Pattern Recognition 142 (2023) 109669

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Person re-identification: A retrospective on domain specific open challenges and future trends

Asmat Zahra[a], Nazia Perwaiz[a], Muhammad Shahzad[a,b], Muhammad Moazam Fraz[a,1,*]

[a] National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[b] Technical University of Munich, Munich D-80333, Germany

ARTICLE INFO

Article history:
Received 7 December 2022
Revised 12 April 2023
Accepted 30 April 2023
Available online 13 May 2023

Keywords:
Person re-Identification
Literature survey
Deep learning
Open challenges
Specific application-driven

ABSTRACT

Person Re-Identification (Re-ID) is a critical aspect of visual surveillance systems, which aims to automatically recognize and locate individuals across a multi-camera network with non-overlapping fields-of-view. Despite significant progress in recent years through the use of deep learning-based approaches, there remain many vision-related challenges, such as occlusion, pose, background clutter, misalignment, scale, viewpoint, low resolution & illumination, and cross-domain generalization across camera modalities, that hinder the accurate identification of individuals. The majority of the proposed approaches directly or indirectly aim to solve one or multiple of these existing challenges. To further advance the development of Re-ID solutions, a comprehensive review of the current approaches is necessary. However, no focused review currently exists that analyses and highlights specific aspects for further development. To fill this gap, we present a systematic challenge-specific literature survey of about 300 papers published between 2015 and 2022, which reviews Re-ID approaches from a solution-oriented perspective. This survey is the first of its kind to provide an in-depth analysis of the different approaches used to address the various challenges in Re-ID. Furthermore, our review highlights several prominent and diverse research trends in the Re-ID domain. These trends offer a visionary perspective regarding ongoing person Re-ID research, and they may eventually lead to the development of practical real-world solutions. We highlighted the AI ethics that must be followed while developing a Re-ID solution, and recently being practiced as well. Another exciting future dimension of person Re-ID research is the long-term Re-ID, which is still under evolution. Overall, our survey aims to serve as a valuable resource for researchers and practitioners working in the field of Re-ID and to inspire the development of innovative and effective Re-ID solutions.

1. Introduction

In recent years, person Re-ID has received much interest owing to its widespread application prospects in numerous fields including intelligent video surveillance [1], robotics [2] and human-computer interaction [3] *etc*. Specifically, it is one of the fundamental components of an automated visual surveillance system where for public safety and security in a smart environment, an individual person may be automatically identified and tracked in videos (or images) acquired through multiple non-overlapping cameras installed on public places like airports, banks, cantonments, parks, streets, educational institutes *etc*. Since it is simply not feasible to rely on manual human intervention to identify a person of interest in a huge amount of video data collected on a daily basis, therefore a plethora of approaches have been proposed by vision researchers that aim to automate this highly challenging problem.

Methodologically, person Re-ID refers to identifying and tracking a person in multi-network non-overlapping cameras installed in indoor and outdoor environments. Given an image of a person captured from one camera, the task of person Re-ID is to identify this person from a pre-stored gallery set captured by other multiple cameras.

Despite the growing trend in the number of publications appearing in top venues achieving increasingly higher accuracy on the existing benchmark datasets, the problem is still far from being solved to be translated into real-world settings. This can be attributed to the number of challenges (*e.g.* occlusion, variations in person pose, viewpoint variations, misalignment, poor resolution *etc*.) that makes the problem extremely hard and needs to be resolved to bridge the performance gap between research (benchmark specific) and real-world environments.

* Corresponding author.
E-mail address: moazam.fraz@seecs.edu.pk (M.M. Fraz).
[1] [orcid=0000-0003-0495-463X]

https://doi.org/10.1016/j.patcog.2023.109669
0031-3203/

Fig. 1. Graphical view of enlisted challenges. (From left to right) (a) Occlusion, (b) Illumination variance, (c) Pose variance, (d) Background clutter, (e) Misalignment, (f) Scale difference, (g) Viewpoint variance, and (h) Low Resolution.

1.1. Scope & objective of the review

In this paper, we have targeted the most popular challenges in person Re-ID to perform a systematic challenge-wise review of the published approaches. In this context, we have collected papers from top conferences and journals for the years 2015 to 2022. The progress in papers addressing each challenge and its influence on published results is comprehensively reviewed. The specific challenges in Re-ID that are mainly considered in this review include: occlusion, pose variance, background clutter, misalignment, scale difference, viewpoint variance, low resolution & illumination variance, and cross-domain or generalization. Particularly, the proposed review makes many-fold contributions. For instance, we provide an in-depth analysis of the impact of the most popular Re-ID challenges by discussing the work on each challenge in top computer vision conferences and journals. This provides insights into complexities that arise due to each challenge in the whole Re-ID process. Moreover, based on reviewed progress, the best performing architectures achieving state-of-the-art (SOTA) results on each challenge (as shown in Fig. 1) are highlighted and critically analyzed. Furthermore, we attempt to make future directions for researchers by comprehensively reviewing publications relevant to each challenge and discussing the limitations and benefits to lessen the gap between close-world and real-world implementations. Lastly, in addition to reporting trends and highlighting interesting approaches, we distill our analysis into a few recommendations in the hope of fostering reproducible and efficient research in the field. For the readers from any of these scenarios, this survey also presents comprehensive information on how the challenges have been addressed in past and how various components of deep learning (DL) can be utilized to contribute to improving the person's Re-ID considering the influence of each individual challenge.

1.2. Comparison with previous reviews

There already exist a few review articles on person Re-ID [4–9]. Each one of them focuses on a different aspect of the Re-ID problem. For instance, in [4] the advantages and disadvantages of the traditional and initial DL-based approaches are analyzed. In [5] representation and metric learning-based approaches were discussed. Publically available datasets and their role in deep learning were analyzed in[6]. In [7] six different learning methods including identification, verification, distance metric learning, part-based, video-based, and data augmentation-based deep models are comprehensively reviewed. In [8] focused on application-driven methods designed for specific applications and defined as generalized open-world Re-ID. While in [9] person Re-ID is reviewed using open and closed world settings while keeping challenges of Re-ID perspectives in view. These different perspectives for close-world settings include feature representation, metric learning, and ranking optimization. For open-world settings, heterogeneity, end-to-end, semi or unsupervised learning, robust model learning with noisy data annotations, and open-set person Re-ID are the considered perspectives.

All the aforementioned reviews have comprehensively provided the shortcomings and benefits of considered methods and settings. Moreover, they have also provided insightful future directions. However, none of them have systematically reviewed the influence of popular challenges on performance results and the role of datasets in resolving these challenges. Specifically, no review exists that comprehensively addressed the challenges (mentioned in Section 1) of the person Re-ID and their proposed DL-based solutions. From 2015 to 2022 numerous articles have been published and each of these addressed the person Re-ID challenge in a specific way. Some of the challenges are open to the Re-ID world and still not addressed properly. This motivated us to write this survey article which comprehensively reviewed how all open challenges are addressed in past and how results getting improved with respect to each challenge using DL-based methods. The difference between our survey from the existing surveys can be visualized in Fig. 2.

2. Survey methodology

This survey paper is organized into six sections, Section 1 introduces the domain, discusses the scope, objective, and rationale of this paper, and highlights the main contribution of this review by providing a detailed comparison with recent survey articles on person Re-ID. Section 2 illustrates the data collection methodology used for selecting the articles included in this review. A comparative account of the publicly available datasets and performance metrics used to report results of person Re-ID methodologies is given in Section 3. In Section 4, the open challenges in person Re-ID are discussed and the methodologies proposed to address these challenges are critically analyzed. Section 5 overviews the impact

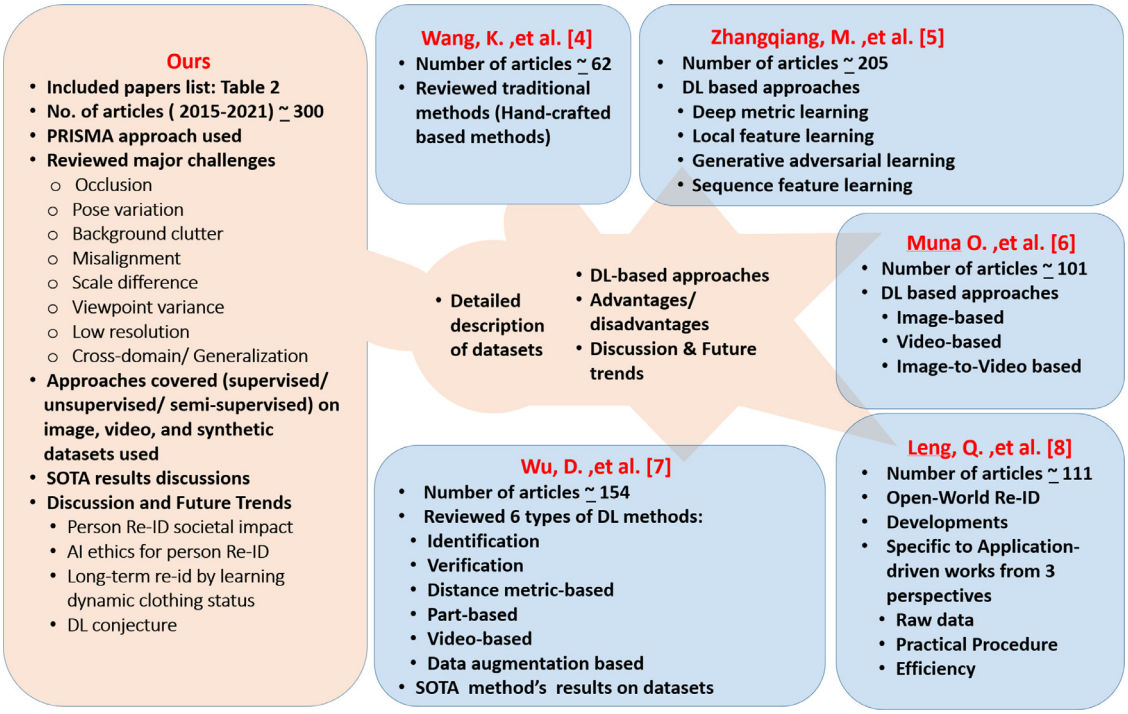

Fig. 2. Comparison with recently published reviews on person Re-ID.

of challenges on results. Role of datasets in resolving the identified challenges along with limitations and benefits. And then the paper concludes with possible future directions.

2.1. Study selection

For study selection, we used PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [10]. PRISMA was initially developed for reporting medical studies but is still useful for writing a computer science domain review paper because it provides a systematic and standardized framework for reporting systematic literature reviews [11–13]. By following the PRISMA guidelines, authors can improve the transparency and rigor of their systematic review process, and ensure that they include all relevant information, such as the search strategy, selection criteria, and quality assessment. This can enhance the credibility of the review and its conclusions, regardless of the specific field or discipline being studied.

Figure 3 shows the summary of the paper selection process. As per PRISMA guidelines, our selection of articles for this review consists of two stages. In the first stage, all the articles that did not meet the eligibility criteria were excluded. While in the second stage we studied full-text reports to find relevant articles. After shortlisting all the articles, we have excluded those papers which have reported results only through graphs and on such performance measures and datasets that were not so widely adopted by the research community. In a complex scenario where paper selection becomes difficult due to some ambiguity, a discussion with senior members is organized to reach a mutual final decision.

2.2. Data extraction methods

We have collected an initial list of articles from Springer, Google Scholar, IEEE Xplore, and Elsevier. We have considered the top seven journals and three conferences held from 2015 to 2022 for review. We have used the following terms (or matching to these) to search relevant articles:

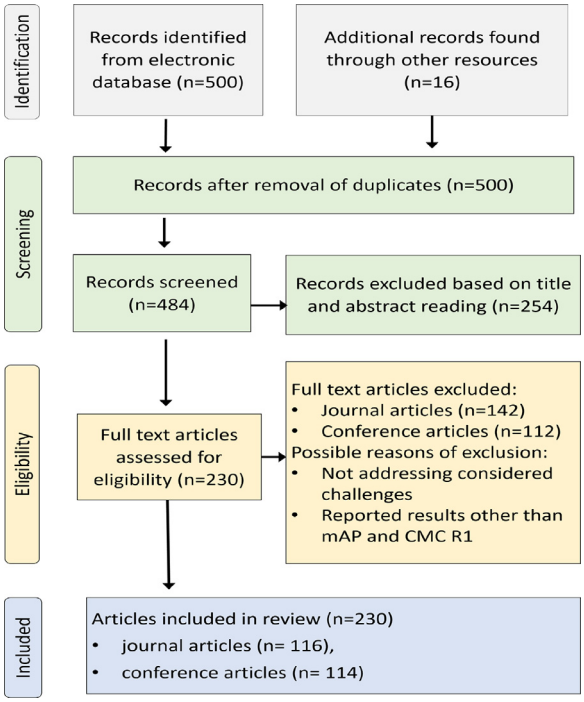

Fig. 3. Summary of paper selection process (PRISMA).

a Person Re-Identification
b Deep learning
c Supervised person Re-ID
d Semi-supervised person Re-ID
e Unsupervised person Re-ID
f Pose variations
g Body misalignment
h Attention-based approach
i Camera View(Viewpoint)

Table 1
Inclusion and exclusion criteria of selected articles.

Inclusion criteria	Exclusion criteria
-Articles that address the identified challenges enlisted in Table 2	-Articles in which qualitative evaluation of results are missing
-Provide a detailed summary of the proposed architecture including training parameters	Papers that do not address the identified challenges enlisted in Table 2
-Articles that are based on deep learning techniques	-Survey papers
-Articles that are published in journals and conferences enlisted in Table 3 between January 2015 to March 2022	-Reported results on metrics other than Rank-1 and CMC
	-Papers that have used datasets that were not so widely adopted by the research community

j End-to-end learning

Search results are further enhanced by combining the mentioned terms using logical operators in a way like ('a') AND ('b' OR 'c' OR 'd' OR 'e' OR 'f' OR 'g' OR 'h' OR 'i' OR 'j'). The papers were then excluded or included based on the criteria listed in Table 1. We first read the title of the article for final selection. In case the title does not clearly fall in our inclusion and exclusion criteria then we also read the abstract and conclusion section of the article as well. After that, we started full reading for data collection. And articles that do not match our criteria were not included in this review.

2.3. Data synthesis

To make our review more useful in the sense that other researchers can contribute to it in the future for the purpose of extension in the review in multiple perspectives, a data extraction sheet was developed that describes the multiple relevant data items to be extracted from the articles. Around 30 data items were used for metadata extraction from each article. These data items were classified into seven categories: Origin of the article, The challenges addressed, Details on the proposed methodology, Implementation detail, Reproducibility and code availability, Performance metrics, and the Datasets used. Table 2 shows the category-wise distribution and description of each of the data items. The results will be stored in a spreadsheet which will be made public for interested researchers intended to extract more information or analyze a different perspective.

The first category extracts the details about the origin of the article whether the article is published in a conference or a journal, the year of publishing, the title of the article, which problem the particular article is addressing, image-based or video-based and finally which methodology is used *i.e.* CNN, Vision Transformer (ViT), GANs or Graph CNNs *etc.* In the second category, the details of enlisted person Re-ID challenge (Occlusion, Pose variation *etc.*) addressed by the article are explored. In the third category, a brief extract of the proposed methodology in the article is analyzed. The fourth category comprehensively extracts the implementation detail including implementation framework, deep learning-based methodology, base model, training approach, and batch normalization. The fifth category tells about the public availability of implementation. The sixth category gathers information about the quantitative performance measures of the proposed work. Finally, the details about the datasets used for evaluation are reported in the last category. The Taxonomy of our review is shown in Fig. 4.

2.4. Results

There were 516 articles collected using the selection criteria. After removing irrelevant articles database contains 500 articles. According to our inclusion and exclusion criteria, more than 254 articles were excluded and therefore we selected 230 articles (116 Journal & 114 Conference articles) for review. Table 3 depicts the publication venue-wise summary of the articles selected for this review, and Fig. 5 illustrates the growing trend of person Re-ID publications each year in reputed venues. Moreover, we also included a few relevant person Re-ID papers from AAAI, IJCAI, NIPS, ICLR, ACM MM, and others, but the venues listed in the Table 3 are the only ones for which all Re-ID papers are included for analysis in this work.

3. Datasets and performance evaluation

Several image-based and video-based datasets for person Re-ID models have been released. In the past few years, several review papers review the available datasets. But they have not reported the progress on each challenge against each dataset. We review several image and video-based datasets by reporting the progress of each challenge on each dataset and how the dataset supports the challenge to be addressed effectively. The attributes of each of the public datasets are summarized in the Table 4.

Few of these datasets have been turned off from the open public access i.e. CUHK01, CUHK03, ViPeR, MSMT17. However, these datasets can be obtained by directly contacting the datasets' authors. It should be noted that the public access to the dataset DukeMTMC-reID was discontinued in 2019 by the original authors. The dataset was reportedly used for commercial purposes with the violation of human rights and individual's privacy, as reported by the Exposing.ai [14] and the Financial Times [15].

3.1. Performance evaluation metrics

For evaluation purposes, person Re-ID models use two widely known measurements named Cumulative Matching Characteristics (CMC) also known as rank-k accuracy, and mean Average Precision (mAP). CMC represents the probability that an accurate match of the image appears in the top-k ranked retrieved results. CMC will be considered accurate when only one ground truth exists for each query since it only considers the first match in the evaluation process. However, the gallery set usually contains multiple ground truths in a large camera network, and CMC cannot completely reflect the discriminability of a model across multiple cameras. The other widely used metric, mAP measures the average retrieval performance with multiple ground truths. It is originally widely used in the image retrieval process.

