IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 4, 2020, accepted February 24, 2020, date of publication March 4, 2020, date of current version March 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978344

A Comprehensive Overview of Person Re-Identification Approaches

HONGBO WANG, HAOMIN DU, YUE ZHAO, AND JIMING YAN
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
Corresponding author: Hongbo Wang (hbwang@bupt.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61002011.

ABSTRACT Person re-identification, identifying and tracking pedestrians in cross-domain monitoring systems, is an important technology in the computer vision field and of real significance for the construction of smart cities. With the development of deep learning techniques, especially convolutional neural networks, this technology has received more extensive attention and improvement in recent years and a large number of noteworthy achievements have emerged. This paper provides a comprehensive overview of person re-identification approaches to assist researchers in quickly understand this field with preference as well as to provide a more structured framework. By reviewing more than 300 re-identification related papers, the focus of these studies is summarized as information extraction, metric learning, post-processing, efficiency improvement, labeling cost reduction, and data type expansion. This classification is then organized based on different technologies, and on this basis, the pros and cons of each technology are analyzed. Moreover, this overview summarizes the difficulties and challenges of re-identification and discusses the possible research directions for reference.

INDEX TERMS Computer vision, convolutional neural networks, distance learning, feature extraction, person re-identification.

I. INTRODUCTION

With the construction of smart cities, a large number of high-definition surveillance cameras have been installed in cities around the world, which generate numerous surveillance videos. The pedestrian data' analyzing and utilizing of among them is a practical subject. These results can not only help the police quickly get clues to the suspect or find missing children and the elderly, but also enable the merchants to obtain the rules of pedestrian behavior and make more reasonable use of commercial value for resource allocation. Therefore, person re-identification [1] emerges as the times require, which is the key technology to be solved in the construction of intelligent surveillance systems. This technology has also been applied in many practical scenarios, including human traffic statistics, street event detection, and person behavior analysis.

Person re-identification [1] refers to using computer vision technology to determine the identity of a particular person in the monitoring system with multiple non-overlapping cameras. Given an interesting query, the person re-identification task devotes to retrieve other images or video sequences from the gallery collected by the multi-camera system and determine whom the person is, as illustrated in Fig. 1 (c).

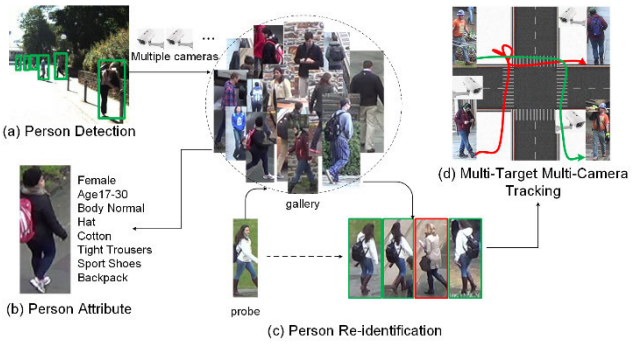


FIGURE 1. Differences and connections between (a) person detection, (b) person attribute, (c) person re-identification, and (d) multi-target multi-camera tracking. The person images among this paper are from the DukeMTMC-reID [82] dataset.

The concept of person re-identification was first defined in 2006 [5]. So far, this field has experienced rapid development, and scholars around the world have done extensive research on this task and have developed many outstanding approaches. But these approaches are numerous and have different focuses. Therefore, for enabling researchers preparing

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Shao.

TABLE 1. Conclusion of several person re-identification reviews in recent years.

Year	Survey	Contributions
2013	People reidentification in surveillance and forensics: A survey [179]	1. summarized the person re-identification research from the five dimensions of camera setting, sample set, signature, body model, machine learning algorithms, and application scenario. 2. focused on the feature space and its connections to human body models
2014	A survey of approaches and trends in person re-identification [180]	1. comprehensively concluded the research works on person re-identification before 2016 from task scenarios, current work, public datasets, and evaluation metrics.
2016	Person re-identification: Past, present and future [1]	1. reviewed person re-identification from the datatype of image and video, as well as the hand-crafted and deep learning system. 2. made an accurate judgment on the future research trends, which is consistent with current.
2017	A study on deep convolutional neural network based approaches for person re-identification [9]	1. briefly summarized the researches and datasets of person re-identification based on deep learning for image and video data type.
2018	Survey on person re-identification based on deep learning [127]	1. focused on the related papers about different person re-identification frameworks based on deep learning, such as CNN, RNN, GAN.
2019	A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets [178]	1. presented a performance evaluation of single-shot and multi-shot re-identification algorithms that were implemented using a unified code library for directly comparing re-identification algorithms. 2. published a new, real-world, large-scale dataset collected at the airport.
2019	Beyond intra-modality discrepancy: A comprehensive survey of heterogeneous person re-identification [181]	1. reviewed the state-of-the-art heterogeneous person re-identification methods comprehensively concerning four main application scenarios low-resolution, infrared, sketch and text.
2019	Deep learning-based methods for person re-identification: A comprehensive review [10]	1. compared the effects of four loss functions with two common baseline models. 2. summarized the person re-identification datasets in three categories: RGB image-based datasets, video sequence-based datasets, and cross-modality datasets.
2019	A Survey on Deep Learning-Based Person Re-Identification Systems [191]	1. discussed the deep learning-based person re-identification systems from the perspective of data type. 2. compared the details of existing state-of-the-art models and their accuracy on common datasets.

to enter this field to quickly understand its development status and to provide researchers in this field with an overview and valuable research directions, we investigate the person re-identification field in depth and summarize the related concepts and achievements so far.

Prior to this survey, some researchers [1], [9], [10], [127], [178]–[187], [189]–[191] have also reviewed re-identification field. In Table 1, we summarize the major contributions of these reviews. However, there are still some improvements to be made in these reviews. Some of these surveys [179], [180] summarized the person re-identification before the outbreak of deep learning, which has developed rapidly after 2014 and become the main research means. Other surveys summarized these outcomes based solely on datatype [1], [9], [191], framework [127], or performance evaluation [178], or reviewed the heterogeneous person re-identification methods [181]. Recently, Wu *et al.* [10] have outlined the person re-identification datasets, loss functions and methods. But this survey lacked logical classification in methods and missed many characteristics and important parts for re-identification, such as re-ranking and camera network-based methods.

On the basis of predecessors, this paper supplemented the missing parts of the above surveys and the rapidly developing means in recent years, as well as generalized re-identification approaches from a more purposeful dimension. Specifically, our work's contributions come from three aspects:

1) We summarize the main contributions of several person re-identification reviews and explain the relevance of the four human-related studies that are critical to person re-identification. However, no one has summed these up yet as far as we know.
2) We fully investigate the research approaches of person re-identification. Different from the previous surveys, these approaches are divided into seven categories according to their purposes. This classification is more suitable for researchers to explore these approaches from their actual needs.
3) We condense the main constraints in the person re-identification field and consider that there is still enough necessity to research person re-identification. Furthermore, we summarize six possible research directions in two aspects for person re-identification researchers.

The remaining parts of this overview are structured as follows. Section II presents the preliminary on person re-identification, including development course, datasets, and evaluation metrics. Section III summarizes the person re-identification approaches according to the research purpose. Section IV discusses performance comparison, research challenges and possible research directions for reference. Section V outlines the work of this overview.

II. PRELIMINARY

In this section, we review the development, datasets, and evaluation metrics of person re-identification as a support for the subsequent content, including the relationship of four kinds of human-related researches and the characteristics of several common datasets.

A. DEVELOPMENT COURSE

Person, one of the most concerned parts of the data collected by city cameras, is getting a lot of attention in the field of computer vision. So there are various person-related researches, such as detection [280], attributes [11],

re-identification, tracking [283], action recognition [284], semantic segmentation [282], and face recognition [281]. Person re-identification is closely connected with other person-related researches. Person detection, a target detection technology, requires determining whether a person appears in the video picture of a single camera, positioning this person, and then giving the cropped image [2]. Later, the fast development of person re-identification technology is based on target detection technology, but it is different from person detection. The person re-identification datasets should include identities captured from multiple cameras, preferably covering different scenes and regions. Generally, these identities are manually labeling or cropped using person detection technology. If there has mistaken in cropping, the accuracy of the person re-identification will be affected. Pedestrian attributes, such as gender, age, and clothing, are conducive to person re-identification [11]. Generally, based on the person re-identification result and combined with geographical location and regional distribution, the trail of the target person can be mapped, which is multi-target multi-camera (MTMC) tracking technology [12]. Fig. 1 presents the relationships and differences between these four person-related studies.

Person re-identification is originated from MTMC technology and its development can be divided into two distinct stages. Before 2014, person re-identification mainly relied on the traditional hand-designed features [6]–[8]. In 2003, Porikli [3] used the correlation coefficient matrix to establish a non-parametric model between camera pairs to obtain the color distribution of the target between different cameras and achieved cross-view target matching. In 2005, Zajdel *et al.* [4] first proposed re-identification in the multi-camera tracking task. In 2006, Gheissari *et al.* [5] used color and salient edge histograms for person re-identification, which was the first time to raise the concept of ''person re-identification''. In 2010, Farenzena *et al.* [6] published the first person re-identification article at the Conference on Computer Vision and Pattern Recognition (CVPR), a top conference in the computer vision field.

In 2014, deep learning [67], [68] was first used in the person re-identification field. Since then, the Convolutional Neural Network [63] (CNN), Recurrent Neural Network (RNN), and other deep learning technologies have been used to design end-to-end models for automatic feature extraction [52]. Fig. 2 illustrates that there has been a significant increase in the proportion of collected person re-identification papers since 2014. At the same time, researchers had achieved lots of breakthroughs in person re-identification performance and greatly promoted the development of person re-identification technology.

Basically, means on person re-identification have been divided into two categories. The simplest idea is to treat it as a binary task and input two images into a model at the same time. This model does not need the identity information of the two images, only output judgment result. If they were the same person, then output ''1'', otherwise output ''0'' [67], [68].

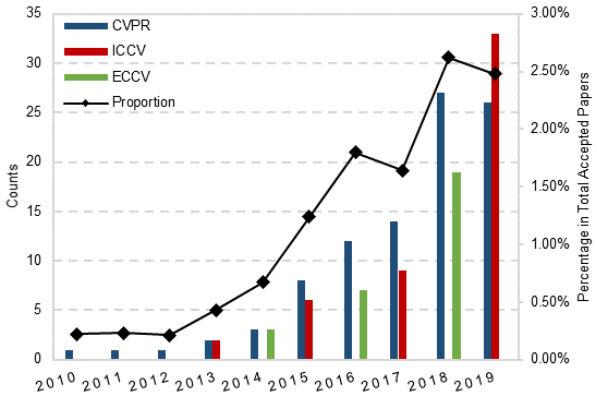


FIGURE 2. The number of accepted person re-identification papers in the three top computer vision conferences, i.e., CVPR, International conference on computer vision (ICCV), and European conference on computer vision (ECCV), and its total proportion. Their percentage and counts are basically increasing year by year.

This method is easy to operate but unable to recognize identity. To make full use of the identity tag, the image and the corresponding tag can be input into the classification model, and finally, the prediction probability of belonging to a certain identity can be output to give the identity. This method classifies the person re-identification task into recognition (multi-classification) task. By learning the correspondence between the characteristics of the person and label, the output of the fully connected layer is set as the number of identities, and finally, the requirement for predicting the identity of different samples is reached. By conducting comparative experiments on large-scale datasets, the results [69] show that the classification model can still show good performance without careful selection of training samples. The disadvantage is that a large number of training samples are required and only applies to closed set problems.

B. DATASETS

For different applications in different scenarios, dozens of person re-identification datasets have appeared from 2007 to 2019. And for eliminating the influence of person detection or tracking algorithm on the re-identification accuracy, person images or video sequences are usually marked by hand-crafted boundary boxes at the start of person re-identification. With the improvement of person detection technology, at present, automatic detection and tracking algorithms, such as deformable parts model (DPM) [40] and neural network, are usually used. Moreover, it is common practice to apply pre-trained models on other datasets to person re-identification tasks, the so-called fine-tuning strategies [98]. For example, the models such as GoogLeNet [64], ResNet [65], and DenseNet [66] are trained on ImageNet [53] dataset, and then experiments are performed based on the person re-identification datasets.

Table 2 (a) and (b) compare several benchmark person re-identification datasets based on RGB image and video, respectively. The following parts mainly describe the details and characteristics of several datasets.

TABLE 2. (a) Image-based datasets comparison [41], [68], [42], [82], [44], [125]. (b) Video-based datasets comparison [45]–[49], [141].

(a)

Dataset	Year	#boxes	#identity	#cam.	Detector	Scene	Crop Size	Evaluation	URL
VIPeR	2007	1,264	632	2	Hand	Outdoor	128×48	CMC	https://vision.soe.ucsc.edu/node/178
CUHK03	2014	13,164	1,467	5	DPM/hand	Indoor	Varied	CMC	http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html
Market1501	2015	32,668	1,501	6	DPM/hand	Outdoor	128×64	CMC+mAP	http://www.liangzheng.com.cn/Project/project_reid.html
DukeMTMC-reID	2017	36,411	1,812	8	Hand	Outdoor	Varied	CMC+mAP	http://vision.cs.duke.edu/DukeMTMC/
MSMT17	2018	126,441	4,101	15	Faster R-CNN	Indoor/ Outdoor	Varied	CMC+mAP	http://www.pkuvmc.com/publications/msmt17.html
KnightReid	2019	315,354	937	2-3	Hand	Outdoor	Varied	CMC+mAP	-

(b)

Dataset	Year	#boxes	#identity	#cam.	Detector	Scene	#sequences	Crop Size	Evaluation	URL
3DPeS	2011	1,011	192	8	Hand	Outdoor	1,000	Varied	CMC	http://www.openvisor.org/3dpes.asp
PRID2011	2011	24,541	934	2	Hand	Outdoor	400	128×64	CMC	https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/PRID11/
iLIDS-VID	2014	43,800	300	2	Hand	Outdoor	600	Varied	CMC	http://www.eecs.qmul.ac.uk/~xiatian/downloads_qmul_iLIDS-VID_ReID_dataset.html
MARS	2016	1,067,516	1,261	6	DPM	Outdoor	20,715	256×128	CMC+mAP	http://www.liangzheng.com.cn/Project/project_mars.html
LPW	2018	590,547	2,731	2-4	NN/hand	Outdoor	7,694	-	CMC	http://liuyu.us/dataset/lpw/index.html
LS-VID	2019	2,982,685	3,772	15	Faster R-CNN	Indoor/ Outdoor	14,943	-	CMC+mAP	-

1) RGB IMAGE-BASED DATASETS

CUHK03 [68] is the first large-scale person re-identification dataset that is enough for deep learning and was collected on campus. There are currently two single-shot setting protocols for this dataset. One is the training set containing 1,160 identities and the test set containing 100 identities [68]. The other is the training set containing 767 identities and the test set containing 700 identities [35], called CUHK03-NP.

Market1501 [42] also includes 2,793 false data from the DPM, which simulates the interference of real scenes, but its boundary frame quality is inferior to that of the CUHK03 dataset. Later, in 2015, the Market1501 dataset was integrated with 500,000 disturbing images. Besides, this dataset has a fixed training/testing split.

DukeMTMC-reID [82] is a subset of the high-definition DukeMTMC [38] tracking dataset in which every identity comes with ground-plane world coordinates across all cameras that appear in it. This dataset also contains pedestrians with occlusions and has 408 identities who appear in only one camera.

The Market1501 dataset and DukeMTMC-reID dataset are commonly used at present. Each identity in these two datasets has multiple queries, which are consistent with practical applications. Each image in both datasets contains the camera identifier. Moreover, both of these datasets and the VIPeR dataset [41] include attribute annotations.

MSMT17 [44]. Probably because this dataset has just been published, there are not many studies comparing their effects on it. The data of this dataset was collected from multiple indoor and outdoor cameras at three different periods over four consecutive days.

KnightReid [125] was collected on the campus at night and remedy for the defect that there is no dataset for the night scene in the current person re-identification field.

2) RGB VIDEO-BASED DATASETS

PRID2011 [46] has more than 900 identities, but only 200 of them appear in both cameras. That means other identities are only single camera's frame segments and the lighting and shooting angles of these identities in this dataset may not change much.

MARS [48], an extended version of the Market1501 dataset, is the first large-scale video-based person re-identification dataset. Because all bounding boxes and tracklets were automatically generated, this dataset contains several natural detection/tracking errors, and each tag may have multiple tracklets.

The PRID2011, iLIDS-VID [47], and MARS dataset are commonly used at present.

DukeMTMC-VideoReID [51] is another commonly used video-based person re-identification dataset, which is also a subset of the DukeMTMC [38] tracking dataset. In total there are 2,196 videos for training and 2,636 videos for testing. Each video contains person images sampled every 12 frames. During testing, a video for each identity is used as the query and the remaining videos are placed in the gallery.

LPW [49] is characterized by its good cleanliness and closer to real-world conditions, which include three different crowded scenes with a large span of age, frequent occlusion, and various postures.

