



Adaptive weight part-based convolutional network for person re-identification

Xiu Shu¹ · Di Yuan²  · Qiao Liu² · Jiaqi Liu³

Received: 3 July 2019 / Revised: 21 April 2020 / Accepted: 1 May 2020 /

Published online: 11 June 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

While part-based methods have been shown effective in the person re-identification task, it is unreasonable for most of them to treat each part equally, due to the retrieved image may be affected by deformation, occlusion and other factors, which makes the feature information of some parts unreliable. Instead of using the same weight of each part for the final person re-ID, we consider using an adaptive weight based on the part image information for each part for precise person retrieval. Specifically, we aim at learning discriminative part-informed features and propose an adaptive weight part-based convolutional network (AWPCN) for the person re-ID task. The core component of our AWPCN framework is an adaptive weight model, in which the part-based convolutional network and the adaptive weight model are used for feature refinement and feature-pair alignment, respectively. Given an image input at first, it outputs a convolutional descriptor consisting of several part-level features by the part-based convolutional network. And then, the corresponding weights of each part are determined by the adaptive weight model. Finally, we can use the adaptive weight part-based convolutional network joint to train each part loss and simultaneous optimization of its feature representations. We evaluate the proposed AWPCN model on Market-1501, DukeMTMC-reID and CUHK03 datasets. In extensive experiments, the AWPCN model outperforms most of the state-of-the-art methods on these representative datasets which clearly demonstrates the effectiveness of our proposed method. Our code will be released at <https://github.com/deasonyuan/AWPCN>.

Keywords Person re-identification · Adaptive weight · Part-based convolutional network

1 Introduction

Person re-identification (re-ID), also called pedestrian retrieval, across different cameras is a very fundamental problem in the computer vision community. The main task of person re-ID is finding the target person in an image gallery by comparing the query image

Xiu Shu and Di Yuan are contributed equally to this work and should be considered co-first authors

✉ Di Yuan
yuandi@stu.hit.edu.cn

Extended author information available on the last page of the article.

of this person with all other images in the same gallery, which has extensive applications in intelligent video surveillance, human-computer interaction, home robotics, especially in public safety. Despite its great progress in recent years, some realistic complex pedestrian retrieval environmental factors still make the re-ID task as a very challenging problem, such as background clutter, occlusion, deformation, and etc (Fig. 1).

Initially, the research of person re-ID mainly focused on design hand-crafted features and learn similarity metrics [6, 14, 18, 23, 31]. In the similarity metric aspect, besides Euclidean distance, cosine distance and Mahalanobis distance, some improved methods based on nearest neighbor are also proposed to reduce intra-class distance and increase inter-class distance [25, 42]. In order to reduce the computational cost of a transformation matrix in Mahalanobis distance, the regularization constraints are relaxed in [10]. Although some improvements have been achieved based on the design of some suitable hand-crafted features and learn matched similarity metrics, the performance of pedestrian recognition is still at a low level. With the emergence of the large-scale dataset and the development of deep learning, deep learning-based methods have been proposed for the person re-ID task [37, 44]. Compared with the hand-crafted feature-based methods, the deep learning-based methods can achieve unprecedented re-identification performance only by using some simple deep features. However, only based on deep features without jointly learning similarity metric holistically, these methods still cannot meet the requirements of the actual environment. Existing deep learning-based person re-ID methods either assume the availability of well-aligned person bounding box images as model input [3, 30, 58] or rely on constrained segmentation mechanisms to calibrate misaligned images [5, 36, 43, 52]. They are not very suitable for re-ID matching in arbitrarily aligned person images potentially with large human pose variations and part occlusions. There are some attention-based methods that attempt to use attention maps to indicate the importance of different pedestrian

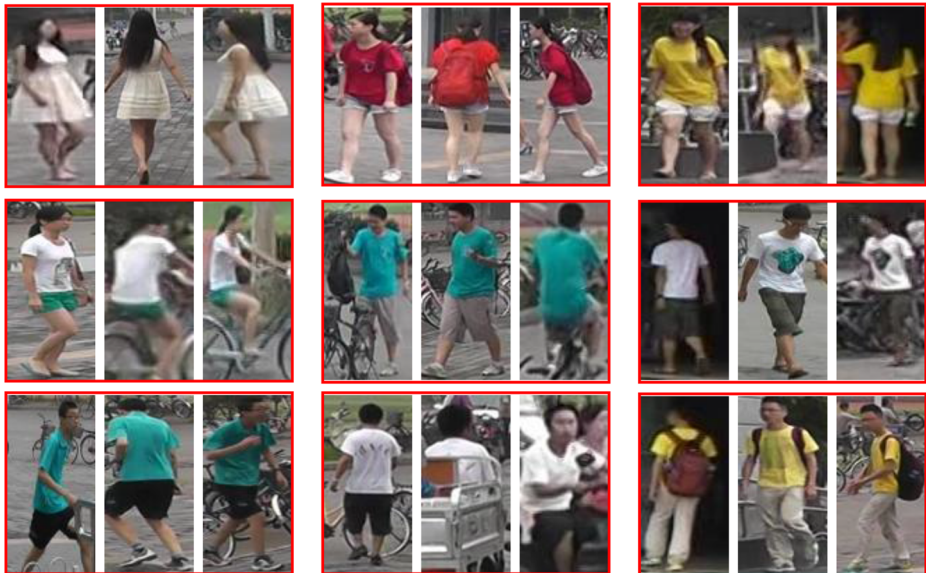


Fig. 1 Examples of person cross-view matched tasks from Market1501 [53]. Each bounding box shows the same person under different cameras

locations solving the re-ID problem [13, 17, 39]. Additionally, a part-based convolutional network as a powerful baseline is also applied to person re-ID tasks [28]. Nevertheless, these deep learning-based methods just simply adopt the existing deep architectures which have high complexity in model design. In addition, the part based methods often give the same weight of each part whilst ignoring the integrity and importance of the information contained in each image part. Hence, these techniques are ineffective when the target person is not filled with the whole image or when the target person is seriously deformed or occluded.

