Research Article

Pose-Guided Part-Based Adaptive Pyramid Features for Occluded Person Reidentification

Xiaobing Lin (ID)**, Jilin Li** (ID)**, Zengxi Huang** (ID)**, and Xiaoqin Tang** (ID)

School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

Correspondence should be addressed to Zengxi Huang; huangzx001@mail.xhu.edu.cn

Reidentifying an occluded person across nonoverlapping cameras is still a challenging task. In this work, we propose a novel pose-guided part-based adaptive pyramid neural network for occluded person reidentification. Firstly, to alleviate the impact of occlusion, we utilize pose landmarks to generate pose-guided attention maps. The attention maps will help the model focus on the nonoccluded regions. Secondly, we use pyramid pooling to extract multiscale features in order to address the scale variation problem. The generated pyramid features are then multiplied by attention maps to achieve pose-guided adaptive pyramid features. Thirdly, we propose a pose-guided body part partition scheme to deal with the alignment problem. Accordingly, the adaptive pyramid features are divided into partitions and fed into individual fully connected layers. In the end, all the part-based matching scores are fused with a weighted sum rule for person reidentification. The effectiveness of our method is clearly validated by the experimental results on two popular occluded and holistic datasets, i.e., Occluded-DukeMTMC and the Market-1501.

1. Introduction

Person reidentification (Re-ID) aims to retrieve a probe person/pedestrian from nonoverlapping cameras [1–3]. It has become increasingly popular in the community due to its application potentials in video surveillance, smart retailing, activity analysis, and so on. Although person Re-ID has achieved a great progress in the recent years, it is still a challenge to reidentify a person who is partially occluded. For example, in the surveillance scenario, a person may be occluded by walls, transformations, and other pedestrians on the road.

The occlusions lead to not only the loss of target information but also the interference of occluded information [4–6]. The holistic person Re-ID methods that only consider the global person feature are easily misled by the occluded body part. As shown in Figure 1, the holistic methods may mistakenly match a person when a person image in another view shares similar obstacle with the probe person image. Therefore, it is significant to seek efficient ways to solve this occluded reidentification problem. One well-known solution is to use part-based models. Part-based models perform part-to-part matching and are much more robust to occlusions than the holistic models [7, 8]. However, they partition a person into a fixed number of parts and heavily depend on precise person detection.

In the recent years, there is a trend to leverage external cues like person mask and pose estimation for occluded person Re-ID [9–12]. Mask-guided models use person mask to help remove the background clutters, including the occluded body parts, and then perform matching using the clean body parts. A major limitation of mask-guided models is that currently even the state-of-the-art person mask segmentation models perform inferiorly in the scenarios with person occlusion, especially when the target person is occluded by other persons [13, 14]. On the other hand, owing to the advancement of [5, 6] pose estimation models, pose-guided models have attracted considerable attention recently [4, 5]. The pose-guided models utilize the skeleton as an external cue to effectively relieve the part misalignment problem by locating each part using person landmarks.

In this paper, we propose a novel pose-guided part-based adaptive neural network for occluded person

FIGURE 1: Illustration of occluded person Re-ID. Red and green rectangles indicate error and correct match results.

reidentification. Firstly, we use the human pose estimator [40] to generate pose landmarks, which are then used to create pixel-wise attention maps. Compared with mask-guided methods [14–17], the pose-guided method can better detect pedestrian's location in a crowd. And then, we use pyramid pooling to extract multiscale features in order to address the scale variation problem. To adapt the pyramid features, we also generate multiscale attention maps. The pyramid features are finally multiplied by the attention maps with the corresponding scales to achieve pose-guided adaptive pyramid features.

In the matching stage, we propose a pose-guided dynamic body part partition scheme to deal with the alignment problem, in which the human parts are defined dynamically based on pose landmarks. In the end, all the part-based matching scores are fused with a weighted sum rule for reidentification. Experimental results on occluded and holistic Re-ID datasets, i.e., Occluded-DukeMTMC and the Market-1501, validate the effectiveness of our proposed method.

The main contributions of our work are summarized as follows: (1) we propose a pose-guided part-based adaptive pyramid neural network (PPAPN) that uses the pose-guided attention map and pyramid pooling so as to deal with the occlusion and scale problems; (2) we propose a pose-guided dynamic body part partition scheme to tackle the alignment problem; and (3) we compare the proposed method with many well-known methods on two popular occluded and holistic Re-ID datasets; the experimental results validate that our proposed method is effective.

