolution Representation Learning for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43:3349-3364.

[22] Huang H J, Li D W, Zhang Z, et al. Adversarially occluded samples for person re-identification[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, New York: IEEE Press, 2018:5098-5107.

[23] Miao J, Wu Y, Yang Y. Identifying visible parts via pose estimation for occluded person re-identification[J]. IEEE Transactions on Neural Networks and Learning Systems (2021).https://doi.org/10.1109/TNNLS.2021.3059515.

[24] Gao S, Wang J Y, Lu H C, et al. Pose-guided Visible Part Matching for Occluded Person ReID[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York:IEEE Press,2020:11741-11749.

[25] Zhou S R,Wu J, Zhang F, et al. Depth occlusion perception feature analysis for person re-identification[J].Pattern Recognition Letters, 2020, 138:617-623.

[26] Yang, J., Zhang, C., Tang, Y. et al. PAFM: pose-drive attention fusion mechanism for occluded person re-identification[J]. Neural Comput & Applic (2022). https://doi.org/10.1007/s00521-022-06903-4.

[27] Jin H Y, Lai S Q , Qian X M. Occlusion-sensitive Person Re-identification via Attribute-based Shift Attention[J].IEEE Transactions on Circuits and Systems for Video Technology (2021).https://doi.org/10.1109/TCSVT.2021.3088446.

[28] Madhu K, Gnana P R, Le T N, et al. Holistic Guidance for Occluded Person Re-Identification[EB/OL]. (2021-4-13)[2022-1-5]. https://arxiv.org/abs/2104.06524.

[29] Sun Y F, Xu Q, Li Y L, et al. Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),June 15-20, Long Beach, CA, USA.New York: IEEE Press,2019:393–402.