In this paper, we consider the problem of jointly part weight selection and deep feature representation for optimizing person re-ID in an adaptive weight part-based convolutional network. When some image patches influenced by occlusion or deformation, the representation ability of these patches will be reduced and the identification ability of the patch-based re-ID model will be affected. Only by adaptively reducing the weight of these corresponding patches can improve the re-ID performance of the model. Compared with the same weight of part-based methods and the attention-based methods, the proposed method has an adaptive weight model that can adjust the corresponding weight of each part adaptively. The adaptive weight model based on the similarity of the same image part of the same person in different images and the location information of each part in the whole image, which has two advantages: 1) Adjusting the weight of each image part based on the location information of each part in the whole image can effectively alleviate the defect of whether pedestrians fill the whole image. 2) Adjusting the weight of each image part based on the similarity of the same image part of the same person in the different images can effectively solve the problem of whether pedestrians have occlusion, deformation or not. Combining with these two aspects, our AWPCN method has achieved state-of-the-art re-ID performance. The main contributions of this paper are as follows:

- We formulate a novel method of jointly learning part weight selection and deep feature representation for optimizing person re-ID in the deep learning-based framework.
- We propose an adaptive weight part-based convolutional network (AWPCN) which can simultaneously divide the image into several parts, extract deep features and learn the corresponding weight of each part.
- Extensive comparative evaluations demonstrate the superiority of the proposed AWPCN model over a wide range of state-of-the-art re-ID models on three large benchmarks CUHK03 [37] Market-1501 [53], and DukeMTMC-ReID [56].

The rest of this paper is structured as follows. We first introduce some related works in Section 2. Next, we propose the adaptive weight part-based convolutional network for person re-ID, including the introduction of the baseline method and the adaptive weight model for the part-based convolutional network in Section 3. Subsequently, we introduce the implementation details and the evaluation criterion, evaluate and discuss our approach on some comprehensive benchmark datasets in Section 4. Finally, we briefly present the conclusion of this work in Section 5.

2 Related works

In this section, we introduce some re-ID methods closely related to our work in the proper context. A comprehensive review of re-ID methods is beyond the scope of this paper, and some survey papers can be found in [1, 15, 54].

2.1 Hand-crafted features-based re-ID methods

Previous to the popularity of deep learning technology, the earlier research on pedestrian retrieval mainly focused on how to design better hand-crafted features and how to learn better similarity measures. Different feature representations are suitable for different recognition scenarios [6, 14, 18, 23, 31, 47, 49, 50]. Common hand-crafted features was used for pedestrian image representation, mainly including color names [14], texture [23], scale invariant feature transform (SIFT) [18], histogram of oriented gradient (HOG) [6], etc. In [31], Tao et al. used the color histogram and texture features to characterize the image and proposed a regularly smoothed KISS metric to retrieve pedestrians in low-dimensional space through principal component analysis (PCA) dimensionality reduction. Pedagadi et al. [21] proposed a measurement method using PCA and local Fisher discriminant analysis in order to preserve the local neighborhood structure of the projected image. In [51], Zhang et al. proposed null Foley-Sammon transformation, which is a feature vector space with good learning and distinguishing ability. It satisfies zero intra-class divergence and positive inter-class divergence. In addition, there are many studies on pedestrian recognition from how to learn better similarity measure [7, 10, 12, 38, 42].

2.2 Deep learning-based re-ID methods

In recent years, with the development of deep learning methods [32–34, 46, 48], deep learning technology has been widely used in person re-ID tasks [4, 16, 20, 59]. Different from traditional re-ID methods, deep learning-based methods can automatically extract better pedestrian image features, while learning to obtain better similarity measures. When deep learning-based methods are adopted to person re-ID [37, 44], it is becoming more and more popular for this retrieval task. According to the different loss types, these deep learning-based re-ID methods can be divided into representation learning and metric learning. As we know the representation learning is a very common person re-ID method [20, 54–56, 59]. When the ultimate goal is to learn the similarity between two images, the representation learning-based method does not directly consider the similarity between images when training the network, but regards re-ID as a classification task [20]. The characteristic of this method is that the output of the last fully connected layer of the network is not the final image eigenvector, but a Softmax activation function to calculate the representation learning loss [55]. Using person ID or attributes as training labels to train the model, and the network can learn whether these two input images belong to the same pedestrian. Among the deep learning based person re-ID methods, the commonly used metric learning loss include contrastive loss [35, 36], triplet loss [9, 19] and quadruplet loss [3]. The comparative loss [35] is used to train the Siamese network. The input of the Siamese network is a pair of person images, which can be the same or different person. In [9], Hermans et al. verified that using a variant of the triplet loss to perform end-to-end deep metric learning can output some better re-ID performance. However, the triplet loss pays the main attention to obtaining correct orders on the training set. It still suffers from a weaker generalization capability from the training set to the testing set, thus resulting in inferior performance. In [3], a quadruplet loss has been proposed to lead the model output with a larger inter-class variation and a smaller intra-class variation compared to the triplet loss. In particular, a quadruplet deep network using a margin-based online hard negative mining is proposed based on the quadruplet loss for the person re-ID.

2.3 Patch-based re-ID methods

These methods described above is susceptible to noise interference and reduces its stability. In order to make the identification method more robust, several patch-based or attention-based methods have been proposed [5, 13, 17, 26, 28, 29, 36, 39, 41, 43, 52]. Figure 2 shows several partition strategies based models in person retrieval. Liu et al. [17] propose an attention-based deep neural network, which can explore the multi-scale selectiveness of attentive features to enrich the final feature representations for a pedestrian image.

In [28], Sun et al. proposes a part-based convolutional baseline network and uses a refined part pooling method to re-assigns outliers to the closest parts, resulting in refined