3.2. Image based datasets

Several image-based person Re-ID datasets are available. We have considered the top 6 image-based person Re-ID datasets *i.e.* Market-1501, DukeMTMC-Re-ID, CUHK03, CUHK01, VIPeR, and MSMT-17 for our review on which person Re-ID models achieved SOTA performance. Major person Re-ID challenges are addressed by using these mentioned datasets. A summary of the number of papers on major challenges is shown in Fig. 6.

3.2.1. Market-1501

The Market-1501 [16] dataset is specifically for person Re-ID and was proposed in 2015. It was collected in front of a supermarket at Tsinghua University. A total of six cameras were used, including five high-resolution cameras, and one low-resolution camera. Overlap exists among different cameras. Overall, this dataset

Table 2
Summary of data-items extracted from each article.

Category	Data Item & Description
Origin of article	Conference & Journal Name
	Publication year & venue
	Article Title
	Image & Video dataset or both
	CNN, ViT, GCN, GAN based approaches used in the paper
Challenges	Occlusion: Blockage or hiding of target person in image
	Pose Variation: Particular person appeared in different positions
	Background clutter: A pattern present in the background resembles the pattern of person's wearing in image
	Body Misalignment: Person not aligned according to viewing angle in image
	Scale: Person appearing in different sizes in an image
	Illumination: Light variations in an image
	Viewpoint variation: Change in position of capturing camera
	Resolution : Clarity & detail in input image
	Cross-domain & Generalization: Images belong to multiple domains
Proposed Methodology	Description of the proposed method in article
Implementation Details	Implementation Framework & Platform used
	Base Model: DL baseline model used as backbone
	Training approach of model
	Take single & multiple images as query
	Data-set used in a pre-trained model
	Batch Normalization Scheme used
	Batch size considered for training
	Type of pooling used in the article
	Learning rate-decay
	Data Augmentation technique used
Reproducible	Code Availability for public use or not
Results	Performance Metric used for reporting the results
Data-sets	Name(s) of data-sets used

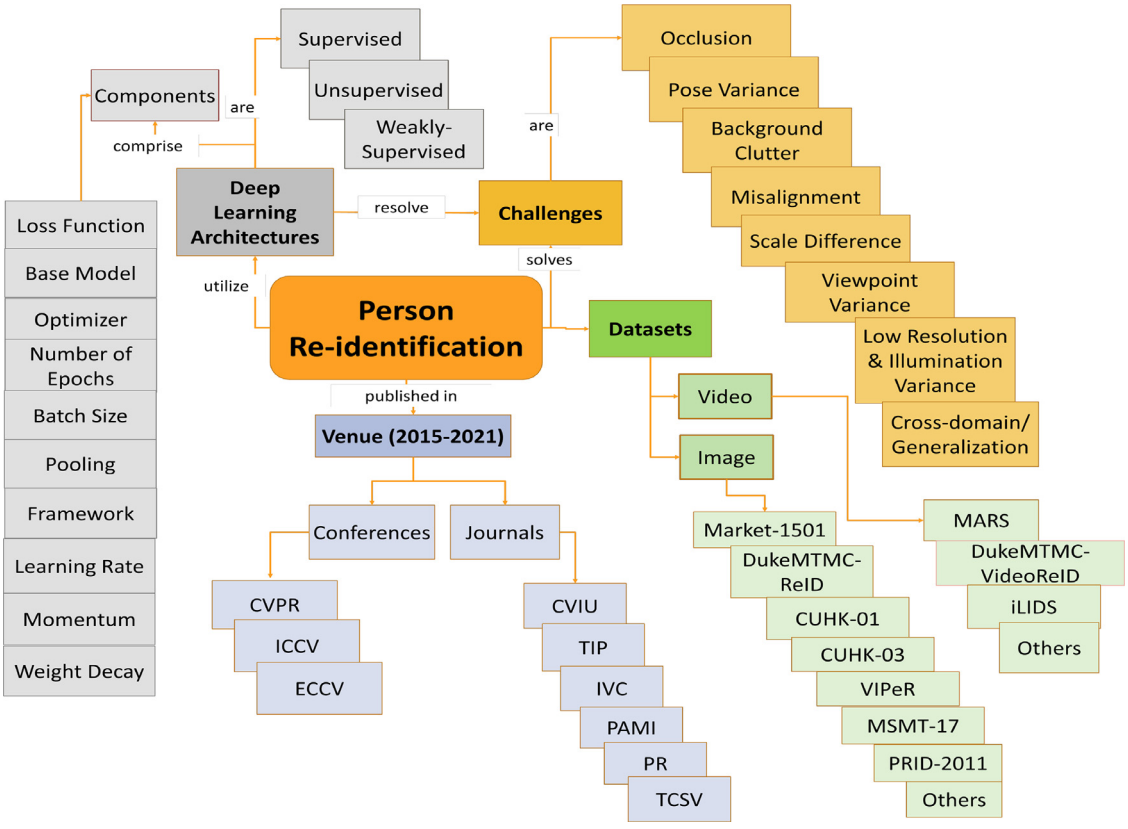

Fig. 4. Taxonomy of the Review.

Table 3
Summary of articles collected.

SN.	Venues (2015–2022)	Found	Filtered	Reviewed
1	Computer Vision and Pattern Recognition (CVPR)	131	72	59
2	IEEE International Conference on Computer Vision (ICCV)	50	18	32
3	European Conference on Computer Vision (ECCV)	45	22	23
4	Elsevier Computer Vision & Image Understanding (CVIU)	6	0	6
5	IEEE Transactions on Image Processing (TIP)	91	46	45
6	Image and Vision Computing (IVC)	8	3	5
7	IEEE transactions on pattern analysis & machine intelligence (PAMI)	42	27	15
8	International Journal of Computer Vision (IJCV)	8	4	4
9	Pattern Recognition (PR)	42	27	15
10	IEEE Transactions on Circuits & Systems for Video Technology (TCSVT)	51	31	20
	Total	484	254	230

Table 4
Properties of person Re-ID datasets.

SN.	Dataset	Year	Environment	IDs	Cams	Resolution	Label	BBoxes	Challenging features
1	VIPeR [21]	2007	Campus	632	2	4828	Hand	1264	VV[1], IV[2]
2	PRID-2011 [22]	2011	Outdoor	200	2	6428	DPM/ GMMCP/ Hand	40,000	IV,VV,BC[3]
3	CUHK01 [19]	2012	Campus	971	2	6060	Hand	3884	VV,OCC[4]
4	CUHK03 [20]	2014	Campus	1360	10	Vary	DPM/ Hand	13,164	VV,OCC
5	iLIDS-Vid [23]	2014	Airport	300	2	Vary	Hand	42,495	VV,IV,BC, OCC
6	Market1501 [16]	2015	Campus	1501	6	6428	DPM/ Hand	32,688	VV, PV[5], RES[6]
7	MARS [24]	2016	Campus	1261	6	25628	DPM/ GMMCP	1,067,516	PV,IV,RES
8	DukeMTMC-ReID [17]	2017	Campus	1404	8	Vary	Doppia/ Hand	36,411	VV,IV,BC, OCC
9	DukeMTMC-Video ReID	2017	Campus	1404	8	Vary	Hand	36,411	VV,IV,BC, OCC
10	MSMT-17 [25]	2018	Campus	4101	15	Vary	Faster RCNN	126,441	VV,IV

[1]Viewpoint Variation. [2]Illumination Variation. [3]Background Clutter. [4]Occlusion. [5]Pose Variation. [6]Resolution. [7]Deformable Part Model [26] (A Pedestrian detector). [8]GMMCP-Generalized Maximum Multi Clique problem [27] (A tracker).

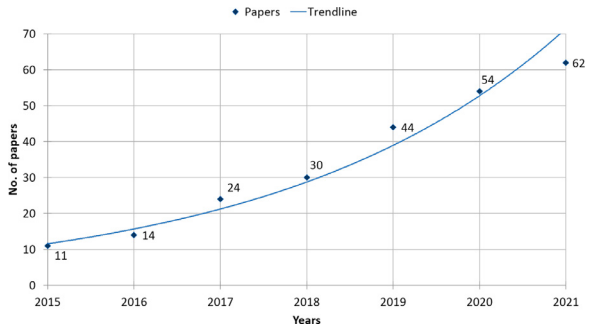

Fig. 5. The yearly frequency of selected Re-ID papers for this review.

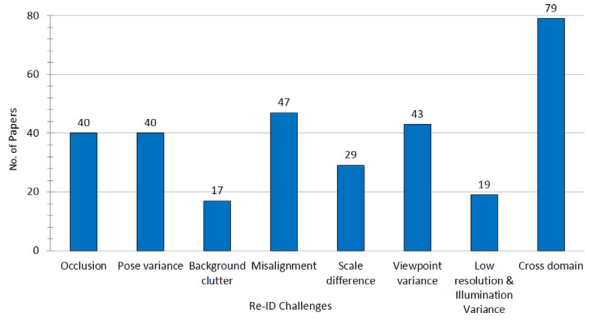

Fig. 6. Count of papers on each enlisted challenge.

contains 32,668 annotated bounding boxes of 1501 identities. Each annotated identity is present in at least two cameras so that a cross-camera search can be performed.

3.2.2. DukeMTMC-Re-ID

DukeMTMC-Re-ID [17] has significant potential, it provides access to details like frame level, ground truth, full frames, and calibration information *etc.*). It corresponds to images of 1852 people existing across all 8 cameras. It covers 1413 unique identities with 22,515 bounding boxes that appear in more than one camera. It also consists of 439 distractor identities with 2195 bounding boxes that appear in only one camera. The size of the bounding box varies from 72*34 pixels to 415*188 pixels.

Occluded-DukeMTMC-Re-ID: The Occluded-DukeMTMC dataset [18] is a new benchmark for occluded person re-identification (re-ID). This dataset uses pose landmarks to guide the learning of non-occluded regions and considers only the visible shared regions for image matching. This approach helps to overcome the challenges posed by occlusions and improves the accuracy of re-ID algorithms in these conditions.

3.2.3. CUHK01

The image quality of CUHK01 [19] datasets is relatively good and this benefits the person Re-ID models to achieve good results and perform well in real-world scenarios. This dataset was published in 2012 and consists of 3884 images of 971 people. Two disjoint cameras were used to capture different views. Each camera captures two images for each person, a total of four images per person.

3.2.4. CUHK03

CUHK03 [20] dataset is one of the largest datasets in 2014 and proved good for person Re-ID and deep learning models to report effective results. The dataset comprised of 13,164 images of 1360 people. Images are captured by using six cameras. Each identity appears in two disjoint camera views (*i.e.* in each view there are 4.8 images on average). In CUHK03 bounding boxes are manually labeled and detected from Deformable Part Models (DPM).

3.2.5. VIPeR

VIPeR [21] is one the most challenging dataset, several researchers tested it and reported very wide and interesting results especially by addressing the viewpoint variation challenges. As this dataset is one of the oldest datasets and is still trending for person Re-ID models. It contains around 632 identities and images are

captured by two cameras one image per person. VIPeR also facilities with the viewing angle of each image. In our review, we have observed that the in last five years most journal paper on person Re-ID test their model on VIPeR dataset to address pose variation and viewpoint challenges.

3.2.6. MSMT-17

MSMT-17 [25] is one of the largest image-based datasets containing 126,441 images and 4101 identities. Images are captured in the morning, noon, and afternoon on campus. In our review, we have observed that MSMT-17 is a widely used dataset, although it contains a similar viewpoint to the Market-1501 dataset but this data is commonly known for capturing the most complicated scenarios that why several and almost all recent person Re-ID models test their models on this dataset and reported SOTA results. It is mostly adopted to address pose variation, viewpoint, and body misalignment challenges.

3.2.7. Partial-ILIDS

The Partial-iLID dataset [28], collected at foreign airports, contains 238 pictures of 119 individuals, with each person having both a partial image (upper body, with mostly suitcases as occlusions) and a full image. The partial images are used for searching and the full images are used for testing, similar to the Partial-REID dataset. The dataset covers a diverse range of resolutions, viewpoints, and occlusions, making it a challenging dataset for person re-identification tasks.

3.3. Video based datasets

Several video-based person Re-ID datasets are available. We have considered the top 3 video-based person Re-ID datasets MARS, DukeMTMC-Video Re-ID, and iLIDS for our review on which person Re-ID models achieved the best performance. Major person Re-ID challenges are also addressed by using these video-based datasets.

3.3.1. MARS (Motion analysis and Re-Identification set)

MARS [24] dataset is an extended version of the Market-1501 dataset. It is the largest video-based dataset having 1,191,003 images and with a maximum crop size of 256*128 among all other video-based datasets. In our review, we have found that MARS is specifically used to deal with pose variation, viewpoint, and similarity measure challenges. Paper published in almost all top journals based on video-based person Re-ID very few of them used this dataset to test their models but these effective results explain that the MARS is a significant dataset because all the tracklets and bounding box are generated automatically. This automatic generation makes learning faster.

3.3.2. iLIDS-VID

iLIDS [23] contains 42,495 images with 300 identities. iLIDS dataset contains heavy occlusion in the captured images that why it is mostly used in person Re-ID models which address occlusion challenges. In our survey, we have found that iLIDS majorly help to achieve good results in pose variation, viewpoint, and similarity measure challenges. Therefore, in the last five years, very few journal papers that address these challenges used the iLIDS dataset.

4. Person Re-ID challenges and state-of-the-art results

In recent years, person Re-ID has gained impact full attention in the community of intelligent systems and computer vision for various decisive applications. Although person Re-ID is comprehensively studied by researchers globally still it is a challenging issue. When images captured by non-over-lapping cameras under dynamic-environment are of low quality and in some images face or some other important features are not covered comprehensively. Conventional methods based on hand-crafted algorithms and small-scale evaluation are not feasible because of their limited applications. In past, results based on subtracting the background from frame to frame in multi-camera tracking are not enough due to variations in viewpoint, domain, and illumination *etc.* However, the reliable Re-ID mainly involves an accurate response that is near to real-time. This requires the availability or selection of good-quality images to cope with the challenges. Therefore, a bunch of challenges are still not fully addressed *i.e.* occlusion, pose variations, background clutter, misalignment, scale, illumination, viewpoint changes, poor resolution, and cross-domain or generalization. A few of these challenging factors occur together (like pose variations and misalignment), however, we do not combine these as the root causes of these challenges are quite different. For instance, pose variations occur when a person appears in different poses in various cameras of a surveillance network. Whereas, the misalignment occurs due to camera orientation. Generally, the Re-ID solutions address multiple challenges however, we keep them segregated, by analyzing which of the challenge is being addressed most optimally by the key strength of a given solution.

In our review, we have provided detailed progress on each of the challenges in the last six years as shown in Fig. 6. How the results are getting better yearly and which datasets are mostly used to solve these challenges. How deep learning techniques have improved results on each challenge.

We have categorized the papers of each challenge on the basis of adopted algorithmic techniques *i.e.* CNN, and attention and self-attention based approaches.

- **CNN based Re-ID Approaches:** CNNs are widely used because of their in-depth learning. They consist of the number of convolutional layers. Mainly used for image processing, segmentation, classification, and other correlated data. A sliding filter is used to convolve over the entire image to gradually learn the portions of the image and its surroundings.
- **Attention based Re-ID Approaches:** Generally the attention-based approaches are a subset of CNN-based solutions and focus to learn specific attributes. Their use enhances the focus on an important part of data while ignoring unused background information.
- **Self-Attention based Approaches:** While transformers are newly introduced in vision processing tasks to boost their performance at next level. They use a multi-head self-attention mechanism to learn image embedding. Transformers are also used in conjunction with CNNs however pure transformers applied directly to image patches produce better results. Since the self-attention based Re-ID approaches are rare therefore we discussed these under the sub-heading of Attention-based approaches in subsequent sub-sections.

4.1. Occlusion

4.1.1. Problem specification

Occlusion is caused by any overlapping object that may lead to wrong results. Persons are occluded by various environmental objects (traffic sign boards, trees in parks, vehicles in parking) or by other pedestrians in person retrieval scenarios which make it difficult to track the movement of people. Visual illustration of occlusion is illustrated in Fig. 1. When a person is occluded, the features extracted from the whole image may contain distracting information and lead to wrong results if the model cannot differentiate between the occluded region and the person region.

4.1.2. Existing methods

CNN-based approaches: Some of the works on occlusion [29,30] have achieved better performance by using part-based models via part-to-part matching. Recently attention mechanism has been introduced in person Re-ID models to pay more attention to the non-occluded part.

Ma et al. [31] handled the occluded regions by introducing a pose-guided relational transformer to create a part-aware solution. Jia et al. [32] encoded the images into a pattern set to form a global vector to handle the occlusions. He and Liu [33] presented an occlusion-aware approach by creating a salience heat map through mask-guided and pose-guided layers, which guided the spatial feature learning. He et al. [30] designed an occlusion and alignment free framework to obtain spatial pyramid features at multiple levels & scales. The semantic information is extracted by Wang et al. [34] using an adaptive graph-based topology and alignment scheme.

Recently a large-scale occluded person Re-ID dataset was introduced by Yan et al. [35] and a novel bounded distance loss that learned discriminative features from occlusion-based augmented data. Yang et al. in [36] addressed the occlusion challenge by suppressing the occluded body regions and discretizing the pose information. Similarly, [37] presented an Incremental Generative Occlusion Adversarial Suppression (IGOAS) network, to suppress the occluded parts over non-occluded parts.

An encoder-decoder based approach [38] builds correlation among occluded and non-occluded regions, in addition, to correlating the spatial and temporal features to learn long-term temporal contexts. Zhang et al. [39] augmented the semantic segmentation with the Re-ID task to learn part level and global features robust to occlusion through foreground-background mask generated by the semantic branch. In [40] a matching framework was proposed containing local and global branches to learn patch-level and part-level features explicitly and used partial person images.