The rest of paper is organized as follows. Some related works are reviewed at first. The structure of our proposed PPAPN and implementation details are presented in the next. Furthermore, we present the detail of the experimental results. And, the conclusions of our work are described in the end.

2. Related Works

2.1. Person Reidentification.
Person reidentification aims to match a probe pedestrian image in candidate pedestrian images across disjoint cameras [1–3]. In recent years, deep learning algorithms [18–23] for person reidentification indicate distinct superiority on matching accuracy. Most of these methods pay attention to extract distinguishable features from holistic pedestrian images [7–9, 22–25]. The

methods in [5–7, 9, 24–26] employ the part-based features leaning method and significantly improve the global-based person Re-ID method performance on the holistic person Re-ID datasets. However, they do not take occlusions into considerations. Besides, human pose estimation and landmark detection having achieve impressive progress, several recent works in Re-ID employ these tools to acquire aligned subregion of person images [6, 7, 24, 25]. However, the occlusion cannot be ignored especially in the crowd scenes like school or metro station.

2.2. Occluded Person Reidentification.
Occluded person Re-ID [26–28] aims to retrieval occluded probe images in occluded gallery pedestrian images in disjoint cameras. Due to the images suffer from occluded noise, occluded person Re-ID is a more challenging task than holistic person Re-ID. To solve this problem, Zhuo et al. [26] defined the problem of occluded reidentification for the first time; they employed multitask losses that lead the model to distinguish simulated occluded samples and nonoccluded samples, in this way to learn a robust feature representation from occluded person images. In this following works, the model employs a cosaliency network to endeavor to pay attention to the person visible part. More recently, Miao et al. [5] utilized semantic key-points to extract the useful information from the occluded person images. And, they used a predefined threshold of key-points confidence to determine whether the part is occluded or not. He et al. [29] used a spatial transform module to transform the holistic image to align with the partial ones and then calculated the distance of the aligned pairs.

2.3. Partial Person Reidentification.
Compared with occluded person Re-ID [30], partial ones often occur due to imperfect detection and outliers of camera views. Like occluded person Re-ID, partial person Re-ID aims to retrieval partial probe images to gallery holistic images. Zheng et al. [30] proposed a local matching strategy based on dictionary learning and called it ambiguity-sensitive matching classifier (AMC), and they introduced a sliding window matching (SWM) model to solve global-part-based matching problems. He et al. [28] proposed a matching model based on sparse reconstruction learning and called it deep spatial feature reconstruction (DSR). DSR can automatically match images of different sizes, thus avoiding the time-consuming spatial alignment step. He et al. [29] further proposed a spatial feature reconstruction model, which generates multiscale features by a fully convolutional network to deal with the scale change of feature maps. Sun et al. [31] proposed a visibility-aware part model (VPM) in [7], which perceives visible areas by self-supervised learning to avoid the noise effect of the occluded regions.