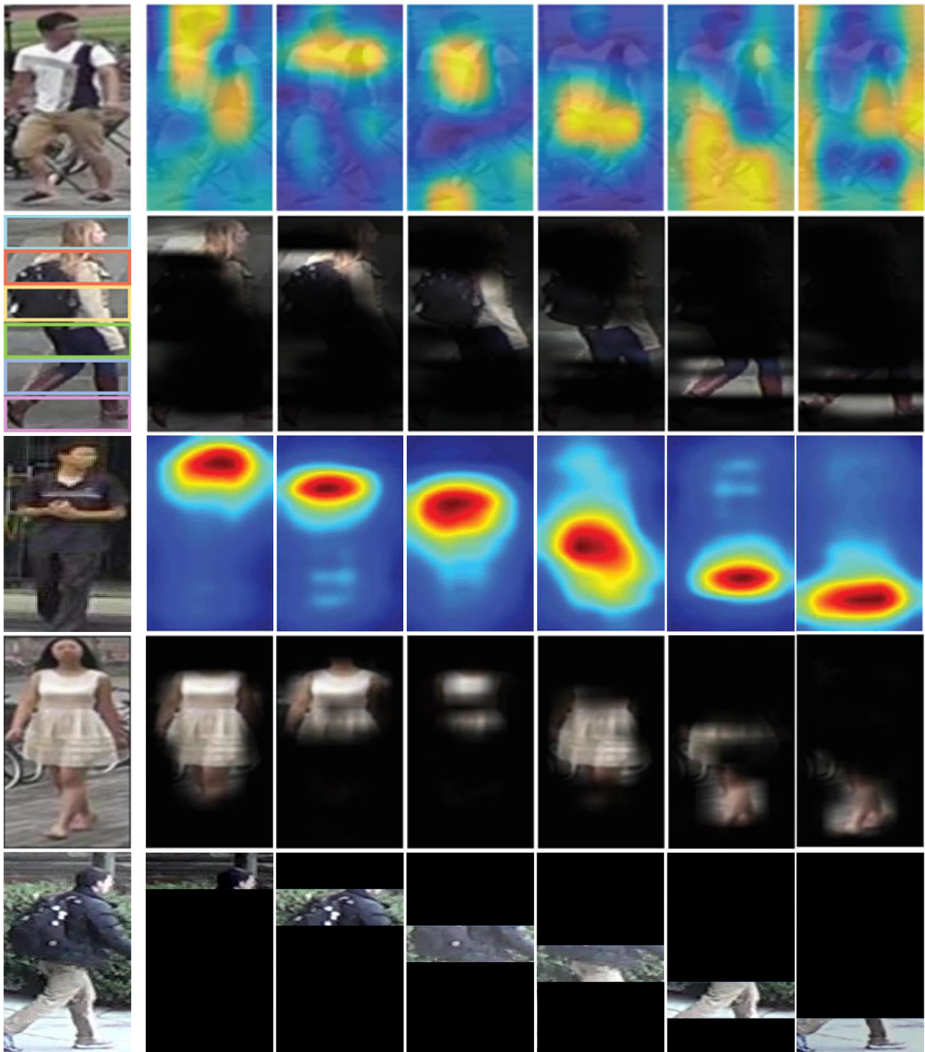


Fig. 2 Partition strategies of several deep part models in person re-ID. From top to bottom are Hydra-plus [17], PCB [28], PL-Net [43], PAR [52], and the proposed AWPCN, respectively

parts with enhanced within-part consistency, which achieved some good identification performance. The PL-Net [43] method can minimize both the empirical classification risk on training person images and the representation learning risk on unseen person images. The representation learning risk is evaluated by the proposed part loss, which automatically detects human body parts and computes the person classification loss on each part separately. In [52], Zhao et al. present a part-aligned representation (PAR) model for handling the body part misalignment problem. The PAR model decomposes the human body into parts which are discriminative for person matching. These part-based or attention-based algorithms all improve the robustness of the identification in varying degrees. Nevertheless, the integration strategy of these methods often uses the same weight of each patch or uses an attention strategy, which cannot show the advantage of the part-based method well. In this paper, we develop an adaptive weight model to mitigate this defect, which can adaptively apply the part similarity information and the part location information to determine the corresponding weight of each local part.

3 Adaptive weight part-based convolutional network

In this section, we introduce the adaptive weight part-based convolutional network for person re-ID. Firstly, we give a brief introduction to the part-based person re-ID methods. Then, we present the part-based convolutional network. Furthermore, we describe the adaptive weighting strategy to adjust the weight of different parts effectively. Finally, we show the re-ID pipeline of the proposed method, which is shown in Fig. 3.

3.1 Part-based convolutional network

Although the PCB [28] method achieves good identification performance, its part-based model trained with the same weight of each part is less effective for capturing the discriminative local structures of the target person. Different parts play different roles in the target representation, especially in the realistic pedestrian retrieval environment (maybe

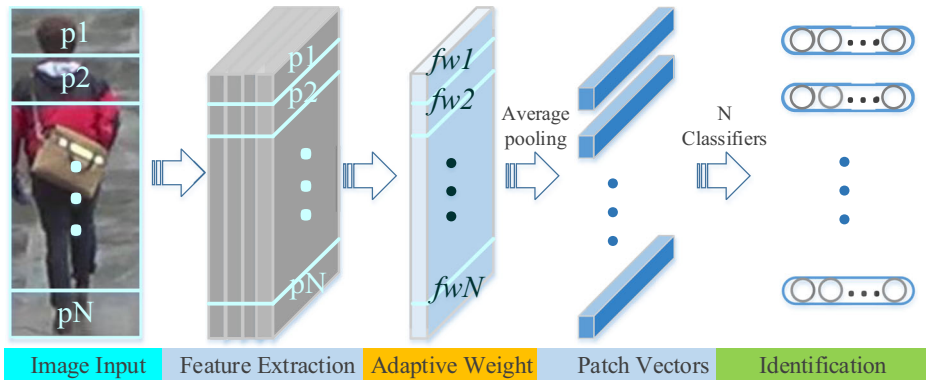


Fig. 3 Structure of the proposed AWPCN. First, the input image through the convolutional layers to extracted patch features. Then, the adaptive weight layer adjusts the importance of different patches in the identification task. Finally, each classifier is implemented with a fully-connected layer and a sequential Softmax layer to predict the identity (ID) of the input image

exist deformation, occlusion, etc.). In order to enhance the accuracy of the part-based convolutional framework, we propose an adaptive weight part-based convolutional network for the person re-ID task.

Intuitively, the target can be divided into several local parts, and the distance (or similarity) of each part is calculated separately. If the occlusion or interference occurs in some local areas, we can still retrieval the same person accurately through other unobstructed or undisturbed parts. The PCB model reshapes the backbone network (ResNet50) with some modifications. Specifically, it retained the structure before the global average pooling layer and removed it and followed. When an image undergoes the PCB network, it becomes several column vectors in the same stripe into a single part-level column vector. Each vector can predict the identity (ID) of the input image through a fully-connected layer and a Softmax function.

$$\text{Softmax}(W_i^T f) = \frac{\exp(W_i^T f)}{\sum_{j=1}^p \exp(W_j^T f)}, \quad (1)$$

where p denotes the number of pre-defined parts, f denotes a single part-level column vector, and W denotes the trainable weight matrix of the part classifier.

The PCB is optimized by minimizing the sum of Cross-Entropy losses over p pieces of ID predictions in the training stage.