Attention-based approaches:

Miao et al. [18] generated a new occluded dataset named Occluded-DukeMTMC, used pose landmarks as guidance to learn the non-occluded regions and the visible shared regions are considered for image matching. Another robust framework [41] resolved partial occlusions by introducing a temporal attention layer to recover the occlusions through spatio-temporal information of the neighboring frames.

To handle occlusions in crowded situations, a Query-Guided Attention network PISNet [42] provides improved location accuracy and attention-based distinctive features for person Re-ID. Wang et al. [43] targeted the challenge of occlusion. In the proposed novel framework both short and long-term temporal information was learned using the attention mechanism in a hierarchical manner.

In [44] a non-parametric attention mechanism was proposed that takes the video pairs as input, refines intra and inter sequence representation & provides self and collaborative feature representations Moreover, in [45], Chen et al. sliced the videos into different spatial-temporal units using the attention model to preserve the body-structure information. Aich et al. [46] proposed a 3D convolution network architecture to learn spatio-temporal feature maps, similarly [47] introduced Spatial and Temporal Memory Networks (STMN) to handle the occlusion challenge. An end-to-end part aware transformer was proposed in [48] in a weakly supervised manner that learned self-attentions in an image for pixel-based context and diverse part-level discovery.

4.1.3. Summary of the state-of-the-art methods

The top three high-performing methods, [43,46] and [47] optimized the attention-based learning of temporal and spatial information to handle occlusions. In [43] both short and long-term temporal information was learned using the attention mechanism in a hierarchical manner. In [46], a factorization attention unit was introduced to learn the spatio-temporal representation using distinctive pathways. Each path-way was then further divided to explicitly learn the person-specific static and dynamic features *i.e.* appearance and motion. In another dominant approach, [47], the spatial and temporal memory networks were introduced to refine the person-level features. The spatial memory module saves information about frequently occurring spatial distractors while the temporal memory module stores attention for typical temporal patterns information. An attention mechanism was adopted to aggregate and save the representation. Fig. 7 gives a brief overview of the published work done on the occlusion challenge.

4.2. Pose variance

4.2.1. Problem specification

Despite hundreds of research papers, person Re-ID is still challenging to be solved mainly due to complex view variations and pose variations in the person images as shown in Fig. 1.

4.2.2. Existing methods

CNN-based approaches: [49] proposes a Global-Local-Alignment Descriptor (GLAD) and an efficient hierarchical indexing and retrieval framework to address pose variations for person Re-ID. Spatial Interaction and Aggregation (SIA) module and Channel Interaction and Aggregation (CIA) were introduced in [50] to optimize the adaptability of the receptive field to learn robust person features. In [51], key-points localization and a part-based Re-ID model are proposed to improve the long-range predictions. In [52] presented a novel pose-driven Re-ID network comprising of a local feature embedding network and a pose transfer network to learn the affine transformation and to relocate the regions. Kocabas et al. [53] proposed a 4x faster multi-person pose estimation approach to jointly learn person detection, segmentation and pose estimation. Cho and Yoon [54] analyzed the camera viewpoints and person poses by calibrating the captured images and estimating their respective poses.

In [55] challenge of extreme changes in pose and view point is handled by novel unified framework that combines the saliency and semantic parsing to enhance the performance results. In [56], an attribute based approach was proposed to handle the spatial shifts caused by different pose variations and camera views. The hand-crafted LOMO features were fused with CNN embeddings for viewpoint invariant representations. Mid-level multi-type human attributes were learned in [57] in weakly supervised manner. McLaughlin et al. in [58] proposed unified framework invariant feature extraction by mapping the diverse images of same person onto same location in feature space, to learn the subtle cues from diverse set of images. In [59] Wang et al. proposed a novel part-based adaptive margin list-wise loss by replacing image pairs with ranking lists and handled the hard negative samples. Tan et al. [60] was another rank-based pose-invariant Re-ID method. Barbosa et al. [61] handled pose variations by learning structural aspects of human body. The ensemble of invariant features comprised of pose-invariant features of specific regions and of holistic image was proposed by Lee et al. [62], where as [63] proposed the Pose Invariant Features (PIF) via pose estimation and normalization and Local Descriptive Features (LDF) to resolve the misalignment errors. Teacher-student model is proposed in [64], where a part prediction alignment student module is guided by a global teacher to estimate person poses. SImilarly, [65] focused on motion-attentive local dynamic pose feature and joint-specific local dynamic pose feature. A channel parse block was proposed by Zhou et al. [66] to extract the pose information at pixel level. Wang et al. [67] proposed a high order Re-ID framework to handle the pose alignment

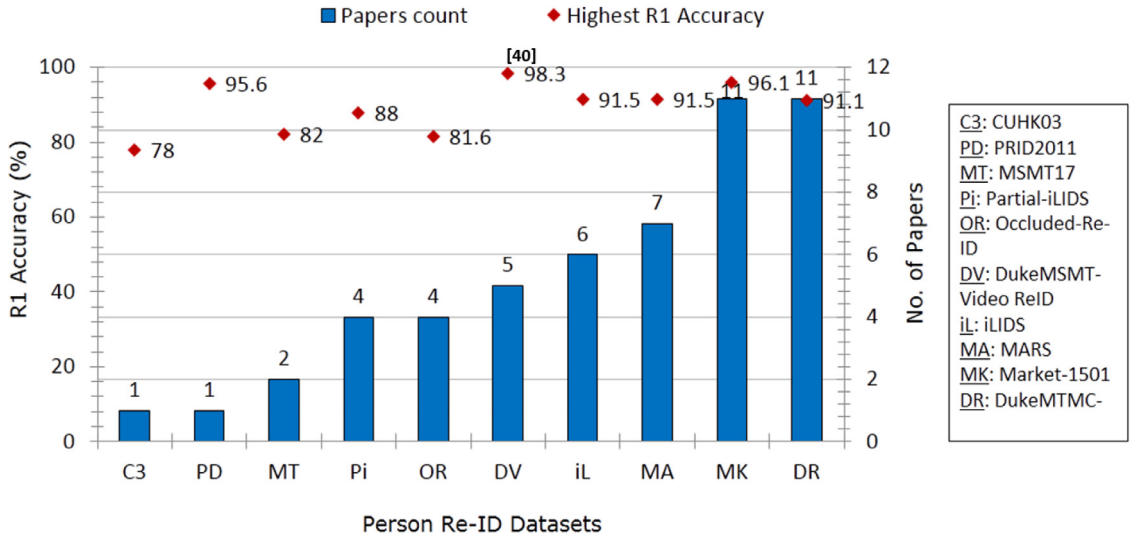

Fig. 7. Progress on the challenge of occlusion for person Re-ID benchmarks.

problem using semantic fine-grained part details of multi-level feature maps.

A unified end-to-end trainable framework is proposed in [68] to accurately rank the probe to gallery and gallery to gallery affinities. In [69] the pose variations are handled by learning fine and coarse information by a novel unsupervised re-ranking framework. Pose Invariant and Embedding (PIE) [70] resolved the challenge of pose changes and errors in pose estimation through PoseBox alignment and Pose Box Fusion (PBF) module. The GAN framework is opted in [71] and [72] proposed a joint framework to simultaneously model the detection and joint partition for pose estimation. An adaptive structure aware graph based approach is proposed for pose alignment in [73]. Another graph based solution [74] learns walking postures for person Re-ID. Chen et al. [75] and Chen et al. [76] stripped the image horizontally to learn the geometric structure in stripped parts. An et al. [77] projected the gallery and probe images onto RCCA subspace to handle pose variations. Sun et al. in [78] proposed a Refined Part Pooling (RPP) method to locate the parts and reallocating the pixel to its closest matching part. A pose-aware multi-shot matching strategy was presented in [79] to cluster the given images.

Attention-based approaches: [80] proposed an end-to-end attention approach by defining activation penalty to learn less activated diverse regions. Similarly, [81] proposed an attention-aware pose guided framework to handle misalignment and noise removal. The Global structural information is preserved using attention mechanism for contextual information learning in [82]. An end-to-end Comparative Attention Network (CAN) based framework is presented in [83], which adaptively finds the multiple local regions via set of glimpses that formulate dynamical pooling feature for enhanced Re-ID.

An end-to-end pose-guided framework was introduced in [84] to jointly learn the visibility prediction model, pose-guided attention and part visibility models and used the graph matching technique.

4.2.3. Summary of the state-of-the-art methods

The study shows that the differences in people's poses can be handled efficiently by specifically learning the part-based information and their alignment. The top three Re-ID solutions to address the pose variations include [67,68], and [82], all of these emphasize on learning the person parts, structure, and their alignment. Wang et al. [67] learns the semantic fine-grained part details of multi-level feature maps. These are then used to find the difference of similarity among aligned and misaligned parts of person images. However, in [68], an end-to-end matching and shuffling operations were used to fully utilize the affinities of both probe and gallery images that had helped to accurately rank the images. Another Re-ID solution [82] among the top performers is based on a global attention module that is used to capture the relationship between global structure and local appearance.

Figure 8 shows the progress on pose variation in literature since 2015.

4.3. Background clutter

4.3.1. Problem specification

Complex background is a critical Re-ID challenge, as depicted in Fig. 1. Generally, the Re-ID datasets are prepared under the controlled environment with less variations in the background, the Re-ID solutions for a given dataset perform well. However, the Re-ID solutions don't perform good for cross datasets due to the huge difference in the backgrounds (Fig. 9).

4.3.2. Existing methods

CNN-based approaches: [85] handled background variations by calculating human parsing maps and proposed a dataset with different backgrounds. Pham et al. [86] addressed the background clutter by removing the casting shadows through density based score matching. Considering both the chromatic and physical based features of shadow regions improved Re-ID in the presence of clustered background with occlusion. Bai et al. [87] emphasized on the foreground person image by utilizing the prior information of body structure, aligned the body parts and multi-stream framework is opted to learn the global and local part-based Re-ID features. Another part-based solution [88] proposed a self-paced constraint and a regularization technique to learn the part features in lower convolutional layers, which are then embedded into higher layers for discriminative feature extraction.

Li et al. [89] used long-short temporal spatial cues by combining the motion appearance by suppressing the background clutter and motion refinement features to activate the person specific features.

Attention-based approaches:

The attention-based Re-ID approaches generally handle the background clutter either by extracting the foreground person fea-

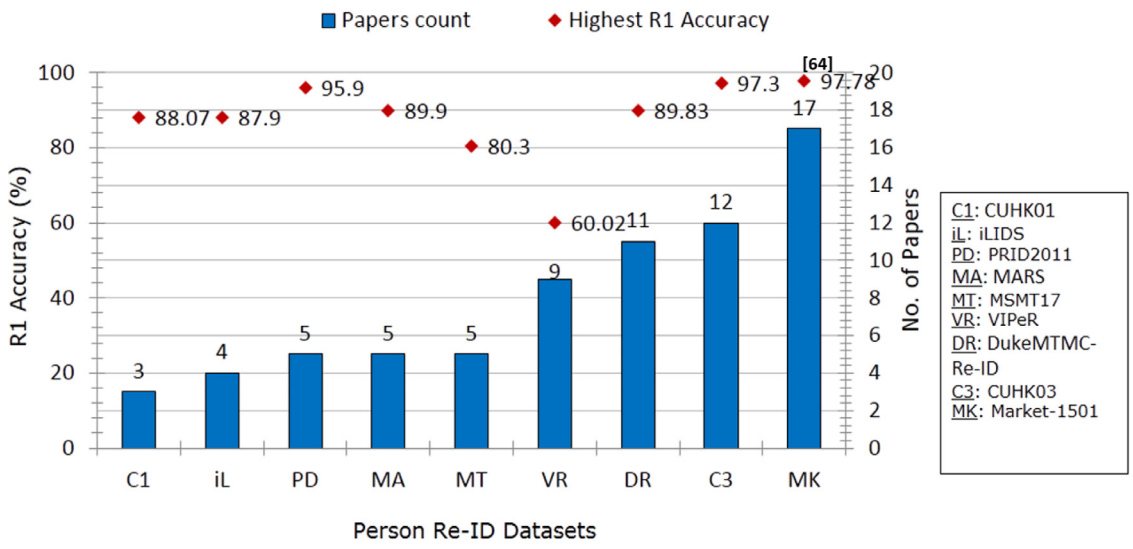

Fig. 8. Progress on the challenge of pose variation for person Re-ID benchmarks.

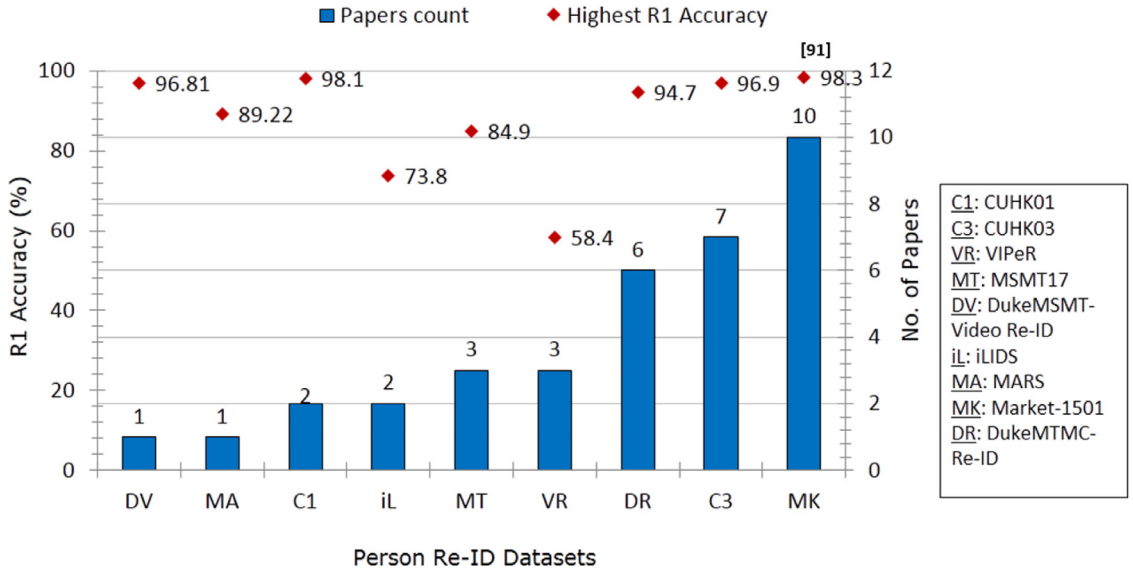

Fig. 9. Progress on the challenge of background clutter for person Re-ID benchmarks.

tures or by removing the background. In [90], the attention features extracted by attention encoder are further refined by the processing of decoder that develops binary masks to segregate the foreground from background. The extended version of [90] was presented in [91] where the encoder-decoder architecture comprised of three sub-network *i.e.*, Foreground attentive network, Body-part sub-network, and Feature fusion network, optimized by the symmetric triplet loss function.

Another soft mask based end-to-end foreground aware network was presented in [92] that modeled the background using pedestrian and camera IDs and reduced the negative impact of background changes. Sun et al. [93] proposed a multi-level attention and fusion model to learn the foreground global and fine granular features. Similarly, in [94], Ning et al. proposed a multi-branch attention network to design a feature refinement approach by removing the background interference.

4.3.3. Summary of the state-of-the-art methods

Background clutter Re-ID challenge is generally addressed by removing the cluttered background from the foreground person image and by learning the multi-level features. The SOTA method [94] removes the background interference from the person features by introducing multi-branch attention network. Another solution [90] among the top performers is an encoder-decoder foreground aware attention subnetwork that focuses to learn the body part information by constructing the binary mask using novel loss function. Its extended version [91] is comprised of three sub-networks, the foreground attentive sub-network, consists of the encoder-decoder baseline; the body-part sub-network, inputs the encodes feature maps, slices them, and learns the body-part features to learn the body-part; and the feature fusion sub-network, to fuse all learned features.

4.4. Misalignment

4.4.1. Problem specification

Body misalignment (the problem in which body parts of person are spatially misaligned) is a critical Re-ID challenge, as shown in Fig. 1. The progress of relevant solutions on various Re-ID datasets is given in Fig. 10. Generally, to resolve this challenge conventional neural network based approaches have been proposed and

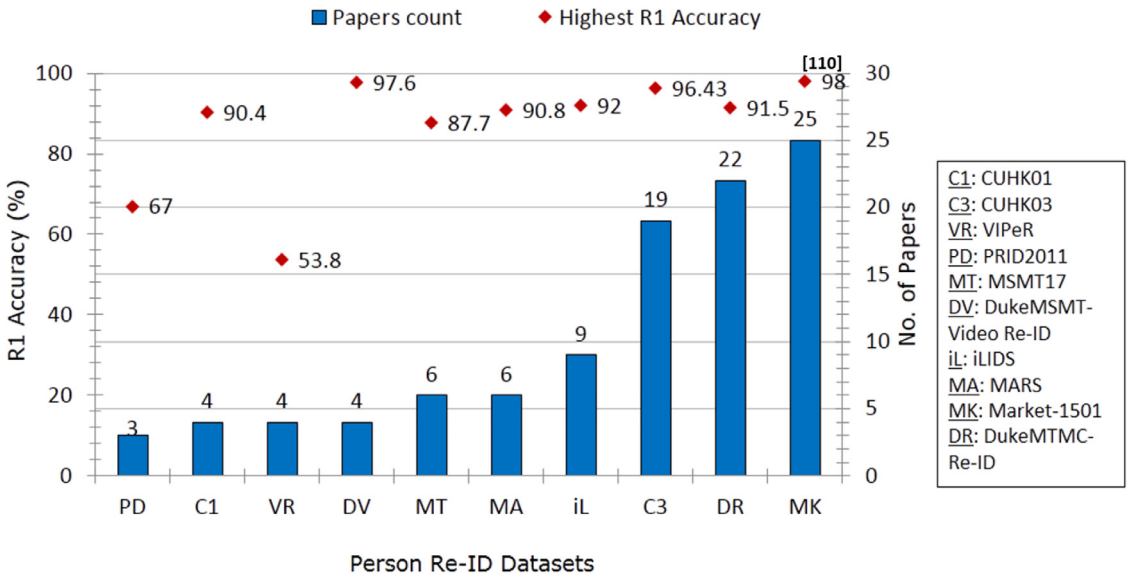

Fig. 10. Progress on the challenge of misalignment for person Re-ID benchmarks.

addressed it by using part based alignment and by focusing on shared regions in the images.