Except for these two types, there are other multi-modal datasets mentioned in Section III, such as those based on

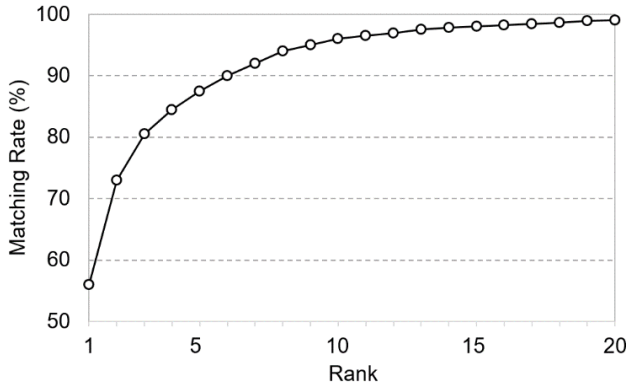


FIGURE 3. Example of the cumulative match characteristic curve.

RGB depth (RGB-D) and RGB infra-red (RGB-IR) data. Another person re-identification survey [10] described these datasets in detail and showed some sample images. The labeling of large person re-identification datasets is extremely laborious, but those works have great significance to the research and development of this community. Appreciate the founders of these datasets and their contributions to person re-identification research.

C. EVALUATION METRICS

To measure the performance of different person re-identification models, the researchers proposed many metrics.

Rank-n measures ranking accuracy. According to the definition of the person re-identification, when a query image is given, the model must retrieve the predicted matching image from the data collected by the multi-camera system to help determine the person's identity in this query image. The first image in the returned search results is usually selected for comparison, namely rank-1. It is difficult to find specific similar samples from a large amount of gallery, and few algorithms can achieve the highest precision. Therefore, as long as the correct picture is retrieved from the top n result, the accuracy can be calculated as rank-n. Researchers often compare rank-1, rank-5, rank-10 and rank-20.

CMC [50]. By artificially setting the size of n, the result can be retrieved quickly and the accuracy can be calculated to measure the performance of algorithms. This method is expressed in terms of the Cumulative Match Characteristic (CMC) curve, shown as Fig. 3, and is defined as follows,

$$cmc\,(N) = \sum_{n=1}^{N} r\,(n) \tag{1}$$

where $r\,(n)$ represents rank-n. The CMC curve is used extensively in the person re-identification task to measure the performance of different algorithms.

AUC and ROC. If the CMC curves of different methods are not much distinctive, it is not easy to compare them. So [88] used the Receiver Operating Characteristic (ROC) curve for performance comparison. Generally, the ROC is a well-accepted measure to express the performance of matchers, and the CMC is used to measure the performance of identification systems [128]. Several researchers [29] also employed the normalized area under curve (AUC) scores for the CMC curves. AUC means the area limited between the ROC curve and the x-axis.

Mean AP. In the case where the dataset is single-query (SQ) in the gallery, the CMC curve can effectively measure the performance of one algorithm. However, in many public datasets, each person usually contains multiple-query (MQ) truth values. If two models are applied to the same dataset and both give the same query results, but the order of appearance of the truth images is different, the performance of the two models is also different. However, the CMC curve only considers the situation where each identity only has one true value match and leads to consistent evaluation results, which is not consistent with reality.

To this end, Zheng *et al.* [42] introduced mean average precision (mAP) to consider the comprehensive indicators of all query pictures. The mAP takes the person re-identification as a retrieval task, and every query image has a matching cross-camera ground truth. For each query, the area under the precision-recall (PR) curve, which is the average precision, is calculated first, and then take the mean of all the precision. There are many ways to calculate the area under the PR curve, one of which is shown as follows,

$$AP = \frac{1}{2}\sum_{i=2}^{M}\left(\left(recall_i - recall_{i-1}\right)\left(precise_i + precise_{i-1}\right)\right) \tag{2}$$

where M is the number of identities in the test set. This evaluation method considers the accuracy and recall of the person re-identification algorithm and providing a more comprehensive assessment criterion.

III. SUMMARY OF VARIOUS APPROACHES

In this section, we summarize the person re-identification approaches into seven categories according to different research purposes. Each category has been divided more meticulous, which basically covers the main research methods of person re-identification, shown as Fig. 4. Further, we discuss the pros and cons of some key methods for reference.

A. LOCAL INFORMATION EXTRACTION

The local information extraction mainly performs identity matching at the image level, which is the basis of re-identification, while the global information extraction constraints and optimizes from the entire camera network.

1) FEATURE DESIGN

To reduce the impact of chaotic backgrounds, it is common practice to extract only the local features of the personal area. By dividing the human body into different blocks, the underlying features are extracted separately, and then the feature vectors are fused to represent this person. Finally, the vectors' similarity between different identities is calculated using a simple Euclidean distance. Farenzena *et al.* [6] proposed a characterization method based on the symmetry of human

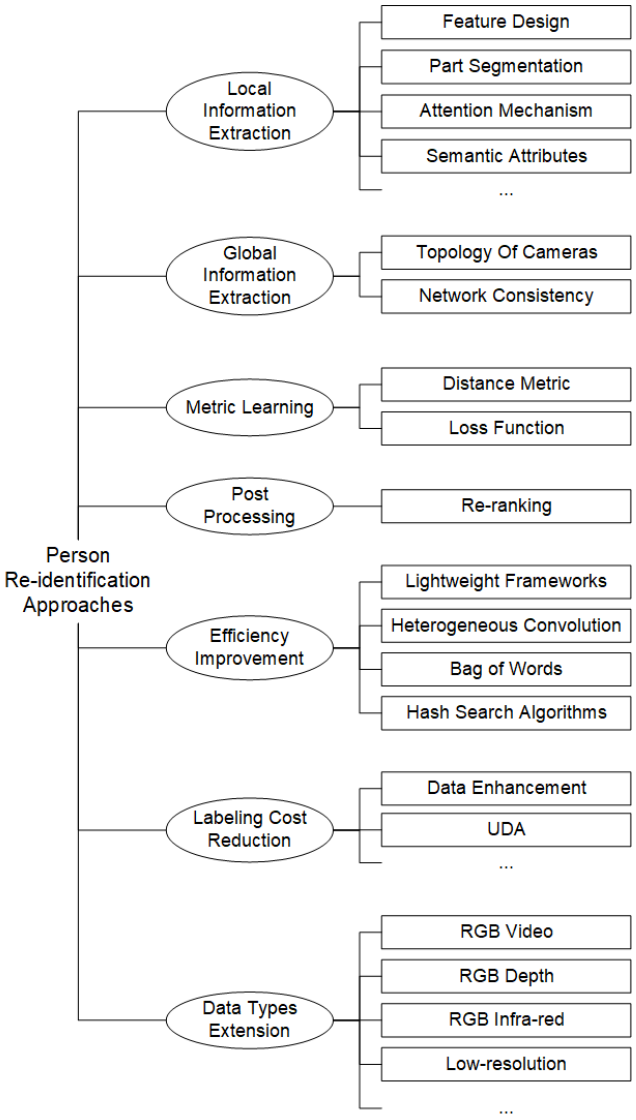


FIGURE 4. **Person re-identification approaches are divided into seven categories according to their research purposes, each of which lists their research methods.**

body parts. Cheng *et al.* [7] proposed a similar approach, the difference being the introduction of the painting structure.

It is not enough to distinguish persons by local features. Integrating the whole and local features together can be considered to better improve the performance of person re-identification. In 2012, Bazzani *et al.* [8] adopted the above idea to extract the color HSV cumulative histogram of multi-frame images to describe the global features. Then the body part was divided into the upper body and the lower body, to respectively extract the local area features with a higher occurrence rate, and finally combined the whole and the part features.

Most of the characterization methods are based on the underlying features. However, in the actual cross-camera scene, the appearance characteristics of persons change greatly. Two different persons may have similar visual characteristics, which can only be distinguished by the details such as the presence or absence of a hat and stripes of shoes, and the type of clothes, which are also semantic features. In 2012, Layne *et al.* [54] defined 15 semantic features, including the hairstyle, shirt type, shoes, etc., using a support vector machine (SVM) to obtain various attribute features, and fusion with the underlying features, to obtain a better distinction. Liu *et al.* [55] detailed the various attributes of the clothes. Su *et al.* [56] embed the binary semantic attributes at the same person under different cameras into a continuous attribute space, which is more discriminating for matching. Shi *et al.* [57] suggested learning attributes from existing fashion photography datasets, including color, texture, and category labels. These attributes are directly transferred to the person re-identification task under the surveillance video, and the experiments have achieved good results. Li *et al.* [58] collected a large-scale dataset with richly annotated pedestrian attributes. Recently, Li *et al.* [13] further expanded this dataset to facilitate attribute-based person re-identification methods.

Most of the early research works used a similar approach, focusing research on images. These image-level approaches usually focused on the similarity of the global image, and in most cases ignored the details and often did not distinguish between two identities whose appearance looks very similar. So many researchers began to study the local areas with discrimination in an image.

2) PART SEGMENTATION

Most existing person re-identification methods focus on learning supervised identity discrimination information, but both feature extraction and distance metric learning, and even deep learning methods assumed that person images are strictly aligned, which would be almost impossible in a real-world scenario. Because there is a deviation in the person detection algorithm, it is likely to cut out part of the body, and the position of the person in the whole image is also different and may be biased. These imperfect person detection boundary boxes can change the posture of the human body and cause great interference to the person re-identification task. Cheng *et al.* [72] proposed a part-based multi-channel network. Zhao *et al.* [263] chose the most distinguishing one from the split patch. Wu *et al.* [73] divided the image into five fixed-length regions. Histograms are extracted for each region and blended with global depth features. Rahul *et al.* [74] also adopted a similar local feature extraction method, which introduced the long short term memory (LSTM) network [111]. First, divide the image vertically into several parts, and then extract the whole and local features separately. Although this cascading method extracted effective local features and achieved better performance, it ignored the problem of image misalignment and easily leads to judgment errors. Several works have been similar [265]–[267], [268].

Earlier re-identification works that considered body parts were [264], [7], [269]. Later, in response to the above problems, many researchers first used bone key points and post

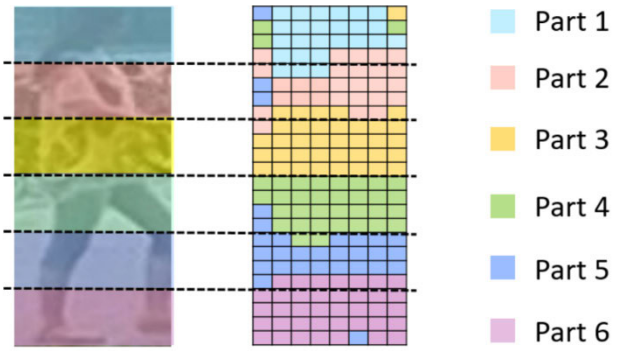


FIGURE 5. Visualization of within-part inconsistency. Left: The image is equally partitioned to six horizontal stripes (parts) during training. Right: Every column vector in the image is denoted with a small rectangle and painted in the color of its closest part [76].

estimation algorithms to align each body part before segmentation. Wei *et al.* [75] used Deeper Cut [112] to detect three overlapping body parts and then learned the feature vectors. Li *et al.* [77] employed the Spatial Transform Network (STN) [78] to extract potential body parts. Reference [79] first used the pose estimation model to locate the 14 key points of the human body. These key points divide the human body into several parts and then align the local areas with affine transformation. To extract local features at different scales, the author designed three different pose frame combinations and then entered them into the network. In 2017, Zhao *et al.* [80] proposed Spindle Net, which divides seven human body parts, including three large regions (head, upper body, and lower body) and four small regions of limbs. Then the original picture and local area are input into a parameter sharing CNN network extraction feature. This research also creatively introduces a tree-level hierarchical fusion strategy, which combines the eight features in a gradient. The experimental results also demonstrate the effectiveness of global and local fusion methods. At the same time, the author also presents a dataset named SenseReID for performance testing only. Subsequently, [122] also studied global and local features together, and automatically aligned local body parts by local distance computed by finding the shortest path. Finally, the author proves that this method has exceeded the average level of human beings. Similar to part alignment, Sun *et al.* [76] devised a new stronger baseline network, named part-based convolutional baseline (PCB). Compared with the traditional image for uniform horizontal slices of the same part alignment problem, this paper put forward by using the refined part pooling to specify and distribution of the PCB part to make it more close to the real part of the section (see Fig. 5). Furthermore, Zheng *et al.* [290] proposed a coarse-to-fine pyramid model, which can grasp the gradual clues from global to local. They also presented a dynamic training strategy to seamlessly combine ID loss and triplet loss. This research greatly improved re-identification accuracy.

A large number of experimental results prove that the part segmentation method can extract more discriminating and regional features, thus improving the re-identification accuracy of the model. However, the above methods also have some problems. When segmenting local regions, these methods based on part segmentation cannot be randomly cropped, and can only be combined with other models, such as human joint point positioning. Some also require additional training posture assessment algorithms to ensure accurate positioning of the human body during the experiment.

3) ATTENTION MECHANISM

In the cognitive process, the human visual system usually selectively focuses on things we are interested in, rather than on the whole. This method of focusing on one part and ignoring other information is the attention strategy [117]. As long as you input an image, the attention model can return to the local interesting area. Through this mechanism, person re-identification models obtain more local and significant features, thereby improving the cognitive ability of the machine, and it is possible to have a brain like a human. Hence, several works mimic the attentional mechanisms of human re-identification processing to improve the performance through local block matching [87], [88] and significance weighting [89], [90]. However, these methods require a strict bounding box for the entire person and are highly sensitive to hand-designed features.

To improve the above methods, some researchers have proposed a deep learning-based attention mechanism to deal with person non-alignment challenges in person re-identification tasks. The common strategy of these methods is to add an additional local attention module to the subnet in the depth recognition model. For example, Su *et al.* [91] integrated the separately trained posture detection model into the part-based person re-identification model. Li *et al.* [77] designed an end-to-end site-aligned CNN network for locating potential discriminative regions and extracting these region features for re-identification. Wang *et al.* [270] exploited the attention mechanism to solve human misalignment. Zhao *et al.* [93] used the STN to obtain several attention regions and then extract local features from the regions. The HA-CNN model [95] focuses on noise problems in pixel-level regions by establishing a joint learning scheme for simulating soft and hard attention in a single model. This is the first attempt at multi-level attention modeling in the person re-identification field. Besides, cross-interaction learning has been introduced to enhance the complementary effects of different levels of attention, which is an important addition to the previous method. However, such multi-task learning also adds to the task's training burden. Therefore, Chen *et al.* [139] proposed a new regularization term on features and weights and combined with complementary channel-wise information and body part spatial awareness to make the attention more extensive and reduce the overfitting of a certain attention area. Zheng *et al.* [104] combined attention consistency and Siamese learning to learn the spatial orientation of people, as shown in Fig. 6. And the Siamese module in its network architecture ensured that

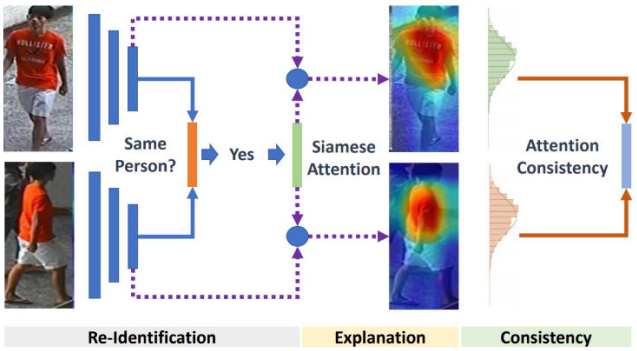


FIGURE 6. Framework for re-identification that provides mechanisms to make attention and attention consistency end-to-end trainable in a Siamese learning architecture as well as explaining the reasoning for why the model predicts that the two images belong to the same person [104].

the same person could obtain spatially consistent attention areas, which enhanced the robustness of the network. This is the first research that combines attentional consistency with Siamese learning, which greatly enhances the generalization ability of data. Guo *et al.* [138] firstly focused on the non-human part's contextual cues, such as backpack, reticule, and umbrella. They extracted the human part and non-human part masks by designing a dual part-aligned representation scheme, its potential branches for extracting non-human masks use the self-attention mechanism. This method achieved 95% rank-1 accuracy on the Market1501 dataset but relied heavily on the invariance of personal belongings. Recently, high-order attention modules have also been designed to model the interaction between attention modules [271]. This module was decoupled from the framework and could be combined with any person re-identification framework.

The essence of attention mechanism is a weight distribution strategy. The parts with high weight are more decisive to the result. Therefore, the correctness of the feature weights obtained by attention mechanism is more important, otherwise, there will be a greater deviation. For person re-identification, the attention mechanism can greatly improve the performance and also enhance the efficiency of data if it extracts more discriminative features. But most of the methods based on attention mechanism need to use pairs of person images as input. Moreover, due to the possible need to extract the attention area and aggregate the features within the multi-level attention area, the attention mechanism modules usually have great computational complexity and video memory capacity.

4) SEMANTIC ATTRIBUTES

Pedestrian attributes task is one of the human-related studies that can lay the foundation for person re-identification and retrieval. Structured pedestrian attributes usually include but are not limited to gender, age, hairstyle, clothing, and accessories. The person re-identification based on semantic attribute is more consistent with the human mind in the ideal scenario.

Earlier studies directly integrated attributes into the person re-identification framework for attribute consistency matching [272], fine-tuned feature extraction using attribute datasets [279], or assisted semi-supervised person re-identification with attribute information [273]. However, due to the huge work of attribute annotation, Chang *et al.* [274] proposed an attribute-based framework without the hand-crafted labels. This method decomposed human visual representation into multiple potential differentiating factors and models them dynamically. In addition, multi-task learning combining attribute recognition and re-identification tasks was favored because it can improve the performance of both tasks [276]–[278]. Among them, Tay *et al.* [276] designed an Attribute Attention Network (AANet), which integrates part segmentation, attribute recognition, and re-identification and performed these tasks sequentially. Reference [278] systematically studied how person re-identification and attribute recognition benefit each other and manually labeled attribute tags for two large-scale re-identification datasets, Market-1501 and DukeMTMC-reID. In the research of video-based re-identification, Zhao *et al.* [275] used attribute information to measure the weight of video frames. Zhang *et al.* [223] mined the dependency relationship between attributes and converted them to human re-identification.

