Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12716

Research Article Open Access

A Review of Research on Person Re-identification in Surveillance Video

Yunzuo ZHANG[*], Weiqi LIAN

School of Information science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, 050043, China

[*]***Corresponding Author:*** *Yunzuo ZHANG, E-mail: zhangyunzuo888@sina.com*

Abstract

Person re-identification has emerged as a hotspot for computer vision research due to the growing demands of social public safety requirements and the quick development of intelligent surveillance networks. Person re-identification (Re-ID) in video surveillance system can track and identify suspicious people, track and statistically analyze persons. The purpose of person re-identification is to recognize the same person in different cameras. Deep learning-based person re-identification research has produced numerous remarkable outcomes as a result of deep learning's growing popularity. The purpose of this paper is to help researchers better understand where person re-identification research is at the moment and where it is headed. Firstly, this paper arranges the widely used datasets and assessment criteria in person re-identification and reviews the pertinent research on deep learning-based person re-identification techniques conducted in the last several years. Then, the commonly used method techniques are also discussed from four aspects: appearance features, metric learning, local features, and adversarial learning. Finally, future research directions in the field of person re-identification are outlooked.

Keywords: *Person re-identification; Deep learning; Metric learning; Local features; Adversarial learning*

1 Research Background and Significance of Person Re-identification

In recent years, with the acceleration of urbanization and the growth of population size, urban traffic, public safety and other issues have become incrSeasingly prominent. Traditional manual patrols and monitoring means have been difficult to meet the demand, and the public's demand for public safety has been increasing. A significant number of surveillance cameras have been placed in public areas including streets and neighborhoods as a result of the ongoing advancements in surveillance video intelligent analysis technologies. And these cameras cover a wide geographic range, generating a large amount of surveillance video data [1]. Person re-identification (Re-ID) [2] technology research has become an inevitable trend in order to study the behavioral traits and activity patterns of individuals in these surveillance recordings in a timely and efficient manner.

Person re-identification can realize automatic tracking and identification of persons, give full play to the role of surveillance video data, and improve the efficiency of the city in maintaining public safety and traffic order. The purpose of person re-identification is to find the image of a person with a given identity in a large-scale image library, and to determine whether the persons in different viewpoints, different cameras, and different video clips are the same person [3]. This technique is commonly used in applications like crowd behavior analysis, multi-camera target tracking, and target detection.

The key to person re-identification is to learn the discriminative features of individuals to differentiate between images of the same person and those of different individuals. However, in reality, people can appear in several cameras in various locations, and it is difficult to learn person discriminative features due to variations in different camera viewpoints and resolutions, person poses, and ambient lighting. Traditional methods for person re-identification primarily rely on manually extracting discriminative features or learning better similarity measures. However, these methods are prone to errors and are time-consuming, which significantly impacts the accuracy and real-time performance of the person re-identification task.

Automated person re-identification methods have significantly advanced the development of intelligent surveillance [4-5]. With automated person re-identification methods, surveillance systems can more accurately recognize the same person appearing in different cameras, reduce the rate of misrecognition, and improve the overall accuracy of surveillance systems [6]. This helps to detect abnormal behaviors or suspicious people in time and

Received on October 25, 2023; Accepted on November 29, 2023

improve the security of the surveillance system. In addition, automated person re-identification methods can improve the processing speed and real-time performance of the surveillance system, enabling the surveillance system to respond more quickly to unexpected events or security threats [7]. The automated person re-identification method can provide richer data and feature information for the surveillance system, supporting the system to perform deep learning and data analysis, so as to realize intelligent decision-making and prediction [8]. By analyzing data such as person activity patterns and paths, the surveillance system can better understand and predict potential risks and improve the intelligence of the surveillance system.

With the development of deep neural network (DNN) based techniques [9], person re-identification techniques are able to extract significant person features more efficiently. Deep neural networks can learn a high-level abstract representation of the data, which allows it to extract more robust pedestrian features under different lighting, angle and occlusion conditions. With an end-to-end training approach, DNNs can automatically learn feature representations that are best suited for person re-identification tasks [10]. Compared to traditional feature extraction methods, DNN can better perform nonlinear modeling, thus better capturing complex features and structural information in pedestrian images [11]. In addition, with a pre-trained deep neural network model, generalized features learned on large-scale image data can be migrated to the person re-identification task. This type of migration learning can accelerate the training process of the model and improve the performance of the model on small sample data [12].