4.4.2. Existing methods

CNN-based approaches: Dividing person images into predefined spatial regions or strips is one of the most famous techniques to align the local body parts. Zhang et al. [95] divided a person image into semantically aligned 24 parts to learn the canonical surface based representation in UV space. Sun et al., in [29] proposed an alignment scheme using two sub-network, one to generate convolutional features and other to align the similar parts.

In [96] proposed a multi-stream part-aligned mechanism by joining the part-maps with appearance features to embed the human poses information into person representations. Zhao et al. [97] proposed a tree structured fusion network is proposed to align the structure of human body parts obtained from ROI pooling framework with the semantic features.

A weakly supervised semantic parsing approach is proposed by Zhu et al. [98] to address the misalignment, where the pseudo-labels of human body parts were generated to refine the part-base features through clustering. Sun et al. [99] proposed a self-supervised Visibility-aware Part Model (VPM) inspired from holistic person Re-ID [29] and [100], and learns the features of shared region among two images using region locator and region extractor. A multi-branch Dynamically Matching Local Information (DMLI) scheme is proposed in [101] to align the horizontal strips without any explicit information of pose estimation.

A patch and saliency matching based computational model, RankSVM is proposed in [102] to handle the misalignment issue. The patch size is taken as 10x10 to extract the dense color histogram and dense SIFT features. Li et al. in [103] addressed misalignment issue by proposing a multi-level cross-view dictionary learning model to capture multi-level characteristic of an image. Another method [104] focused on stripe-based approach to learn the features at multi-granular level. Ding et al. [105] presented a multi-task part-aware network to handle the challenge of misalignment by aligning the semantically aligned part features through part specific channels. Li et al. [106] learnt the patch wise attributes at local level and fused them for final representations.

Poor spatial alignment for video sequences was addressed in [107] by temporal residual module and spatial-temporal modules. Shi et al. in [108] fine-grained features are learnt from Re-ID videos using a three dimensional semantic appearance alignment module to align the local features and weakening the influence of occluded body parts. An unsupervised video matching algorithm is proposed in [109] to address the alignment issue. Patch wise metric learning approach [110] learnt the patch features and aligned them using deformable models.

A spatial and temporal alignment method for video Re-ID was proposed in [111], where a video chunk is divided in walking cycles to extract the temporal sequence, the body parts are obtained for the spatial domain and fisher vector and bag-of-words features are used for person representations. Shen et al. [112] handled the spatial misalignment by matching the local and global patches among the given pair of cameras through one-to-one and one-to-many graph structures. Similarly, [113] proposed a part-guided graph convolution network to handle misalignment issue.

Attention-based approaches:

Li et al. [114] proposed a framework to resolve the challenge of alignment by joint learning of hard region-level attention along with soft pixel-level attention. In [115], a decoder and encoder based approach learned multi-grained spatio-temporal and positional spatio-temporal features to handle the misalignment. Chen et al. [116] and Fang et al. [117] handled misalignment by extracting channel wise and position wise attention information through specialized channel and position attention modules. Guo et al. [118] presented a supervised human body parts parsing approach to segregate the human and non-human parts from given images and self-attention mechanism was used to group alike pixels. Wang et al. [119] introduced a novel fully attention based block into any convolution neural network to obtain the channel-wise and spatial-wise attention information. Zhao et al. [120] decomposed the body parts into regions and proposed an attention feature mapping mechanism.

In [121] performed pose estimation in a multi-branch network architecture through inter and intra-attention feature extraction of the body-part and whole-body. [122] divided the image vertically and horizontally to obtain the structural information and performed parts detection and representation through attention mechanism.

In order to learn semantically aligned part-level features a simple batch-driven approach was proposed in [123], comprising two modules *i.e.* a guided attention channel to highlight channel attention and a pair of regularization term to maintains the consistency

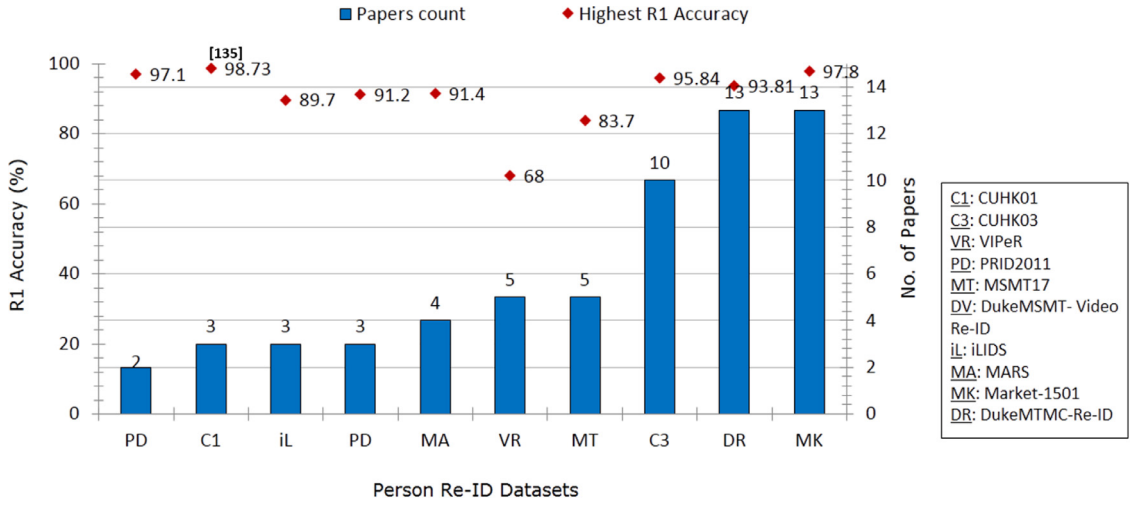

Fig. 11. Progress on the challenge of scale variations for person Re-ID benchmarks.

among batches. In [124] a novel triplet loss was proposed to align the distributions of local parts generated through attention mechanism.

An attention mechanism was adopted in [125] that exhibit its focus on body parts rather than background. Gu et al. in [126] handled the temporal appearance misalignment in video based person Re-ID by proposing an appearance preserving framework to preserves the appearance at pixel level and to model the temporal information.

4.4.3. Summary of the state-of-the-art methods

In order to handle the misalignment issues, generally the local parts associations are built among person images so that various local parts could be aligned with their respective local parts of corresponding images. Another popular technique is to extract spatial and temporal information. In [113] used part-guided graph convolution network to describe the relationship among adjacent parts. The inter-local graph learns local associations among the same parts of given person images and the intra-local graph learns associations among variant parts of a person image. He et al. [115] and Aich et al. [46] extracted spatio-temporal features using encoder decoder technique and through explicit pathways respectively.

4.5. Scale difference

4.5.1. Problem specification

In open surveillance, scene images are captured at an arbitrary scale, which makes it challenging to learn correlations among features of different scales. Therefore, the re-id solutions should be capable to extract the person representations at various scales. Generally, most of the Re-ID solutions operate on a single scale, hence not providing optimal Re-id results. In contrast, multi-scale representations can reidentify people at different scales as shown in Fig. 1. In the past five years progress achieved on this challenge is shown in Fig. 11.

4.5.2. Existing methods

CNN-based approaches: [127] presented a Siamese network-based Re-ID solution to perform cross-camera matching and extracted person features at different scales. A coarse-to-fine model [128] addressed the scalability challenge by learning 3D sub-maps of discriminative features at different scales. In [129], an omni-scale feature extraction method is proposed based on different receptive fields and introduced point-wise and depth-wise convolutions instead of standard convolution.

In [130] multi-scale feature learning problem is resolved using an end-to-end Deep Pyramidal Feature Learning CNN architecture and learnt correlation among scale specific multiple branches. In [131] issue of multi-scale along with pose estimation was resolved using the novel Spatial-Temporal Correlation and Topology Learning framework (CTL) to obtain the diverse discriminative semantics local features at multi-granularity levels.

Li et al. [132] presented a video-based Re-ID solution by introducing a 3D multi-scale convolution layer that can be inserted into any existing 2D convolution network to learn the spatial and temporal cues. In [133] a novel similarity learning model was presented that used multi-view visual words and metric optimization through k-means clustering. Another unsupervised cross-dataset omni scale Re-ID approach was presented in [134] to learn the features at multiple spatial scales enhanced by channel-wise weights. Wu et al. [135] proposed a scalable adaptive framework to address the challenge of scalability in unsupervised manner, where the source dataset was trained on labelled dataset and teacher-student transfer learning technique was adopted.

Attention-based approaches: In [136], an attention based spatial transformer Siamese network is proposed to learn the visual similarities among images at different scales. A cross attention-based multi-scale model was presented in [137] to learn the discriminative information of different body parts of specific identity from multiple views. A new large scale Bird-View dataset was also proposed in this work.

Qian et al. [138] handled multi-scale challenge by introducing a multi-scale layer to learn the multi-scale features at local and global level, and a leader based attention learning layer to selectively learn the optimal weightage assigned to each scale. Another weight assignment based solution is proposed by Zhang et al. in [139] where the weights are assigned to local regions using attention mechanism at both spatial and temporal levels. In [140], the spatial and temporal representations were obtained using two branches *i.e.* pyramid dilated convolution and pyramid attention pooling to extract mulit-scale features. In [141], multi-scale pooled regions are fed into a novel deep architecture to extract the discriminative features at multiple scale semantic levels.

Another multi-scale attention pyramid method was presented in [142] comprising channel-wise and spatial attention modules, where attention features were learnt at multiple local levels and stacked to form attention pyramids. Zhong et al. [143] introduced a two stage attention was also introduced to filter the noisy feature maps. An end-to-end Re-ID solution, a multi-scale Omni-bearing

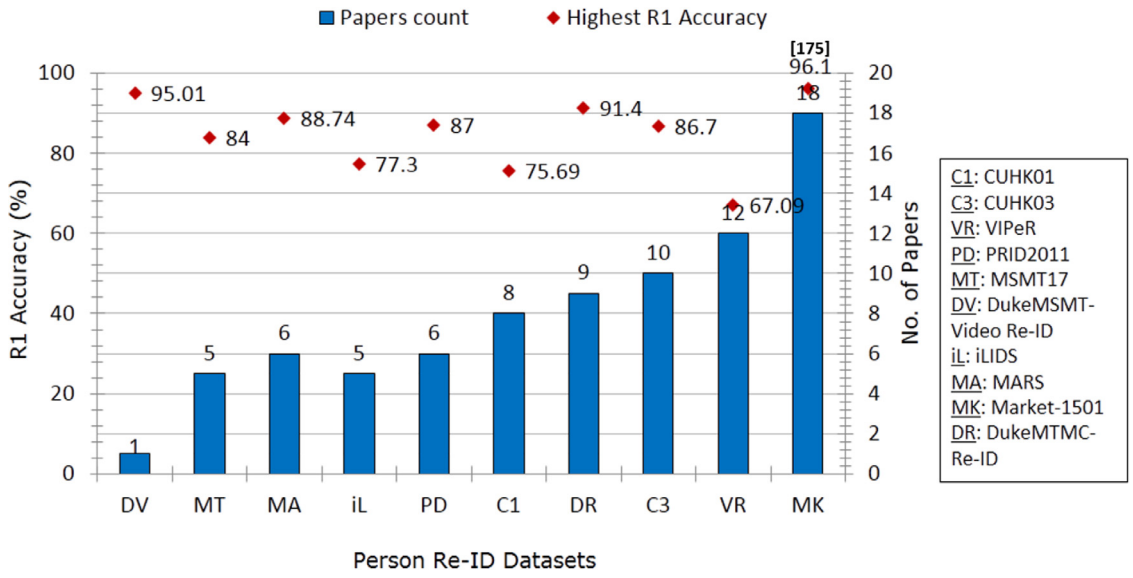

Fig. 12. Progress on the challenge of viewpoint variation for person Re-ID benchmarks.

attention network is proposed in [144] to extract holistic and local feature maps.

4.5.3. Summary of the state-of-the-art methods

To handle the scale differences, the multi-scale approaches extracted person features at various scales. Learning of spatial and channel wise attention features at various scales performed well to handle multi-scale variations. Qian et al. [138], is among the top performing Re-ID solutions that presented an attention based multi-scale Re-ID solution and introduced optimal weighting schemes. Moreover, [141] and [142] are based on multi-scale attention pyramid method to mitigate the scale challenges. Martinel et al. [141] introduced the attention mechanism in pyramid-based multi-scale network and [142] extended this work by dividing the features into multiple local parts to learn the channel-wise and spatial-wise attention features. Similarly, another technique [144] extracted the holistic and local feature maps using multi-scale omni-bearing attention network by learning spatial and channel wise attention features.

4.6. Viewpoint variance

4.6.1. Problem specification

Viewpoint is the most important challenge in developing person Re-ID solutions because different views of a pedestrian across non-overlapping cameras contain different information. Either finding the particular view angles/ viewpoints or extracting view-generic/ view invariant person features can play an important role to reduce the intra-class distance and increase the intra-class distance. Fig. 1 demonstrates the viewpoint variation challenge. When the complex changes occur in the visual appearance between non-overlapping cameras, the view generic features become inadequate in solving person Re-ID task. Another bigger problem in resolving this challenge is the lack of data and some datasets contain fixed and insufficient distribution of environmental factors *e.g.* in pedestrian viewpoint, some angles might contain few samples. The progress of work done so far on viewpoint challenge is shown in Fig. 12

4.6.2. Existing methods

CNN-based approaches: Sun and Zheng [145] designed the practical solution by formulating the synthetic dataset PersonX for subjective study of Re-ID challenges. They have used IDE+ with ResNet-50 [146] as backbone with 36 angles and pre-trained ImageNet weights [147]. They have concluded with controlled experimentation that person with side views makes better query. Another CNN-based architecture presented in [148] proposed pixel-wise matching among the obtained representation to produce the promising results. In [149] single dictionary was encoded through constraint on association of sparse representations, for both gallery and probe images simultaneously to obtain the view-invariant feature. Liao et al. [150] analyzed the presence of local features in horizontal direction and the occurrences of local features are maximized to make the approach viewpoint invariant.

In [151] visual variety was handled using teacher-student framework, teacher guides the student regarding multiple views and as a result student resulted in state-of-the-art in Image-To-Video by large margin. Dai et al. [152] proposed a novel cross-view semantic projection learning algorithm and introduced three components, a shared basis matrix to explore intrinsic structure of raw descriptors across camera views; a pair of semantic projection functions to map the original hand crafted features into common semantic space; and an optimal association function to capture the best association between alike semantic representation. Chen et al. proposed a asymmetric distance learning model [153] to tackle the multi-camera view challenge by learning camera-specific projections followed by modeling the correlation among different views. Wu et al. [154] proposed a metric learning based viewpoint invariant algorithm by considering the pose information prior from training data and hand-crafted features for person representation.

In [155] viewpoint specific approach is proposed to adopt camera view features based on cross-view correlation and adaptive feature augmentation for transformation of original features to new augmented space and for controlling the degree of correlation among them. A view-specific model presented in [156] proposed cross-view euclidean constraint to reduces the distance between similar identities through the novel cross-view center loss. [157] proposed a ranking model to build the relationship among input image pairs and their similarity scores via joint representational learning directly from raw image pixels.

In [158] multiple metric learning method was presented that was based on cross-view quadratic discriminant analysis algorithm to find the importance of each feature. Wu et al. in [159] proposed a deep hashing framework and calculated Hamming distance to

rank similar images closer and hash codes to make the process faster. In [160], body parts based multiple convolution feature are extracted and inter-layer interactions were learnt. Borgia et al. presented a metric learning approach in [161] and proposed a novel loss function. Garcia et al. [162] proposed an unsupervised method to obtain the content and context from initial ranked lists and then removed the visual ambiguities from them for re-ranking. A view-invariant video-based few-shot learning method was proposed in [163] using the recurrent neural network. Lin et al. [164] proposed an unsupervised style transfer model to generate the images with transferred-style and different camera styles and then similar images are grouped using cluster approach. Another clustering based deep asymmetric metric learning method [165] proposed a two stream deep neural network to jointly learn the view and feature specific transformations.

A novel descriptor enhanced by metric learning was presented in [166] to learn the structural information and to maximize the horizontal occurrences of multi-granularity to extract the rich features representation even in case of drastic change in viewpoint. A pragmatic semi-supervised framework [167] proposed the view-specific projections against each view and introduced a re-ranking strategy. Another adaptive multi-projection metric-learning method was introduced in [168] to jointly learn the different camera projections into a common feature space and for newly added camera projections are learnt without updating the existing projection matrices. A view-invariant subspace was learnt in [169] using adversarial approach and adaptive weighing was implemented to handle the imbalance of identity pairs. Another multi-view metric learning method [170] exploited the fusion of handcrafted and deep features to produce the discriminative feature representation.