3.2 Adaptive weight strategy

In the training process, there will be inconsonant changes in each part of the target, such as occlusion, deformation, etc. If the model uses the same weight to the p part into the integrated output directly, the reliability of these parts may be inconsistent with the same weight, which can reduce the identification performance. In fact, the corresponding weight should be suppressed if the image part is occluded, and vice versa. Therefore, we propose an adaptive weighting strategy to achieve adaptation.

For person re-ID task, the distance (or similarity) of two images can be used to quantify the reliability of each part and it can be calculated as:

$$\begin{aligned} \text{dis}_{I_1, I_2}^E &= \|f_{I_1} - f_{I_2}\|^2, \\ \text{dis}_{I_1, I_2}^C &= 1 - \frac{f_{I_1} \cdot f_{I_2}}{\|f_{I_1}\|^2 \|f_{I_2}\|^2}, \end{aligned} \quad (2)$$

where I_1, I_2 denotes two different images, dis_{I_1, I_2}^E and dis_{I_1, I_2}^C denotes the Euclidean distance and Cosine distance respectively.

Usually, the information on the center position of the image has the most credibility and discrimination. The closer to the edge, the lower the credibility of the information. The main reason of this phenomenon is that there are some problems in the way of image acquisition, such as pedestrians not filling up the whole image, pedestrians deformation, etc. According to this phenomenon, we suggest that the weight of the edge image part should be lower than that of the image part near the center of the image when the image is partitioned. So we decide the weight of the corresponding image is computed as:

$$\begin{aligned} w_p &= \frac{1}{\|p_c - I_c\|^2 + I_{dis}/p}, \\ w'_p &= \frac{w_p}{\sum w_p}, \end{aligned} \quad (3)$$

where w_p is the original weight of p -th image part, w'_p is the normalized weight of p -th image part, p_c denotes the central coordinates of p -th image part, I_c denotes the central coordinates of the whole image, I_{dis} denotes the length of the image.

In fact, besides the problems mentioned above, pedestrian deformation or occlusion can also significantly affect the performance of retrieving. Therefore, in the training process, we consider the similarity between the corresponding image blocks. The high similarity indicates that the image part is more important in the pedestrian representation, and the corresponding weight should be greater when determining the final pedestrian identification. We consider the similarity from the distance between image parts:

$$Similarity = \frac{1}{\|p_{I_1} - p_{I_2}\|^2 + \varepsilon}, \quad (4)$$

where p_{I_i} denotes p -th image part in different image of the same pedestrian, ε is a regularization parameter to avoid denominator zero (we set $\varepsilon = 0.5$). Combining the similarity measure with the location information mentioned above, we can set the final weight of each image part as follows:

$$fw_p = sw'_p \cdot \frac{Similarity(p)}{\sum Similarity(p)}, \quad (5)$$

where fw_p is the final weight of p -th image part which we used in the model.

As is shown in Fig. 3, we add the adaptive weight layer into the identification network after feature extraction layers directly, which can adaptively adjust the weight of each part. Each adaptive weight image feature feeds into the corresponding classifier which is implemented with a fully-connected layer and a sequential Softmax layer to predict the identity of the input pedestrian image.

4 Experiments

We evaluate the proposed AWPCN compare with other recently published state-of-the-art re-ID methods on three widely used benchmarks Market-1501, DukeMTMC-ReID and CUHK03 [37, 53, 56].

4.1 Datasets and evaluation metrics

There are three databases that are widely used for person re-ID, CUHK03 [37], Market-1501 [53] and DukeMTMC-ReID [56]. Table 1 shows a visualized information about these datasets. And we give a brief introduction for these datasets as bellow:

CUHK03 CUHK03 [37] is the first person re-identification dataset that is large enough for deep learning. It provides the bounding boxes detected from deformable part models and manually labeling, and we use the detected from in this paper. This re-ID dataset with 13,164

Table 1 Person re-ID evaluation datasets

| Datasets | # Image | # Cameras | # ID | # Train | # Test |
|--------------------|---------|-----------|-------|---------|--------|
| Market-1501 [53] | 32,668 | 6 | 1,501 | 751 | 750 |
| DukeMTMC-ReID [56] | 36,411 | 8 | 1,404 | 702 | 702 |
| CUHK03 [37] | 13,164 | 10 | 1,467 | 767 | 700 |

images collected from 10 cameras. There are 1,467 identities which are divided into two parts: 767 identities are used for training and the remaining 700 identities are used for testing.

Market-1501 This dataset [53] is one of the largest person re-identification benchmark datasets. It consists of 5 high-resolution (1280×1080) cameras and 1 low-resolution (720×576) camera with a total of 6 cameras. There are 32,668 bounding boxes of 1,501 identities: 751 identities are used for training and the remaining 750 identities are used for testing.

DukeMTMC-ReID This dataset [56] is also one of the largest person re-identification benchmark datasets collected from campus. There are 8 cameras used to collect image data. It directly uses the manually labeled ground-truth for training and testing. There are 36,411 manually labeled bounding boxes of 1404 identities: half of these identities are used for training and others are used for testing.

Evaluation metrics The evaluation metrics we used in this paper are provided by Market-1501 [53] and DukeMTMC-ReID [56], respectively. All the experiments evaluate the single-query setting. We mainly use Rank-1 and mAP as evaluation metrics for comparative experiments. In order to make the performance comparison clearer, we did not implement the re-ranking [57] mechanism in our model.

4.2 Implementation details

We use ResNet-50 [8] as the backbone network, changing the output size of the classifier to the number of identities in the training set. Cosine distance is used for similarity metrics. These settings are the same as PCB in [28]. All the input person images are resized to 384×192 . We empirically set the part number $p = 9$. The learning rate λ is set to 0.015 and multiplied by 0.1 after every 30 epoch. The network is trained in an end-to-end manner by Stochastic Gradient Descent (SGD) with a momentum of 0.9 and a weight decay of $5e - 4$. During training, we insert a dropout layer before the classifier to regularize the network and the dropout rates are set to 0.5 for all training sets. Figure 4 shows the training and verification losses on three datasets. From this figure we can see that the training and validation errors are almost stable after 30 rounds of training, so we terminated training after 60 epochs.

4.3 Comparison to the state-of-the-art

In this section, in order to verify the effectiveness of our proposed AWPCN method, we compare it with some recently published state-of-the-art methods on Market1501, DukeMTMC-reID and CUHK03 datasets [37, 53, 56].