The importance of semantic attributes, similar to part alignment, are self-evident for person re-identification tasks. However, how to weigh the proportion of attributes, how to deal with occluded attributes and generalization of attributes, how to solve the variable representation of attributes caused by changes in body posture or camera angle, and how to reduce unnecessary labeling costs need to be explored in depth.

5) OTHERS

a: OTHER APPROACHES TO ALIGN HUMAN

Part segmentation, attention mechanism, and semantic attributes are all based on the idea of alignment to find the spatial distribution and correspondence between human images. By fusing the information of the human body and posture, some researches [286]–[289], [300] have greatly improved the performance of re-identification. Reference [300] treated the human body as a whole composed of several components and aligned these components, which are represented as the key point of the human body. Suh *et al.* [286] designed a dual-stream network to extract appearance and body part features separately, where the part sub-stream was pre-trained using the existing pose estimation network. Wu *et al.* [315] proposed a Siamese network that the designed multiplicative integration gating function was first used to strengthen the local and then the four-directional recurrent neural network was used to solve the misalignment problem. Reference [287] proposed to solve the alignment problem with dense semantics. This method densely corresponded 2D person image and 3D surface-based canonical representation of the human body. By establishing a human pose model to

learn pose priors, Wu *et al.* [289] realized a robust online re-identification framework for perspectives transformation. In addition, some works had attempted to find the most distinguishable part to align images from multi-shot queries [288] or restore probabilistic similarities between spatial regions of two images based on the Kronecker Product Matching (KPM) module [285].

b: GRAPHS-BASED PERSON RE-IDENTIFICATION

Moreover, graphs are so suitable for expressing the relationship between objects that some scholars have proposed to use graph structure to deal with the person re-identification task. And recently, graphs have been widely applied in the image analysis field. Borrowing from the greedy algorithm and Ant Colony Optimization, Barman *et al.* [301] formulated an extensible algorithm to solve the re-identified sorting issue and aggregate the results of multiple algorithms. Shen *et al.* [153], [318] integrated the relationships between images in the gallery into the training of the model. Reference [232] suggested using Graph Convolutional Network (GCN) to model the correlation between spatial regions. [154], [155] utilized graph structure to deal with unsupervised cross-domain adaptive problems.

c: CAMERA VIEW-BASED PERSON RE-IDENTIFICATION

Due to the high dependence of the re-identification system on the camera settings, most frameworks cannot fully adapt to various camera perspective changes. In this regard, scholars have proposed several methods to reduce the feature differences caused by cross-view problems. The Camera coRrelation Aware Feature augmenTation (CRAFT) method [308] automatically measured the camera correlation from the cross-vision data distribution and adaptively performed feature enhancement to transform the original features into a new adaptive space for cross-view adaptation. Feng *et al.* [312] designed a view-specific network with a cross-view Euclidean constraint (CV-EC) to reduce the differences between individual view features. The Deep Asymmetric Metric learning (DAM) method [313] was designed to jointly learn viewpoint- specific and feature-specific transformations. Zhu *et al.* [314] fully mined the information of negative samples and proposed a projection-based method to reduce the differences between camera pairs.

d: ANTI-CHALLENGES PERSON RE-IDENTIFICATION

Finally, some studies have proposed processing methods for challenges such as foreground occlusion [297]–[299], lighting changes [294]–[296], and background disturbances [291]–[293].

B. GLOBAL INFORMATION EXTRACTION

Person re-identification is a task of matching and retrieving person with multiple non-overlapping cameras. Those methods mentioned above try to extract the detail information in each image, considering to obtain more discriminative features, so as to match identity further based on these features. Other researchers have used the constraint of ''non-overlapping cameras'' to improve the accuracy of person re-identification in the overall perspective.

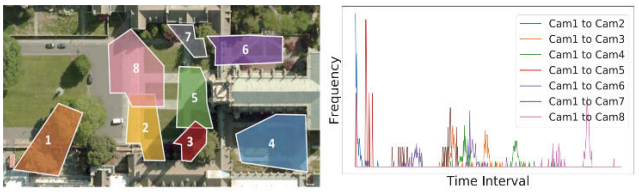


FIGURE 7. (a) Camera topology of DukeMTMC-reID. (b) Spatial-temporal distribution, i.e., frequency of positive image pairs (an image pair with the same person identity denotes a positive pair) with respect to time interval [170].

The ''non-overlapping cameras'' means that the data come from different cameras (domains), which cause the difficulty of intra-category differences, such as an identity under different cameras with different light, different image acquisition quality, and different angles. However, this essential condition also provides some relatively rigid constraints for person re-identification. For example, the distance between the cameras that identities may appear is greatly related to their moving speed and the dress of identity under different cameras is basically the same in a short period.

1) TOPOLOGY OF CAMERAS

Therefore, several works combine the first constraint, depending on the camera's topology, to judge the camera that a pedestrian may appear on. Cai *et al.* [174] hypothesized that the camera network topology was given and then proved the validity of the topology information for inter-camera multiple target tracking. Subsequently, some person re-identification researchers [175], [168], [171] tried to infer the camera's network topology in a variety of ways. With the publication of the Market1501, DukeMTMC-reID and GRID dataset with camera topology (see Fig. 7(a)), work to improve the person re-identification's accuracy based directly on the given camera topologies began to emerge [170]. Wang *et al.* [170] combined spatial-temporal information and visual semantic information to eliminates lots of irrelevant images and alleviate the problem of appearance ambiguity in person (see Fig. 7(b)).

In conclusion, the advantages of camera topology information are obvious. In reality, this information is easy to obtain, and there are now several large datasets that include this information. There is a strong correlation between the locations where an identity appears. Furthermore, it is necessary to exclude some unreasonable data with this information as data volumes grow. However, we believe that it's research still needs to overcome at least the three problems. First, this constraint relationship is not absolute, and easy to produce negative effects if it is not handled well. So how to better combine with the detail features for performance optimization needs further study. Second, the range and speed of human activities are highly uncertain, for example, there is a great

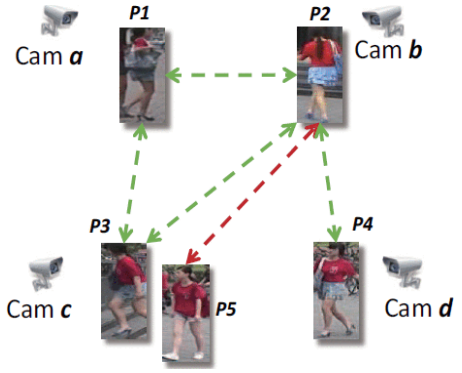


FIGURE 8. An illustration of person re-identification in a camera network. The green lines refer to correct matches and the red line indicates wrong matches. If P1 is considered to be the same person as both P2 and P3, then P2 and P3 must be the same person. Otherwise, the inconsistency will arise. This constraints results in the upper left green triangle [169].

speed difference between the runner and the disabled. Third, this information may not work for remote re-identification.

2) NETWORK CONSISTENCY

On the one hand, the person re-identification involves a growing number of cameras that do not overlap, making pedestrian movements unpredictable. On the other hand, the appearance of pedestrians under different cameras can vary greatly, making it difficult to completely distinguish pedestrians using only visual information. Hence, others work to improve accuracy by maintaining consistency of the results across the network according to another constraint. Das *et al.* [173] firstly explored the consistency in re-identification across a network of cameras and used the consistency information from additional cameras to improve the otherwise standard camera pairwise re-identification. Compared to the above paper, which only maintained consistency during the matching part, Lin *et al.* [169] also exploited consistency information in the training stage. They sought the globally optimal matching by maximizing the sum of all matching similarity for all camera pairs while satisfying all the consistency constraints (see Fig. 8). To extract full-scale features, Yang *et al.* [172] combined the visual information, temporal-spatial information, and consistent information to re-rank person re-identification results.

The method based on network consistency converts person re-identification into a global optimization problem under the consistency constraints and maintains the consistency of the camera's re-identification results. But this method did not apply to any paired camera similarity score.

Finally, it must be admitted that this information, if properly exploited, can improve the performance. Nevertheless, such information can only be used as supplemental factors, and ultimately needs to be confirmed by the person's detail features.

C. METRIC LEARNING

Person re-identification methods based on classical machine learning generally design standard distances between descriptor vectors that are not affected by light and background. The closer the distance, the higher the ranking, and the more likely it is the same person. Those methods based on neural networks update the weights in the network every iteration to minimize discrimination loss.

1) DISTANCE METRIC

The main idea of those methods based on distance metric is to design a new measurement function, project all the eigenvectors, pull the eigenvectors belonging to the same persons closer, and push the eigenvectors belonging to different persons farther to effectively distinguish identities [60], [61], [106], [107]. In general, human eyes judge whether two objects are the same, mainly to determine whether two objects look similar, and how similar they are. For judging whether two samples are the same person, the person re-identification system should measure the similarity between images, that is, the distance, which represents the difference between two images. The feature description method generally uses the Euclidean distance when calculating the similarity of the feature vectors. But in the multi-camera monitoring environment, descriptors are affected by factors such as viewing angle and illumination. The Euclidean distance method does not consider the effect change of different features in different environments, and treats the effect equally poorly, thus causing bias. Therefore, the researchers attempt to obtain a new metric subspace by training learning methods, making it easier to distinguish eigenvectors in this space. Different from simple similarity calculation, metric learning can be trained with fully annotated data to learn more discriminating subspaces. In general, metric learning is based on Mahalanobis [60],

$$d\left(x_i, x_j\right) = \left(x_i - x_j\right)^T M \left(x_i - x_j\right) \tag{3}$$

where x_i and x_j are samples and M is a semi-definite matrix. Nowadays, metric learning methods based on Mahalanobis distance have been widely used in person re-identification tasks. Xing *et al.* [59] divided two different pairing data based on identity tags. Weinberger *et al.* [60] added triad samples to the network and added constraints on positive and negative sample pairs to limit training. In the end, the images of the same person gradually approach, and the gap is smaller than different persons. This method was therefore named as the large interval nearest neighbor classification. Assuming that the input triplet samples are (x_i, x_j, x_k), where the labels $\mathrm{y}\left(x_i\right) = \mathrm{y}\left(x_j\right)$, $\mathrm{y}\left(x_i\right) \neq \mathrm{y}\left(x_k\right)$, learn the optimal matrix M that satisfies the following constraints,

$$d\left(x_i, x_k\right) \geq d\left(x_i, x_j\right) + 1 \tag{4}$$

In 2012, Kostinger *et al.* [61] conducted supervised linear metric learning based on keeping simple and straightforward metric (KISSME) strategies. The main difference between this work and other methods is to use the log-likelihood ratio to measure the divergence between samples. In essence, it is learning the matrix M in Mahalanobis distance, which is

defined as follows,

$$\delta\left(x_i, x_i\right) = \log\left(\frac{p\left(x_i, x_j|H_0\right)}{p\left(x_i, x_j|H_1\right)}\right) \tag{5}$$

where H_0 assumes that the sample pair is negative and H_1 assumes that the sample pair is positive. The smaller δ, the more similar the samples. Finally, the Mahalanobis distance metric reflecting the properties of the log-likelihood ratio test is obtained as,

$$d_M^2\left(x_i, x_j\right) = \left(x_i - x_j\right)^T M\left(x_i - x_j\right) \tag{6}$$

Liao *et al.* [62] added a branch to the KISSME algorithm, which can learn both Mahalanobis matrix and feature extraction. This crossover learning method is called cross-view quadratic discriminant analysis.

2) LOSS FUNCTION

In the person re-identification datasets, there are a large number of person samples that look very similar, which are caused by environmental and subjective factors. The difference between the feature vectors of the entire image may be small, and generally can only be distinguished by local minute details, for which different weights need to be assigned to these features. If a standard distance is used, the model will ignore these details and treat all features equally, resulting in poor accuracies at the end of the model. To this end, the more discriminating measurement algorithm must be designed, that is, the metric learning. Throughout the training process, the researchers also designed the loss function as the target of the entire depth model and optimized by continuous learning to achieve the requirements of the person re-identification task. Such as the comparison loss function of receiving pair samples as input, the triplet loss function of receiving triplet samples as input, the Binary Cross-Entropy (BCE) objective loss function [104], and the identification loss of receiving the identity's tag as the input.

Contrast loss function [70] used to train the Siamese network, receiving sample pairs and labels as inputs. The tag is a binary value. When it is "1", it means that two pictures belong to the same person, that is, positive sample pairs, and in reverse, they belong to different persons that are, negative sample pairs. Given two images and labels, the contrast loss function is defined as,

$$Loss_c = y d_{x,y}^2 + (1 - y)\left(\alpha - d_{x,y}\right)_+^2 \tag{7}$$

where $(z)_+$ demotes $max(z, 0)$, y is the label, α is a parameter, $d_{x,y}$ is the Euclidean distance, which is used to measure the degree of the difference between the features. Minimizing the loss function also means pulling the positive sample pair closer, pushing the negative sample pair farther away, and learning continuously to achieve optimality during the training process.

Triplet loss function [71], [116] is similar to the contrast loss function, except that its input data is a manually designed triplet sample. Among them, two images are from the same identity, and the other image is from different identities. Then the three images are combined to form a positive and negative sample pair. Assuming the input samples (a, p, n), a and p belong to the same person, a and n belong to different persons, then the triplet loss function is formulated as,

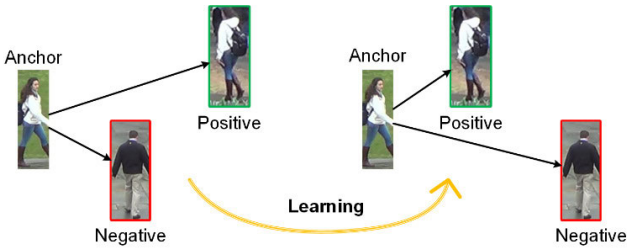


FIGURE 9. Triplet loss diagram.

$$Loss_{tri} = \left(d_{a,p} - d_{a,n} + \alpha\right)_+ \tag{8}$$

Triplet loss is one of the most basic objective functions in the person re-identification task. It also enables the distance between positive sample pairs to be shortened and the distance between negative sample pairs to be expanded, illustrated in Fig. 9. This loss function can well solve the problems of inter-class similarity and intra-class differences existing in the person re-identification task and can be used to train samples with small differences. When the dataset is small, its over-fitting problem can be effectively alleviated by constructing triplets that can generate much more than training data. Although simple but very practical, this idea has high application value and has been widely used by researchers.

Quadruplet loss [109] *and TriHard loss* [108] are two improved versions of the triplet loss function. Quadruplet loss considers the absolute distance between positive and negative sample pairs compared to Triplet loss, while TriHard loss [108] introduces the idea of hard sample mining. Margin Sample Mining Loss (MSML) [114] introduces the idea of difficult sample mining into Quadruplet loss, absorbs the advantages of the above several loss functions, further improves the generalization ability of the model, and becomes a common measurement learning method in the field of image identification. In [114], several performance indicators of these loss functions under different datasets and baseline conditions were compared, and the experiment proved that TriHard and MSML had better performance in terms of accuracy.

Identification loss function [14] used to train the neural network of identification tags whose input data is the picture and its identification. The last layer of the network is the full connection layer, whose dimension N is the number of identities in the dataset. For an input sample (x_i, y_i), where x_i is the input person image y_i is the real identification tag, p_i is the prediction junction probability. Therefore, the cross-entropy

FIGURE 10. Diagram of person re-identification based on re-ranking.

loss of the identification task is expressed as,

$$Loss_{ID} = \sum_{i=1}^{N} -q_i \log(p_i) \begin{cases} q_i = 0, & y_i \neq i \\ q_i = 1, & y_i = i \end{cases} \quad (9)$$

The above are several common loss functions in the field of person re-identification. Besides, many studies combine several loss functions to fully utilize the advantages and avoid the disadvantages to obtain better experimental results [14], [98], [110].

D. POST PROCESSING

Previously, re-ranking is used to improve the accuracy of target retrieval and has achieved certain success [123], [124]. Subsequently, some person re-identification researchers put their attention on re-ranking [29], [37], [208]. Re-ranking of person re-identification, also known as post-ranking, mainly refers to that after obtaining initial ranking by some other methods, researchers expect to add a re-ranking step to make the image related to the target image get a higher ranking than the initial ranking, as shown in Fig. 10. Reference [29] summarized three characteristics that should be noted in the person re-identification re-ranking research. (1) Each identity has multiple target images in the context. (2) The input of this task is the results generated by the person re-identification method and need to be re-ranked. (3) Its task is to output a re-ranked list with higher accuracy.

Nguyen *et al.* [29] were believing that if an image has been ranked high in one result, then this identity should be ranked low in the other results. So, based on this commonsense restriction, high-ranked gallery images will be assigned a penalty to update gallery images' scores in other lists. However, the limitation of this method is that it is only effective if there have other rankings (at least three) except the current probe ranking. In addition to the above methods based on content and context information [29], [30], [37], [204], there are many re-ranking methods in person re-identification, such as handcrafted annotation or labeling supervision [31]– [33], common nearest neighbors [34], [122], [207], as well as k-reciprocal method [35], [36], [39], [122], [203].