An unsupervised framework for video-based person Re-ID was presented in [171] to learn the relation of frame with its first neighbour and to form the group in each camera. Cross-view matching strategy then find the matching relationship among them. Similarly, [83] used tracklet as query search for nearest neighbour best match after possible iterations through KNN search. Another video-based Re-ID approach [172] addressed the variations exist in same video and a novel loss function is proposed using intra-video loss and Siamese loss. [173] generated the synthetic data using already available grouping information data to minimize the discrepancy among multiple views through multi-level learning framework in iterative manner.

An unsupervised approach to learn asymmetric learning of cross-view person images was presented in [174]. For each view model learns the specific projection, based on asymmetric clustering. In order to achieve the better matching performance, model finds the shared space with low view-specific bias. Another unsupervised approach [175] proposed Clustering-based Asymmetric Metric Learning (DECAMEL) to learn an initial asymmetric metric using a linear unsupervised model and by jointly learning metric and features in the shared space in an end to end manner.

Attention-based approaches: Zheng et al. [176] proposed a Siamese-based flexible attention mechanism architecture to address the challenge of viewpoint to achieve the attention consistency among the images of same person. [177] proposed attention-based view-invariant features through the adversarial learning, drawing same features towards center and SIFT guidance. In [178] attention aligned network was presented to learn view-invariant foreground information using channel wise multi-scale attention aware mechanism, which are optimized by an improved triplet loss.

To learn the view-invariant features GAN and another contrastive learning module was combined into one training framework in [179], where the novel views are generated using mesh based view generator. The proposed flexible method did not rely on labeled source domain.

4.6.3. Summary of the state-of-the-art methods

Generally the viewpoint variations are handled by maximizing the inter-class distance and minimizing the intra-class distance at holistic level as well as at the local-parts levels through various alignment schemes. In [178] attention aligned network was presented that uses channel wise multi-scale attention aware mechanism using triplet loss to learn the invariant views obtained from different cameras.The Re-ID solution [180] handled the viewpoint differences uniquely by involving a lightweight and labelled part segmentation head to the backbone of Re-ID during training process and obtained diverse set of features. A training strategy adopted in [151] where multiple views are learned representing the target object to produce better results.

4.7. Low resolution

4.7.1. Problem specification

The low resolution images captured from the distant surveillance cameras is a big challenge as shown in Fig. 1. And the low resolution probe images and high resolution gallery images make the Re-ID more challenging.

Fig. 13 shows the progress of available papers on low resolution between 2015 to 2021.

4.7.2. Existing methods

CNN-based approaches: A supervised framework combined embedding of multiple layers into single layer to resolve the challenge of resolution [181]. In order to achieve efficient results lower layers with higher resolution are combined with higher layers having semantic information. Li et al. [182] addressed the low resolution that was beyond the scope of re-scaling and interpolation mechanism. The same scale image features are presented in a latent space and then distance metric modeling is performed at each scale, this resulted in formation of shared space among low resolution and normal images of same person to obtain the effective results.

In [183] presented procedure pair of high resolution and low resolution dictionaries and mappings functions are learned during training, due to this learned dictionary and mapping function low resolution images are converted to high resolution discriminant features. Extension of this work [184] supported multi-view, and learnt different mappings to convert low resolution images into discriminative high resolution features. Han et al. in [185] handled low resolution by learning the content aware details through a self supervised approach. The model assigned soft labels that are automatically generated according to the optimal scale factor. The probability of generated labels indicates the optimality of assigned scale level. Hence increasing the level of confidence against optimal scale for optimal resolution with context aware prediction.

In [186] an end-to-end adaptive feature fusion framework was proposed that has proved effective in handling resolution at different recovered body regions and at multiple scales. An adaptive feature integration module balances the relative importance of super resolved image content. In this manner adaptive weights were assigned to input features with super resolution. In [187] resolution-aware framework used the knowledge transfer technique to minimize the variations in resolution of images. Teacher knowledge was exploited and transferred to the low resolution student network to narrow down the resolution differences.

Cross resolution problem was considered in [188], where an existing GAN architecture was improved using high resolution images in an end-to-end fashion. Problem of low resolution was resolved by forming the association between super resolution images and Re-ID task. The formulation was due to parameterized sharing while training in end-to-end manner.

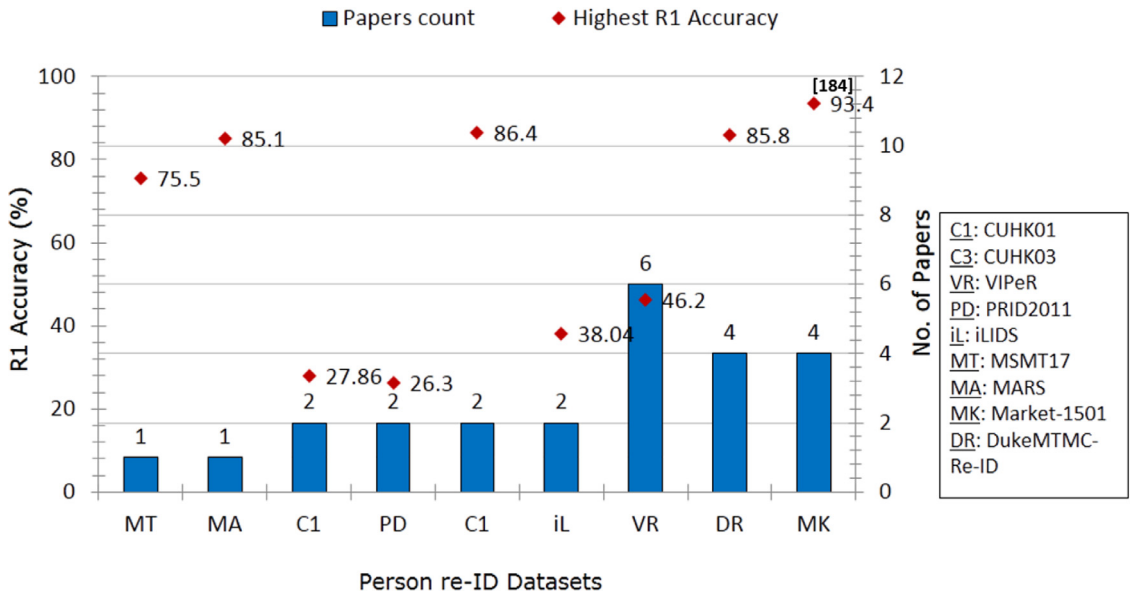

Fig. 13. Progress on the challenges of low resolution for person Re-ID benchmarks.

In [189], self-supervised strategy extracted the identity related information to resolve the challenge of degradation in real time. Here, using GAN based approach, the gap between gallery images and real world images was reduced. Model effectively preserves the features related to identity information and remove the features that were related to degradation to achieve the effective results.

4.7.3. Summary of the state-of-the-art methods

Most of the Re-ID benchmarks comprise of low resolution person images, therefore, handling low resolution is mostly the prime feature of almost all Re-ID solutions. However [187] handles this issue efficiently by using cross-resolution knowledge transfer technique to obtain the discriminative representation even in case of low resolution images. In another technique [181] effective embedding are combined at multiple layers using skip connections resulted to obtain the unique features. In [188] an association scheme is learnt from training data having super resolution images, this has helped to obtain the discriminative features even from low resolution images.

4.8. Cross-Domain/Generalization

4.8.1. Problem specification

Person Re-ID models produce improved results when trained and tested on same dataset but perform poor when tested on different dataset due to different scenarios *e.g.* changes in viewpoint, place, background, resolution and different visual appearance. For this, recently clustering, domain adaptation and image-to-image translation-based approaches reported SOTA results. One such approach is shown in Fig. 14.

The aims of image-to-image translation are to develop a mapping function between two domains and it required paired training data which is difficult to manage. For domain adaptation, labeling a dataset is an expensive and time-consuming task. Due to these factors improving generalization is still challenging in person Re-ID. Progress of Papers that have addressed the generalization challenge in top conferences and journals is shown in Fig. 15.

4.8.2. Existing methods

CNN-based approaches: Generally, the concept of pseudo labels is used to transfer the knowledge of data for cross-domain adaptation by unsupervised solutions. However, the resultant noise from this procedure reduces the efficiency of the proposed models.

Ge et al. [190] refined the pseudo labels by proposing an unsupervised framework, Mutual Mean-Teaching (MMT). It defines the off-line and online refined hard pseudo labels alternative training method through a novel soft softmax-triplet loss. Dai et al. [191] proposed a dual-refinement approach to minimize the influence of noisy labels, and refined the pseudo labels through off-line clustering, while feature refinement was performed at online training phase. Zheng et al. [192] introduced label banks in an iterative manner and formed hierarchical clusters to split the computed and propagated label information. Huang et al. [193] suppressed the background inferences, focused on the foreground person information, and trained the model using virtual labels of the target domain. For pseudo-label's generation [194] proposed two modules *i.e.* inter-camera and intra-camera computation, to compute a new feature vector for different cameras and to measure the intra-camera computation similarity respectively. In [195] a Group-aware label transfer algorithm was proposed to promote the pseudo-labels via online interaction and clustering algorithm.

The dictionary-based approaches [196–198] transferred from the labeled source to the unlabelled target domain, and then fine-tuned the pseudo-labels. Similarly, in [199] context representation was learned and transferred to the target Re-ID dataset via Bayesian adaptation by handling both labeled and weakly labeled data. To tackle the challenge of pseudo label noise [200] refined the pseudo labels based on clustering consensus.

Bai et al. in [201] generated pseudo labels and proposed hierarchical scheme graph convolutions to learn the structure of each cluster and refined them. In video person Re-ID [202] a joint global and local framework estimated the pseudo labels and a novel loss term chose the pseudo label with higher confidence iteratively. A self-training scheme, in [203] trained the encoder on the basis of guessed labels for unlabeled target data to achieve effective results.

Jin et al. [204] introduced a style normalization and restitution module to filter out the variations in style and color using instance normalization for identity-specific feature refinement through a novel feature disentanglement dual loss. Chen et al. in [205] presented novel quadruplet ranking loss and used four hard samples of the dataset to propose a generalized Re-ID solution. Khatun et al. [206] introduced an additional term in the quartet

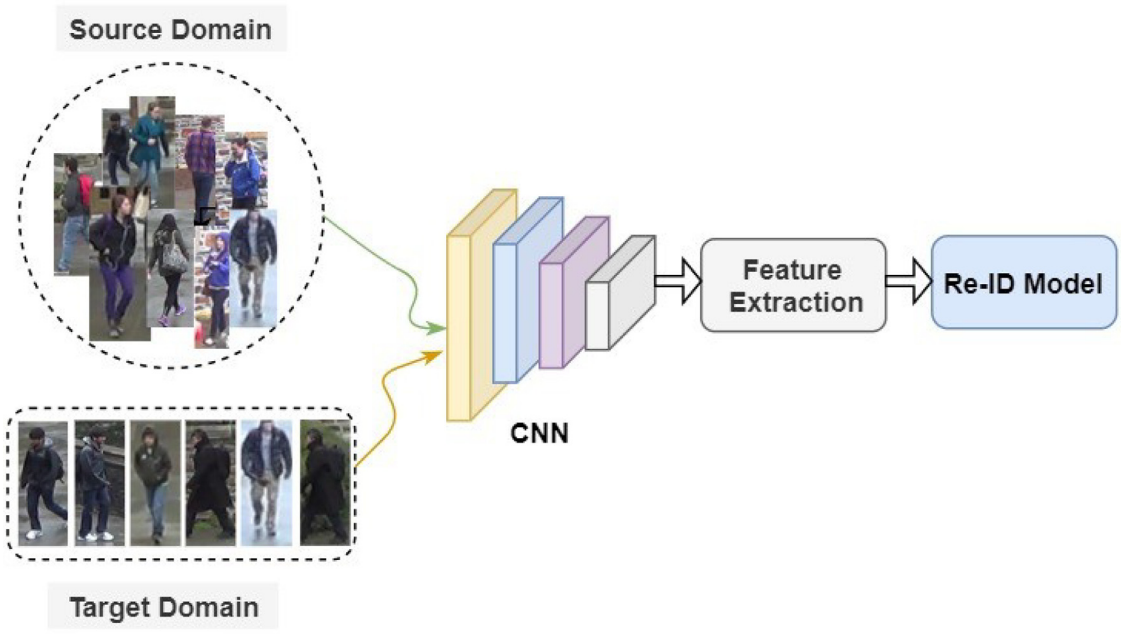

Fig. 14. An approach to achieve generalization.

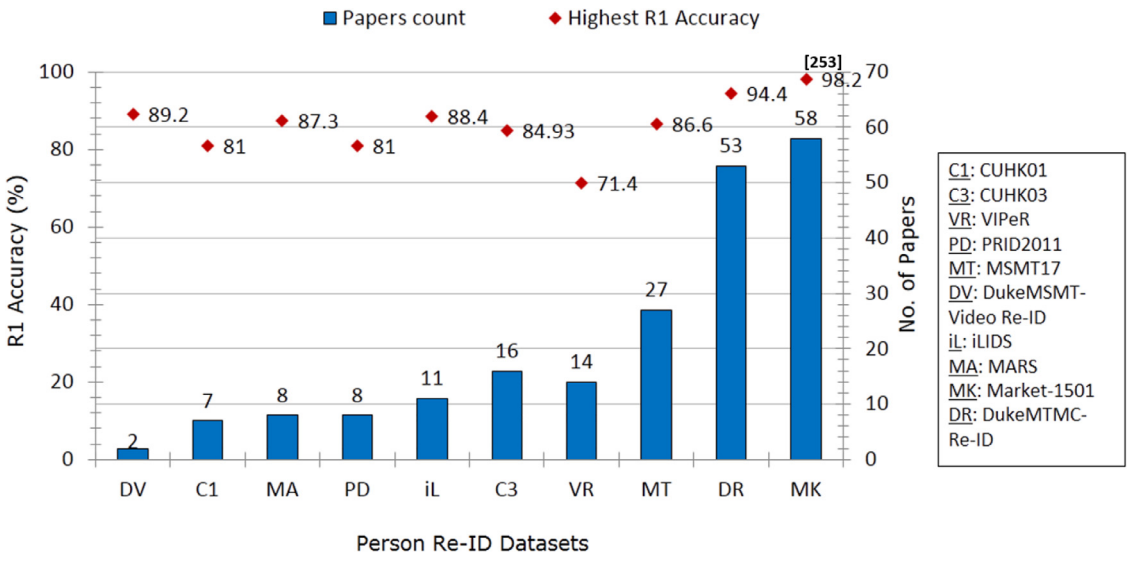

Fig. 15. Progress on the challenge of cross-domain generalization for person Re-ID benchmarks.

loss to ensure the distance between matching pairs is less than the specified threshold. In [207] Meta batch instance normalization (MetaBIN) is attained through a novel meta-train loss and a cyclic inner-updating manner without using conventional augmentation mechanisms. Feng et al. [208] formed subgroups by merging neighbors and introduced similarity-aggregating loss to pull similar images closer. Chen et al. [209] adaptively learned the credible samples for training by avoiding noisy labels and an instance margin loss was introduced that has increased the margin of instance.

Mekhazni et al. in [210] presented pair-wise distances optimization by using gradient descent and introduced novel dissimilarity-based discrepancy loss to make the source and target distribution similar. In [211] proposed a Dynamic and Symmetric Cross-Entropy (DSCE) loss to cope with the challenge of negative effect caused due to noisy labels and the camera shift issue was also addressed. Xiao et al. [212] simply merged multiple person Re-ID datasets and trained a CNN using single softmax loss, and replaced the standard dropout layer with domain guided dropout layer to introduce a generalized Re-ID model. Zhao et al. [213] proposed a meta-learning strategy and introduced a memory-based generalization loss and a meta-batch normalization layer to diversify the advantages of meta-learning.

Kalayeh et al. [100] is a semantic segmentation-based approach to learn local part-based features. Quan et al. [214] introduced a flexible part-aware module to capture the structure information of the human body. PatchNet [215], introduced part-based learning in an unsupervised manner. In [216], Part Loss Network (PL-Net) was proposed to minimize the empirical classification risk (ECR) on training images and representation learning risk (RLR) on unseen images via part loss. A voting-based approach, relevance-aware

mixture of experts network [217] dynamically integrated the diverse characteristics of source domains. Second order information bottleneck was integrated into the network by Zhang et al. [218] to introduce domain invariance. Zhang et al. [219] provided a cost effective solution by presenting synthetic data and introduced generalization via transfer knowledge technique on real datasets. domains. [220] introduced patch-based metric learning via CNN and positive patch-pairs and local salience learning through k-mean clustering and patch weights.

[221] introduced collaborative learning through multiple branches, one to learn spatial contextual information along the temporal range and the other branch to learn the identity information. In [222], a multiple expert brainstorming network was built using multiple expert models, and their heterogeneity was adjusted through a novel regularization scheme for their feature distribution in the target domain[223]. reduced the domain gap by assigning a common subspace to all camera images of a network. A camera-aware approach [224] introduced an interpolation mechanism to obtain cross-domain generality.

In [225] a joint learning framework learned id-related feature space and the disentangling module encoded cross-domain images into appearance space and structure space. Another feature-centric solution [226], DMG-Net worked for Re-ID generalization and proposed a new dataset Person30k. An end-to-end domain generalization network presented in [227], is a domain invariant solution based on adversarial feature learning and identity similarity learning. Another domain adaptive network [228] comprised of graph consistency constraint applied on different teacher and student networks to extract discriminative features.