3. Methodology

As shown in Figure 2, the proposed pose-guided part-based adaptive pyramid neural network (PPAPN) consists of three
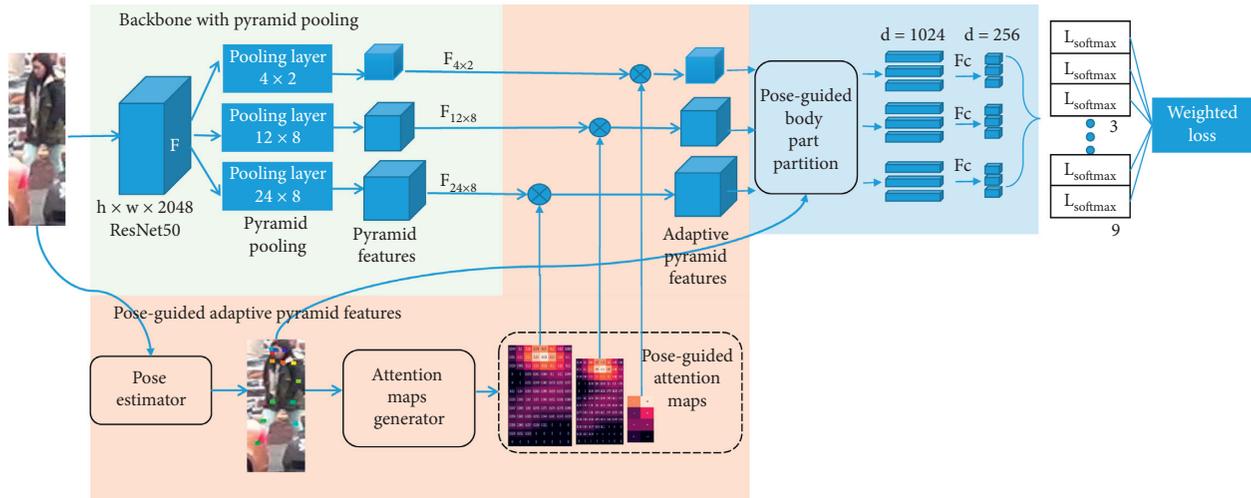
FIGURE 2: Pipeline of the proposed pose-guided part-based adaptive pyramid neural networks. It consists of three key components: backbone with pyramid pooling, pose-guided adaptive pyramid features module, and pose-guided body part partition.

components: backbone with pyramid pooling, a pose-guided adaptive pyramid features module, and a pose-guided body part partition. Firstly, the backbone with pyramid pooling could match multiscales of person images in occluded scenarios and get multiscale pyramid features. Secondly, we use the human pose estimator OpenPose [32] to generate pose landmarks, which is used to create pose-guided attention maps of pedestrians. And then, we multiply the features from pyramid pooling by pose-guided attention maps which have the corresponding scales to achieve pose-guided adaptive pyramid features. Finally, pose-guided body part partition splits the pose-guided pyramid adaptive features into three local region features for the matching stage, and all the part-based matching scores are fused with a weighted sum rule for reidentification.

3.1. Backbone and Pyramid Pooling.

Our proposal PPAPN mainly uses ResNet-50 [33] which pretrained on ImageNet as backbone to extract global feature from given images, and we make a minor modification on it. Specifically, we remove the average pooling layer and the fully connected layer of ResNet-50 to obtain the refined ResNet-50. The detail of the backbone is shown in Table 1; it contains 1 convolutional layer, namely, conv1 and 4 Resblock layers (conv2_x, conv3_x, conv4_x, and conv5_x). The output size of each layer is 384×128, 192×64, 96×32, 48×16, and 24×8, respectively. And, the last Resblock conv5_x outputs the spatial feature map $F \in \mathbb{R}^{h \times w \times c}$, in which h, w, and c denotes the height, width, and channel number, respectively. Then, the feature map F is fed into the pyramid pooling.

Due to the different distances between pedestrians and cameras, the detected pedestrians may have various scales. In the one hand, if we resize the person images to the same size, it will fail to align their spatial features and bring errors to their distance measure. In the other hand, if we employ the original multiscales person image from the pedestrian detector, the ResNet-50 backbone will output

TABLE 1: A refined ResNet-50.

Layer name	Output_size	Refined ResNet-50
conv1	384×128	7×7, 64, stride 2
		3×3 max pool, stride 2
conv2_x	192×64	$\begin{pmatrix} 1 & 1, 64 \\ 3 & 3, 54 \\ 1 & 1, 256 \end{pmatrix} \times 3$
conv3_x	96×32	$\begin{pmatrix} 1 & 1, 128 \\ 3 & 3, 128 \\ 1 & 1, 512 \end{pmatrix} \times 4$
conv4_x	48×16	$\begin{pmatrix} 1 & 1, 256 \\ 3 & 3, 256 \\ 1 & 1, 1024 \end{pmatrix} \times 6$
conv5_x	24×8	$\begin{pmatrix} 1 & 1, 512 \\ 3 & 3, 512 \\ 1 & 1, 2048 \end{pmatrix} \times 3$