Evaluation on market-1501 Table 2 shows comparison results of our AWPCN with some state-of-the-art on Market-1501 [53] dataset. In the top section of the table, we compare some approaches without any attention mechanism or partitioning strategy to our adaptive weight part-based convolutional network. In the second section of the table, we compare some part-based or attention-based approaches with our approach. Compare with these state-of-the-art methods, we can see that our method achieves the best accuracy on Rank-1 and the second-best accuracy on mAP metrics. Compare with the baseline PCB [28],

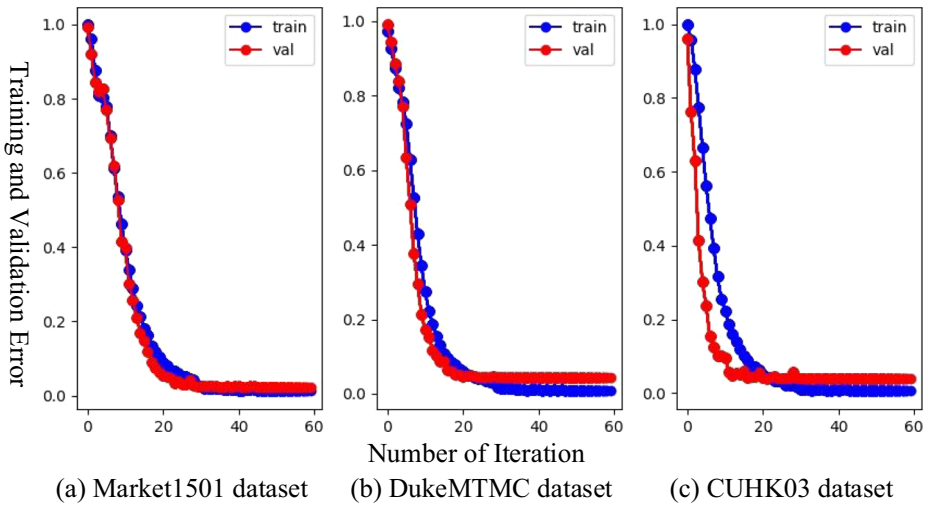


Fig. 4 Training and validation errors on three datasets

Table 2 Rank-1 and mAP results of the proposed AWPCN and some state-of-the-art methods on Market-1501 [53] dataset. The first and second best scores are highlighted in red and blue colors, respectively

| Methods | Reference | Rank-1 | mAP |
|---------------|-----------|-------------|-------------|
| GaSiam [35] | ECCV2016 | 65.9 | 39.6 |
| TriNet [9] | arXiv2017 | 84.9 | 69.1 |
| SVDNet [30] | ICCV2017 | 82.3 | 62.1 |
| DLCE [55] | TOMM2017 | 79.5 | 59.9 |
| MSCAN [16] | CVPR2017 | 80.3 | 57.5 |
| Re-rank [57] | CVPR2017 | 77.1 | 63.6 |
| SSSM [27] | CVPR2017 | 82.2 | 68.8 |
| CamStyle [58] | CVPR2018 | 89.5 | 71.6 |
| PSE [22] | CVPR2018 | 87.7 | 69.0 |
| AOSPR [11] | CVPR2018 | 86.5 | 70.4 |
| DGRW [24] | CVPR2018 | 92.7 | 82.5 |
| MAR [45] | CVPR2019 | 67.7 | 40.0 |
| SiamLSTM [36] | ECCV2016 | 61.6 | 35.2 |
| PDC [5] | ICCV2017 | 84.1 | 63.4 |
| PAR [52] | ICCV2017 | 81.0 | 63.4 |
| PL-Net [43] | ICCV2017 | 88.2 | 69.3 |
| PCB [28] | ECCV2018 | 92.3 | 77.4 |
| PCB+RPP [28] | ECCV2018 | 93.8 | 81.6 |
| AACN [13] | CVPR2018 | 85.9 | 66.9 |
| HA-CNN [39] | CVPR2018 | 91.2 | 75.7 |
| DuATM [26] | CVPR2018 | 91.4 | 76.6 |
| PAUL [41] | CVPR2019 | 68.5 | 40.1 |
| VPM [29] | CVPR2019 | 93.0 | 80.8 |
| AWPCN(Ours) | This work | 94.0 | 82.1 |

the proposed method adding an adaptive weight model on the base part-based convolutional network, which increases 1.7%/4.7% on Rank-1 and mAP metrics. It's better than the refined part pooling (PCB+RPP) method. Compare with the patch-based PAR [52] method, the proposed AWPCN method achieved more than 10% improvement on both Rank-1 and mAP metrics. The above experimental comparisons verify the effectiveness of our adaptive weighting method for the patch-based re-ID model. Figure 5 shows some re-ID samples on the Market-1501 dataset. The images in the first column are the query images. The re-ID images are sorted according to the level of similarity from the left to the right. Most candidate images are correctly re-identified. Although the network re-identifies some