A memory module was proposed in [229] with the purpose to make the system invariant in terms of camera viewpoint and neighborhood changes. They have used the specifically unlabelled dataset to learn unsupervised discriminative representation and to make model domain invariance. An Adversarial auto-encoder-based domain invariant approach was designed in [230], which aligned the distributions across variant domains by using the maximum mean discrepancy measure.

For domain adaptation, a 3D guided network, [231] performed image-to-image translation to form the possible synthetic poses and viewpoints of a person while preserving the identity information. An unsupervised approach [232] managed the domain gaps using Gaussian distribution and applying a momentum update mechanism.

Multi-source domain optimization was presented in [233] that first introduced domain-specific batch normalization and then infused multi-domain information. The limited scale issue of existing Re-ID datasets was addressed in a new dataset named Large scale Unsupervised Person Re-ID (LUPerson), introduced in [234].

In another domain adaptation approach [235] a clustering algorithm optimized the features in an iterative manner for progressive domain adaptation and the Fourier augmentation was deployed to enhance the class separability. In [236] patch-wise features were learned from unlabelled patches of a person's image. And a novel loss function guided the model to mine the discriminative information.

An end-to-end self-supervised approach [237] minimized the domain discrepancy using the agent learning mechanism. Zhang et al. presented an unsupervised view-invariant approach [238] to tackle the person Re-ID at multi-scale levels by improving pseudo-labels progressively. For unsupervised domain adaptation a multi-loss optimization learning model was proposed in [239], where the pseudo labels are estimated via a clustering mechanism. And for the intra-domain relation similarity and adversarial learning, two losses were introduced.

In [240], the proposed framework learnt the discrimination and association within-camera and cross-camera respectively. [241] bridged the gap between source and target by introducing attribute-identity embedding and fine-tuned them to make them adaptive to the target domain. In [24] video tracklets were used to achieve the generalized results at a large scale, hence proposed a new ideal dataset named MARS.

The challenge of cross-view and intra-bag alignment was addressed using a weakly supervised approach [242]. A domain adaptive, density-based clustering technique is adopted by Zhai et al. [243] that used GAN min-max strategy and refined the clusters iteratively.

Liao1 et al. in [244] bridged the domain gap using the "divide-and-conquer" strategy for factor-wise style transfer [245] based on CycleGAN [246]. Similarly, [247] also used CycleGAN to transfer the label information to the unlabeled domain and used the maximum mean discrepancy as a base to pull the alike distribution closer. Another Cycle-GAN-based approach [248] proposed "learning via translation", and generalized the domain in an unsupervised manner.

In [249] generalization was targeted using a lifelong learning scenario. The proposed framework maintained the learnable knowledge graph that updates the previously learned knowledge in an adaptive manner.

Attention-based approaches: A self-critical attention-based learning mechanism [250] guided the attention agent to learn the correct attention maps through unified modules, the self-critic and self-correctness modules. In [251], the high-order attention module learnt high-order relationship among human parts using a polynomial predictor of high-order. In this way, discriminative attention maps are obtained with subtle differences.

Tay et al. [252] proposed an attention mechanism to identify the specified body parts in a unified learning framework and the identity information is integrated with attribute features and body parts. An end-to-end supervised approach [253] proposed a Siamese network intra-sequence and inter-sequence attention mechanism for feature refinement and alignment accordingly. A novel cross-correlated attention module was presented in [254] to learn the inherent spatial relation of different regions of a person's image.

In [255], a novel harmonious attention network was deployed to jointly learn the attention-based pixel representation of soft and hard regions. Another attention aggregation formulation was designed in [256] to handle the changing representation of identity in a query image. Videos captured from different cameras might end up a video with different camera views, and this view variation is a challenge handled in [257]. The end-to-end framework handled the inter-dependencies among video sequences using RNNs. These similarity scores are then used to form the attention network in spatial as well as temporal dimensions.

4.8.3. Summary of the state-of-the-art methods

To perform person Re-ID across different domains is quite challenging, however, it is mostly handled by incorporating generalization in a Re-ID solution. However, few top-ranked Re-ID solutions emphasized on integrating variant modalities in the default input features or presented the synthetic data and introduced transfer knowledge techniques to generalize their solution for unseen domain experiments[256]. introduced generalization by in solution by developing an attentional aggregation formulation, which flexibly incorporates the similarity metrics along with multiple representations. Fu et al. [234] enhanced Re-ID by data augmentation and temperature usage in a contrastive learning framework. In another framework, [250] quality of attention maps is measured at both spatial and channel levels, which has guided the learning process to obtain unique representation via reinforcement learning mechanism.

Table 5
SOTA results obtained on each challenge against each dataset.

Sr.No	Dataset	SOTA R1-results	Paper cited	Challenge addressed	Venue
1	Market-1501	98.3	[94]	Background	TCSV-2020
2	DukeMTMC-Re-ID	94.7	[94]	Background	TCSV-2020
3	CUHK01	98.73	[138]	Scale	PAMI-2020
4	CUHK03	97.3	[68]	Pose	PAMI-2021
5	VIPeR	71.4	[258]	Generalization	CVIU-2020
6	MSMT-17	87.7	[113]	Misalignment	PR-2021
7	PRID-2011	95.9	[74]	Pose	PR-2021
8	MARS	91.5	[43]	Occlusion	ICCV-2021
9	DukeMTMC-Video Re-ID	98.3	[43]	Occlusion	ICCV-2021
10	iLIDS	92.0	[115]	Misalignment	ICCV-2021

5. Discussion & future trends

In this systematic review, 230+ articles are reviewed that were published from January 2015 to October 2021 focused on the challenges faced by person Re-ID. In all these papers, the specified challenges have been addressed and the achieved state-of-the-results have been summarized in this review article. We have grouped the articles into a few categories and have critically analyzed their impact on the obtained results. The limitations along with the datasets used in the published articles have also been reviewed against each challenge. Table 5 presents a comprehensive overview of the state-of-the-art Re-ID results achieved on each dataset, including citations for the venue and corresponding papers that presented the SOTA results.

5.1. Impact of automated person Re-ID on society

Generally, due to the scarcity of security resources as well as the lack of technological advancements in third-world countries, the traditional law and order system does not meet the needs of the public, hence failing to build the people confidence. The automated surveillance systems *i.e.* person Re-ID aims to improve the quality of the lives of common people by providing a sustainable living environment to them. Since the assurance of security and implementation of law and order are the basic needs of human beings, the person Re-ID solutions can assist law enforcement authorities in providing enhanced preemptive security and quick implementation of the law and order. Moreover, in case of any adverse happening, the Re-ID solutions can greatly assist the security officials in rapid response and quick resolution of security issues and can prevent the delays caused by manual video forensic analytics. In order to trace people in a camera network, faces can only be used for recognition only if the subject is close enough and facing towards the camera. But usually, in CCTV footage this is not the case, people are captured in variant poses and viewpoints where their faces are not clear. Therefore, the features found in a person's entire body (like clothing, height, *etc*) are more useful to identify a person across different cameras of the network.

5.2. AI Ethics for person Re-Identification

Person re-identification (re-ID) research raises several ethical concerns due to its potential for misuse and impact on privacy. AI ethics involves taking a multi-faceted approach that considers both the technical and non-technical aspects of the technology.

The research relating to ethical AI for person re-identification is actively evolving, few relevant research dimensions include the exploration of ethical considerations, such as informed consent and data governance, and the use of explainable AI techniques in the context of person re-identification. The explainable AI techniques make the decisions made by person re-identification systems more transparent and understandable.

The privacy-preserving person re-identification methods are recently proposed [259], such as differential privacy and generative adversarial networks, that aim to reduce the risk of privacy violations while still allowing re-identification to occur. RichardWebster et al. [260] proposes a new approach for person re-identification (matching individuals across multiple camera views) that is designed to be more ethical. The authors argue that current person re-identification methods can be used maliciously and raise privacy concerns, so they propose a new method that aims to reduce the amount of information disclosed while still allowing re-identification to occur. The new approach, called "Doppelganger Saliency," adds a random "disturbance" to the person's appearance in each camera view, making it harder to extract and use personal information while still allowing re-identification to be possible.

Chen et al. [261], Goyal et al. [262], Specker [263], Zhao et al. [264], Bekele et al. [265] propose new approaches for person re-identification that aims to be more explainable and interpretable. A new method is proposed in [261] that incorporates information about the person's attributes, such as gender and age, into the re-identification process. The approach uses a "metric distillation" process to transfer knowledge from a teacher network that is trained on attributes to a student network that is responsible for re-identification. In [262], a reinforcement learning agent is used in the deep baseline to guide the training process. The reinforcement learning agent provides feedback on the performance of the deep neural network, which helps to make the re-identification process more explainable.

Specker [263] and Bekele et al. [265] proposed a "ranking mechanism" and incorporated an "explanatory bias" respectively, to provide explanations for the re-identification process. The ranking mechanism and explanatory bias are based on the person's attributes, such as height, weight, and clothing, and it provides information about why a particular person was identified as a match. Another similar approach [264] proposed a new method called Deep Semantic Structured Hashing (DeepSSH) that integrates deep learning with semantic hashing to perform person re-identification. The semantic hashing component provides interpretable explanations for the re-identification process by representing the person's attributes as binary codes.

Moreover, the investigation of the potential for person re-identification to be used for malicious purposes, such as surveillance and targeted advertising, and the development of countermeasures to prevent such abuses is another prospective research dimension.

Recently, Dietlmeier et al. [266] investigated the impact of face blurring on person re-identification performance. This study found that the performance of different state-of-the-art methods remains relatively consistent after anonymization, enabling safe comparisons using anonymized data, hence provided valuable guidance for future Re-ID research. Similarly, Ahmad et al. [267] proposed a new privacy-preserving approach to person re-identification which utilizes event cues rather than facial features. A new dataset, Event-

Driven Person Re-Identification (ED-PRID) is also introduced where the individuals' identities are preserved through blurring.

5.3. Long-term Re-ID by learning dynamic representations

Long-term person re-identification (LT Re-ID) is a challenging problem that is gaining popularity among the research community due to recently introduced large-scale LT Re-ID datasets including the Celeb-ReID dataset [268], long-term cloth changing (LTCC) dataset [269], Cloth-Changing Video Re-ID (CCVID) dataset [270] etc.

Generally, the LT Re-ID has been approached with biometrics-based and data adaptation-based methods. Biometrics-based approaches aim to use features such as motion, body contour/shape, and face to identify individuals, but these approaches are limited by the quality of the footage and difficulties in image segmentation and tracking. Zhang et al. [271] proposed a novel spatial-temporal model that learns a feature representation for each walking tracklet, taking into account the spatial and temporal information, and the skeleton motion. Similarly, [272] also learn soft biometric information of person's gait in addition to the appearance to address long-term Re-ID.

Alternatively, the data adaptation-based approaches use fine-tuning mechanisms to adjust models trained on short-term Re-ID datasets to handle diverse clothing-change cases. However, these methods do not explicitly consider actual clothing status during training, which can lead to sub-optimal performance in no-clothing-change cases. A successful LT Re-ID method should be able to dynamically regulate ID features by correctly sensing clothing status, allowing it to handle clothing-change cases while maintaining discrimination ability in no-clothing-change cases. There is still a need for further research and development to improve LT Re-ID methods, as current approaches are still in preliminary stages and face limitations in practical implementation.

Huang et al. [273] proposed a clothing status awareness Re-ID approach to improve the robustness of ID discriminative features for Long-Term Re-Identification and introduced an additional feature regularization process to handle both the clothing-change cases and no-clothing-change cases, and outperformed existing long-term Re-ID methods for Celeb-ReID dataset. However, there is still a huge room for improvement in the performance of Re-ID.

In addition to proposing the long-term Re-ID dataset [268,274] handled the cloth changing challenge by using vector-neuron capsules instead of scalar neurons by perceiving the cloth changes information of the same person. Additionally, through capsule integration and the use of Soft Embedding Attention (SEA) and Feature Sparse Representation (FSR) mechanisms, the Re-ID performance is boosted[270]. proposed a novel loss function for long-term person Re-ID to learn clothes-irrelevant features by penalizing its predictive power w.r.t. clothing.

Qian et al. [269] proposed a long-term cloth changing (LTCC) Re-ID data set that contains different changes in clothing as well as diverse poses, lighting, and occlusion. Moreover, a task-driven method is proposed to learn identity-sensitive and cloth-insensitive representations by utilizing the relationship between the human key points to extract biological structural features and apply attention mechanisms to disentangle the identity-relevant features from clothing-related information.

Zhang et al. [275] proposed a novel framework for person retrieval and verification, where the verification network compares the query with the initially shortlisted candidate images to obtain similarity scores by contrasting local details, a ranking scheme keeps the balance between retrieval and verification results. [8,276] have provided comprehensive reviews about the long-term Re-ID approaches.

5.4. Deep learning conjecture

Since the evolution of deep learning methods in 2015, the computer vision research community immediately shifted from handcrafted machine learning research to deep learning-based algorithms. Soon after the availability of a very large-scale vision-based dataset *i.e.* ImageNet and its pre-trained weights, like many other research domains, deep person Re-ID solutions came into existence. Meanwhile, medium to large-scale Re-ID benchmarks was proposed so that specialized custom Re-ID solutions could be developed. Beginning from the transfer learning methods using the generic deep architectures like AlexNet, ResNet *etc*, the Re-ID research quickly got independence in designing much more sophisticated solutions due to the availability of large-scale person Re-ID benchmarks. For many years, convolution neural networks served as a strong backbone for Re-ID solutions. The CNNs are used to learn a variety of person representations,*i.e.* the global person representations, local parts-based person representations, semantics-based person representations, attributes driven representations *etc*. The CNN architectures with single stream & branch were common in the start, however, with the passage of time, multi-streamed architectures are proposed for person Re-ID, where each stream targets a different perspective.

Later on, with the development of attention-based mechanisms for vision problems, the same were extensively explored to develop the Re-ID solutions. Most of the attention-based Re-ID solutions are developed on the backbone of CNNs and are multi-steam architectures to capture the various types of attention features *i.e.* spatial attention, temporal attention, channel-wise attention *etc*. The attention-based Re-ID solutions performed really well for various Re-ID challenges in comparison with the methods that do not involve computation of the attention.

Since the deep-learning is still evolving and new SOTA backbone architectures are being developed by the research community with every passing year, it drives the whole research community to new dimensions. Transformers and its variants are SOTA for language problems for a long but due to certain limitations, these were not used for vision problems in a holistic way. Recently, the development of a vision transformer (in the year 2021) was a great breakthrough in vision research and opened up new ways to vision research. The vision transformer outperforms the counterpart CNN baseline deep architectures with a great margin for various vision problems, and this fact has been validated and reported for various similar kinds of computer vision-based research problems [277–280].

Similarly, the use of vision transformer backbones in person Re-ID problems has been on the rise due to its success in capturing contextual information and global features.

This comprehensive study on person Re-ID research unfolds a few interesting aspects for future Re-ID research. While exploring the Re-ID solutions for the foremost common Re-ID challenge,*i.e.*, the occlusion, it has been observed that different perspectives of the spatial and temporal features are learned to capture the dynamic and static information of a person. The temporal features compensate for the occluded spatial regions and enhance the performances of the Re-ID solutions. The top three Re-ID solutions [43,46,47] that optimally Reidentify the occluded persons are all attention-based approaches and learn both the spatial and temporal features. These attention-based mechanisms are strengthened by the use of 3D convolutional architecture, pyramid architecture, and memory units respectively.

The pose and viewpoint variations make the person Re-ID quite challenging, especially in the case of inter-class differences (where people of different classes appear similar under different camera acquisitions) due to the similar appearances. This study highlights that for this particular challenge, the CNN architectures, with and

without attention modules, work well. For instance, among the top performing Re-ID solutions [67,68,68] and [82], only [67] presented the approach which is based on learning attention weights while [66] and [68] employed the higher level semantic information to generate multi-level feature maps and an end-to-end matching & shuffling operations, respectively. Zhang et al. [82] presented a novel Kronecker product matching operation to perform the Re-ID. Generally, these approaches work to handle the misalignment caused by pose variances, by learning global features along with the alignment of local parts based on dynamic feature learning.

The cluttered background, if not efficiently removed or suppressed, inversely affects the performance of the re-ID solution. Since the attention-based mechanism inherently focuses and highlights the attentive parts & regions of a person's image, the top performing Re-ID solutions even in the case of cluttered background also employed attention-based mechanisms to exclude the background information. Generally, these methods address this issue by removing the cluttered background from the foreground person image and by learning the multi-level features. The SOTA and other high performers, [94] and [90,93] learn multi-level attention in different ways and fuse the information learned from various levels and branches. Another top performer [91] is an encoder-decoder foreground aware attention subnetwork that focuses to learn the body part information by constructing the binary mask. Generally, learning the binary masks is common practice to exclude the cluttered background from the person's foreground. However, these approaches do not aid the scenario where the background information might play a vital role in the identification of a person.

The orientation of cameras and different viewing angles result in the misalignment of person images. Generally, to tackle this issue, the local parts associations are built among person images so that various local parts could be aligned with their respective local parts of corresponding images for robust person Re-ID. The same has opted in the top few approaches whether these are CNN architectures with or without the attention module. He et al. [115], encoder-decoder-based architectures, is a hybrid approach that takes the advantage of both CNN and attention-based mechanisms to handle the misalignment efficiently and outperformed all other solutions for video-based Re-ID benchmarks. Aich et al. [46] learns both the static and dynamic features by spatial and temporal factorization branches. Chen et al. [116] is an attention mechanism that handled the misalignment by learning the channel wise and position & spatial information and [113] handled the misalignment by part-guided graph convolution network. Therefore, the local parts & patches based learning in addition to global feature learning aids to address the misalignment issues in person Re-ID.