features of variable sizes. However, the fully connected layers require fixed size features. In this work, we propose to use pyramid pooling after the backbone to address the scale variation problem. The pyramid pooling contains several max-pooling layers with different output size to generate multiscale pyramid features. The main purpose of each max-pooling layer is to produce a fixed output for any size of input, while several max-pooling layers could preserve global information in different scales as much as possible. Specifically, for any size feature map, firstly, it is divided into some blocks of several scales, such as a feature map is divided into 1×1, 2×2, and 4×4 blocks, and then the maximum pooling is performed on each block to obtain multiscale pyramid pooling features. In this work, we divide the feature map into 4×2, 12×8, and 24×8 blocks. As shown in Figure 2, finally, we obtain the basic multiscale pyramid features which are denoted as $F_{pi} \in \{F_{4 \times 2}, F_{12 \times 8}, F_{24 \times 8}\}$, and the output size of the pyramid features is $4 \times 2 \times 1024$, $12 \times 8 \times 1024$, and $24 \times 8 \times 1024$. The output features from pooling layers of small output size obtain the appearance information of a large local region, such as F_{4x2}; the feature map is divided

into 4×2 blocks, and it is obvious that each block contains a larger region than $F_{12 \times 8}$. On the contrary, the output features from the pooling layers of large output size obtain the appearance information from relatively small regions in the image like $F_{24 \times 8}$. Thus, the scale variation problem has been well addressed.

3.2. Pose-Guided Adaptive Pyramid Features Module.

In occlusion person Re-ID, it is necessary to pay more attention on the features that are more closely related to the pedestrians. So, we use pose landmarks to generate attention maps, which are used to separate the foreground information of pedestrians from the occlusion noise and dig out useful background information, such as pedestrian attachments.

3.2.1. Pose-Guided Attention Maps Generator.

To alleviate the impact of occlusion, we design a pose-guided attention maps generator to guide the model extract more useful and robust features. Firstly, we employ the human pose estimator OpenPose [32] to detect N pose landmarks from an input person image, where N is 18 in this work. Each landmark contains its coordinates and confidence score. Inspired by [34], we employ a sigmoid form of decay function to better distinguish between occluded regions and nonoccluded regions. The decay function increases with the increase in the confidence of the landmark, and the sigmoid form is defined as follows:

$$S(x) = \frac{1}{1 + \exp^{-\beta(x - \alpha)}},$$

$$f_d = \frac{s(x) - s(0)}{s(1) - s(0)},$$ (1)

$$LM_j = f_d\big(S\big(C_j\big)\big).$$

We set a threshold α to suppress values less than α; in this paper, α is fixed at 0.5. As β increases, the decay function becomes steeper near 0.5. LM_j denotes the sigmoid form resulted the j-th landmark, and C_j is the confidence score. Similar to [5], we use the landmarks to generate the pose-guided attention map M consisting of a 2D Gaussian centred on the ground truth location, and it is defined as follows:

$$M_j = \frac{1}{2\pi\sigma^2} e^{-\left(c_{xj}^2 + c_{yj}^2\right)/2\sigma^2} \times LM_j,$$

$$M = \sum_j^N M_j,$$ (2)

where M_j denotes the generated pose-guided attention map of the j-th landmark, $j = 1, 2, \ldots, N$ and c_{xj} and c_{yj} denote the coordinates of the j-th landmark. Each attention map M_j is set to the same size of $h \times w$. To adapt the multiscale pyramid

features, we also employ bilinear interpolation to get multiscale pose-guided attention maps $M_i' \in \{M_{4 \times 2}, M_{12 \times 8}, M_{24 \times 8}\}$.

3.2.2. Adaptive Pyramid Features.

The pose-guided adaptive pyramid feature module aims to integrate the feature maps information and the pose information from the target person. As shown in Figure 2, firstly the feature map F is pyramid pooled as the multiscale features F_{pi}. Then, the pose-guided attention maps M_i', $i = 1$, 2, 3, multiply the feature maps F_{pi} element-wisely and output the pose-guided adaptive pyramid feature map F_i'. Since each attention map M_i' has explicitly encoded the information of different regions on the target person, i.e., which region is occluded, the pose-guided attention maps M_i' can focus on nonoccluded parts of the target person and depress the information from occluded regions.

$$F_i' = M_i' \otimes F_{pi}.$$ (3)