Given fuzzy vision, García *et al.* [37] proposed using an unsupervised post-ranking framework to find out the parts that caused fuzzy vision according to the content information extracted from gallery persons and the context information extracted from the feature extractor and delete them. Finally, its result obtained was greatly improved compared to the baseline. Reference [31] used supervised embedding manifold to represent the feature vector of person image with dimensionality reduction, estimated the similarity value between two instances under the background of other instance pairs, and ensured the constraint of dimensionality reduction on the intra-class distance and inter-class distance. The difference between extensible manifolds and ordinary manifolds lies in the fact that in the process of embedding a manifold, in addition to preserving the relation of its neighbors, the relation of global embedding can be extended. In order to improve efficiency, associative learning is carried out only when the dataset instance is offline. Zheng *et al.* [35] believed that if a gallery image is similar to probe in k-reciprocal neighbors, it is more likely to be a true match. Therefore, they proposed a k-reciprocal encoding method to compare a person's k-reciprocal feature and finally realize the reordering of person re-identification results. Inspired by this idea, Su *et al.* [39] proposed a k-reciprocal attention module to integrate informative context into frame-level features and applied non-local attention to the video-based person re-identification' space and channel field to improve performance. Fei *et al.* [202] further explored the relationships between each gallery and other queries for the problem that [35] are not suitable for the single query scenario. Recently, Luo *et al.* [205] suggested avoiding the extra inferential costs by representing the data points in the training batch as a graph. Bai *et al.* [206] proposed a more robust Unified Ensemble Diffusion (UED) method, which is generally applicable to object retrieval.

After re-ranking, person re-identification results can achieve a significant performance improvement, even directly improving accuracy by five percentages. However, the consumption of calculation and time is not to be neglected in re-ranking. Therefore, future re-ranking method should try to avoid excessive computational cost and label cost.

E. EFFICIENCY IMPROVEMENT

The emphasis of the above several research methods is to improve the accuracy of the person re-identification task to ensure more correct results. However, the efficiency issue cannot be ignored for person re-identification task. In most practical applications, such as criminal investigation, the retrieval time is even more important than the accuracy. Unfortunately, most of the previous studies focused on improving re-identification accuracy and ignoring time consumption. At present, no evaluation strategy for the efficiency of the person re-identification model has been proposed.

For deep convolutional neural networks, accuracy and computational cost are often difficult to achieve concurrently.

Some person re-identification researchers have designed lightweight neural network frameworks to reduce computational complexity and model memory. Guo *et al.* [209] used

the multitasking model, in which ranking loss was calculated from the attended regions extracted from the feature mapping of the image using STN. This method not only kept the accuracy but also reduced the computational complexity. Wang *et al.* [210] reduced the amount of computation and model size based on the lightweight strategy of yolov3-tiny, and eventually increased the speed of computation for real-time monitoring. Zhou *et al.* [211] designed a lightweight architecture of CNN, namely OSNet, using pointwise kernel and depthwise kernel to replace standard 3×3 kernel and modifying the residual bottleneck. This architecture is suitable for many computer vision tasks, including re-identification. Different from the above studies, Yao *et al.* [212] considered a coarse-to-fine framework that would extract the global descriptors and local descriptors of images. The global descriptor was responsible for selecting the Top-M similar images from the gallery to narrow down the scope. The local descriptor performed a fine-level search on the M images.

Otherwise, the research community has been exploring ways to solve this problem of computational cost by model compression or designing efficient architectures [97]. Recently, some researchers [96] proposed a new deep learning architecture that uses heterogeneous cores for convolution operations. The so-called heterogeneous convolution refers to the convolution using heterogeneous convolution filters, including convolution kernels of different sizes. This method could reduce the amount of calculation and the number of parameters compared to standard convolution operations. Improve the efficiency of existing models without sacrificing accuracy. Experimenting on standard convolutional neural networks, such as VGG [118] and ResNet, by replacing the standard convolution filter with a heterogeneous convolution filter, it was found that three to eight times the computational speed can be achieved while maintaining or even improving accuracy.

Person re-identification tasks need to extract person features and then feature matching. For the former stage, the above can be referenced to shorten the calculation time and maintain the quality of the representation. At present, many researchers also use the Bag of Words (BoW) to quickly describe the characteristics of the image [119]–[121], [123]. First, each image needs to be divided into small blocks, and the features of each region are extracted separately, and then the vocabulary is constructed by the k-means algorithm. K-means algorithm is an indirect clustering method based on the similarity measure between samples. This algorithm takes k as the parameter and divides n objects into k clusters, so that objects within the cluster have a higher similarity, while objects of different clusters have a lower similarity. By setting the parameters of the dictionary and algorithm, all the small blocks are clustered. When the k-means algorithm converges, the center of each cluster is obtained, that is, the corresponding words are obtained. Finally, the words in the vocabulary are used to represent the image. When all the images' features have been calculated, classification training and prediction can be performed. In the field of person re-identification, Zheng *et al.* [42] used the BoW model and simulated and tested the feature extraction time and retrieval time of the SDALF [6] and SDC [90] methods on Matlab. The experimental results, shown in Table 3, showed that the method can effectively shorten the feature extraction time.

TABLE 3. Comparison of average query time of different methods based on the market-1501 dataset.

Time(s)	BoW	SDALF	SDC
Feature extraction	0.62	2.92	0.76
Feature matching	0.98	2644.80	437.97

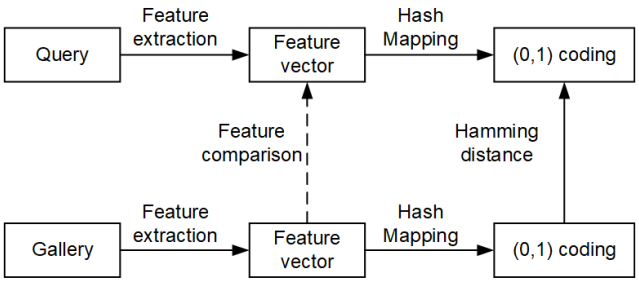


FIGURE 11. Hash-based neighbor search process. The dashed line indicates that the extracted feature is directly used for linear search. The hashed nearest neighbor search algorithm maps the extracted feature to a low-dimensional space, generates binary code, and then uses the binary code for fast search.

In the person re-identification field, given a query picture, the similarity is calculated from the picture library and the matching process of the ranking can be regarded as an image retrieval problem. The comparison experiment results present that in the re-identification problem of persons, the matching retrieval time is far more than the feature extraction time. If the efficiency of the person re-identification algorithm should be optimized, the query time can be shortened. Recently, hash search algorithms for fast approximate nearest neighbors have attracted many researchers [166], [167], [213]–[215] because of their superior performance under large-scale data. The purpose of this algorithm is to establish a data structure that can find the nearest point of any query in a short time, and sacrifice a bit of accuracy for faster search speed than violent search. Instead of using the high-dimensional features extracted by the model to represent the image, researchers project the vector into a subspace-based on the hash map, reducing the vector dimension. The process is shown in Fig. 11. In simple terms, this method is to encode the original data into 0-1 code, which not only achieves high efficiency in time but also achieves high efficiency in storage. To solve the problem of massive image retrieval speed in the future, researchers still need to explore a better indexing method.

F. LABELING COST REDUCTION

A severe problem with person re-identification is that there is less training data and the model is easy to over-fit. Meantime,

the labeling of the person re-identification dataset is tedious, and the training process of deep learning needs a lot of data to support the training process. Currently, the largest image dataset contains more than 4,000 identities, a total of more than 100,000 images. In addition to addressing the challenge of large-scale training data needs from the perspective of building larger and more realistic datasets, there are multiple ways to reduce labeling costs.

1) DATA ENHANCEMENT

Data generation can also be used to augment training data [115]. A severe problem with person re-identification is that there is less training data and the model is easy to over-fit. Currently, the largest image dataset contains more than 4,000 identities, a total of more than 100,000 images. In addition to addressing the challenge of large-scale training data needs from the perspective of building larger and more realistic datasets, data generation methods can also be used to augment training data [115]. Data enhancement methods are traditional and based on deep learning. There are many traditional data generation methods widely used, such as performing flipping, cropping [99] and random erasing augmentation [98], [100], [316] on images and building pyramid input [94].

In recent years, there have been many methods for deep learning, and a more representative one is generative adversarial networks (GAN). GAN is a generative model proposed by Goodfellow *et al.* [101] in 2014, which produces a fairly good output through game learning between (at least) two modules in the framework: the generative model and the discriminant model. It has been favored by many researchers since its introduction. Nowadays, GAN has gradually become a hot research direction in industry and academia [81]. GAN can generate new data samples based on the potential distribution of data samples, with numerous advantages in super-resolution image generation. For the person re-identification problem of insufficient training samples, data generation can be realized and the training process of the model can be enhanced by using GAN and its deformation.

Zheng *et al.* [82] first proposed the study of GAN in the field of person re-identification, which was published at the International Conference on Computer Vision (ICCV) in 2017. Although the paper's thinking is relatively simple, it confirms the effectiveness of the data enhancement method. This paper uses a deep convolutional generative adversarial network to generate samples and then directly into the training data. While the performance of the training model is improved, the quality of the generated samples is not high, and the images are randomly generated and no tags can be used. To solve this problem, this paper proposed a method of label smoothing regularization to assign labels for the sample generation. The actual operation is also very simple, that is, each element in the label vector takes the same value and satisfies the condition. The generated image was added to the training process as training data to avoid model over-fitting. Huang *et al.* [102] proposed to supplement the generated tags with real data dynamically by applying the multi-pseudo-regularized tags (MpRL) method. These tags reflect the different contributions' weight of predefined training classes to the generated new data in GAN training. Considering the open-world person re-identification problem where the similarity between non-target and target people makes the framework difficult to judge, Li *et al.* [243] proposed the method of generating target people with generators and improving the robustness of the framework with discriminators. Zheng *et al.* [247] first integrated discriminative and generative learning into a unified framework for end-to-end training and generated high-quality cross-id composed images with changed appearance and structure.

In the field of person re-identification, there are still dataset bias, camera bias, and person posture change, which leads to poor generalization ability of the model. Applying a trained model in one scenario may be very poor in another one. Therefore, many researchers have proposed using GAN's derivative model and combining it with other algorithms to migrate images [83], [84]. Reference [85], [242] used GAN to generate a series of images with standard poses extracted from the MARS dataset and covered all angles. And Zhong *et al.* [86] focused on the style deviation between different cameras, used CycleGAN to transfer the image style of one camera to another one, and tried to retain the identity information of a person during the migration. In particular, Label Smooth Regularization (LSR) was further applied to the sample of style transfer in order to reduce the noise generated during the style transfer process by making the labels soft and distributed. Similarly, Deng *et al.* [103] focused on how to ensure that the person area with potential relationship and tag information remained unchanged after style migration. The author constructed an unsupervised self-similarity and domain-dissimilarity relationship to constrain the learning of source-target translation. Differently, Liu *et al.* [244] proposed a Similarity Preserved Camera-to-Camera GAN (SPCGAN) framework to replace the background of person images. Huang *et al.* [252] suppressed backgrounds by suppression of Background Shift Generative Adversarial Network (SBSGAN).

Person re-identification research based on GAN research illustrates that the data generated by GAN can indeed expand the training data and prevent overfitting by combining with other datasets or migrating datasets style. Compared with all other models, GAN can produce clearer and more real samples. However, GAN's training needs to achieve a Nash balance, and there is no good way to achieve this balance. In addition, GAN is prone to problems such as training instability, gradient disappearance, and mode collapse. Further studies are needed to make the training of GAN more stable and generate more diverse samples.

2) UNSUPERVISED DOMAIN ADAPTIVE

When there is little or no tag data, the performance of supervised person re-identification methods will be greatly reduced. Moreover, traditional machine learning methods

FIGURE 12. Illustration of the domain disparity among Market1501, DukeMTMC-reID, and PRID2011 benchmarks, presenting significant variances in illumination, resolution and camera viewpoint etc [255].

assume that training data and test data are distributed independently and identically, but this assumption is basically not satisfied in practical scenarios (see Fig. 12). Given these above problems, the unsupervised domain adaptive (UDA) method emerged. UDA is a kind of transfer learning technology, which refers to learning the discriminative representation of the unlabeled target domain through the labeled source domain. It can minimize the distribution difference between domains and effectively solve the change of data distribution between domains. Therefore, a few researchers focus on the UDA-based person re-identification.

In 2013, Ma *et al.* [249] proposed a person re-identification method based on domain transfer learning, which simultaneously learned the optimal model of source task and target task. In 2015, the earliest UDA-based person re-identification method was proposed by Ma *et al.* [129], which used SVMs for cross-domain ranking to estimate the positive information of the target. However, this method was essentially weakly supervised because the SVM-based model cannot be completely learned in unlabeled datasets. In 2016, Peng *et al.* [130] employed the dictionary learning model to transfer source label knowledge without any labels, which is the first UDA-based person re-identification research. Also to extract initialization features from a labeled source domain, Yu *et al.* [28] found shared space between the source and target domains by learning the view-specific projections of each camera view. Li *et al.* [253] combined the pose features from the source domain and target domain and the content features from the source domain to generate the target domain data with changing the pose.

With the popularity of deep learning, some researchers have respectively used k-reciprocal nearest neighbor [131], k-means [27] and self-training clustering target domain [132], [139], [255] to estimate labels of the unlabeled target domain. Zhong *et al.* [133] used an encoding and decoding model to learn the invariant characteristics between the domains. Recently, some scholars have focused on using GAN to narrow the gap between the images in the source domain and the target domain [44], [103], [134]. In addition, [135], [136], [255] proposed using auxiliary attributes and identity tags for UDA-based person re-identification, which provided a new idea for follow-up researches. These methods achieved great results but ignored the intra-domain changes in the target domain, that is, the differences between cameras in the target domain. Given this problem, Zhong *et al.* [137] proposed a style transfer model using LSR, which could transfer the labeled image to each camera style. And Liu *et al.* [26] used GAN to generate images with the target camera domain style. Qi *et al.* [254] developed an unsupervised online in-batch triplet generation method to explore the discriminative information in the target domain. Then, an alternating optimization algorithm [259] is designed to jointly solve the cross-view and cross-domain problems. Combining the defined semantic attributes and potential attributes in the source domain, Peng *et al.* [258] decomposed the dictionary space into three parts for the UDA problem. In 2019, Zhong *et al.* [113] employed three underlying invariances of the target domain, i.e., exemplar-invariance, camera-invariance and neighborhood-invariance, which can improve the transferable ability of person re-identification model.

For cross-domain adaptation, the differences in light, resolution, and camera views between the two domains are so large that the difficulty of direct cross-domain conversion increases. Therefore, several studies [251], [256] have attempted to gradually narrow the difference between the source domain and the target domain. Thereinto, for incrementally optimize the classifiers based on the unlabeled target domain, Lv *et al.* [251] presented an unsupervised incremental learning algorithm. Liu *et al.* [256] decomposed complex cross-domain transmissions, focusing on one factor in each sub-transfer, such as illumination, resolution, and camera view.

The UDA-based person re-identification technologies reduced the deviation between the source domain and the target domain at the image or feature level, making the person re-identification model extendable and transferable. However, their accuracies were still significantly lower than that of supervised learning. Adopting UDA in the person re-identification field has a problem, which is different from other fields. As an open-set problem, person re-identification tasks not only have different camera views between the source domain and target domain but also have different identification tasks, that is, the target domain needs to match the identities who are different from the source domain. This unique characteristic makes it difficult for person re-identification researchers who are dedicated to this method to directly refer to other fields' UDA outcomes. Nevertheless, the UDA method has excellent research value because of the explosion of big data and the huge cost of labeling dataset.

3) OTHERS

a: UNSUPERVISED RE-IDENTIFICATION

Generally, unsupervised re-identification researches mainly start from several perspectives: transfer learning [303],

feature representation extraction [304], [307], dictionary learning [302], clustering [305], [306], and distance measurement. Reference [257] proposed a state-of-the-art method, which looked for distinguishable patch features from unlabeled data sets. Lin *et al.* [305] designed a simple and intuitive bottom-up clustering approach. For the realistic challenge of newly added cameras in open-world re-identification, Panda *et al.* [250] managed to find the most suitable source camera and applied a transitive inference algorithm to adapt to this change.

b: OTHER MEANS OF LEARNING

To reduce the cost of labeling data and improve the scalability of the model, researchers have made a lot of attempts. Semi-supervised learning [143], [144], [146], weakly supervised learning [145], meta learning [262], small sample learning [147], [148], active learning [33], [149], [150], online learning [151], knowledge transfer [152], reinforcement learning [309]–[311], and learning without forgetting [260], [261] have also attracted the attention of person re-identification researchers. These studies have greatly promoted the exploration of open-world person re-identification.

G. DATA TYPES EXTENSION

1) RGB VIDEO DATA TYPE

Before 2016, the person re-identification research methods were mainly used CNN to extract spatial features on images. However, due to the small size of the datasets in the early stage and the limited available information of a single image, the advantages of deep learning cannot be given full play. In 2016, Zheng *et al.* [48] proposed a large-scale dataset MARS for motion analysis and person re-identification. Each identity in this dataset contains several video sequences, and many researchers began to study the video-based person re-identification methods using deep learning. The idea of image-based person re-identification methods can usually be borrowed or further developed by the video-based re-identification framework because video sequences are extensions of images in the time dimension. These data have more available information, such as inter-frame temporal relationship, pedestrian motion characteristics, and occlusion completion. But compared with image-based re-identification, continuous video frames often contain more noise and redundant information.