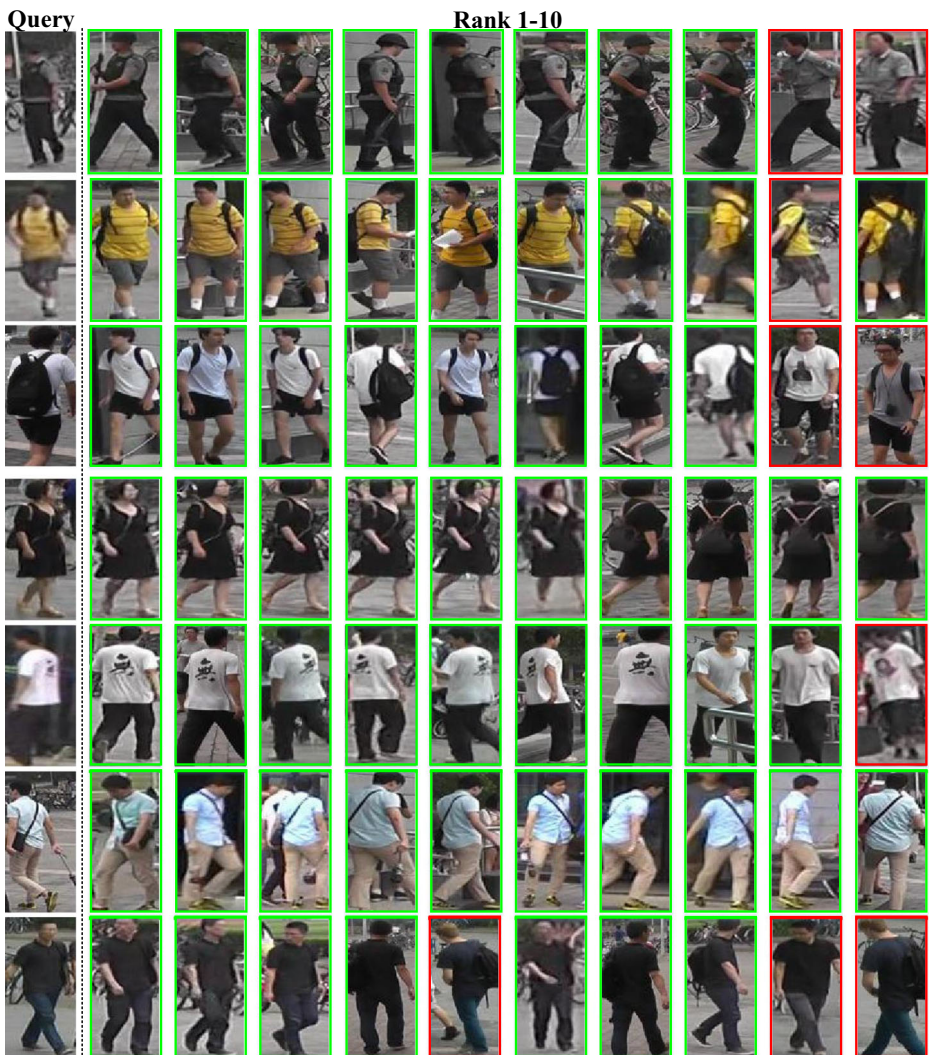


Fig. 5 Person re-ID samples on the Market-1501 [53]. The first column are query images. The re-ID images are sorted according to the level of similarity from left to right. The correct and false matches are in the green bounding-box and the red bounding-box, respectively

incorrect candidates in some rows, we find that most of the candidate samples for incorrect re-identification are in the latter position, which also shows that our model has a perfect re-ID performance.

Evaluation on DukeMTMC-ReID We report some competitive results in Table 3 on DukeMTMC-ReID [56] dataset. Compared to recently proposed part-based methods, PCB (baseline) and PCB+RPP [28], our AWPCN achieves 4.0% and 2.4% rank-1 accuracy improvements and 8.0% and 4.9% mAP improvements respectively. Compared to recently proposed attention-based methods, HA-CNN [39] and DuATM [26], our AWPCN achieves 5.2% and 3.9% rank-1 accuracy improvements and 10.3% and 9.5% mAP improvements respectively. All these improvements are benefit from the adaptive weight part-based convolutional network.

Evaluation on CUHK03 Table 4 shows comparison results of the proposed AWPCN and some state-of-the-art methods on CUHK03 [37] dataset. As the Table shows that the rank-1 and the mAP of AWPCN are 63.7%/62.8% with detected bounding boxes, which is the best or close to the best identification performance. The performance of our method is superior to that of the baseline [28], which benefits from the proposed adaptive weight model. Compare with the harmonious attention network-based HA-CNN [39] method, the proposed AWPCN method achieved more than 20% improvement on both Rank-1 and mAP metrics. These experimental comparisons verify the effectiveness of our adaptive weighting method in the patch-based model for the person re-ID task.

4.4 Discussion

Due to our AWPCN is an adaptive weight patch-based re-ID method, in this section, we mainly discuss the number of parts p which is essential to our re-ID performance. Table 5

Table 3 Rank-1 and mAP results of the proposed AWPCN and some state-of-the-art methods on DukeMTMC-ReID [56] dataset. The first and second best scores are highlighted in red and blue colors, respectively

| Methods | Reference | Rank-1 | mAP |
|---------------|-----------|-------------|-------------|
| SVDNet [30] | ICCV2017 | 76.7 | 56.8 |
| CamStyle [58] | CVPR2018 | 78.3 | 57.6 |
| PSE [22] | CVPR2018 | 79.8 | 62.0 |
| AOSPR [11] | CVPR2018 | 84.1 | 78.2 |
| DGRW [24] | CVPR2018 | 80.7 | 66.4 |
| MAR [45] | CVPR2019 | 67.1 | 48.0 |
| PCB [28] | ECCV2018 | 81.7 | 66.1 |
| PCB+RPP [28] | ECCV2018 | 83.3 | 69.2 |
| AACN [13] | CVPR2018 | 76.8 | 59.3 |
| HA-CNN [39] | CVPR2018 | 80.5 | 63.8 |
| DuATM [26] | CVPR2018 | 81.8 | 64.6 |
| PAUL [41] | CVPR2019 | 72.0 | 53.2 |
| VPM [29] | CVPR2019 | 83.6 | 72.6 |
| AWPCN(Ours) | This work | 85.7 | 74.1 |

Table 4 Rank-1 and mAP results of the proposed AWPCN and some state-of-the-art methods on CUHK03 [37] dataset. The first and second best scores are highlighted in red and blue colors, respectively

| Methods | Reference | Rank-1 | mAP |
|---------------|-----------|-------------|-------------|
| SiamLSTM [36] | ECCV2016 | 57.3 | 46.3 |
| GaSiam [35] | ECCV2016 | 68.1 | 58.8 |
| Re-rank [57] | CVPR2017 | 34.7 | 37.4 |
| DPFL [4] | ICCVW2017 | 40.7 | 37.0 |
| AOSPR [11] | CVPR2018 | 54.6 | 56.1 |
| PSE [22] | CVPR2018 | 27.3 | 32.0 |
| HA-CNN [39] | CVPR2018 | 41.7 | 38.6 |
| DaRe+RE [40] | CVPR2018 | 63.3 | 59.0 |
| MLFN [2] | CVPR2018 | 52.8 | 47.8 |
| PCB [28] | ECCV2018 | 61.3 | 54.2 |
| PCB+RPP [28] | ECCV2018 | 63.7 | 57.5 |
| AWPCN(Ours) | This work | 67.3 | 62.8 |

shows the rank-1 and mAP results of the proposed method under different p value. In fact, the value of p directly determines the granularity of the part feature. $p = 1$ means the part-based convolutional network learn global feature. As p from 1 increases to 9, the identification accuracy gradually improves. However, the accuracy does not always increase with the value of p . As show in Table 5, when p increases from 9 to 15, the re-ID performance decreases significantly. As a comparison, we also report the performance of $p = 20, 25, 30$. From Table 5, we know that an over-increased value of p actually compromises the discriminative ability of the part features. This is mainly because too much segmentation of the image will reduce the representation ability of each image patch, thus reducing the re-ID ability of the model. Therefore, we suggest using $p = 9$ in practical applications.