3.3. Pose-Guided Body Part Partition.

Previous methods [22, 23, 31] horizontal cut the probe images and gallery images and extract their horizontal features, as demonstrated in Figure 3(a). Obviously, it is difficult to match the body parts in the gallery images when the pedestrians' posture changes, and there is occlusion. So, we propose a pose-guided body part partition. As shown in Figure 3(b), we employ the landmarks detected by the human pose estimator to explicit positioning person body parts, and we define three body parts (R1, R2, and R3) based on these landmarks. In order to extract robust features, each adjacent part has an overlapping region. Specifically, R1 contains these landmarks $\{p_0, p_1, p_2, p_5, p_{14}, p_{15}, p_{16}, p_{17}\}$, R2 contains these landmarks $\{p_1, p_2, p_5, p_3, p_4, p_6, p_7, p_8, p_{11}\}$, and R3 contains these landmarks $\{p_8, p_{11}, p_9, p_{12}, p_{10}, p_{13}\}$. The corresponding body part bounding box $B_i \in \{B_1, B_2, B_3\}$ can be obtained based on the location coordinates of all body joints.

$$B_j = [x_{\min} - 10, x_{\max} + 10, y_{\min} - 10, y_{\max} + 10].$$ (4)

Given landmarks from the pose estimator, firstly, we obtain the original corresponding body parts bounding box based on the coordinates of body joints. Secondly, to process the different size of adaptive pyramid features, we adjust the original corresponding body parts bounding box matching the adaptive pyramid feature size. We follow the part alignment pooling [10] method to extract these part features on query images and gallery images. We fed the output features from pose-guided body part partition into the fully connect layer to obtain the partial features $F_{pi} \in \{F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}\}$, $F_{pi} \in \mathbb{R}^{256}$.

For basic discrimination learning, we regard the identification task as a multiclass classification problem. The learned feature softmax loss is formulated as follows:
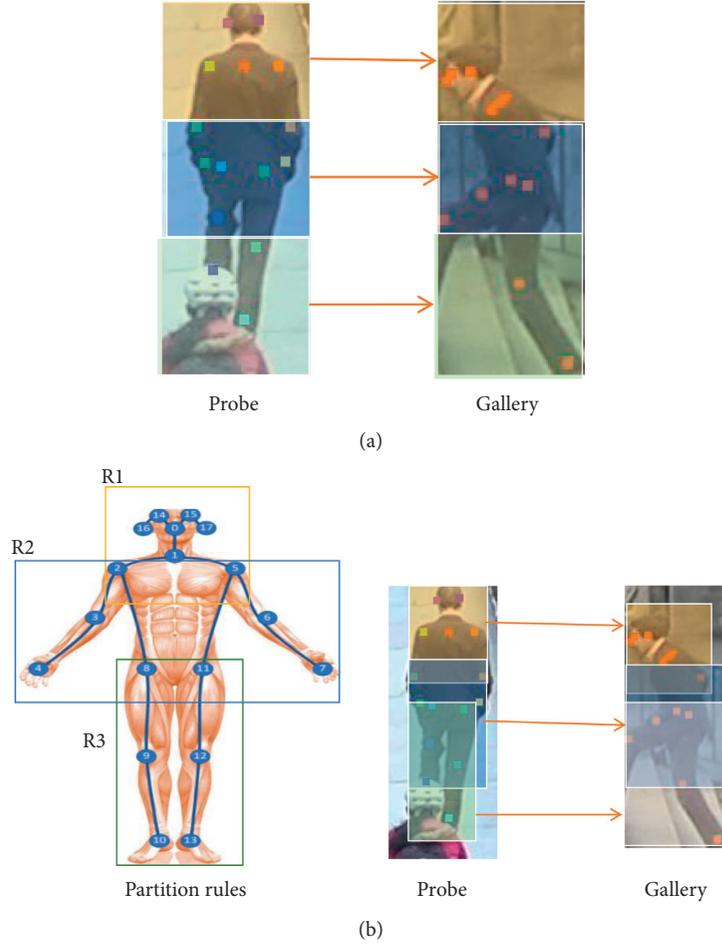
FIGURE 3: (a) PCB [35] divides the person image evenly without overlap. (b) The proposed pose-guided body part partition and examples.

$$L_{\text{weighted}} = -\sum_{s=1}^{N} \lambda_s \times \log \frac{e^{F_{\text{pi}}}}{\sum_{i=1}^{L} e^{F_{\text{pi}}}} \ s.t \ \sum \lambda_s = 1, \qquad (5)$$

where N denotes the number of branches; we set N is 9 in our method. The parameter $\lambda_s (0 < \lambda_s < 1)$ is a coefficient to balance of each branch. L denotes the number of classes in the training dataset.