Although the person re-identification method based on video sequence has early machine learning methods [16]–[18], [47], the recent mainstream methods utilize RNN to extract temporal features while using CNN to extract spatial features [52]. The model input image sequence, each image through a shared Siamese network to extract the image spatial features. Then these feature vectors are input into an RNN network to extract the temporal features, as shown in Fig. 13. The final feature representation of the identities integrates the features of the single-frame image and the motion features between frames to replace the feature vectors of the previous single-frame method to train the network.

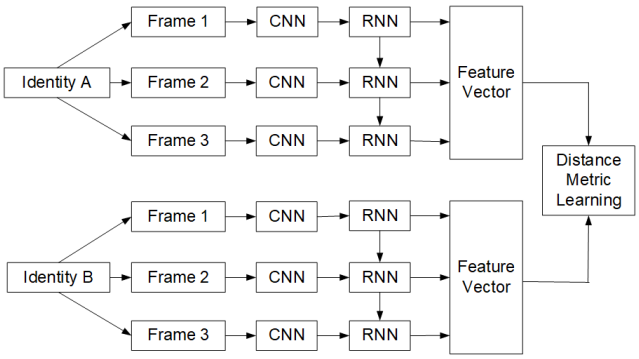


FIGURE 13. RGB Video-based person re-identification process.

References [19], [20] proposed to use the above process for the video-based person re-identification task. These papers are the earliest papers that use CNN and RNN to complete this task. Among them, McLaughlin *et al.* [20] inputted three-channel color images and two-channel images processed by the optical flow method. Subsequently, this network structure became the baseline of many video-based frameworks [21], [22]. These frameworks refer to other methods on this basis and use ResNet, GoogLeNet, and other common neural network structures as the base network. In [22], the author combined the attention mechanism based on the above structure, taking into account the emphasis of the human brain's attention on different frames and the interaction between video frame sequences. This paper proposed a joint spatial and temporal attention pooling network, which is mainly composed by a spatial pyramidal pooling and a temporal attention pooling in the natural language processing field. This method not only improved accuracy but also reduced computation. Chung *et al.* [219] gave different weights to spatial and temporal features. Zhang *et al.* [233] used Bidirectional Recurrent Neural Networks (BRCNN) in the temporal branch to better extract the information before and after frames.

The attention mechanism is also used in [23], [24], [141], [224]. Liu *et al.* [24] referred to the non-local attention layer [105], generated attention masks according to the characteristics of different frames and different spatial positions. And due to applying the non-local attention layer to multiple feature extraction layers, they utilized the redundancy in space and time to reduce the total floating-point operations (FLOP) and better balance the performance and complexity. Rao *et al.* [229] designed a new non-local temporal attention model, which could capture the long-term and global dependencies between video frames.

Part of the works [23], [227] took advantage of the fact that video sequences have more information than images to try to complete the occluded part. Among them, Hou *et al.* [23] proposed a Spatio-temporal completion network to restore the occludes in the video. In addition to using other parts of the person body to predict the occludes, this method also restored the appearance of persons with the help of time series information in a video, and finally used the unoccluded frame to train the re-identification network.

However, except for the advantages, video sequences also contain more noise. Several works [25], [141], [217], [220], [222], [227] are centered on this shortcoming. Li *et al.* [141] used a Temporal Self-Attention (TSA) model to deal with spot occlusions and noises, which could capture the global temporal cues and combined an improved Dilated Temporal Pyramid (DTP) convolution to distinguish two pedestrians, which focus on the local temporal context learning. Its structure design is simple, but the effect is better than the others. Recent results also suggest that current researches were more likely to use concise but effective designs. Liu *et al.* [25] proposed a network consisting of two branches, namely, the feature generation branch and the quality generation branch. This network mainly aimed at the fact that noise caused by jitter or blur in the video will affect the mutual supplement of a group of image information. The second branch in this network can predict the mass fraction of frames, and finally, extract the features of the first branch according to the fraction of each frame. This method focused on the set to set recognition and the training is simple, but the convergence speed is very slow due to the use of triplet loss.

Another problem with video-based re-identification is that video sequences often contain too many frames, which have redundant information. So the focuses of several works are to pick out frames worth noting [234]–[238] or assign different weights to different frames [218], [225], [228], [230], [231]. Zhou *et al.* [218] learned the weight of each frame utilizing the temporal attention model. Wu *et al.* [238] proposed a deep Siamese attention architecture to learn together spatial parts and temporal frames that require extra attention.

There are some works [217], [221], [235], [240] on the alignment of people in the video. Among them, Dai *et al.* [240] designed a spatial-temporal transformer network (ST2N) module to spatially align the people in the video, which leverage the context knowledge of consecutive frames.

In the video-based computer vision community, except using optical flow information and RNN to extract the features of the time dimension, another common method is to use a 3D convolutional neural network [241], which takes continuous frames as extra channels. This technology is also used in the video-based person re-identification [226], [235], [236]. Wu *et al.* [235] encoded the Spatio-temporal signals, extracted from video frame using 3D convolution, into a global compact video descriptor (see Fig. 14).

In the video-based re-identification dataset, each identity collects a set of video sequences instead of a single image. Such a set of video sequences complement each other, and also provides more information than a single image, such as a person's different positions. At the same time, video sequences include the walking posture and walking speed of a human, and light has less impact on the re-identification result, which is helpful for the neural network to extract more discriminant and robust features. However, video sequences also contain more noise, making the problem of human alignment more difficult. And most video-based person re-identification is still based on the attention mechanism, with few unique innovations in the temporal dimension. We believe that other time modeling methods need to be explored in the future.

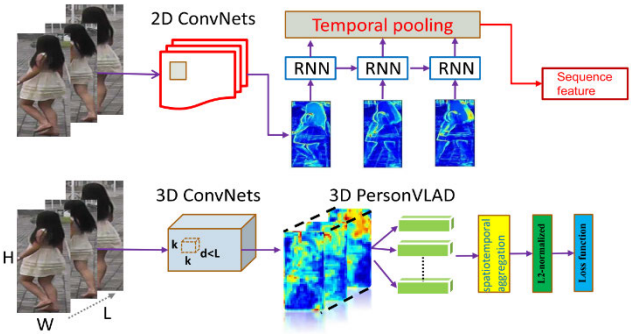


FIGURE 14. Illustration of video-based 2D convolution (top) and 3D convolution (bottom) [234].

2) OTHERS

a: DEPTH-BASED PERSON RE-IDENTIFICATION

Compared with the RGB camera with only color information, the RGB-D sensors can attach the spatial location information, and using RGB-D data for person re-identification is not easily affected by light. Specifically, the RGB-D sensors can sample the environment, and people within the scene to generate dense point clouds, which are arrangements of points that simulate the scene in three-dimensional space [156]. Given this advantage, some research outcomes on the person re-identification of the RGB-D sensor have emerged to fill this gap [160], [161]. Wu *et al.* [182] used the skeleton-based features to identify a person and, in particular, proposed an implicit feature transfer scheme when depth information was unavailable. Karianakis *et al.* [183] and Hafner *et al.* [184] mainly studied the cross-modal implementation from RGB-based image datasets to the RGB-D dataset because of their scale differences. There are also some several RGB-D datasets, such as RGBD-ID [157], BIWI RGBD-ID [158], and KinectREID [159], have been proposed to measure the performances of these studies. With the development of fields such as self-driving vehicles, computer vision research based on depth information will become more meaningful.

b: INFRARED-VISIBLE PERSON RE-IDENTIFICATION

Most person re-identification studies were done during the daytime and using visible images, and few studies on infrared images. Infrared images, mainly for the nighttime, are created by a thermal infrared scanner that receives and records thermal radiation energy emitted by a target. Dual-cameras capture different images depending on the brightness of the environment. However, People in infrared images have few obvious color features, but these color features are often key features for person re-identification, which causes this situation is more difficult than others (see Fig. 15). Therefore, in 2017, Wu *et al.* [162] first

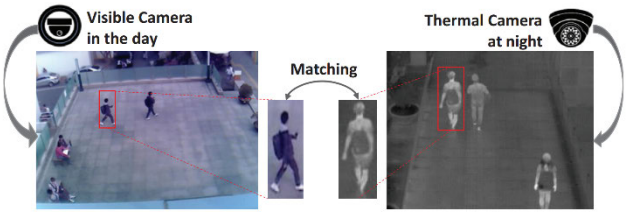


FIGURE 15. Illustration about the visible thermal person re-identification (VT-REID). Person images from different modalities should be matched [165].

presented the cross-modal re-identification problem between RGB images and infrared images. The following studies [164], [165], [185]–[190] mainly used dual-flow networks to extract features from each mode, then fused or shared features, and finally optimizes the scheme by specific metric learning. Among them, Wang *et al.* [164] first decomposed and treated the mixed modal and appearance differences separately. Some researchers also proposed some datasets, such as SYSU-MM01 [162] and RegDB [163].

c: LOW-RESOLUTION PERSON RE-IDENTIFICATION

The resolution difference, which caused by different camera resolutions or different distances between pedestrians and the cameras, is another focus in the person re-identification community. It is difficult to extract discriminative features from low-resolution images. Therefore, the general re-identification methods, which directly unified the low-resolution images and high-resolution images to the same size, are inconvincible. Li *et al.* [192] first studied the low-resolution problem in person re-identification through cross-scale image domain alignment. Subsequent studies focus on methods such as low-resolution information recovery [194], [197], dictionary learning [193], [195], and resolution adaptation [196], [198].

In addition, there are some special person re-identification works [239]. Text-based retrieval [199], [200] is similar to pedestrian attribute recognition and person search. Image to video person re-identification [201], [216] is also one of the research directions. These methods are better for helping the police collect evidence and solve crimes. Particularly, some studies [245], [246], [248] have attempted to construct 3D models of the human body to assist person-related researches. Chakrabortyn *et al.* [317] used face data collected by wearable devices for re-identification.

IV. DISCUSSION

In this section, we analyze and discuss the preferable performance results of person re-identification methods on each dataset and the main challenges of this task. We also propose our views on possible research directions in this field.

A. PERFORMANCES COMPARISON

In Table 4, we collect accuracy rates, including mAP and rank-1, 5, 10, and 20, reported by state-of-the-art methods on four kinds of datasets. In these datasets, the CUHK03 dataset is divided into handcrafted labeling parts and DPM detected parts. Generally speaking, the accuracy of DPM annotation is lower because it is closer to the real scene and influenced by the person detection algorithm. The Market1501 dataset usually has two comparison methods: single-query and multiple-query. The data in Table 4 are the results of studies with high accuracy that we know of at present, most of which are published in 2019.

Table 4 shows that among these research results with high accuracy, the most accurate one is the attention mechanism-based research methods. This is because, on the one hand, the attention mechanism can effectively reduce the impact or other interfering information in image or video on the accuracy. On the other hand, by determining the most noticeable part of the image or video, it can accurately locate the most differentiated part of the multiple similar probes to find the real target person. In addition to these methods, other methods, such as feature fusion and designing more effective distance measurement can also achieve better re-identification performance. Considering that re-ranking often leads to a significant improvement in accuracy, usually around 5% or even 10%, the results presented in Table 4 are compared without re-ranking.

According to the accuracy listed in Table 4, we can find that the accuracy of some early datasets has been very high with the contribution of researchers. Rank-20 has basically reached 95%. Some datasets, such as PRID2011 and iLIDS-VID, have reached 100% accuracy. Moreover, the rank-1 accuracy of most datasets is over 90%. Many of these datasets listed in Table 4 are manually annotated. The feature of these datasets is that the boundary boxes of person are well handled and there is no error condition existing in the automatic person detection algorithm. Therefore, it is easier to achieve high accuracy compared with the datasets detected by the automatic detection algorithm. However, other challenging datasets, such as MSMT17 and PRW, or data types, i.e., RGB-D and RGB-IR, remain less accurate because they are more realistic or cross-modal.

Finally, comparing the accuracy results before 2015, in terms of re-identification accuracy, the deep learning method greatly surpasses the traditional level of manual design features. However, the deep learning method is too dependent on the quality of data. If the training data can cover different application scenarios, the model will have generalization ability. Otherwise, better algorithms should be designed to make up for the lack of data.

B. CONSTRAINTS AND CHALLENGES

Compared with other visual image fields, person re-identification is a very challenging task, and there are still many difficulties, which lead to the current technology cannot be applied on a large scale and put higher requirements on the person re-identification algorithm. So the more discriminating and realistic model must be designed to distinguish two different persons. In the real world like Fig. 16, person re-identification is not a simple query-search task,

TABLE 4. The state-of-the-art methods with high accuracy on several datasets.

Data type	Datasets	Methods	Accuracy					Notes
			mAP	R-1	R-5	R-10	R-20	
RGB Image	VIPeR	BRE[15]	-	69.6	90.6	96.2	98.8	1.Ensemble different distance metrics. 2.Utilized rectifier loss to prevent overfitting.
	CUHK03 (detected)	[142]	-	93.2	99.2	-	-	1.A novel feedforward attention network and consistent attention regularizer. 2.Designed an improved triplet loss.
	CUHK03 (Labeled)	[142]	-	96.9	99.6	-	-	1.A novel feedforward attention network and consistent attention regularizer. 2.Designed an improved triplet loss.
	Market1501(SQ)	SCAL[319]	89.3	95.8	98.7	-	-	A self-critical attention learning method with reinforcement learning.
	Market1501(MQ)	HA-CNN [95]	82.8	93.8	-	-	-	Multilevel attention modeling.
	DukeMTMC-reID	SCAL[319]	79.6	89.0	95.1	-	-	A self-critical attention learning method with reinforcement learning.
	MSMT17	ABD-Net [139]	60.8	82.3	90.6	-	-	1.A novel spectral value difference orthogonality regularization. 2.Combined with channel-wise information and body part spatial awareness.
	KnightReid	[125]	10.2	14.3	22.5	26.7	31.4	A comprehensive benchmark result that is evaluated on the dataset.
RGB Video	3DPeS	FANN[43]	-	78.9	92.3	95.7	99.4	1.Emphasize the foreground persons adaptively. 2.Symmetric triplet loss function.
	PRID2011	GLTR [141]	-	95.5	100	100	100	Combine the long-term relations to alleviate the occlusions and noises and the short-term temporal cues to distinguish two pedestrians.
	iLIDS-VID	SCAN[92]	-	88.0	96.7	98.0	100.	Align the discriminative frames from two videos.
	MARS	NVAN[24]	82.8	90.0	-	-	-	Generate the attention mask by using the features of different frames and different spatial positions.
	DukeMTMC-VideoReID	NVAN[24]	94.9	96.3	-	-	-	Generate the attention mask by using the features of different frames and different spatial positions.
	PRW	[126]	33.4	73.6	-	-	-	1.A relative attention module to search and filter useful context information. 2.A graph learning framework to employ context pairs to update target similarity.
	LS-VID	GLTR[141]	44.3	63.1	77.2	83.8	88.4	Experimental result of the dataset presenters.
RGB-D	RGBD-ID	UVDL [320]	-	76.7	92.0	98.2	-	A uniform and variational deep learning method to exploit the correlations between RGB and depth images.
	KinectREID	UVDL [320]	-	99.4	100	100	-	A uniform and variational deep learning method to exploit the correlations between RGB and depth images.
RGB-IR	SYSU-MM01 (SQ, all-search)	MSR[189]	38.1	37.3	-	83.4	93.3	A separate network for each mode to extract a modal-specific representation and using DGD[69] as the baseline model.
	RegDB	D-HSME [188]	47.0	50.8	-	73.3	81.6	A dual-stream hypersphere manifold embedding network with decorrelation using Sphere Softmax.

FIGURE 16. Challenges in the person re-identification task. (a) Cropping Errors. (b) Indistinct Image. (c) Occlusion. (d) Illumination Variation. (e) Different identities with a similar appearance. (f) One identity with different clothes and hats.

and there are still many challenges to solve these problems completely. We summarize the main influencing factors into four categories.

1) INHERENT FACTORS

This factor causes the non-ideal scene. The commonly used person detection technology has problems such as low precision and cropping errors, which often leads to the person missing part or at the edge of the entire image. This problem brings the non-aligned challenge. In the early years, the resolution and image quality of the monitoring system is relatively low, so it is difficult to distinguish different pedestrians from blurred images. Besides, different shooting angles and heights [245] can also lead to huge personal visual differences.

2) ENVIRONMENTAL FACTORS

The current person re-identification approaches will segment the foreground human body to reduce the impact of the chaotic background. People who are not all in the foreground are covered when the human traffic in the shooting scene is intensive. In natural scenes, some sensitive features such as color will significantly deviate when the lighting change, which would affect the robust feature selection.

3) SUBJECTIVE FACTORS

Subjective factors cause matching difficulty and mainly reflected in the dressing and walking posture of people. Human beings will present various postures under different cameras, such as calling phones, turning around and changing the position and angle of their belongings. A human with a similar appearance to others or having different clothes, hats, glasses, or hairstyles will also bring huge challenges to confirm its identity.

4) DATA FACTORS

The data factor will cause the model to be easily over-fitting. Judging from the current person re-identification datasets, its collection and labeling are difficult, and the privacy issues have also hindered data access. Moreover, the spatial and temporal distribution of the collected data relative to the real world is very limited, incomplete, and circumscribed. Geographical location, weather changes, night and day, indoor and outdoor, and other scenes cannot be fully covered. The number of identities and cameras cannot meet the actual needs.