Table 5 Rank-1 and mAP results of the proposed AWPCN method under different value of p on Market-1501 [53] and DukeMTMC-ReID [56] datasets

| Part number | Market-1501 | | DukeMTMC-ReID | |
|-------------|-------------|-------------|---------------|-------------|
| | Rank-1 | mAP | Rank-1 | mAP |
| $p = 1$ | 85.4 | 68.5 | 77.6 | 58.7 |
| $p = 3$ | 89.4 | 74.4 | 80.3 | 64.1 |
| $p = 5$ | 92.7 | 79.9 | 83.7 | 72.0 |
| $p = 7$ | 93.2 | 80.2 | 84.6 | 72.6 |
| $p = 9$ | 94.0 | 82.1 | 85.7 | 74.1 |
| $p = 11$ | 92.4 | 77.9 | 82.3 | 72.2 |
| $p = 13$ | 91.3 | 74.2 | 80.5 | 63.1 |
| $p = 15$ | 90.2 | 70.3 | 79.1 | 62.0 |
| $p = 20$ | 89.7 | 69.8 | 78.5 | 61.6 |
| $p = 25$ | 88.5 | 67.7 | 77.4 | 60.7 |
| $p = 30$ | 87.4 | 64.6 | 76.5 | 60.2 |

5 Conclusions

In this work, we show the advantage of the part-based convolutional network for feature representation and the necessity of adaptive weight for each part combination. Specifically, we formulate a novel adaptive weight model for joint training each part loss along with simultaneous optimization of its feature representations, dedicated to optimizing person re-id in misaligned/deformation images. Extensive comparative evaluations validate the superiority of this new adaptive weight part-based convolutional network for person re-ID over a wide variety of state-of-the-art methods on three large-scale benchmarks including CUHK03, Market-1501, and DukeMTMC-ReID.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant No. 61672183), by the Natural Science Foundation of Guangdong Province (Grant No. 2015A030313544), by the Shenzhen Research Council (Grant No. JCYJ20170413104556946, JCYJ20170815113552036), and by the project "The Verification Platform of Multi-tier Coverage Communication Network for Oceans (PCL2018KP002)". Di Yuan is supported by a scholarship from China Scholarship Council (CSC).

References

1. Bedagkar-Gala A, Shah SK (2014) A survey of approaches and trends in person re-identification. *Image Vis Comput* 32(4):270–286
2. Chang X, Hospedales TM, Tao X (2018) Multi-level factorisation net for person re-identification. In: *Computer vision and pattern recognition*, pp 2109–2118
3. Chen W, Chen X, Zhang J, Huang K (2017) Beyond triplet loss: a deep quadruplet network for person re-identification. In: *Computer vision and pattern recognition*, pp 1320–1329
4. Chen Y, Zhu X, Gong S (2017) Person re-identification by deep learning multi-scale representations. In: *International conference on computer vision workshop*, pp 2590–2600
5. Chi S, Li J, Zhang S, Xing J, Wen G, Qi T (2017) Pose-driven deep convolutional model for person re-identification. In: *International conference on computer vision*, pp 3960–3969
6. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: *Computer vision and pattern recognition*, pp 886–893
7. Geng S, Yu M, Liu Y, Yu Y, Bai J (2019) Re-ranking pedestrian re-identification with multiple metrics. *Multimed Tools Appl* 78(9):11631–11653
8. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Computer vision and pattern recognition*, pp 770–778
9. Hermans A, Beyer L, Leibe B (2017) Defense of the triplet loss for person re-identification. [arXiv:1703.07737](https://arxiv.org/abs/1703.07737)
10. Hirzer M, Roth PM, K?stinger M, Bischof H (2012) Relaxed pairwise learned metric for person re-identification. In: *European conference on computer vision*, pp 780–793
11. Huang H, Li D, Zhang Z, Chen X, Huang K (2018) Adversarially occluded samples for person re-identification. In: *Computer vision and pattern recognition*, pp 5098–5107
12. Jia J, Ruan Q, An G, Yi J (2017) Multiple metric learning with query adaptive weights and multi-task re-weighting for person re-identification. *Comput Vis Image Underst* 160:87–99
13. Jing X, Rui Z, Feng Z, Wang H, Ouyang W (2018) Attention-aware compositional network for person re-identification. In: *Computer vision and pattern recognition*, pp 2119–2128
14. Kuo CH, Khamis S, Shet V (2013) Person re-identification using semantic color names and rankboost. In: *IEEE workshop on applications of computer vision*, pp 281–287
15. Lavi B, Serj MF, Ullah I (2018) Survey on deep learning techniques for person re-identification task. [arXiv:1807.05284](https://arxiv.org/abs/1807.05284)
16. Li D, Chen X, Zhang Z, Huang K (2017) Learning deep context-aware features over body and latent parts for person re-identification. In: *Computer vision and pattern recognition*, pp 384–393
17. Liu X, Zhao H, Tian M, Lu S, Wang X (2017) Hydraplus-net: attentive deep features for pedestrian analysis. In: *International conference on computer vision*, pp 350–359
18. Lowe DG (1999) Object recognition from local scale-invariant features. In: *International conference on computer vision*, pp 1150–1157