2 Current Status of Domestic and Overseas Research

The traditional method of person re-identification is to manually extract a single fixed person feature, which is not only error-prone but also consumes a lot of time. The emergence of deep learning has greatly promoted the research and development in the field of person re-identification. Person re-identification technology based on deep learning has been applied in many fields.

Research on feature extraction has progressively moved from conventional manual feature extraction techniques to deep learning-based techniques [13]. In order to be able to extract more robust and detailed features, Sun et al. [14] introduced local feature learning and Tay et al. [15] used an attention mechanism to concentrate on the important details of body parts. By merging the global and local features of individuals, Zhang et al. [16] improved the final feature representation. Because Generative Adversarial Networks (GANs) are good at creating images and learning features, generative adversarial learning has been applied extensively in person Re-ID tasks [17-19]. Several loss functions have been created by some researchers to maximize the ability of network models to learn discriminative features [20-22].

When compared to still photos, video or image sequences provide additional information about human attributes. Therefore, examining these more comprehensive attribute signals could be a more effective approach to address the person re-identification challenge. Important identifying information lost in a single image due to occlusion, changes in field of view, and other causes can be compensated for by the temporal correlation between frames of a person sequence or trajectory. In view of these results, several researchers have begun to efficiently merge more information in video sequences by utilizing complementing spatial and temporal cues [23-25]. Furthermore, by modeling graph relations on person photos, several researchers have employed graph convolutional network-based techniques to learn more robust and discriminative features [26-28].

Three-dimensional shape information is an important cue for understanding person pose and shape. With the exception of the true 3D human anatomy and the spatial interaction between people and interfering objects, the majority of person Re-ID techniques now in use learn person feature representations from photos. The robustness of person Re-ID models can be improved by utilizing information about the 3D shape of the person [29-30].

3. Datasets and Evaluation Metrics for Person Re-ID

This section contains datasets that are frequently used in person Re-ID techniques that are based on deep learning. Furthermore, a brief description of popular evaluation measures for person Re-ID is provided.

3.1. Commonly used datasets

DukeMTMC-VideoReID [31] consists of approximately 4,832 videos from 1,812 identities, using Duke as its acronym. There are 702 identities in the dataset for testing and training, and 408 identities for distraction. Manual annotations are made to its bounding boxes.

The first extensive dataset for video-based re-identification was introduced in 2016 and is called MARS [32]. It comprises over 20,000 video sequences of person trajectories that were obtained from six different cameras, as well as 1,261 identity sequences. Sixty-five of these twelve hundred IDs were utilized for training, while the remaining six hundred and thirty-six were used for testing.

CUHK01 [33] comprises 3,884 manually cropped photos and 971 IDs, with each individual having at least two images captured using two separate cameras. In the CUHK01 dataset, the images from one of the cameras have more different viewpoints and pose variations, while the images from the other camera mainly consist of front and back views.

With 1,816 subjects and 7,264 manually cropped photos, CUHK02 [34] includes ten camera perspectives total—five pairs of views. At least two pictures in two separate discontinuous camera viewpoints have been taken of each participant. This dataset has more person identities and camera views than CUHK01, and more person picture attributes are available for retrieval.

CUHK03 [13] is a large Re-ID dataset. Ten cameras provide data for CUHK03, which detects person bounding boxes using hand labeling and a deformable part model (DPM) detector. It has 13,164 photos total, each of varied sizes, and 1,360 distinct people. CUHK03 captures more photographs and has more cameras in order to record people from a variety of angles.

Table 1 Commonly used datasets for person re-identification

Datasets	Data type	ID	Boxes/ Tracks	Cameras	Labeled
DukeMTMC-VideoReID	Video	1,812	4,832	8	DPM
MARS	Video	1,261	20,715	6	DPM+GMMCP
CUHK01	Image	971	3,884	2	Handcrafted
CUHK02	Image	1,816	7,264	10	Handcrafted
CUHK03	Image	1,360	13,164	10	DPM+ Handcrafted
Market-1501	Image	1,501	32,217	6	DPM+ Handcrafted
MSMT17	Image	4,101	126,441	15	Faster RCNN

Five high-resolution cameras and one low-resolution camera were used to collect Market-1501 [35]. Market-1501 automatically detects the human bounding box using the person detector DPM. It comprises 32,668 photos total, each measuring 128 by 64, and it features 1,501 distinct people. Market-1501 features more annotated photos and an interference factor, making the graphics more realistic than CUHK03.