C. POSSIBLE DIRECTION FOR FUTURE WORKS

We consider that the research on person re-identification will still focus on deep learning in the next few years and analyze the possible research directions from the perspective of data and algorithms.

1) DATA AUGMENTATION

a: MORE REALISTIC DATASETS

Compared with other research fields, the current data of person re-identification datasets can only be used for academic research, and it is far from the actual application requirements. In addition to increasing the amount of data, it is more important to provide more diverse scenes, richer identities, more comprehensive human perspectives, and more realistic person re-identification datasets. For example, the elderly and children are easy to get lost, and accidents are likely to happen at night or on rainy or snowy days. These real situations most require person re-identification technology to locate the missing and suspect, but currently, there are few datasets included elderly, children, night cases, and extreme weather.

b: ADVERSARIAL EXAMPLES

The current person re-identification datasets have higher accuracy, but the approaches are not robust enough. The study of adversarial examples, a research hotspot, is very helpful to the system's robustness. However, Zheng *et al.* [176] demonstrated that the problem of person re-identification, an open set problem, requires different attack methods from other communities. For this issue, there is still a need for further in-depth study.

c: SYNTHETIC HUMANS

There are many methods for synthetic humans, such as GAN or 3D modeling, but the gap between the use of synthetic human data and real samples, as well as the proportion of synthetic data in all training data, on the effect of person re-identification still needs further analysis.

2) ALGORITHM DESIGN

a: MORE SUITABLE FOR LARGE, REAL-WORLD SYSTEMS

There are more identities in the real world than in the dataset, and there are plenty of instances where there are no matches in the gallery. Therefore, the optimization of matching algorithms and the screening of non-matching identities are important. Then, cross-modal algorithms may be emphasized, such as infrared-based re-identification. In addition, the real-world systems need not only result matching but also reasonable use of other information to assist the recognition task. Finally, neural architecture search can be attempted to be used for re-identification to reduce design complexity and experience dependence.

b: NOT JUST FULL-SUPERVISED LEARNING

Active learning can not only improve data utilization but also reduce data annotation. Self-supervised learning, more in line with the law of biological learning, has been very successful at Natural Language Processing (NLP). Recently, Wang *et al.* [177] have used self-supervised learning to achieve accurate 3D human posture estimation.

c: DRAWING ON IDEAS OF OTHER COMMUNITIES

Person re-identification can refer to the research method in vehicle re-identification, which makes full utilization of the Wi-Fi network, vehicle GPS, and regional electronic map to quickly eliminate persons who do not meet the requirements according to the camera distribution.

V. CONCLUSION

As an important technology in the smart city, person re-identification still needs to be perfected and further applied in many aspects. Therefore, we reviewed the research on person re-identification. Firstly, this paper has outlined some necessary premises of the person re-identification task. Then, given the diversity and complexity of re-identification approaches, we have summarized them according to different research purposes and have expressed some views on these techniques' pros and cons. Finally, we have analyzed the performance and the challenges of current re-identification research as well as proposed our ideas on the possible development direction in the future.

REFERENCES

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," 2016, *arXiv:1610.02984*. [Online]. Available: http://arxiv.org/abs/1610.02984

[2] A. Angelova, A. Krizhevsky, and V. Vanhoucke, "Pedestrian detection with a large-field-of-view deep network," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Seattle, WA, USA, May 2015, pp. 704–711.

[3] F. Porikli, "Inter-camera color calibration by correlation model function," in *Proc. ICIP*, Barcelona, Spain, 2003, p. II-133.

[4] W. Zajdel, Z. Zivkovic, and B. J. A. Krose, "Keeping track of humans: Have I seen this person before?" in *Proc. ICRA*, Barcelona, Spain, 2005, pp. 2081–2086.

[5] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Proc. CVPR*, New York, NY, USA, 2006, pp. 1528–1535.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Denver, CO, USA, Jun. 2010, pp. 2360–2367.

[7] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *Proc. Brit. Mach. Vis. Conf.*, London, U.K., 2011, pp. 68.1–68.11.

[8] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 898–903, May 2012.
[9] H. Chahar and N. Nain, "A study on deep convolutional neural network based approaches for person re-identification," in *Proc. PReMI*, Kolkata, India, vol. 10597, 2017, pp. 543–548.
[10] D. Wu, S.-J. Zheng, X.-P. Zhang, C.-A. Yuan, F. Cheng, Y. Zhao, Y.-J. Lin, Z.-Q. Zhao, Y.-L. Jiang, and D.-S. Huang, "Deep learning-based methods for person re-identification: A comprehensive review," *Neurocomputing*, vol. 337, pp. 354–371, Apr. 2019.
[11] X. Wang, S. Zheng, R. Yang, B. Luo, and J. Tang, "Pedestrian attribute recognition: A survey," 2019, *arXiv:1901.07474*. [Online]. Available: http://arxiv.org/abs/1901.07474
[12] R. Iguernaissi, D. Merad, K. Aziz, and P. Drap, "People tracking in multi-camera systems: A review," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 10773–10793, Sep. 2018.
[13] D. Li, Z. Zhang, X. Chen, and K. Huang, "A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1575–1590, Apr. 2019.
[14] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned CNN embedding for person reidentification," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, pp. 1–20, Dec. 2017.
[15] Z. Li, Z. Han, J. Xing, Q. Ye, X. Yu, and J. Jiao, "High performance person re-identification via a boosting ranking ensemble," *Pattern Recognit.*, vol. 94, pp. 187–195, Oct. 2019.
[16] K. Liu, B. Ma, W. Zhang, and R. Huang, "A spatio-temporal appearance representation for video-based pedestrian re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3810–3818.
[17] S. Karanam, Y. Li, and R. J. Radke, "Sparse re-id: Block sparsity for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Boston, MA, USA, Jun. 2015, pp. 33–40.
[18] Y. Yan, B. Ni, Z. Song, C. Ma, Y. Yan, and X. Yang, "Person re-identification via recurrent feature aggregation," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 701–716.
[19] L. Wu, C. Shen, and A. van den Hengel, "Deep recurrent convolutional networks for video-based person re-identification: An end-to-end approach," 2016, *arXiv:1606.01609*. [Online]. Available: http://arxiv.org/abs/1606.01609
[20] N. McLaughlin, J. M. D. Rincon, and P. Miller, "Recurrent convolutional network for video-based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1325–1334.
[21] H. Liu, Z. Jie, K. Jayashree, M. Qi, J. Jiang, S. Yan, and J. Feng, "Video-based person re-identification with accumulative motion context," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2788–2802, Oct. 2018.
[22] S. Xu, Y. Cheng, K. Gu, Y. Yang, S. Chang, and P. Zhou, "Jointly attentive spatial-temporal pooling networks for video-based person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4743–4752.
[23] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "VRSTC: Occlusion-free video person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 7183–7192.
[24] C.-T. Liu, C.-W. Wu, Y.-C. F. Wang, and S.-Y. Chien, "Spatially and temporally efficient non-local attention network for video-based person re-identification," 2019, *arXiv:1908.01683*. [Online]. Available: http://arxiv.org/abs/1908.01683
[25] Y. Liu, J. Yan, and W. Ouyang, "Quality aware network for set to set recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4694–4703.
[26] J. Liu, W. Li, H. Pei, Y. Wang, F. Qu, Y. Qu, and Y. Chen, "Identity preserving generative adversarial network for cross-domain person re-identification," *IEEE Access*, vol. 7, pp. 114021–114032, 2019.
[27] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 4, pp. 1–18, Oct. 2018.
[28] H.-X. Yu, A. Wu, and W.-S. Zheng, "Cross-view asymmetric metric learning for unsupervised person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 994–1002.
[29] V. Nguyen, T. D. Ngo, K. M. T. T. Nguyen, D. A. Duong, K. Nguyen, and D. Le, "Re-ranking for person re-identification," in *Proc. SoCPaR*, Hanoi, Vietnam, 2013, pp. 304–308.
[30] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen, "Person re-identification with content and context re-ranking," *Multimedia Tools Appl.*, vol. 74, no. 17, pp. 6989–7014, 2015.
[31] S. Bai, X. Bai, and Q. Tian, "Scalable person re-identification on supervised smoothed manifold," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3356–3365.
[32] C. Liu, C. C. Loy, S. Gong, and G. Wang, "POP: Person re-identification post-rank optimisation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 441–448.
[33] H. Wang, S. Gong, X. Zhu, and T. Xiang, "Human-in-the-loop person re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 405–422.
[34] M. Ye, J. Chen, Q. Leng, C. Liang, Z. Wang, and K. Sun, "Coupled-view based ranking optimization for person re-identification," in *Proc. MMM*, Sydney, NSW, Australia, 2015, pp. 105–117.
[35] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with K-reciprocal encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1318–1327.
[36] H. Liu and J. Cheng, "Gallery based K-reciprocal-like re-ranking for heavy cross-camera discrepancy in person re-identification," *Neurocomputing*, vol. 333, no. 14, pp. 64–75, Mar. 2019.
[37] J. Garcia, N. Martinel, C. Micheloni, and A. Gardel, "Person re-identification ranking optimisation by discriminant context information analysis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1305–1313.
[38] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. ECCV Workshops*, Amsterdam, The Netherland, 2016, pp. 17–35.
[39] X. Su, X. Qu, Z. Zou, P. Zhou, W. Wei, S. Wen, and M. Hu, "K-reciprocal harmonious attention network for video-based person re-identification," *IEEE Access*, vol. 7, pp. 22457–22470, 2019.
[40] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[41] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, Munich, Germany, 2008, pp. 262–275.
[42] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1116–1124.
[43] S. Zhou, J. Wang, D. Meng, Y. Liang, Y. Gong, and N. Zheng, "Discriminative feature learning with foreground attention for person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4671–4684, Sep. 2019.
[44] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 79–88.
[45] D. Baltieri, R. Vezzani, and R. Cucchiara, "3DPeS: 3D people dataset for surveillance and forensics," in *Proc. J-HGBU*, New York, NY, USA, 2011, pp. 59–64.
[46] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. SCIA*, Berlin, Germany, 2011, pp. 91–102.
[47] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *Proc. ECCV*, Zürich, Switzerland, 2014, pp. 688–703.
[48] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "MARS: A video benchmark for large-scale person re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 868–884.
[49] G. Song, B. Leng, Y. Liu, C. Hetang, and S. Cai, "Region-based quality estimation network for large-scale person re-identification," presented at the AAAI, New Orleans, USA, Feb. 2018.
[50] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1846–1855.
[51] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Ouyang, and Y. Yang, "Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 5177–5186.
[52] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[53] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[54] R. Layne, T. Hospedales, and S. Gong, "Person re-identification by attributes," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 8. [Online]. Available: http://www.bmva.org/bmvc/2012/BMVC/paper024/index.html

[55] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: What features are important?" in *Proc. ECCV*, Berlin, Germany, 2012, pp. 391–401.

[56] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao, "Multi-task learning with low rank attribute embedding for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3739–3747.

[57] Z. Shi, T. M. Hospedales, and T. Xiang, "Transferring a semantic representation for person re-identification and search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 4184–4193.

[58] D. Li, Z. Zhang, X. Chen, H. Ling, and K. Huang, "A richly annotated dataset for pedestrian attribute recognition," 2016, *arXiv:1603.07054*. [Online]. Available: http://arxiv.org/abs/1603.07054

[59] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," in *Proc. Adv. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 2002, pp. 521–528.

[60] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.

[61] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 2288–2295.

[62] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 2197–2206.

[63] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[64] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[66] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.

[67] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 34–39.

[68] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 152–159.

[69] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1249–1258.

[70] R. R. Varior, M. Haloi, and G. Wang, "Gated Siamese convolutional neural network architecture for human re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 791–808.

[71] H. Liu, J. Feng, M. Qi, J. Jiang, and S. Yan, "End-to-End comparative attention networks for person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3492–3506, Jul. 2017.

[72] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1335–1344.

[73] S. Wu, Y.-C. Chen, X. Li, A.-C. Wu, J.-J. You, and W.-S. Zheng, "An enhanced deep feature representation for person re-identification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Lake Placid, NY, USA, Mar. 2016, pp. 1–8.

[74] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang, "A Siamese long short-term memory architecture for human re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 135–153.

[75] L. Wei, S. Zhang, H. Yao, W. Gao, and Q. Tian, "Glad: Global-local-alignment descriptor for pedestrian retrieval," in *Proc. MM*, Mountain View, CA, USA, 2017, pp. 420–428.

[76] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proc. ECCV*, Munich, Germany, 2018, pp. 501–518, doi: 10.1007/978-3-030-01225-0_30.

[77] D. Li, X. Chen, Z. Zhang, and K. Huang, "Learning deep context-aware features over body and latent parts for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 7398–7407.

[78] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. NIPS*, Montreal, QC, Canada, Dec. 2015, pp. 2017–2025.

[79] L. Zheng, Y. Huang, H. Lu, and Y. Yang, "Pose-invariant embedding for deep person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4500–4509, Sep. 2019.

[80] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang, "Spindle net: Person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1077–1085.

[81] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative adversarial networks: Introduction and outlook," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 4, pp. 588–598, Sep. 2017.

[82] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 3754–3762.

[83] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5967–5976.

[84] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. NIPS*, Barcelona, Spain, Dec. 2016, pp. 469–477.

[85] X. Qian, Y. Fu, and W. Wang, "Pose-normalized image generation for person re-identification," in *Proc. ECCV*, Munich, Germany, 2018, pp. 661–678.

[86] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camera style adaptation for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 5157–5166.

[87] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, and J. Wang, "Person re-identification with correspondence structure learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3200–3208.

[88] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong, "Partial person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 4678–4686.

[89] H. Wang, S. Gong, and T. Xiang, "Unsupervised learning of generative topic saliency for person re-identification," presented at the BMVC, Nottingham, U.K., 2014. [Online]. Available: http://dx.doi.org/10.5244/C.28.48

[90] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 3586–3593.

[91] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Pose-driven deep convolutional model for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 3980–3989.

[92] R. Zhang, J. Li, H. Sun, Y. Ge, P. Luo, X. Wang, and L. Lin, "SCAN: Self-and-Collaborative attention network for video person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 4870–4882, Oct. 2019.

[93] L. Zhao, X. Li, Y. Zhuang, and J. Wang, "Deeply-learned part-aligned representations for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 3239–3248.

[94] M. Zeng, Z. Wu, C. Tian, L. Zhang, and L. Hu, "Efficient person re-identification by hybrid spatiogram and covariance descriptor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Boston, MA, USA, Jun. 2015, pp. 48–56.

[95] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2285–2294.

[96] P. Singh, V. K. Verma, P. Rai, and V. P. Namboodiri, "HetConv: Heterogeneous kernel-based convolutions for deep CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 4835–4844.

[97] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," 2017, *arXiv:1710.09282*. [Online]. Available: http://arxiv.org/abs/1710.09282