The differences in scales of captured images is generally seen in CCTV footage due to the variations in the distances of cameras from the targets. Generally, the multi-scale approaches are introduced to learn the person's representations at various scales. Additionally, it is observed that the attention-based approaches perform better than the rest of the Re-ID solutions. The top few solutions either used the multi-scale attention pyramid [138] or divided the image into multiple local parts and then learned the attention [142], or [144] extracted the holistic and local feature maps using multi-scale omni-bearing attention network. The Re-ID solutions that support multi-scale Re-ID learn the person's features at multiple scales through multi-sized convolutional layers or branches. And then aggregate the learned information into the final person descriptor. For video-based benchmarks, the temporal information alleviates the multi-scale Re-ID solutions.

Viewpoint variations occur in the surveillance data as different views of pedestrians are captured across non-overlapping cameras of a surveillance network. In order to address the viewpoint variations, it is needed to maximize the inter-class distance and minimize the intra-class distance, whether it is at the holistic level i.e. image, or the local-parts level. Additionally, handling the image alignment issues aids to address the viewpoint differences. The top-performing approaches opt the similar kinds of schemes. Lian et al. [178] presented a channel-wise multi-scale attention-aware mechanism using triplet loss to learn the invariant views obtained from different cameras. The Re-ID solution [180] handled the viewpoint differences in a unique way by involving a lightweight and labeled part segmentation head to the backbone of Re-ID during the training process and obtained a diverse set of features. Similarly, in [151], multiple views are learned representing the target object to produce better results.

Since the surveillance cameras work 247 and capture the person images from a distance, generally low image resolution is observed in CCTV footage. This results in another Re-ID challenge i.e. to identify a person correctly in low-resolution images. Although, handling low resolution is mostly the prime feature of almost all Re-ID solutions, however, few Re-ID solutions focus on this problem explicitly by opting for specialized architectures for this. Feng et al. [187] proposed a resolution-aware Re-ID framework that works very well and follows the teacher-student learning mechanism. In another technique, [181] effective embedding is combined at multiple layers using skip connections resulting to obtain the unique features. In [188] an association scheme is learned from training data having super-resolution images, this has helped to obtain the discriminative features even from low-resolution images. Recently, [187] proposed a resolution-aware Re-ID framework that works very well and follows the teacher-student learning mechanism.

Lastly, the cross-domain person Re-ID is quite challenging with a huge room for improvement. Due to numerous Re-ID challenges, already discussed in previous sections, it is difficult to re-identify the people of totally different camera networks. A well-generalized Re-ID solution is highly desirable for cross-domain Re-ID. Few top-ranked Re-ID solutions emphasized on integrating variant modalities in the default input features or presented the synthetic data and introduced transfer knowledge techniques to generalize their solution for unseen domain experiments. Fu et al. [256] introduced generalization in the Re-ID solution by developing an attention-based aggregation formulation, which flexibly incorporates the similarity metrics along with multiple representations. Fu et al. [234] introduced the person Re-ID specific pre-training framework for the first time and enhanced Re-ID by data augmentation and temperature usage in a contrastive learning framework and proposed an unlabeled Re-ID dataset named "LUPerson".

In [250], the quality of attention maps is measured at both spatial and channel levels through reinforcement learning mechanisms. To address this problem, a significant diversity is observed in the Re-ID solutions.

5.5. Concluding remarks

Generally, the Re-ID solutions address multiple Re-ID challenges simultaneously, however, this study highlights the specialized schemes of the deep architectures, designed to address the particular Re-ID challenge(s). The detailed specifications are discussed in the respective section as well. On the basis of deep analysis, it can be emphasized that attention-based Re-ID solutions are gaining more interest in the research community with their promising performances. Moreover, since the self-attention-based vision research is in its beginning phase, the full strength of these approaches is yet to be researched and analyzed. Additionally, this study emphasized the ongoing research for AI ethics for person Re-ID.

During recent years, the Re-ID research is at its best for a few initially proposed Re-ID datasets *i.e.* Market1501, DukeMTMC-Re-ID *etc*, which were captured from the public places with a controlled environment, hence do not depict the real world scenario. The customized Re-ID algorithms addressed most of the Re-ID challenges effectively due to the medium level of complexity for these benchmarks. Among various Re-ID challenges, the pose variations remained most popular among the research community, as it is the most common Re-ID issue with a significant impact on the performance of Re-ID solutions. The newly proposed Re-ID benchmarks (*i.e.* MSMT17 *etc*), captured from the complex scenes with a large number of indoor and outdoor cameras and closer to the real-world complex scenarios, therefore need more sophisticated Re-ID solutions to solve the real world problems.

Since the start of the Re-ID research era, the majority of Re-ID solutions propose their effectiveness using various Re-ID benchmarks, however, the evaluations are performed using the same domain benchmarks for their unseen identities. In contrast, rare Re-ID solutions exist in the literature that could perform better for the cross-domain Re-ID benchmarks, especially for the recently proposed complex and large-scale Re-ID benchmarks, hence this needs more sophisticated research solutions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Z. Wang, R. Hu, C. Liang, Y. Yu, J. Jiang, M. Ye, J. Chen, Q. Leng, Zero-shot person re-identification via cross-view consistency, IEEE Trans. Multimedia 18 (2) (2015) 260–272.

[2] Y. Wang, J. Shen, S. Petridis, M. Pantic, A real-time and unsupervised face re-identification system for human-robot interaction, Pattern Recognit. Lett. 128 (2019) 559–568.

[3] H. Wang, S. Gong, X. Zhu, T. Xiang, Human-in-the-loop person re-identification, in: European Conference on Computer Vision, Springer, 2016, pp. 405–422.

[4] K. Wang, H. Wang, M. Liu, X. Xing, T. Han, Survey on person re-identification based on deep learning, CAAI Trans. Intell. Technol. 3 (4) (2018) 219–227.

[5] Z. Ming, M. Zhu, X. Wang, J. Zhu, J. Cheng, C. Gao, Y. Yang, X. Wei, Deep learning-based person re-identification methods: a survey and outlook of recent works, Image Vis. Comput. 119 (2022) 104394.

[6] M.O. Almasawa, L.A. Elrefaei, K. Moria, A survey on deep learning-based person re-identification systems, IEEE Access 7 (2019) 175228–175247, doi:10.1109/ACCESS.2019.2957336.

[7] D. Wu, S.-J. Zheng, X.-P. Zhang, C.-A. Yuan, F. Cheng, Y. Zhao, Y.-J. Lin, Z.-Q. Zhao, Y.-L. Jiang, D.-S. Huang, Deep learning-based methods for person re-identification: a comprehensive review, Neurocomputing 337 (2019) 354–371.

[8] Q. Leng, M. Ye, Q. Tian, A survey of open-world person re-identification, IEEE Trans. Circuits Syst. Video Technol. 30 (4) (2019) 1092–1108.

[9] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, S.C. Hoi, Deep learning for person re-identification: a survey and outlook, IEEE Trans. Pattern Anal. Mach. Intell. 44 (6) (2021) 2872–2893.

[10] M. Liberati, J. Tetzlaff, D. Altman, Group tp. preferred reporting items for systematic reviews and meta analyses: the prisma statement, PLoS Med. 6 (7) (2009) 1–6.

[11] S. Xu, J. Wang, W. Shou, T. Ngo, A.-M. Sadick, X. Wang, Computer vision techniques in construction: a critical review, Arch. Comput. Methods Eng. 28 (2021) 3383–3397.

[12] S. Katoch, S.S. Chauhan, V. Kumar, A review on genetic algorithm: past, present, and future, Multimed. Tools Appl. 80 (2021) 8091–8126.

[13] A. Budrionis, D. Plikynas, P. Daniušis, A. Indrulionis, Smartphone-based computer vision travelling aids for blind and visually impaired individuals: a systematic review, Assistive Technol. 34 (2) (2022) 178–194.

[14] Exposing.ai, Duke mtmc, 2019, https://exposing.ai/duke_mtmc/, Last accessed on 2023-04-07.

[15] Financial Times, Whos using your face? The ugly truth about facial recognition, 2019,. https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e, Last accessed on 2023-04-07.

[16] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: a benchmark, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1116–1124.

[17] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: European Conference on Computer Vision, Springer, 2016, pp. 17–35.

[18] J. Miao, Y. Wu, P. Liu, Y. Ding, Y. Yang, Pose-guided feature alignment for occluded person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 542–551.

[19] W. Li, R. Zhao, X. Wang, Human reidentification with transferred metric learning, in: Asian Conference on Computer Vision, Springer, 2012, pp. 31–44.

[20] W. Li, R. Zhao, T. Xiao, X. Wang, DeepReID: deep filter pairing neural network for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152–159.

[21] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, in: Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), Vol. 3, Citeseer, 2007, pp. 1–7.

[22] M. Hirzer, C. Beleznai, P.M. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification, in: Scandinavian Conference on Image Analysis, Springer, 2011, pp. 91–102.

[23] T. Wang, S. Gong, X. Zhu, S. Wang, Person re-identification by video ranking, in: European Conference on Computer Vision, Springer, 2014, pp. 688–703.

[24] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, Q. Tian, MARS: a video benchmark for large-scale person re-identification, in: European Conference on Computer Vision, Springer, 2016, pp. 868–884.

[25] L. Wei, S. Zhang, W. Gao, Q. Tian, Person transfer GAN to bridge domain gap for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 79–88.

[26] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2009) 1627–1645.

[27] A. Dehghan, S. Modiri Assari, M. Shah, GMMCP tracker: globally optimal generalized maximum multi clique problem for multiple object tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4091–4099.

[28] W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by probabilistic relative distance comparison, in: CVPR 2011, 2011, pp. 649–656, doi:10.1109/CVPR.2011.5995598.

[29] Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline), in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 480–496.

[30] L. He, Y. Wang, W. Liu, H. Zhao, Z. Sun, J. Feng, Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 8450–8459.

[31] Z. Ma, Y. Zhao, J. Li, Pose-guided inter-and intra-part relational transformer for occluded person re-identification, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 1487–1496.

[32] M. Jia, X. Cheng, Y. Zhai, S. Lu, S. Ma, Y. Tian, J. Zhang, Matching on sets: conquer occluded person re-identification without alignment, in: Proc. AAAI Conf. Artif. Intell., 2021, pp. 1673–1681.

[33] L. He, W. Liu, Guided saliency feature learning for person re-identification in crowded scenes, in: European Conference on Computer Vision, Springer, 2020, pp. 357–373.

[34] G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou, J. Sun, High-order information matters: learning relation and topology for occluded person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6449–6458.

[35] C. Yan, G. Pang, J. Jiao, X. Bai, X. Feng, C. Shen, Occluded person re-identification with single-scale global representations, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11875–11884.

[36] J. Yang, J. Zhang, F. Yu, X. Jiang, M. Zhang, X. Sun, Y.-C. Chen, W.-S. Zheng, Learning to know where to see: a visibility-aware approach for occluded person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11885–11894.

[37] C. Zhao, X. Lv, S. Dou, S. Zhang, J. Wu, L. Wang, Incremental generative occlusion adversarial suppression network for person ReID, IEEE Trans. Image Process. 30 (2021) 4212–4224.

[38] Hou Ruibing, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, Xilin Chen, Feature completion for occluded person re-identification, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2021) 4894–4912.

[39] X. Zhang, Y. Yan, J. H. Xue, Y. Hua, H. Wang, Semantic-aware occlusion-robust network for occluded person re-identification, IEEE Transactions on Circuits and Systems for Video Technology 31 (2020) 2764–2778.

[40] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, S. Gong, Partial person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4678–4686.

[41] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, X. Chen, VRSTC: occlusion-free video person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7183–7192.

[42] S. Zhao, C. Gao, J. Zhang, H. Cheng, C. Han, X. Jiang, X. Guo, W.-S. Zheng, N. Sang, X. Sun, Do not disturb me: Person re-identification under the interference of other pedestrians, in: European Conference on Computer Vision, Springer, 2020, pp. 647–663.
[43] Y. Wang, P. Zhang, S. Gao, X. Geng, H. Lu, D. Wang, Pyramid spatial-temporal aggregation for video-based person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 12026–12035.
[44] R. Zhang, J. Li, H. Sun, Y. Ge, P. Luo, X. Wang, L. Lin, SCAN: self-and-collaborative attention network for video person re-identification, IEEE Trans. Image Process. 28 (10) (2019) 4870–4882.
[45] G. Chen, J. Lu, M. Yang, J. Zhou, Spatial-temporal attention-aware learning for video-based person re-identification, IEEE Trans. Image Process. 28 (9) (2019) 4192–4205.
[46] A. Aich, M. Zheng, S. Karanam, T. Chen, A.K. Roy-Chowdhury, Z. Wu, Spatio-temporal representation factorization for video-based person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 152–162.
[47] C. Eom, G. Lee, J. Lee, B. Ham, Video-based person re-identification with spatial and temporal memory networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 12036–12045.
[48] Y. Li, J. He, T. Zhang, X. Liu, Y. Zhang, F. Wu, Diverse part discovery: occluded person re-identification with part-aware transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2898–2907.
[49] L. Wei, S. Zhang, H. Yao, W. Gao, Q. Tian, GLAD: global-local-alignment descriptor for pedestrian retrieval, in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 420–428.
[50] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, X. Chen, Interaction-and-aggregation network for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9317–9326.
[51] G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson, K. Murphy, PersonLab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 269–286.
[52] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, Q. Tian, Pose-driven deep convolutional model for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3960–3969.
[53] M. Kocabas, S. Karagoz, E. Akbas, MultiPoseNet: fast multi-person pose estimation using pose residual network, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 417–433.
[54] Y.-J. Cho, K.-J. Yoon, Improving person re-identification via pose-aware multi-shot matching, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1354–1362.
[55] R. Quispe, H. Pedrini, Improved person re-identification based on saliency and semantic parsing with deep neural network models, Image Vis. Comput. 92 (2019) 103809.
[56] Y. Chen, S. Duffner, A. Stoian, J.-Y. Dufour, A. Baskurt, Deep and low-level feature based attribute learning for person re-identification, Image Vis. Comput. 79 (2018) 25–34.
[57] C. Su, S. Zhang, J. Xing, W. Gao, Q. Tian, Multi-type attributes driven multi–camera person re-identification, Pattern Recognit. 75 (2018) 77–89.
[58] N. McLaughlin, J.M. del Rincon, P.C. Miller, Person reidentification using deep convnets with multitask learning, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2016) 525–539.
[59] J. Wang, Z. Wang, C. Gao, N. Sang, R. Huang, DeepList: learning deep features with adaptive listwise constraint for person reidentification, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2016) 513–524.
[60] S. Tan, F. Zheng, L. Liu, J. Han, L. Shao, Dense invariant feature-based support vector ranking for cross-camera person reidentification, IEEE Trans. Circuits Syst. Video Technol. 28 (2) (2016) 356–363.
[61] I.B. Barbosa, M. Cristani, B. Caputo, A. Rognhaugen, T. Theoharis, Looking beyond appearances: synthetic training data for deep CNNs in re-identification, Comput. Vision Image Understanding 167 (2018) 50–62.
[62] Y.-G. Lee, S.-C. Chen, J.-N. Hwang, Y.-P. Hung, An ensemble of invariant features for person reidentification, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2016) 470–483.
[63] J. Li, S. Zhang, Q. Tian, M. Wang, W. Gao, Pose-guided representation learning for person re-identification, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2) (2019) 622–635.
[64] Z. Li, J. Lv, Y. Chen, J. Yuan, Person re-identification with part prediction alignment, Comput. Vision Image Understanding 205 (2021) 103172.
[65] J. Yin, A. Wu, W.-S. Zheng, Fine-grained person re-identification, Int. J. Comput. Vis. 128 (6) (2020) 1654–1672.
[66] Q. Zhou, B. Zhong, X. Lan, G. Sun, Y. Zhang, B. Zhang, R. Ji, Fine-grained spatial alignment model for person re-identification with focal triplet loss, IEEE Trans. Image Process. 29 (2020) 7578–7589.
[67] P. Wang, Z. Zhao, F. Su, X. Zu, N.V. Boulgouris, HOReID: deep high-order mapping enhances pose alignment for person re-identification, IEEE Trans. Image Process. 30 (2021) 2908–2922.
[68] Y. Shen, T. Xiao, S. Yi, D. Chen, X. Wang, H. Li, Person re-identification with deep kronecker-product matching and group-shuffling random walk, IEEE Trans. Pattern Anal. Mach. Intell. 43 (5) (2021) 1649–1665.
[69] M. Saquib Sarfraz, A. Schumann, A. Eberle, R. Stiefelhagen, A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 420–429.
[70] L. Zheng, Y. Huang, H. Lu, Y. Yang, Pose-invariant embedding for deep person re-identification, IEEE Trans. Image Process. 28 (9) (2019) 4500–4509.
[71] J. Liu, B. Ni, Y. Yan, P. Zhou, S. Cheng, J. Hu, Pose transferrable person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4099–4108.
[72] X. Nie, J. Feng, J. Xing, S. Yan, Pose partition networks for multi-person pose estimation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 684–699.
[73] Y. Wu, O.E.F. Bourahla, X. Li, F. Wu, Q. Tian, X. Zhou, Adaptive graph representation learning for video person re-identification, IEEE Trans. Image Process. 29 (2020) 8821–8830.
[74] X. Hu, D. Wei, Z. Wang, J. Shen, H. Ren, Hypergraph video pedestrian re-identification based on posture structure relationship and action constraints, Pattern Recognit. 111 (2021) 107688.
[75] D. Chen, Z. Yuan, B. Chen, N. Zheng, Similarity learning with spatial constraints for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1268–1277.
[76] D. Chen, Z. Yuan, G. Hua, N. Zheng, J. Wang, Similarity learning on an explicit polynomial kernel feature map for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1565–1573.
[77] L. An, M. Kafai, S. Yang, B. Bhanu, Person reidentification with reference descriptor, IEEE Trans. Circuits Syst. Video Technol. 26 (4) (2015) 776–787.
[78] Y. Sun, L. Zheng, Y. Li, Y. Yang, Q. Tian, S. Wang, Learning part-based convolutional features for person re-identification, IEEE Trans. Pattern Anal. Mach. Intell. (2019).
[79] Y.-J. Cho, K.-J. Yoon, PaMM: pose-aware multi-shot matching for improving person re-identification, IEEE Trans. Image Process. 27 (8) (2018) 3739–3752.
[80] W. Yang, H. Huang, Z. Zhang, X. Chen, K. Huang, S. Zhang, Towards rich feature discovery with class activation maps augmentation for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1389–1398.
[81] J. Xu, R. Zhao, F. Zhu, H. Wang, W. Ouyang, Attention-aware compositional network for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2119–2128.
[82] Z. Zhang, C. Lan, W. Zeng, X. Jin, Z. Chen, Relation-aware global attention for person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3186–3195.
[83] H. Liu, J. Feng, M. Qi, J. Jiang, S. Yan, End-to-end comparative attention networks for person re-identification, IEEE Trans. Image Process. 26 (7) (2017) 3492–3506.
[84] S. Gao, J. Wang, H. Lu, Z. Liu, Pose-guided visible part matching for occluded person ReID, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11744–11752.
[85] M. Tian, S. Yi, H. Li, S. Li, X. Zhang, J. Shi, J. Yan, X. Wang, Eliminating background-bias for robust person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5794–5803.
[86] T.T.T. Pham, T.-L. Le, H. Vu, T.K. Dao, et al., Fully-automated person re-identification in multi-camera surveillance system with a robust kernel descriptor and effective shadow removal method, Image Vis. Comput. 59 (2017) 44–62.
[87] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, Y. Xu, Deep-person: learning discriminative deep features for person re-identification, Pattern Recognit. 98 (2020) 107036.
[88] S. Zhou, J. Wang, D. Meng, X. Xin, Y. Li, Y. Gong, N. Zheng, Deep self-paced learning for person re-identification, Pattern Recognit. 76 (2018) 739–751.
[89] S. Li, W. Song, Z. Fang, J. Shi, A. Hao, Q. Zhao, H. Qin, Long-short temporal-spatial clues excited network for robust person re-identification, Int. J. Comput. Vis. 128 (12) (2020) 2936–2961.
[90] S. Zhou, F. Wang, Z. Huang, J. Wang, Discriminative feature learning with consistent attention regularization for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 8040–8049.
[91] S. Zhou, J. Wang, D. Meng, Y. Liang, Y. Gong, N. Zheng, Discriminative feature learning with foreground attention for person re-identification, IEEE Trans. Image Process. 28 (9) (2019) 4671–4684.
[92] Y. Liu, W. Zhou, J. Liu, G.-J. Qi, Q. Tian, H. Li, An end-to-end foreground-aware network for person re-identification, IEEE Trans. Image Process. 30 (2021) 2060–2071.
[93] J. Sun, Y. Li, H. Chen, B. Zhang, J. Zhu, MEMF: multi-level-attention embedding and multi-layer-feature fusion model for person re-identification, Pattern Recognit. 116 (2021) 107937.
[94] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, S. Tian, Feature refinement and filter network for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 31 (9) (2020) 3391–3402.
[95] Z. Zhang, C. Lan, W. Zeng, Z. Chen, Densely semantically aligned person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 667–676.
[96] Y. Suh, J. Wang, S. Tang, T. Mei, K. Mu Lee, Part-aligned bilinear representations for person re-identification, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 402–419.
[97] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, X. Tang, Spindle net: person re-identification with human body region guided feature decomposition and fusion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1077–1085.