19. Ma F, Zhu X, Zhang X, Yang L, Zuo M, Jing X-Y (2019) Low illumination person re-identification. *Multimed Tools Appl* 78(1):337–362
20. Matsukawa T, Suzuki E (2016) Person re-identification using cnn features learned from combination of attributes. In: International conference on pattern recognition, pp 2428–2433
21. Pedagadi S, Orwell J, Velastin S, Boghossian B (2013) Local fisher discriminant analysis for pedestrian re-identification. In: Computer vision and pattern recognition, pp 3318–3325
22. Sarfraz MS, Schumann A, Eberle A, Stiefelwagen R (2018) A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: Computer vision and pattern recognition, pp 420–429
23. Satta R (2013) Appearance descriptors for person re-identification: a comprehensive review. arXiv:1307.5748
24. Shen Y, Li H, Xiao H, Yi S, Chen D, Wang X (2018) Deep group-shuffling random walk for person re-identification. In: Computer vision and pattern recognition, pp 2265–2274
25. Si J, Zhang H, Li CG, Guo J (2018) Spatial pyramid-based statistical features for person re-identification: a comprehensive evaluation. *IEEE Trans Sys Man Cybern Sys* 48(7):1140–1154
26. Si J, Zhang H, Li CG, Kuen J, Gang W (2018) Dual attention matching network for context-aware feature sequence based person re-identification. In: Computer vision and pattern recognition, pp 5363–5372
27. Song B, Xiang B, Qi T (2017) Scalable person re-identification on supervised smoothed manifold. In: Computer vision and pattern recognition, pp 2530–2539
28. Sun Y, Liang Z, Yi Y, Qi T, Wang S (2018) Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: European conference on computer vision, pp 501–518
29. Sun Y, Xu Q, Li Y, Zhang C, Li Y, Wang S, Sun J (2019) Perceive where to focus: learning visibility-aware part-level features for partial person re-identification. arXiv:1904.00537
30. Sun Y, Zheng L, Deng W, Wang S (2017) Svdnet for pedestrian retrieval. In: International conference on computer vision, pp 3820–3828
31. Tao D, Jin L, Wang Y, Yuan Y, Li X (2013) Person re-identification by regularized smoothing kiss metric learning. *IEEE Trans Circ Sys Video Technol* 23(10):1675–1685
32. Tian C, Fei L, Zheng W, Xu Y, Zuo W, Lin C-W (2019) Deep learning on image denoising: an overview. arXiv:1912.13171
33. Tian C, Xu Y, Li Z, Zuo W, Fei L, Liu H (2020) Attention-guided cnn for image denoising. *Neural Networks*. <https://doi.org/10.1016/j.neunet.2019.12.024>
34. Tian C, Xu Y, Zuo W (2020) Image denoising using deep cnn with batch renormalization. *Neural Netw* 121:461–473
35. Varior RR, Haloi M, Gang W (2016) Gated siamese convolutional neural network architecture for human re-identification. In: European conference on computer vision, pp 791–808
36. Varior RR, Shuai B, Lu J, Xu D, Wang G (2016) A siamese long short-term memory architecture for human re-identification. In: European conference on computer vision, pp 135–153
37. Wei L, Rui Z, Tong X, Wang XG (2014) Deepreid: deep filter pairing neural network for person re-identification. In: Computer vision and pattern recognition, pp 152–159
38. Wei L, Yang W, Li J (2017) Re-identification by neighborhood structure metric learning. *Pattern Recogn* 61:327–338
39. Wei L, Zhu X, Gong S (2018) Harmonious attention network for person re-identification. In: Computer vision and pattern recognition, pp 2285–2294
40. Yan W, Wang L, You Y, Xu Z, Weinberger KQ (2018) Resource aware person re-identification across multiple resolutions. In: Computer vision and pattern recognition, pp 8042–8051
41. Yang Q, Yu H-X, Wu A, Zheng W-S (2019) Patch based discriminative feature learning for unsupervised person re-identification. In: Computer vision and pattern recognition
42. Yang Y, Liao S, Zhen L, Li SZ (2016) Large scale similarity learning using similar pairs for person verification. In: Thirtieth AAAI conference on artificial intelligence, pp 3655–3661
43. Yao H, Zhang S, Hong R, Zhang Y, Tian Q (2019) Deep representation learning with part loss for person re-identification. *Trans Image Process* 28(6):2860–2871
44. Yi D, Lei Z, Liao S, Li SZ (2014) Deep metric learning for person re-identification. In: International conference on pattern recognition, pp 34–39
45. Yu HX, Zheng WS, Wu A, Guo X, Lai JH (2019) Unsupervised person re-identification by soft multilabel learning. arXiv:1903.06325
46. Yuan D, Fan N, He Z (2020) Learning target-focusing convolutional regression model for visual object tracking. *Knowledge-Based Systems*. <https://doi.org/10.1016/j.knsys.2020.105526>
47. Yuan D, Kang W, He Z (2020) Robust visual tracking with correlation filters and metric learning. *Knowledge-Based Systems*. <https://doi.org/10.1016/j.knsys.2020.105697>

48. Yuan D, Li X, He Z, Liu Q, Lu S (2020) Visual object tracking with adaptive structural convolutional network. *Knowledge-Based Systems*. <https://doi.org/10.1016/j.knosys.2020.105554>
49. Yuan D, Lu X, Li D, Liang Y, Zhang X (2019) Particle filter re-detection for visual tracking via correlation filters. *Multimed Tools Appl* 78(11):14277–14301
50. Yuan D, Zhang X, Liu J, Li D (2019) A multiple feature fused model for visual object tracking via correlation filters. *Multimed Tools Appl* 78(19):27271–27290
51. Zhang L, Xiang T, Gong S (2016) Learning a discriminative null space for person re-identification. In: *Computer vision and pattern recognition*, pp 1239–1248
52. Zhao L, Xi L, Wang J, Zhuang Y (2017) Deeply-learned part-aligned representations for person re-identification. In: *International conference on computer vision*, pp 3219–3228
53. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: *International conference on computer vision*, pp 1116–1124
54. Zheng L, Yang Y, Hauptmann AG (2016) Person re-identification: past, present and future. [arXiv:1610.02984](https://arxiv.org/abs/1610.02984)
55. Zheng Z, Zheng L, Yang Y (2017) A discriminatively learned cnn embedding for person re-identification. *ACM Transactions on Multimedia Computing Communications and Applications* 14(1):13:1–13:20
56. Zheng Z, Zheng L, Yang Y (2017) Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: *International conference on computer vision*, pp 3774–3782
57. Zhong Z, Zheng L, Cao D, Li S (2017) Re-ranking person re-identification with k-reciprocal encoding. In: *Computer vision and pattern recognition*, pp 384–393
58. Zhong Z, Zheng L, Zheng Z, Li S, Yang Y (2018) Camera style adaptation for person re-identification. In: *Computer vision and pattern recognition*, pp 5157–5166
59. Zhu F, Kong X, Wu Q, Fu H, Li M (2018) A loss combination based deep model for person re-identification. *Multimed Tools Appl* 77(3):3049–3069

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Xiu Shu¹ · Di Yuan²  · Qiao Liu² · Jiaqi Liu³

Xiu Shu
shuxiuhit@gmail.com

Qiao Liu
liuqiao.hit@gmail.com

Jiaqi Liu
jiaqi_liu1993@163.com

¹ School of Science, Harbin Institute of Technology, Shenzhen, China

² School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China

³ College of Finance and Statistics, Hunan University, Changsha, China