MSMT17 [36] from which 15 cameras provided the images. For automatic detection, it makes use of the person detector Faster R-CNN labeled frames. This is one of the larger person and annotated image datasets for the current person Re-ID endeavor, with 4,101 distinct person details and 126,441 photos. The MSMT17 dataset has a larger range of perspectives, considerable illumination differences, and a greater coverage of scenes.

Table 1 shows the details of the above datasets. Many methods are challenging to apply in real-world applications due to uncertainty caused by occlusion, lighting changes, camera view switching, position changes, and related clothing. In person re-identification research, it is imperative to investigate large-scale person Re-ID datasets covering more real-world events.

3.2. Commonly used evaluation metrics

The commonly used evaluation indexes for evaluating the algorithm model in the person re-identification technology are CMC [37] curve and mAP [38].

CMC curve is a very important and commonly used evaluation index in pattern recognition, which can comprehensively reflect the performance of person Re-ID model. CMC uses the top-k approach, which involves sorting the results based on how similar the query target and the target image are to one other. Top-k is the likelihood that, upon sorting, the first k photos returned have the right query result. To arrive at the final result, the hit probability of each query image that was previously collected is added up and divided by the total number of query images.

The accuracy and recall of each category are combined by mAP. In a target detection task, objects are usually categorized according to categories, and for each category, its AP [39] value can be calculated. The mAP value is then calculated by averaging these AP values. In general, the detection method performs better when the mAP is higher. For each category, the detection results are sorted from high to low confidence, and then the number of True Positive (TP), False Positive (FP) and False Negative (FN), i.e., the confusion matrix, is determined based on the threshold value. The confusion matrix for binary classification is shown in Table 2. Accuracy and recall are calculated based on the number of TP, FP and FN, and AP values are calculated for different recall rates. Lastly, the mAP value is calculated by averaging the AP values across all categories. The mAP evaluation metric is commonly used in person re-identification to comprehensively assess algorithm performance while accounting for differences between multiple categories.

Table 2 Binary confusion matrix

	Prediction=1	Prediction=0
Reference=1	TP	FN
Reference=0	FP	TN

4 Deep Learning Based Person Re-identification Methods

In this section, deep learning-based person re-identification methods are classified into four categories. As shown in Figure 1, these categories include appearance-based features, metric-based learning, localized feature-based, and adversarial learning-based methods. In addition, the above four methods are introduced and their respective advantages and disadvantages are discussed.

4.1 Methods based on appearance feature

This type of method mainly performs re-identification by extracting appearance features of persons, such as color, texture and shape. These are shallow visual characteristics, and convolutional neural networks (CNN), local binary patterns (LBP), and the histogram of oriented gradients (HOG) are frequently used feature extraction techniques. Liao et al. [40] presented a person re-identification technique with

LOMO+XQDA, which improved the RANK-1 accuracy on four datasets.

In addition to this, person re-identification can also be performed based on information such as backpacks, glasses, hair, and so on. And this information belongs to the mid-level visual features, i.e., semantic attributes. The semantic attributes of the same person change very little under different camera shots. Therefore, these semantic attributes can be combined with the attribute weights and shallow features to finally describe the person image.

Feature extraction techniques will directly affect the accuracy of identification, and establishing high-level visual features is a challenge in appearance feature extraction. Matsukawa et al. [41] proposed GOG, which divides an image into horizontal strips and local blocks. Each strip is modeled with a Gaussian distribution and establishes high-level visual features by extracting image regions with predefined block or stripe shapes in color or texture histograms.