4. Experiments

4.1. Datasets and Evaluation Metrics. In our experiments, an occluded person Re-ID dataset and a holistic person Re-ID dataset are employed, i.e., Occluded-DukeMTMC [18] and Market-1501. For performance evaluation, we use the standard metrics as in the most person Re-ID literature, the cumulative matching curve (CMC), and the mean Average Precision (mAP). For the CMC curve, we report the Rank-1 accuracy.

(i) *Occluded-DukeMTMC.* This dataset is a subset of the DukeMTMC [15] used for person reidentification by images. It contains 15618 training images, 17661 gallery images, and 2210 occluded query images.

Occluded-DukeMTMC is selected from DukeMTMC-Re-ID by leaving occluded images and filtering out some overlap images.

(ii) *Market-1501.* This dataset includes images of 1501 persons captured from 6 different cameras. The pedestrians are cropped with bounding box predicted by the DMP detector [21]. The whole dataset is divided into training set with 12936 images of 751 persons and testing set with 3368 query images and 19732 gallery images of 750 persons.

4.2. Training Details. We use ResNet-50 [33] as our backbone and make a minor modification on it: removing the average pooling layer and fully connected layer and setting the stride of conv4_1 to 1. We initialize our mode by the ImageNet [36] pretrained model. In our experiment setting, we did not force the image resize to a fixed size, but we augmented by random flipping and random erasing. We set the batch size to 32 and the training epoch number to 60. The base learning rate is initialized at 0.1 and decayed to 0.01 after 40 epochs. The coefficient λ is empirically set to be 0.2.

TABLE 2: The person Re-ID results on the Occluded-DukeMTMC dataset.

Method	Rank-1 (%)	mAP (%)
Part-Aligned [9]	42.8	33.2
PCB [35]	42.6	33.7
DSR [21]	40.8	30.4
SFR [22]	42.3	32.0
Ad-occluded [8]	44.5	32.2
PGFA [5]	51.4	37.3
PPAPN(ours)	53.6	38.36

TABLE 3: The person Re-ID results on the Market-1501 dataset.

Method	Rank-1 (%)	mAP (%)
SVDNet [24]	82.3	62.4
PAN [37]	82.8	63.4
Pedestrian [30]	82.0	63.0
DSR [21]	83.5	64.2
TripletLoss [4]	84.9	69.1
AFPB [26]	86.5	78.3
PCB [35]	92.3	66.1
MaskReID [20]	90.0	75.3
Pose-transfer [22]	87.7	68.9
PSE [8]	87.7	69.0
PGFA [5]	91.2	76.8
PAPM(ours)	94.2	83.6

FIGURE 4: Comparison of the PGFA [5] method and our PPAPN method. Green and red numbers indicate correct and error retrieval results.

4.3. Results

4.3.1. Results on Occluded-DukeMTMC.
We evaluate our method PPAPN on Occluded-Re-ID at first. As shown in Table 2, six kinds of methods are compared. As we can see, there is no significant gap between holistic Re-ID methods. For example, PCB [31] and Part-Aligned [9] both achieve approximately 43%

Rank-1 score on the Occluded-DukeMTMC dataset, showing that simply using landmarks information may not significantly improve occluded Re-ID task. Our method PPAPN achieves 53.6% Rank-1 accuracy and 38.36% mAP, which outperforms all the competing methods. Compared with the state-of-the-art method PGFA [5], the PPAPN surpasses it by +2.2% Rank-1 accuracy and +1.06% mAP.

4.3.2. Results on Market-1501. Although some occluded Re-ID methods obtain improvements on occluded datasets, they may not get a satisfying performance on holistic datasets; this is caused by the noise during feature learning and alignment. In this part, we also evaluate our method PPAPN on a holistic person Re-ID dataset Market-1501. As shown in Table 3, our method still outperforms the state-of-the-art methods in terms of both metrics. This result demonstrates that our pose-guided adaptive skills do not bring negative effect.