[98] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proc. CVPR Workshops*, Long Beach, CA, USA, Jun. 2019, pp. 4321–4329.
[99] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural Comput.*, vol. 22, no. 12, pp. 3207–3220, Dec. 2010.
[100] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," 2017, *arXiv:1708.04896*. [Online]. Available: http://arxiv.org/abs/1708.04896
[101] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, Cambridge, MA, USA, Dec. 2014, pp. 2672–2680.
[102] Y. Huang, J. Xu, Q. Wu, Z. Zheng, Z. Zhang, and J. Zhang, "Multi-pseudo regularized label for generated data in person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1391–1403, Mar. 2019.
[103] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 994–1003.
[104] M. Zheng, S. Karanam, Z. Wu, and R. J. Radke, "Re-identification with consistent attentive Siamese networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 5635–5744.
[105] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2018, pp. 7794–7803.
[106] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, Corvallis, OR, USA, 2007, pp. 209–216.
[107] B. McFee and G. R. Lanckriet, "Metric learning to rank," in *Proc. ICML*, Haifa, Israel, 2010, pp. 775–782.
[108] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*. [Online]. Available: http://arxiv.org/abs/1703.07737
[109] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 403–412.
[110] H. Chen, Y. Wang, Y. Shi, K. Yan, M. Geng, Y. Tian, and T. Xiang, "Deep transfer learning for person re-identification," in *Proc. IEEE 4th Int. Conf. Multimedia Big Data (BigMM)*, Xi'an, China, Sep. 2018, pp. 1–5.
[111] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[112] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A deeper, stronger, and faster multi-person pose estimation model," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 34–50.
[113] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: Exemplar memory for domain adaptive person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 598–607.
[114] Q. Xiao, H. Luo, and C. Zhang, "Margin sample mining loss: A deep learning based method for person re-identification," 2017, *arXiv:1710.00478*. [Online]. Available: http://arxiv.org/abs/1710.00478
[115] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*. [Online]. Available: http://arxiv.org/abs/1712.04621
[116] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 815–823.
[117] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proc. NIPS*, Cambridge, MA, USA, Dec. 2014, pp. 2204–2212.
[118] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[119] S. Xu, T. Fang, D. Li, and S. Wang, "Object classification of aerial images with bag-of-visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 366–370, Apr. 2010.
[120] Y. Zhang, R. Jin, and Z.-H. Zhou, "Understanding bag-of-words model: A statistical framework," *Int. J. Mach. Learn. Cybern.*, vol. 1, nos. 1–4, pp. 43–52, Aug. 2010.
[121] S. Malpani, C. S. Asha, and A. V. Narasimhadhan, "Thermal vision human classification and localization using bag of visual word," in *Proc. IEEE Region Conf. (TENCON)*, Singapore, Nov. 2016, pp. 3135–3139.
[122] X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun, "AlignedReID: Surpassing human-level performance in person re-identification," 2017, *arXiv:1711.08184*. [Online]. Available: http://arxiv.org/abs/1711.08184
[123] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu, "Object retrieval and localization with spatially-constrained similarity measure and k-NN re-ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 3013–3020.
[124] Z. Zhong, M. Lei, D. Cao, J. Fan, and S. Li, "Class-specific object proposals re-ranking for object detection in automatic driving," *Neurocomputing*, vol. 242, pp. 187–194, Jun. 2017.
[125] J. Zhang, Y. Yuan, and Q. Wang, "Night person re-identification and a benchmark," *IEEE Access*, vol. 7, pp. 95496–95504, 2019.
[126] Y. Yan, Q. Zhang, B. Ni, W. Zhang, M. Xu, and X. Yang, "Learning context graph for person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA Jun. 2019, pp. 2158–2167.
[127] K. Wang, H. Wang, M. Liu, X. Xing, and T. Han, "Survey on person re-identification based on deep learning," *CAAI Trans. Intell. Technol.*, vol. 3, no. 4, pp. 219–227, Dec. 2018.
[128] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, "The relation between the ROC curve and the CMC," in *Proc. AutoID*, Buffalo, NY, USA, 2005, pp. 15–20.
[129] A. J. Ma, J. Li, P. C. Yuen, and P. Li, "Cross-domain person reidentification using domain adaptation ranking SVMs," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1599–1613, May 2015.
[130] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian, "Unsupervised cross-dataset transfer learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1306–1315.
[131] Z. Liu, D. Wang, and H. Lu, "Stepwise metric promotion for unsupervised video person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2448–2457.
[132] L. Song, C. Wang, L. Zhang, B. Du, Q. Zhang, C. Huang, and X. Wang, "Unsupervised domain adaptive re-identification: Theory and practice," 2018, *arXiv:1807.11334*. [Online]. Available: http://arxiv.org/abs/1807.11334
[133] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero- and homogeneously," in *Proc. ECCV*, Munich, Germany, 2018, pp. 176–192.
[134] S. Bąk, P. Carr, and J. F. Lalonde, "Domain adaptation through synthesis for unsupervised person re-identification," in *Proc. ECCV*, Munich, Germany, 2018, pp. 193–209.
[135] J. Wang, X. Zhu, S. Gong, and W. Li, "Transferable joint attribute-identity deep learning for unsupervised person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2275–2284.
[136] S. Lin, H. Li, C.-T. Li, and A. C. Kot, "Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification," in *Proc. BMVC*, Newcastle, U.K., 2018, p. 9.
[137] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "CamStyle: A novel data augmentation method for person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1176–1190, Mar. 2019.
[138] J. Guo, Y. Yuan, L. Huang, C. Zhang, J.-G. Yao, and K. Han, "Beyond human parts: Dual part-aligned representations for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 3642–3651.
[139] T. Chen, S. Ding, J. Xie, Y. Yuan, W. Chen, Y. Yang, Z. Ren, and Z. Wang, "ABD-Net: Attentive but diverse person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8351–8361.
[140] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, U. Uiuc, and T. Huang, "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 6112–6121.
[141] J. Li, S. Zhang, J. Wang, W. Gao, and Q. Tian, "Global-local temporal representations for video person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 3958–3967.

[142] S. Zhou, F. Wang, Z. Huang, and J. Wang, "Discriminative feature learning with consistent attention regularization for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8040–8049.

[143] X. Xin, J. Wang, R. Xie, S. Zhou, W. Huang, and N. Zheng, "Semi-supervised person re-identification using multi-view clustering," *Pattern Recognit.*, vol. 88, pp. 285–297, Apr. 2019.

[144] Y. Liu, G. Song, J. Shao, X. Jin, and X. Wang, "Transductive centroid projection for semi-supervised large-scale recognition," in *Proc. ECCV*, Munich, Germany, 2018, pp. 70–86.

[145] J. Meng, S. Wu, and W.-S. Zheng, "Weakly supervised person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 760–769.

[146] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Bian, and Y. Yang, "Progressive learning for person re-identification with one example," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2872–2881, Jun. 2019.

[147] T. M. F. Ali and S. Chaudhuri, "Maximum margin metric learning over discriminative nullspace for person re-identification," in *Proc. ECCV*, Munich, Germany, 2018, pp. 123–141.

[148] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1239–1248.

[149] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Temporal model adaptation for person re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 858–877.

[150] S. Roy, S. Paul, N. E. Young, and A. K. Roy-Chowdhury, "Exploiting transitivity for learning person re-identification models on a budget," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7064–7072.

[151] W.-H. Li, Z. Zhong, and W.-S. Zheng, "One-pass person re-identification by sketch online discriminant analysis," 2017, *arXiv:1711.03368*. [Online]. Available: http://arxiv.org/abs/1711.03368

[152] A. Wu, W.-S. Zheng, X. Guo, and J.-H. Lai, "Distilled person re-identification: Towards a more scalable system," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 1187–1196.

[153] Y. Shen, H. Li, S. Yi, D. Chen, and X. Wang, "Person re-identification with deep similarity-guided graph neural network," in *Proc. ECCV*, Munich, Germany, 2018, pp. 508–526.

[154] J. Wu, H. Liu, Y. Yang, Z. Lei, S. Liao, and S. Li, "Unsupervised graph association for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8321–8330.

[155] Q. Zhou, H. Fan, H. Yang, H. Su, S. Zheng, S. Wu, and H. Ling, "Robust and efficient graph correspondence transfer for person re-identification," *IEEE Trans. Image Process.*, to be published, doi: 10.1109/TIP.2019.2914575.

[156] C. Patruno, R. Marani, G. Cicirelli, E. Stella, and T. D'Orazio, "People re-identification using skeleton standard posture and color descriptors from RGB-D data," *Pattern Recognit.*, vol. 89, pp. 77–90, May 2019.

[157] I. B. Barbosa, M. Cristani, A. D. Bue, L. Bazzani, and V. Murino, "Re-identification with RGB-D sensors," in *Proc. ECCV*, Berlin, Germany, 2012, pp. 433–442.

[158] M. Munaro, A. Fossati, and A. Basso, "One-shot person re-identification with a consumer depth camera," in *Springer Person Re-Identification*. London, U.K.: Springer, 2014, pp. 161–181.

[159] F. Pala, R. Satta, G. Fumera, and F. Roli, "Multimodal person reidentification using RGB-D cameras," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 788–799, Apr. 2016.

[160] D. Liciotti, M. Paolanti, E. Frontoni, A. Mancini, and P. Zingaretti, "Person re-identification dataset with RGB-D camera in a top-view configuration," in *Proc. VAAM, FFER*, Cham, Switzerland, 2016, pp. 1–11.

[161] J. Lorenzo-Navarro, M. Castrillón-Santana, and D. Hernández-Sosa, "On the use of simple geometric descriptors provided by RGB-D sensors for re-identification," *Sensors*, vol. 13, no. 7, pp. 8222–8238, Jun. 2013.

[162] A. Wu, W.-S. Zheng, H.-X. Yu, S. Gong, and J. Lai, "RGB-infrared cross-modality person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5390–5399.

[163] D. Nguyen, H. Hong, K. Kim, and K. Park, "Person recognition system based on a combination of body images from visible light and thermal cameras," *Sensors*, vol. 17, no. 3, p. 605, Mar. 2017.

[164] Z. Wang, Z. Wang, Y. Zheng, Y.-Y. Chuang, and S. Satoh, "Learning to reduce dual-level discrepancy for infrared-visible person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 618–626.

[165] M. Ye, X. Lan, J. Li, and P. C. Yuen, "Hierarchical discriminative learning for visible thermal person re-identification," presented at the AAAI, New Orleans, USA, Feb. 2018.

[166] X. Zhu, B. Wu, D. Huang, and W.-S. Zheng, "Fast open-world person re-identification," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2286–2300, May 2018.

[167] J. Chen, Y. Wang, J. Qin, L. Liu, and L. Shao, "Fast person re-identification via cross-camera semantic binary transformation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3873–3882.

[168] Y.-J. Cho and K.-J. Yoon, "Distance-based camera network topology inference for person re-identification," *Pattern Recognit. Lett.*, vol. 125, no. 1, pp. 220–227, Jul. 2019.

[169] J. Lin, L. Ren, J. Lu, J. Feng, and J. Zhou, "Consistent-aware deep learning for person re-identification in a camera network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3396–3405.

[170] G. Wang, J. Lai, P. Huang, and X. Xie, "Spatial-temporal person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, Honolulu, Hawaii, USA, Jul. 2019, pp. 8933–8940.

[171] N. Narayan, N. Sankaran, D. Arpit, K. Dantu, S. Setlur, and V. Govindaraju, "Person re-identification for improved multi-person multi-camera tracking by continuous entity association," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 566–572.

[172] H. Yang, Z. Cheng, and L. Chen, "Reranking optimization for person re-identification under temporal-spatial information and common network consistency constraints," *Pattern Recognit. Lett.*, vol. 127, no. 1, pp. 146–155, Nov. 2019.

[173] A. Das, A. Chakraborty, and A. K. Roy-Chowdhury, "Consistent re-identification in a camera network," in *Proc. ECCV*, Zürich, Switzerland, 2014, pp. 330–345.

[174] Y. Cai and G. Medioni, "Exploring context information for inter-camera multiple target tracking," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Steamboat Springs, CO, USA, Mar. 2014, pp. 761–768.

[175] Y.-J. Cho, J.-H. Park, S.-A. Kim, K. Lee, and K.-J. Yoon, "Unified framework for automated person re-identification and camera network topology inference in camera networks," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Venice, Italy, Oct. 2017, pp. 2601–2607.

[176] Z. Zheng, L. Zheng, Z. Hu, and Y. Yang, "Open set adversarial examples," 2018, *arXiv:1809.02681*. [Online]. Available: http://arxiv.org/abs/1809.02681

[177] K. Wang, L. Lin, C. Jiang, C. Qian, and P. Wei, "3D human pose machines with self-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[178] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke, "A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 523–536, Mar. 2019.

[179] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Comput. Surv.*, vol. 46, no. 2, no. 29, pp. 1–37, Dec. 2013.

[180] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image Vis. Comput.*, vol. 32, no. 4, pp. 270–286, Apr. 2014.

[181] Z. Wang, Z. Wang, Y. Zheng, Y. Wu, and S. Satoh, "Beyond intra-modality: A survey of heterogeneous person re-identification," 2019, *arXiv:1905.10048*. [Online]. Available: http://arxiv.org/abs/1905.10048

[182] A. Wu, W.-S. Zheng, and J.-H. Lai, "Robust depth-based person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2588–2603, Jun. 2017.

[183] N. Karianakis, Z. Liu, and Y. Chen, "Reinforced temporal attention and split-rate transfer for depth-based person re-identification," in *Proc. ECCV*, Munich, Germany, 2018, pp. 737–756.

[184] F. M. Hafner, A. Bhuiyan, J. F. P. Kooij, and E. Granger, "RGB-depth cross-modal person re-identification," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Taipei, Taiwan, Sep. 2019, pp. 1–8.

[185] M. Ye, Z. Wang, X. Lan, and P. C. Yuen, "Visible thermal person re-identification via dual-constrained top-ranking," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 1092–1099.

[186] M. Ye, X. Lan, and Q. Leng, "Modality-aware collaborative learning for visible thermal person re-identification," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 347–355.

[187] G. Wang, T. Zhang, J. Cheng, S. Liu, Y. Yang, and Z. Hou, "RGB-infrared cross-modality person re-identification via joint pixel and feature alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 3623–3632.
[188] Y. Hao, N. Wang, J. Li, and X. Gao, "HSME: Hypersphere manifold embedding for visible thermal person re-identification," presented at the AAAI, Honolulu, Hawaii, USA, Jan. 2019.
[189] Z. Feng, J. Lai, and X. Xie, "Learning modality-specific representations for visible-infrared person re-identification," *IEEE Trans. Image Process.*, vol. 29, pp. 579–590, Jul. 2019, doi: 10.1109/TIP.2019.2928126.
[190] M. Ye, X. Lan, Z. Wang, and P. C. Yuen, "Bi-directional center-constrained top-ranking for visible thermal person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 407–419, Jun. 2019, doi 10.1109/TIFS.2019.2921454.
[191] M. O. Almasawa, L. A. Elrefaei, and K. Moria, "A survey on deep learning-based person re-identification systems," *IEEE Access*, vol. 7, pp. 175228–175247, 2019.
[192] X. Li, W.-S. Zheng, X. Wang, T. Xiang, and S. Gong, "Multi-scale learning for low-resolution person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3765–3773.
[193] X. Y. Jing, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," in *Proc. CVPR Workshops*, Boston, MA, USA, 2015, pp. 695–704.
[194] Z. Wang, M. Ye, F. Yang, X. Bai, and S. Satoh, "Cascaded SR-GAN for scale-adaptive low resolution person re-identification," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3891–3897.
[195] K. Li, Z. Ding, S. Li, and Y. Fu, "Discriminative semi-coupled projective dictionary learning for low-resolution person re-identification," presented at the AAAI, New Orleans, LA, USA, Feb. 2018.
[196] J. Jiao, W. S. Zheng, A. Wu, X. Zhu, and S. Gong, "Deep low-resolution person re-identification," presented at the AAAI, New Orleans, LA, USA, Feb. 2018.
[197] Y.-J. Li, Y.-C. Chen, Y.-Y. Lin, X. Du, and Y.-C.-F. Wang, "Recover and identify: A generative dual model for cross-resolution person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8090–8099.
[198] Y. C. Chen, Y. J. Li, X. Du, and Y. C. F. Wang, "Learning resolution-invariant deep representations for person re-identification," presented at the AAAI, Honolulu, HI, USA, Jan. 2019.
[199] D. Chen, H. Li, X. Liu, Y. Shen, J. Shao, Z. Yuan, and X. Wang, "Improving deep visual representation for person re-identification by global and local image-language association," in *Proc. ECCV*, Munich, Germany, 2018, pp. 56–73.
[200] J. Liu, Z.-J. Zha, R. Hong, M. Wang, and Y. Zhang, "Deep adversarial graph attention convolution network for text-based person search," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 665–673.
[201] X. Zhu, X.-Y. Jing, X. You, W. Zuo, S. Shan, and W.-S. Zheng, "Image to video person re-identification by learning heterogeneous dictionary pair with feature projection matrix," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 717–732, Mar. 2018.
[202] W. Fei, Z. Zhao, and F. Su, "Deep global and local saliency learning with new re-ranking for person re-identification," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, San Diego, CA, USA, Jul. 2018, pp. 1–6.
[203] N. Mansouri, S. Ammar, and Y. Kessentini, "Improving person re-identification by combining Siamese convolutional neural network and re-ranking process," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Taipei, Taiwan, Sep. 2019, pp. 1–8.
[204] Z. Wang, J. Jiang, Y. Yu, and S. Satoh, "Incremental re-identification by cross-direction and cross-ranking adaption," *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2376–2386, Sep. 2019.
[205] C. Luo, Y. Chen, N. Wang, and Z.-X. Zhang, "Spectral feature transformation for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 4976–4985.
[206] S. Bai, P. Tang, P. H. S. Torr, and L. J. Latecki, "Re-ranking via metric fusion for object retrieval and person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 740–749.
[207] W. Li, Y. Wu, M. Mukunoki, and M. Minoh, "Common-near-neighbor analysis for person re-identification," in *Proc. 19th IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 1621–1624.
[208] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, and R. Hu, "Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2553–2566, Dec. 2016.
[209] Y. Guo and N.-M. Cheung, "Efficient and deep person re-identification using multi-level similarity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2335–2344.
[210] C.-Y. Wang, P.-Y. Chen, M.-C. Chen, J.-W. Hsieh, and H.-Y.-M. Liao, "Real-time video-based person re-identification surveillance with light-weight deep convolutional networks," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Taipei, Taiwan, Sep. 2019, pp. 1–8.
[211] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-scale feature learning for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 3702–3712.
[212] H. Yao, "Large-scale person re-identification as retrieval," in *Proc. ICME*, Hong Kong, 2017, pp. 1440–1445.
[213] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4766–4779, Dec. 2015.
[214] F. Zhu, X. Kong, L. Zheng, H. Fu, and Q. Tian, "Part-based deep hashing for large-scale person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4806–4817, Oct. 2017.
[215] L. Wu, Y. Wang, Z. Ge, Q. Hu, and X. Li, "Structured deep hashing with convolutional neural networks for fast person re-identification," *Comput. Vis. Image Understand.*, vol. 167, pp. 63–73, Feb. 2018.
[216] X. Gu, B. Ma, H. Chang, S. Shan, and X. Chen, "Temporal knowledge propagation for Image-to-Video person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 9647–9656.
[217] J. You, A. Wu, X. Li, and W.-S. Zheng, "Top-push video-based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1345–1353.
[218] Z. Zhou, Y. Huang, W. Wang, L. Wang, and T. Tan, "See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6776–6785.
[219] D. Chung, K. Tahboub, and E. J. Delp, "A two stream siamese convolutional neural network for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 1992–2000.
[220] D. Chen, H. Li, T. Xiao, S. Yi, and X. Wang, "Video person re-identification with competitive snippet-similarity aggregation and co-attentive snippet embedding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 1169–1178.
[221] S. Li, S. Bak, P. Carr, and X. Wang, "Diversity regularized spatiotemporal attention for video-based person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 369–378.
[222] A. Subramaniam, A. Nambiar, and A. Mittal, "Co-segmentation inspired attention networks for video-based person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 562–572.
[223] X. Zhang, F. Pala, and B. Bhanu, "Attributes co-occurrence pattern mining for video-based person re-identification," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Lecce, Italy, Aug. 2017, pp. 1–6.
[224] Y. Wu, "Temporal-enhanced convolutional network for person re-identification," presented at the AAAI, New Orleans, LA, USA, Feb. 2018.
[225] W. Huang, "Video-based person re-identification via self paced weighting," presented at the AAAI, New Orleans, LA, USA, Feb. 2018.
[226] J. Li, S. Zhang, and T. Huang, "Multi-scale 3D convolution network for video based person re-identification," presented at the AAAI, Honolulu, HI, USA, Jan. 2019.
[227] Y. Liu, Z. Yuan, W. Zhou, and H. Li, "Spatial and temporal mutual promotion for video-based person re-identification," presented at the AAAI, Honolulu, HI, USA, Jan. 2019.
[228] Y. Fu, X. Wang, Y. Wei, and T. Huang, "STA: Spatial-temporal attention for large-scale video-based person re-identification," presented at the AAAI, Honolulu, HI, USA, Jan. 2019.
[229] S. Rao, P. Cao, T. Rahman, M. Rochan, and Y. Wang, "Non-local attentive temporal network for video-based person re-identification," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Taipei, Taiwan, Sep. 2019, pp. 1–8.
[230] T. Rahman, M. Rochan, and Y. Wang, "Video-based person re-identification using refined attention networks," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Taipei, Taiwan, Sep. 2019, pp. 1–8.