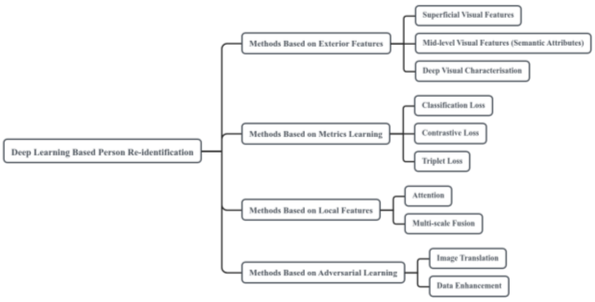

Figure 1 Classification of Deep Learning Based Person Re-identification Methods

4.2 Methods based on metric learning

This type of method mainly measures the similarity between different persons by learning a suitable distance metric. Person re-identification is a common use of deep metric learning (DML).

Among these, marginal distance, cosine similarity, and Euclidean distance are the most often utilized metric learning techniques. While the loss function of the model is typically calculated to constrain the learning of discriminative features in deep metric learning [42]. Commonly used loss functions include classification loss [43-45], contrastive loss [46-47] and triplet loss [48-49]. Because classification loss has the benefits of straightforward training and mining hard samples, it is commonly utilized in person re-identification algorithms based on metric learning. Models that are not sufficiently capable of generalization are not learned by using ID information alone. Thus, in order to limit the model's training, classification loss typically has to be paired with other losses. The contrastive loss is a good representation of sample pair matching.In models for feature extraction, the contrastive loss is frequently employed in conjunction with the classification loss to train the network. Deep learning based Re-ID techniques are frequently used in conjunction with triplet loss and classification loss to enhance model performance. The goal of these techniques is to reduce the intra-class distance and maximize the inter-class distance of the samples.

In addition to this, some studies propose new ways of computing the loss function to improve the modeling approach [50]. Du et al. [51] proposed an adaptive heterogeneous centroid loss, which reduces the distance between samples corresponding to the same identity while increasing the centroid distance for different samples. By eliminating the need for expensive manual feature design, the deep metric learning approach enables the model to automatically learn for feature recognition.

4.3 Methods based on local features

Person re-identification techniques can be classified as global or local based on the features that are taken from person photos for categorization.

Global feature-based methods usually extract only one feature of the persons [52], which makes the feature information of the persons we can get is limited. In order to obtain more detailed, nuanced, and discriminative features, local feature-based methods have gained widespread attention. Learning people's local features and maintaining their alignment are the goals of local feature-based learning approaches. These techniques concentrate on an individual's head, upper body, or lower body, among other local aspects. Person re-identification models can perform better by extracting and matching local features.

Paying attention to every element could be very expensive. An attention method can be used to solve this issue by concentrating just on the local feature regions that have a significant influence. In order to better cluster people's temporally dispersed features and remove contaminated features, Wang et al. [53] introduced a temporal attention mechanism and proposed a new temporal attention module that adaptively evaluates the importance of each extracted feature and combines all the valuable features together. While large-scale feature maps can more accurately depict the image's semantic content and contain the contour information of the image, small-scale feature maps contain more detailed information. The extracted multi-scale person features are combined using a multi-scale fusion technique to get richer person features. In order to overcome the challenges associated with cross-scale feature learning, Chen et al. [54] suggested a deep pyramid feature learning approach for learning multi-scale complimentary features. An all-scales network (OSNet) design was proposed by Zhou et al. [55] in order to learn features. The design encapsulates the synergistic combination of various sizes in addition to capturing aspects at distinct spatial scales.

4.4 Methods based on adversarial learning

The GAN generative adversarial network consists of a generator and a discriminator. The generator is responsible for generating pseudo-labels close to the real situation, while the discriminator is responsible for discriminating the pseudo-labels generated by the generator. The two fight against each other, constantly updating and iterating so that the results generated by the generator are closer and closer to the real value, and the discriminator is able to discriminate and distinguish the results that are closer to the real situation.

There is a domain gap between different datasets,i.e., there is a degradation of model performance when the same model is trained and tested on different datasets. In order to tackle this issue, Ge et al. [56] utilized an image-to-image translation network trained to translate images in two directions using the appropriate generators, utilizing the popular CycleGAN architecture. Additionally, by diversifying the training data, the model's capacity for generalization can be strengthened.An Adaptive Memory with Group Labeling (AdaMG) framework for Unsupervised Person Re-ID was proposed by Peng et al. [57]. It uses adaptive memory to build a multi-branch structure that resists noisy labels and takes use of data diversity. Dual et al. [58] proposed a framework which includes two phases of offline clustering, which refines the pseudo-labels, and online training, which optimizes the features. For person re-identification in UDA, Li et al. [59] suggested a triple adversarial learning and multi-perspective creative inference network that improves the discriminability and robustness of the learnt features while discriminating domain-invariant features.