4.3.3. Visualization. Figure 4 shows some retrieval examples of the PGFA [5] method and our PPAPN method on Occluded-DukeMTMC. The top-10 retrieval results are shown. The retrieval results show that PGFA is prone to mix the information of the target person and obstacles, resulting in retrieving a wrong person with a similar obstacle. On the contrary, our PPAPN makes no mistakes in the same situation.

5. Conclusion

In this paper, we propose a novel pose-guided part-based adaptive pyramid neural network for occluded person reidentification. Our foremost purpose is to let the model focus on nonoccluded body regions and meanwhile alleviate the negative impact of occlusion. We utilize pose landmarks to generate multiscale attention maps. The pyramid pooling is used to extract multiscale features in order to address the scale variation problem. The generated multiscale pyramid features are then multiplied by attention maps to achieve pose-guided adaptive pyramid features. The last but not the least, we propose a part-based matching approach to deal with the alignment problem. The human parts are defined dynamically based on pose landmarks and then the adaptive pyramid features are divided accordingly, which are fed into individual fully connected layers. In the end, all the part-based matching scores are fused with a weighted sum rule for reidentification. Our method is finally evaluated on the Occluded-DukeMTMC and Market-1501 datasets and is compared with the state-of-the-art methods. Experimental results demonstrate that our method can boost the person Re-ID performance in both occluded and nonoccluded scenarios.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] M. Cristani and V. Murino, "Person re-identification," *Academic Press Library in Signal Processing*, vol. 6, pp. 365–394, 2018.

[2] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3820–3828, IEEE, Venice, Italy, October 2017.

[3] W. Li, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 2194–2200, Melbourne, Australia, August 2017.

[4] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," http://arxiv.org/abs/1703.07737.

[5] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, "Pose-guided feature alignment for occluded person re-identification," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 542–551, IEEE, Seoul, South Korea, October 2019.

[6] S. Gao, J. Wang, H. Lu, and Z. Liu, "Pose-guided visible part matching for occluded person ReID," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11741–11749, IEEE, Seattle, WA, USA, June 2020.

[7] G. M. Zakharova, I. Bartoshevich, I. A. Rapoport, Q. Tian, and S. Wang, "Dependence of the mutagenic effect of N-nitroso-N-methylbiuret on the pH index," *Antibiotiki*, vol. 21, no. 9, pp. 795–798, 1976.

[8] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," 2018, https://arxiv.org/abs/1804.01438.

[9] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Pose-Driven Deep Convolutional Model for Person Re-identification," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3980–3989, IEEE, Venice, Italy, October 2017.

[10] H. Huang, W. Yang, X. Chen et al., "EANet: enhancing alignment for cross-domain person re-identification," 2018, https://arxiv.org/abs/1812.11369.

[11] N. Liu, Q. Zhao, N. Zhang, X. Cheng, and J. Zhu, "Pose-guided complementary features learning for amur tiger re-identification," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 286–293, IEEE, Seoul, South Korea, October 2019.

[12] C. Song, Y. Huang, W. Ouyang, and L. Wang, "Mask-guided contrastive attention model for person re-identification," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1179–1188, IEEE, Salt Lake City, UT, USA, June 2018.

[13] J. Zhuo, Z. Chen, J. Lai, and G. Wang, "Occluded person re-identification," 2018, https://arxiv.org/abs/1804.02792.

[14] H. Huang, X. Chen, and K. Huang, "Human parsing based alignment with multi-task learning for occluded person re-identification," in *Proceedings of the 2020 IEEE International*

Conference on Multimedia and Expo (ICME), London, UK, July 2020.

[15] D. Chen, S. Zhang, W. Ouyang, J. Yang, and Y. Tai, "Person search via a mask-guided two-stream CNN model," in *Computer Vision–ECCV*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., pp. 764–781, Springer International Publishing, Berlin, Germany, 2018.

[16] Q. Yang, P. Wang, Z. Fang, and Q. Lu, "Focus on the visible regions: semantic-guided alignment model for occluded person re-identification," *Sensors*, vol. 20, no. 16, p. 4431, 2020.

[17] P. Zhang, J. Lai, Q. Zhang, and X. Xie, "MGD: mask guided de-occlusion framework for occluded person re-identification," *Intelligence Science and Big Data Engineering. Visual Data Engineering*, pp. 411–423, Springer, Berlin, Germany, 2019.