[231] T. Rahman, M. Rochan, and Y. Wang, "Convolutional temporal attention model for video-based person re-identification," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, Jul. 2019, pp. 1102–1107.

[232] L. Bao, B. Ma, H. Chang, and X. Chen, "Preserving structural relationships for person re-identification," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Shanghai, China, Jul. 2019, pp. 120–125.

[233] W. Zhang, X. Yu, and X. He, "Learning bidirectional temporal cues for video-based person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2768–2776, Oct. 2018.

[234] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by discriminative selection in video ranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12, pp. 2501–2514, Dec. 2016.

[235] L. Wu, Y. Wang, L. Shao, and M. Wang, "3-D PersonVLAD: Learning deep global representations for video-based person reidentification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3347–3359, Nov. 2019.

[236] J. Liu, Z.-J. Zha, X. Chen, Z. Wang, and Y. Zhang, "Dense 3D-convolutional neural network for person re-identification in videos," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 1s, pp. 1–19, Jan. 2019.

[237] W. Zhang, X. He, W. Lu, H. Qiao, and Y. Li, "Feature aggregation with reinforcement learning for video-based person re-identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3847–3852, Dec. 2019.

[238] L. Wu, Y. Wang, J. Gao, and X. Li, "Where-and-When to look: Deep siamese attention networks for video-based person re-identification," *IEEE Trans. Multimedia*, vol. 21, no. 6, pp. 1412–1424, Jun. 2019.

[239] F. Ma, X.-Y. Jing, X. Zhu, Z. Tang, and Z. Peng, "True-color and grayscale video person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 115–129, May 2019, doi: 10.1109/TIFS.2019.2917160.

[240] J. Dai, P. Zhang, D. Wang, H. Lu, and H. Wang, "Video person re-identification by temporal residual learning," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1366–1377, Mar. 2019.

[241] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 4489–4497.

[242] J. Liu, B. Ni, Y. Yan, P. Zhou, S. Cheng, and J. Hu, "Pose transferrable person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4099–4108.

[243] X. Li, A. Wu, and W. S. Zheng, "Adversarial open-world person re-identiication," in *Proc. ECCV*, Munich, Germany, 2018, pp. 280–296.

[244] J. Liu, Y. Zhou, L. Sun, and Z. Jiang, "Similarity preserved camera-to-camera GAN for person re-identification," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Shanghai, China, Jul. 2019, pp. 531–536.

[245] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of viewpoint," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 608–617.

[246] J. Wang, Y. Zhong, Y. Li, C. Zhang, and Y. Wei, "Re-identification supervised texture generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 11838–11848.

[247] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 2133–2142.

[248] D. Baltieri, R. Vezzani, and R. Cucchiara, "Mapping appearance descriptors on 3D body models for people re-identification," *Int. J. Comput. Vis.*, vol. 111, no. 3, pp. 345–364, Jul. 2014.

[249] A. J. Ma, P. C. Yuen, and J. Li, "Domain transfer support vector ranking for person re-identification without target camera label information," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 3567–3574.

[250] R. Panda, A. Bhuiyan, V. Murino, and A. K. Roy-Chowdhury, "Unsupervised adaptive re-identification in open world dynamic camera networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1377–1386.

[251] J. Lv, W. Chen, Q. Li, and C. Yang, "Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7948–7956.

[252] Y. Huang, Q. Wu, J. Xu, and Y. Zhong, "SBSGAN: Suppression of inter-domain background shift for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 9527–9536.

[253] Y.-J. Li, C.-S. Lin, Y.-B. Lin, and Y.-C.-F. Wang, "Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 7919–7929.

[254] L. Qi, L. Wang, J. Huo, L. Zhou, Y. Shi, and Y. Gao, "A novel unsupervised camera-aware domain adaptation framework for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8080–8089.

[255] J. Wu, S. Liao, Z. Lei, X. Wang, Y. Yang, and S. Z. Li, "Clustering and dynamic sampling based unsupervised domain adaptation for person re-identification," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, Jul. 2019, pp. 886–891.

[256] J. Liu, Z.-J. Zha, D. Chen, R. Hong, and M. Wang, "Adaptive transfer network for cross-domain person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 7195–7204.

[257] Q. Yang, H.-X. Yu, A. Wu, and W.-S. Zheng, "Patch-based discriminative feature learning for unsupervised person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 3628–3637.

[258] P. Peng, Y. Tian, T. Xiang, Y. Wang, M. Pontil, and T. Huang, "Joint semantic and latent attribute modelling for cross-class transfer learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1625–1638, Jul. 2018.

[259] L. Qi, J. Huo, X. Fan, Y. Shi, and Y. Gao, "Unsupervised joint subspace and dictionary learning for enhanced cross-domain person re-identification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1263–1275, Dec. 2018.

[260] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.

[261] N. Sugianto, D. Tjondronegoro, G. Sorwar, P. Chakraborty, and E. I. Yuwono, "Continuous learning without forgetting for person re-identification," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Taipei, Taiwan, Sep. 2019, pp. 1–8.

[262] J. Song, Y. Yang, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 719–728.

[263] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 144–151.

[264] A. Bhuiyan, A. Perina, and V. Murino, "Person re-identification by discriminatively selecting parts and features," in *Proc. ECCV*, Zürich, Switzerland, 2014, pp. 147–161.

[265] S. B. P. Carr, "Deep spatial pyramid for person re-identification," in *Proc. AVSS*, Lecce, Italy, 2017, pp. 1–6.

[266] E. Ustinova, Y. Ganin, and V. Lempitsky, "Multi-region bilinear convolutional neural networks for person re-identification," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Lecce, Italy, Aug. 2017, pp. 1–6.

[267] Y. Fu and Y. Wei, "Horizontal pyramid matching for person re-identification," presented at the AAAI, Honolulu, HI, USA, Jan. 2019.

[268] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proc. ACM Multimedia Conf. Multimedia Conf. (MM)*, Seoul, Republic Korea, 2018, pp. 274–282.

[269] D. S. Cheng and M. Cristani, "Person re-identification by articulated appearance matching," in *Person Re-Identification*. London, U.K.: Springer, 2014, pp. 139–160.

[270] C. Wang and Q. Zhang, "Mancs: A multi-task attentional network with curriculum sampling for person re-identification," in *Proc. ECCV*, Munich, Germany, 2018, pp. 365–381.

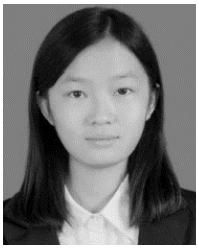
[271] B. Chen, W. Deng, and J. Hu, "Mixed high-order attention network for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 371–381.

[272] S. Khamis and C. H. Kuo, "Joint learning for attribute-consistent person re-identification," in *Proc. ECCV*, Zürich, Switzerland, 2014, pp. 134–146.

[273] C. Su, "Deep attributes driven multi-camera person re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 475–491.

[274] X. Chang, T. M. Hospedales, and T. Xiang, "Multi-level factorisation net for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2109–2118.
[275] Y. Zhao, X. Shen, Z. Jin, H. Lu, and X.-S. Hua, "Attribute-driven feature disentangling and temporal aggregation for video person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 4908–4917.
[276] C.-P. Tay, S. Roy, and K.-H. Yap, "AANet: Attribute attention network for person re-identifications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 7127–7136.
[277] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao, "Multi-task learning with low rank attribute embedding for multi-camera person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1167–1181, May 2018.
[278] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C. Yan, and Y. Yang, "Improving person re-identification by attribute and identity learning," *Pattern Recognit.*, vol. 95, pp. 151–161, Nov. 2019.
[279] T. Matsukawa and E. Suzuki, "Person re-identification using cnn features learned from combination of attributes," in *Proc. ICPR*, Cancún, Mexico, 2016, pp. 2428–2433.
[280] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, "How far are we from solving pedestrian detection?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1259–1267.
[281] D. Sáez Trigueros, L. Meng, and M. Hartnett, "Face recognition: From traditional to deep learning methods," 2018, *arXiv:1811.00116*. [Online]. Available: http://arxiv.org/abs/1811.00116
[282] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille, "Detect what you can: Detecting and representing objects using holistic models and body parts," 2014, *arXiv:1406.2031*. [Online]. Available: http://arxiv.org/abs/1406.2031
[283] S. Ojha and S. Sakhare, "Image processing techniques for object tracking in video surveillance—A survey," in *Proc. ICPC*, Pune, India, 2015, pp. 1–6.
[284] C. J. Dhamsania and T. V. Ratanpara, "A survey on human action recognition from videos," in *Proc. IC-GET*, Coimbatore, India, 2016, pp. 1–5.
[285] Y. Shen, T. Xiao, H. Li, S. Yi, and X. Wang, "End-to-end deep kronecker-product matching for person re-identification," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 6886–6895.
[286] Y. Suh and J. Wang, "Part-aligned bilinear representations for person re-identification," in *Proc. ECCV*, Munich, Germany, 2018, pp. 402–419.
[287] Z. Zhang, C. Lan, W. Zeng, and Z. Chen, "Densely semantically aligned person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 667–676.
[288] T.-Y. Hu and A. G. Hauptmann, "Multi-shot person re-identification through set distance with visual distributional representation," in *Proc. Int. Conf. Multimedia Retr. (ICMR)*, Ottawa, ON, Canada, 2019, pp. 262–270.
[289] Z. Wu, Y. Li, and R. J. Radke, "Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1095–1108, May 2015.
[290] F. Zheng, C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang, and R. Ji, "Pyramidal person re-IDentification via multi-loss dynamic training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 8506–8514.
[291] C. Song, Y. Huang, W. Ouyang, and L. Wang, "Mask-guided contrastive attention model for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 1179–1188.
[292] M. Tian, S. Yi, H. Li, S. Li, X. Zhang, J. Shi, J. Yan, and X. Wang, "Eliminating background-bias for robust person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 5794–5803.
[293] Y. Tang, X. Yang, N. Wang, X. Jiang, B. Song, and X. Gao, "Person re-identification with gradual background suppression," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, Jul. 2019, pp. 706–711.
[294] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Re-identification in the function space of feature warps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1656–1669, Aug. 2015.
[295] T. Yu and H. Jin, "SKEPRID: Pose and illumination change-resistant skeleton-based person re-identification," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 4, Oct. 2018, Art. no. 82.
[296] Y. Huang, Z.-J. Zha, X. Fu, and W. Zhang, "Illumination-invariant person re-identification," in *Proc. ACM MM*, 2019, pp. 365–373.
[297] H. Huang, D. Li, Z. Zhang, X. Chen, and K. Huang, "Adversarially occluded samples for person re-identification," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 5098–5107.
[298] J. Zhuo, Z. Chen, J. Lai, and G. Wang, "Occluded person re-identification," in *Proc. ICME*, San Diego, CA, USA, 2018, pp. 1–6.
[299] L. He and Y. Wang, "Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification," in *Proc. ICCV*, Seoul, South Korea, Oct. 2019, pp. 8450–8459.
[300] Y. Xu, L. Lin, W. Zheng, and X. Liu, "Human re-identification by matching compositional template with cluster sampling," in *Proc. ICCV*, Sydney, NSW, Australia, 2013, pp. 3152–3159.
[301] A. Barman and S. K. Shah, "SHaPE: A novel graph theoretic algorithm for making consensus-based decisions in person re-identification systems," in *Proc. ICCV*, Venice, Italy, 2017, pp. 1124–1133.
[302] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Person re-identification by unsupervised ℓ_1 graph learning," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 178–195.
[303] M. Ye, A. J. Ma, L. Zheng, J. Li, and P. C. Yuen, "Dynamic label graph matching for unsupervised video re-identification," in *Proc. ICCV*, Venice, Italy, 2017, pp. 5152–5160.
[304] Y. Yang and L. Wen, "Unsupervised learning of multi-level descriptors for person re-identification," presented at the AAAI, San Francisco, CA, USA, Feb. 2017.
[305] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, "A bottom-up clustering approach to unsupervised person re-identification," presented at the AAAI, Honolulu, HI, USA, Jan. 2019.
[306] H. Yu, A. Wu, and W. Zheng, "Unsupervised person re-identification by deep asymmetric metric embedding," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2018.2886878.
[307] H. Wang, X. Zhu, T. Xiang, and S. Gong, "Towards unsupervised open-set person re-identification," in *Proc. ICIP*, Phoenix, AZ, USA, 2016, pp. 769–773.
[308] Y.-C. Chen, X. Zhu, W.-S. Zheng, and J.-H. Lai, "Person re-identification by camera correlation aware feature augmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 392–408, Feb. 2018.
[309] X. Yang, M. Wang, R. Hong, Q. Tian, and Y. Rui, "Enhancing person re-identification in a self-trained subspace," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 3, pp. 1–23, Jun. 2017.
[310] Z. Liu and J. Wang, "Deep reinforcement active learning for human-in-the-loop person re-identification," in *Proc. ICCV*, Seoul, South Korea, Oct. 2019, pp. 6122–6131.
[311] J. Zhang, N. Wang and, and L. Zhang, "Multi-shot pedestrian re-identification via sequential decision making," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 6781–6789.
[312] Z. Feng, J. Lai, and X. Xie, "Learning view-specific deep networks for person re-identification," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3472–3483, Jul. 2018.
[313] J. Meng, A. Wu, and W.-S. Zheng, "Deep asymmetric video-based person re-identification," *Pattern Recognit.*, vol. 93, pp. 430–441, Sep. 2019.
[314] X. Zhu, X.-Y. Jing, F. Zhang, X. Zhang, X. You, and X. Cui, "Distance learning by mining hard and easy negative samples for person re-identification," *Pattern Recognit.*, vol. 95, pp. 211–222, Nov. 2019.
[315] L. Wu, Y. Wang, X. Li, and J. Gao, "What-and-where to match: Deep spatially multiplicative integration networks for person re-identification," *Pattern Recognit.*, vol. 76, pp. 727–738, Apr. 2018.
[316] Z. Dai, M. Chen, X. Gu, S. Zhu, and P. Tan, "Batch DropBlock network for person re-identification and beyond," in *Proc. ICCV*, Seoul, South Korea, Oct. 2019, pp. 3691–3701.
[317] A. Chakraborty, B. Mandal, and H. K. Galoogahi, "Person re-identification using multiple first-person-views on wearable devices," in *Proc. WACV*, Lake Placid, NY, USA, 2016, pp. 1–8.
[318] Y. Shen, H. Li, T. Xiao, S. Yi, D. Chen, and X. Wang, "Deep group-shuffling random walk for person re-identification," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 2265–2274.
[319] G. Chen and C. Lin, "Self-critical attention learning for person re-identification," in *Proc. ICCV*, Seoul, South Korea, Oct. 2019, pp. 9637–9646.
[320] L. Ren, J. Lu, J. Feng, and J. Zhou, "Uniform and variational deep learning for RGB-D object recognition and person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 4970–4983, Oct. 2019.

HONGBO WANG received the B.S. degree in computer software from Hebei University, China, in 1998, and the Ph.D. degree in computer application technology from the Beijing University of Posts and Telecommunications (BUPT), China, in 2006. He is currently an Associate Professor with the State Key Laboratory of Networking and Switching Technology, BUPT. His main research interests include computer vision, cloud computing, big data, data center networks, and Internet measurement, in which he has published over 60 technical articles in refereed journals and conference proceedings.



HAOMIN DU is currently pursuing the master's degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. Her research interests are computer vision and machine learning, especially in human attribute recognition and reidentification.

YUE ZHAO received the M.S. degree in computer science and technology from the Beijing University of Posts and Telecommunications, in 2019. Her research interests include computer vision and deep learning.

JIMING YAN is currently pursuing the master's degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His research interests are object detection and person reidentification.