5 Research Outlook

This paper provides an overview and summary of the literature on deep learning-based person re-identification technologies. Firstly, typical datasets and assessment indices for person re-identification technologies are presented, followed by a summary of common research directions in person re-identification both domestically and internationally. Secondly, four perspectives—appearance-based features, metric-based learning, localized features, and adversarial learning—are explored regarding the popular deep learning-based person re-identification techniques.

Some current person re-identification datasets suffer from category imbalance and inaccurate labeling, which may lead to degradation of model performance and generalization ability. Secondly, person re-identification techniques need to cope with the great variation of pedestrian appearance in different scenarios. The appearance of pedestrian images may vary greatly due to a variety of factors such as the angle of the surveillance camera, lighting conditions, pedestrian posture and occlusion. The robustness and generalization ability of current person re-identification techniques still need to be improved. In addition, the computational efficiency and storage space issues when dealing with large-scale data are also challenges for person re-identification techniques. In practical applications, surveillance systems often generate a large amount of pedestrian images and video data, which require efficient algorithms and sufficient storage space to support if they are to be comprehensively analyzed and processed. In real-world scenarios, there are often multiple pedestrian targets that need to be detected and recognized at the same time, which requires algorithms that can accurately locate and recognize multiple pedestrian targets in complex backgrounds.

In summary, although the existing research on deep learning-based person re-identification techniques has achieved very significant results. However, there are still many aspects that need attention and improvement in the future.

(1) Cross-modal person re-identification: Data in the real world may come from different types of cameras. For example, at night, ordinary cameras cannot clearly capture the features of persons, so infrared images from infrared cameras are needed to capture the features. When matching images from different modalities, reducing the differences in feature distribution in different modalities is a future task to be considered.

(2) 3D person re-identification: 3D shape information provides insight into richer person pose and shape information. The majority of current algorithms ignore the true 3D human anatomy and the spatial relationship between people and interferences in favor of just learning person features from photos. By obtaining 3D point cloud data of an image and processing it, data enhancement is performed by aligning 2D and 3D feature information.

(3) Domain adaptive person Re-ID: It becomes more challenging to effectively apply a Re-ID model that was trained on one dataset to another due to domain variations between datasets. While optimizing with pseudo-labels, a variety of unsupervised domain-adaptive techniques tend to transfer the learned knowledge from one domain to another, they are not without limits. In particular, these algorithms introduce a high number of noisy labels by one-time clustering, which hinders the retraining process and limits generalization, and they always generate a pseudo-label for each unlabeled sample, which makes it challenging to accurately describe an individual.

(4) Semi-supervised or unsupervised person re-identification: Usually the size of the dataset used by the model is very large and the overhead of annotating the data is very high. Therefore semi-supervised or unsupervised methods become a focus for future attention. They can extract features from unlabeled or partially labeled datasets. Therefore, in order to enhance the performance of semi-supervised or unsupervised algorithms in the future, more suitable clustering or label assignment procedures must be discovered.

Acknowledgements: This work is funded by the Science and Technology Project of Hebei Education Department (No. ZD2022100).

References

[1] Wang Xiaogang. Intelligent multi-camera video surveillance:

A review [J]. Pattern Recognition Letters, 2013,34:3-19.
[2] Ning Xin, Gong Ke, Li Weijun, et al.. Feature refinement and filter network for person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,31:3391-3402.
[3] Luo Hao, Jiang Wei, Fan Xing, et al.. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019,45:2032-2049.
[4] Bharath Ramesh, Shihao Zhang, Hong Yang,et al.. e-TLD: Event-based framework for dynamic object tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,31:3996-4006.
[5] Fan Heng, Ling Haibin, Siamese cascaded region proposal networks for real-time visual tracking[C]. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019.
[6] M. Saquib Sarfraz, Arne Schumann, Andreas Eberle,et al.. A pose-Sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
[7] Félix Remigereau;Djebril Mekhazni;Sajjad Abdoli, et al.. Knowledge distillation for multi-target domain adaptation in real-time person re-identification[C]. IEEE International Conference on Image Processing (ICIP), 2022.
[8] Wu Fei ,GaoYang ,Liu Jing ,et al.. Multi-stage feature interaction network for masked visible-thermal person re-identification[C]. 2023 35th Chinese Control and Decision Conference (CCDC), 2023.
[9] Pan Honghu ,Chen Yongyong ,He Zhenyu,et al..TCDesc: Learning topology consistent descriptors for image matching[J]. IEEE Trans. Circuits Syst. Video Technol, 2021,32:2845-2855.
[10] Yang Xi , Liu Liangchen, Wang Nannan , Gao Xinbo. A two-Stream dynamic pyramid representation model for video-based person re-identification[J]. IEEE Transactions on Image Processing, 2021,30:6266-6276.
[11] G. Han, M. Lin, Z. Li, H. Zhao and S. Kwong. Text-to-Image Person Re-identification Based on Multimodal Graph Convolutional Network[J]. IEEE Transactions on Multimedia, 2023,47:145-165.
[12] Zhang Xinyu,Cao Jiewei , Shen Chunhua ,et al.. Self-Training With Progressive Augmentation for Unsupervised Cross-Domain Person Re-Identification[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
[13] Wei Li,Rui Zhao,Tong Xiao,Xiaogang Wang. Deepreid: deep filter pairing neural network for person re-identification[C], IEEE International Conference on Computer Vision, 2014.
[14] Sun Tifan, Zheng Liang, Yang Yi,et al.. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]. European Conference on Computer Vision, 2018.
[15] C.P. Tay, S. Roy, K.H. Yap. Aanet: attribute attention network for person re-identifications[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[16] Zhang Hongwei ,Zhang Guoqing,Chen Yuhao,et al.. Global relation-Aware contrast learning for unsupervised person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:8599-8610.
[17] C. Wei, H. Wang, W. Shen,et al.. CO2: Consistent contrast for unsupervised visual representation learning[J]. 2020.
[18] Zhang Demao,Zhang Zhizhong, Ju Ying,et al.. Dual mutual learning for cross-modality person re-identification[J]. IEEE Trans. Circuits Syst Video Technol, 2022(47):550-560.
[19] E. Ristani, F. Solera, R. Zou, et al.. Performance measures and a data set for multi-target, multi-camera tracking[C]. Eur. Conf. Comput. Vis, 2016.
[20] Zheng Meng, S. Karanam, Wu Ziyan, et al.. Re-identification with consistent attentive siamese networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[21] Chai Tianrui, Chen Zhiyuan, Li Annan, et al.. Video person re-identification using attribute-enhanced features[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:7951-7966.
[22] Chen Weihua, Chen Xiaotang, Zhang Jiangguo,et al.. Beyond triplet loss: a deep quadruplet network for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[23] Yang Jinrui, Zheng Wei-Shi, Yang Qize, et al.. Spatial-temporal graph convolutional network for video-based person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.
[24] X. Liao, L. He, Z. Yang,et al.. Video-based person re-identification via 3d convolutional networks and non-local attention[C]. Asian Conference on Computer Vision, 2019.
[25] Yan Yichao, Qin Jie, Chen Jiaxin, et al.. Learning multi-granular hypergraphs for video-based person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.
[26] Bai Zechen, Wang Zhigang, Wang Jian, et al.. Unsupervised multi-source domain adaptation for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2021.
[27] Wu Yiming, O. Bourahla, Li Xi, et al.. Adaptive graph representation learning for video person re-identification[J]. IEEE Trans. Image Process, 2020,29:8821-8830.
[28] Yan Yichao, Zhang Qiang, Ni Bingbing,et al.. Learning context graph for person search[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[29] Huang Yewen, Lian Sicheng, Zhang Suian, et al..Three-dimension transmissible attention network for person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,30:4540-4553.
[30] Wang Changshuo, Ning Xin, Li Weijun,et al.. 3D Person Re-identification Based on Global Semantic Guidance and Local Feature Aggregation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023(55):102-113.
[31] Wu Yu, Lin Yutian , Dong Xuanyi ,et al.. Exploit the unknown gradually:one-shot video-based person re-identification by stepwise learning[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[32] Zheng Liang, Bie Zhi, Sun Yifan,et al.. Mars: a video benchmark for large-scale person re-identification[C]. European Conference on Computer Vision, 2016.