[18] B. Chen, W. Deng, and J. Hu, "Mixed high-order attention network for person re-identification," in *Proceedings of the 2019 IEEE/CVF International Conference On Computer Vision (ICCV)*, pp. 371–381, IEEE, Seoul, South Korea, October 2019.

[19] L. Lin, D. Zhang, X. Zheng, M. Ye, and J. Guo, "Recurrent matching networks of spatial alignment learning for person re-identification," *Multimedia Tools and Applications*, vol. 79, no. 45-46, pp. 33735–33755, 2019.

[20] F. Tan, X. Zhao, K. Liu, and Q. Liao, "Person re-identification across non-overlapping cameras based on two-stage framework," in *Proceedings of the 2018 10th International Conference On Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 235–238, IEEE, Changsha, China, February 2018.

[21] W. Li, X. Zhu, and S. Gong, "Harmonious Attention Network for Person Re-identification," in *Proceedings of the 2018 IEEE/CVF Conference On Computer Vision and Pattern Recognition*, pp. 2285–2294, Salt Lake City, UT, USA, 2018.

[22] L. Wu, Y. Wang, J. Gao, and X. Li, "Deep adaptive feature embedding with local sample distributions for person re-identification," *Pattern Recognition*, vol. 73, pp. 275–288, 2018.

[23] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person Re-identification," in *Proceedings of the 26th ACM International Conference on Multimedia*, pp. 274–282, Seoul, South Korea, October 2018.

[24] H. Chen, B. Lagadec, and F. Bremond, "Learning discriminative and generalizable representations by spatial-channel partition for person re-identification," in *Proceedings of the 2020 IEEE Winter Conference On Applications Of Computer Vision (WACV)*, pp. 2472–2481, Snowmass Village, CO, USA, March 2020.

[25] C. Wan, Y. Wu, X. Tian, J. Huang, and X.-S. Hua, "Concentrated local Part Discovery with fine-grained Part Representation for person Re-identification," *IEEE Transactions on Multimedia*, vol. 22, no. 6, pp. 1605–1618, 2020.

[26] J. Zhuo, Z. Chen, J. Lai, and G. Wang, "Occluded person Re-identification," in *Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME)*, San Diego, CA, USA, July 2018.

[27] Y. Huang, P. Peng, Y. Jin, Y. Li, and J. Xing, "Domain adaptive attention learning for unsupervised person Re-identification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 11069–11076, 2020.

[28] L. He, J. Liang, H. Li, and Z. Sun, "Deep spatial feature reconstruction for partial person Re-identification: alignment-free approach," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7073–7082, Salt Lake City, UT, USA, June 2018.

[29] L. He, Y. Wang, W. Liu, H. Zhao, Z. Sun, and J. Feng, "Foreground-aware pyramid reconstruction for alignment-free occluded person Re-identification," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8449–8458, IEEE, Seoul, South Korea, 2019.

[30] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong, "Partial person Re-identification," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 4678–4686, IEEE, Santiago, Chile, December 2015.

[31] Y. Sun, Q. Xu, Y. Li et al., "Perceive where to focus: learning visibility-aware part-level features for partial person Re-identification," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 393–402, IEEE, Long Beach, CA, USA, November 2019.

[32] Z. Cao, T. Simon, S.-E. Wei, and Y. A. Sheikh, "Realtime multi-person 2D pose estimation using part affinity Ffelds," 2017, https://arxiv.org/abs/1611.08050.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.

[34] R. Lu and H. Ma, "Occluded pedestrian detection with visible IoU and box sign predictor," in *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1640–1644, Taipei, Taiwan, 2019.

[35] T.-Yi Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *Computer Vision – ECCV*, pp. 740–755, Springer International Publishing, Berlin, Germany, 2014.

[36] O. Russakovsky, J. Deng, H. Su, J. Krause, and S. Satheesh, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[37] X. Fan, H. Luo, X. Zhang, L. He, C. Zhang, and W. Jiang, "SCPNet: spatial-channel parallelism network for joint holistic and partial person Re-identification," in *Computer Vision–ACCV*, C. V. Jawahar, H. Li, G. Mori, and K. Schindler, Eds., pp. 19–34, Springer International Publishing, Berlin, Germany, 2019.