[33] Li Wei, Wang Xiaogang. Locally aligned feature transforms across views[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013.
[34] Li Wei, Zhao Rui, Wang Xiaogang. Human re-identification with transferred metric learning[C]. Asian conference on Computer Vision, 2012.
[35] Zheng Liang, Shen Liyue, Tian Lu, et al..Scalable person reidentification: a benchmark[C], IEEE International Conference on Computer Vision, 2015.
[36] Wei Longhui, Zhang Shiliang, Gao Wen, Tian Qi. Person transfer gan to bridge domain gap for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[37] Gray D, Tao H, Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]. European Conference on Computer Vision(ECCV), 2008.
[38] L. Zheng, L. Shen, L. Tian, et al.. Scalable person re-identification: a benchmark[C]. IEEE International Conference on Computer Vision (ICCV) 2015.
[39] Kunho Kim,Min-Jae Kim,Hyungtae Kim,et al..Person re-identification method using text description through CLIP[C]. 2023 International Conference on Electronics, Information, and Communication (ICEIC), 2023.
[40] Liao Shengcai, Hu Yang, Zhu Xiangyu,et al.. Person re-identification by local maximal occurrence representation and metric learning[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[41] MATSUKAWA T, OKABE T, SUZUKI E, et al.. Hierarchical gaussian descriptor for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[42] Ye Mang, Shen Jianbing, Lin Gaojie,et al..Deep learning for person reidentification: a survey and outlook[J]. IEEE Trans. Pattern Anal. Mach. Intell, 2021(37):99-111.
[43] Zhong Zhun, Zheng Liang, Luo Zhiming, et al.. Invariance matters: exemplar memory for domain adaptive person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[44] Zhong Zhun, Liang Zheng, Zheng Zhedong,et al.. Camera style adaptation for person reidentification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[45] Huang Houjing, Li Dangwei, Zhang Zhang,et al.. Adversarially occluded samples for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[46] Wang Yicheng, Chen Zhenzhong, Wu Fen, et al.. Person re-identification with cascaded pairwise convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[47] M. Wang, B. Lai, and J. Huang, Camera-aware proxies for unsupervised person re-identification[C]. AAAI Conf. Artif. Intell. 2021.
[48] E. Ristani, C. Tomasi. Features for multi-target multi-camera tracking and re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[49] Song Chunfeng, Huang Yan, Ouyang Wanli,et al.. Mask-guided contrastive attention model for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[50] Zhang Peng, Xu Jingsong, Wu Qiang,et al..Top-Push constrained modality-adaptive dictionary learning for cross-modality person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,30:4554-4566.
[51] Du Guodong, Zhang Liyan, Enhanced invariant feature joint learning via modality-invariant neighbor relations for cross-modality person re-Identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023(28):66-75.
[52] Wang Faqiang, W. Zuo, L. Lin,et al.. Joint learning of single-image and crossimage representations for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[53] Wang Zhikang,He Lihuo,Tu Xiaoguang,et al.. Robust video-based person re-identification by hierarchical mining[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:8179-8191.
[54] Chen Yanbei, Zhu Xiatian, Gong Shaogang. Person re-identification by deep learning multi-scale representations[C]. IEEE International Conference on Computer Vision Workshops, 2017.
[55] Zhou Kaiyang,Yang Yongxin,Andrea Cavallaro.Learning generalisable omni-scale representations for person re-identification[J]. IEEE Trans. Pattern Anal. Mach. Intell, 2021(26):66-71.
[56] Ge Yixiao, Zhu Feng, Chen Dapeng,et al.. Structured domain adaptation with online relation regularization for unsupervised person Re-ID[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024,35:258-271.
[57] Peng Jinjia, Jiang Guangqi, Wang Huibing. Adaptive memorization with group labels for unsupervised person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023,33:5802-5813.
[58] Dai Yongxing, Liu Jun, Bai Yan,et al.. Dual-refinement: Joint label and feature refinement for unsupervised domain adaptive person re-identification[J]. IEEE Trans. Image Process, 2021,30:7815-7829.
[59] Li Huafeng, Dong Neng, Yu Zhengtao, et al.. Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:2814-2830.