[98] K. Zhu, H. Guo, Z. Liu, M. Tang, J. Wang, Identity-guided human semantic parsing for person re-identification, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, Springer, 2020, pp. 346–363.
[99] Y. Sun, Q. Xu, Y. Li, C. Zhang, Y. Li, S. Wang, J. Sun, Perceive where to focus: learning visibility-aware part-level features for partial person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 393–402.
[100] M.M. Kalayeh, E. Basaran, M. Gökmen, M.E. Kamasak, M. Shah, Human semantic parsing for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1062–1071.
[101] H. Luo, W. Jiang, X. Zhang, X. Fan, J. Qian, C. Zhang, AlignedReID++: dynamically matching local information for person re-identification, Pattern Recognit. 94 (2019) 53–61.
[102] R. Zhao, W. Oyang, X. Wang, Person re-identification by saliency learning, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2) (2016) 356–370.
[103] S. Li, M. Shao, Y. Fu, Person re-identification by cross-view multi-level dictionary learning, IEEE Trans. Pattern Anal. Mach. Intell. 40 (12) (2017) 2963–2977.
[104] G. Wang, Y. Yuan, J. Li, S. Ge, X. Zhou, Receptive multi-granularity representation for person re-identification, IEEE Trans. Image Process. 29 (2020) 6096–6109.
[105] C. Ding, K. Wang, P. Wang, D. Tao, Multi-task learning with coarse priors for robust part-aware person re-identification, IEEE Trans. Pattern Anal. Mach. Intell. 44 (3) (2020) 1474–1488.
[106] S. Li, H. Yu, R. Hu, Attributes-aided part detection and refinement for person re-identification, Pattern Recognit. 97 (2020) 107016.
[107] J. Dai, P. Zhang, D. Wang, H. Lu, H. Wang, Video person re-identification by temporal residual learning, IEEE Trans. Image Process. 28 (3) (2018) 1366–1377.
[108] W. Shi, H. Liu, M. Liu, Image-to-video person re-identification using three-dimensional semantic appearance alignment and cross-modal interactive learning, Pattern Recognit. 122 (2022) 108314.
[109] X. Ma, X. Zhu, S. Gong, X. Xie, J. Hu, K.-M. Lam, Y. Zhong, Person re-identification by unsupervised video matching, Pattern Recognit. 65 (2017) 197–210.
[110] S. Bąk, P. Carr, Deep deformable patch metric learning for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 28 (10) (2017) 2690–2702.
[111] K. Liu, B. Ma, W. Zhang, R. Huang, A spatio-temporal appearance representation for video-based pedestrian re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3810–3818.
[112] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, J. Wang, Person re-identification with correspondence structure learning, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3200–3208.
[113] Z. Zhang, H. Zhang, S. Liu, Y. Xie, T.S. Durrani, Part-guided graph convolution networks for person re-identification, Pattern Recognit. 120 (2021) 108155.
[114] W. Li, X. Zhu, S. Gong, Harmonious attention network for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2285–2294.
[115] T. He, X. Jin, X. Shen, J. Huang, Z. Chen, X.-S. Hua, Dense interaction learning for video-based person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1490–1501.
[116] T. Chen, S. Ding, J. Xie, Y. Yuan, W. Chen, Y. Yang, Z. Ren, Z. Wang, Abd-net: Attentive but diverse person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 8351–8361.
[117] P. Fang, J. Zhou, S.K. Roy, L. Petersson, M. Harandi, Bilinear attention networks for person retrieval, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 8030–8039.
[118] J. Guo, Y. Yuan, L. Huang, C. Zhang, J.-G. Yao, K. Han, Beyond human parts: Dual part-aligned representations for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 3642–3651.
[119] C. Wang, Q. Zhang, C. Huang, W. Liu, X. Wang, Mancs: A multi-task attentional network with curriculum sampling for person re-identification, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 365–381.
[120] L. Zhao, X. Li, Y. Zhuang, J. Wang, Deeply-learned part-aligned representations for person re-identification, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 3219–3228.
[121] F. Yang, K. Yan, S. Lu, H. Jia, X. Xie, W. Gao, Attention driven person re-identification, Pattern Recognit. 86 (2019) 143–155.
[122] K. Wang, C. Ding, S.J. Maybank, D. Tao, Cdpm: convolutional deformable part models for semantically aligned person re-identification, IEEE Trans. Image Process. 29 (2019) 3416–3428.
[123] K. Wang, P. Wang, C. Ding, D. Tao, Batch coherence-driven network for part-aware person re-identification, IEEE Trans. Image Process. 30 (2021) 3405–3418.
[124] Z. Zhang, Y. Xie, D. Li, W. Zhang, Q. Tian, Learning to align via wasserstein for person re-identification, IEEE Trans. Image Process. 29 (2020) 7104–7116.
[125] Z. Zheng, L. Zheng, Y. Yang, Pedestrian alignment network for large-scale person re-identification, IEEE Trans. Circuits Syst. Video Technol. 29 (10) (2018) 3037–3045.
[126] X. Gu, H. Chang, B. Ma, H. Zhang, X. Chen, Appearance-preserving 3d convolution for video-based person re-identification, in: European Conference on Computer Vision, Springer, 2020, pp. 228–243.
[127] X. Qian, Y. Fu, Y.-G. Jiang, T. Xiang, X. Xue, Multi-scale deep learning architectures for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5399–5408.
[128] F. Zheng, C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang, R. Ji, Pyramidal person re-identification via multi-loss dynamic training, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8514–8522.
[129] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Omni-scale feature learning for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 3702–3712.
[130] Y. Chen, X. Zhu, S. Gong, Person re-identification by deep learning multi-scale representations, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 2590–2600.
[131] J. Liu, Z.-J. Zha, W. Wu, K. Zheng, Q. Sun, Spatial-temporal correlation and topology learning for person re-identification in videos, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4370–4379.
[132] J. Li, S. Zhang, T. Huang, Multi-scale temporal cues learning for video person re-identification, IEEE Trans. Image Process. 29 (2020) 4461–4473.
[133] C. Zhao, X. Wang, W. Zuo, F. Shen, L. Shao, D. Miao, Similarity learning with joint transfer constraints for person re-identification, Pattern Recognit. 97 (2020) 107014.
[134] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Learning generalisable omni-scale representations for person re-identification, IEEE Trans. Pattern Anal. Mach. Intell. (2021).
[135] A. Wu, W.-S. Zheng, X. Guo, J.-H. Lai, Distilled person re-identification: Towards a more scalable system, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1187–1196.
[136] Y. Guo, N.-M. Cheung, Efficient and deep person re-identification using multi-level similarity, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2335–2344.
[137] C. Yan, G. Pang, L. Wang, J. Jiao, X. Feng, C. Shen, J. Li, Bv-person: A large-scale dataset for bird-view person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10943–10952.
[138] X. Qian, Y. Fu, T. Xiang, Y. Jiang, X. Xue, Leader-based multi-scale attention deep architecture for person re-identification, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 371.
[139] W. Zhang, X. He, X. Yu, W. Lu, Z. Zha, Q. Tian, A multi-scale spatial-temporal attention model for person re-identification in videos, IEEE Trans. Image Process. 29 (2019) 3365–3373.
[140] X. Yang, L. Liu, N. Wang, X. Gao, A two-stream dynamic pyramid representation model for video-based person re-identification, IEEE Trans. Image Process. 30 (2021) 6266–6276.
[141] N. Martinel, G.L. Foresti, C. Micheloni, Deep pyramidal pooling with attention for person re-identification, IEEE Trans. Image Process. 29 (2020) 7306–7316.
[142] G. Chen, T. Gu, J. Lu, J.-A. Bao, J. Zhou, Person re-identification via attention pyramid, IEEE Trans. Image Process. 30 (2021) 7663–7676.
[143] Y. Zhong, Y. Wang, S. Zhang, Progressive feature enhancement for person re-identification, IEEE Trans. Image Process. 30 (2021) 8384–8395.
[144] Y. Huang, S. Lian, H. Hu, D. Chen, T. Su, Multiscale omnibearing attention networks for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 31 (5) (2020) 1790–1803.
[145] X. Sun, L. Zheng, Dissecting person re-identification from the viewpoint of viewpoint, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 608–617.
[146] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[147] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[148] A. Subramaniam, M. Chatterjee, A. Mittal, Deep neural networks with inexact matching for person re-identification, Adv. Neural Inf. Process. Syst. 29 (2016) 1–9.
[149] S. Karanam, Y. Li, R.J. Radke, Person re-identification with discriminatively trained viewpoint invariant dictionaries, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4516–4524.
[150] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
[151] A. Porrello, L. Bergamini, S. Calderara, Robust re-identification by multiple views knowledge distillation, in: European Conference on Computer Vision, Springer, 2020, pp. 93–110.
[152] J. Dai, Y. Zhang, H. Lu, H. Wang, Cross-view semantic projection learning for person re-identification, Pattern Recognit. 75 (2018) 63–76.
[153] Y.-C. Chen, W.-S. Zheng, J.-H. Lai, P.C. Yuen, An asymmetric distance model for cross-view feature mapping in person reidentification, IEEE Trans. Circuits Syst. Video Technol. 27 (8) (2016) 1661–1675.
[154] Z. Wu, Y. Li, R. Radke, Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features, IEEE Trans. Pattern Anal. Mach. Intell. 37 (5) (2015) 1095.
[155] Y.-C. Chen, X. Zhu, W.-S. Zheng, J.-H. Lai, Person re-identification by camera correlation aware feature augmentation, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2) (2018) 392–408.
[156] Z. Feng, J. Lai, X. Xie, Learning view-specific deep networks for person re-identification, IEEE Trans. Image Process. 27 (7) (2018) 3472–3483.

[157] S.-Z. Chen, C.-C. Guo, J.-H. Lai, Deep ranking for person re-identification via joint representation learning, IEEE Trans. Image Process. 25 (5) (2016) 2353–2367.
[158] J. Jia, Q. Ruan, G. An, Y. Jin, Multiple metric learning with query adaptive weights and multi-task re-weighting for person re-identification, Comput. Vision Image Understanding 160 (2017) 87–99.
[159] L. Wu, Y. Wang, Z. Ge, Q. Hu, X. Li, Structured deep hashing with convolutional neural networks for fast person re-identification, Comput. Vision Image Understanding 167 (2018) 63–73.
[160] X. Chen, X. Zheng, X. Lu, Bidirectional interaction network for person re-identification, IEEE Trans. Image Process. 30 (2021) 1935–1948.
[161] A. Borgia, Y. Hua, E. Kodirov, N.M. Robertson, Cross-view discriminative feature learning for person re-identification, IEEE Trans. Image Process. 27 (11) (2018) 5338–5349.
[162] J. Garcia, N. Martinel, A. Gardel, I. Bravo, G.L. Foresti, C. Micheloni, Discriminant context information analysis for post-ranking person re-identification, IEEE Trans. Image Process. 26 (4) (2017) 1650–1665.
[163] L. Wu, Y. Wang, H. Yin, M. Wang, L. Shao, Few-shot deep adversarial learning for video-based person re-identification, IEEE Trans. Image Process. 29 (2020) 1233–1245.
[164] Y. Lin, Y. Wu, C. Yan, M. Xu, Y. Yang, Unsupervised person re-identification via cross-camera similarity exploration, IEEE Trans. Image Process. 29 (2020) 5481–5490.
[165] J. Meng, A. Wu, W.-S. Zheng, Deep asymmetric video-based person re-identification, Pattern Recognit. 93 (2019) 430–441.
[166] C. Zhao, X. Wang, D. Miao, H. Wang, W. Zheng, Y. Xu, D. Zhang, Maximal granularity structure and generalized multi-view discriminant analysis for person re-identification, Pattern Recognit. 79 (2018) 79–96.
[167] J. Jia, Q. Ruan, Y. Jin, G. An, S. Ge, View-specific subspace learning and re-ranking for semi-supervised person re-identification, Pattern Recognit. 108 (2020) 107568.
[168] H.-M. Hu, W. Fang, B. Li, Q. Tian, An adaptive multi-projection metric learning for person re-identification across non-overlapping cameras, IEEE Trans. Circuits Syst. Video Technol. 29 (9) (2019) 2809–2821.
[169] L. Wu, R. Hong, Y. Wang, M. Wang, Cross-entropy adversarial view adaptation for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 30 (7) (2020) 2081–2092.
[170] D. Tao, Y. Guo, B. Yu, J. Pang, Z. Yu, Deep multi-view feature learning for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 28 (10) (2017) 2657–2666.
[171] X. Wang, R. Panda, M. Liu, Y. Wang, A.K. Roy-Chowdhury, Exploiting global camera network constraints for unsupervised video person re-identification, IEEE Trans. Circuits Syst. Video Technol 10 (2020) 4020–4030.
[172] W. Zhang, Y. Li, W. Lu, X. Xu, Z. Liu, X. Ji, Learning intra-video difference for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 29 (10) (2018) 3028–3036.
[173] L. An, Z. Qin, X. Chen, S. Yang, Multi-level common space learning for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 28 (8) (2018) 1777–1787.
[174] H.-X. Yu, A. Wu, W.-S. Zheng, Cross-view asymmetric metric learning for unsupervised person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 994–1002.
[175] H.-X. Yu, A. Wu, W.-S. Zheng, Unsupervised person re-identification by deep asymmetric metric embedding, IEEE Trans. Pattern Anal. Mach. Intell. 42 (4) (2020) 956–973.
[176] M. Zheng, S. Karanam, Z. Wu, R.J. Radke, Re-identification with consistent attentive siamese networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5735–5744.
[177] L. Zhang, F. Liu, D. Zhang, Adversarial view confusion feature learning for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 31 (4) (2021) 1490–1502.
[178] S. Lian, W. Jiang, H. Hu, Attention-aligned network for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 8 (2020) 3140–3153.
[179] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, F. Bremond, Joint generative and contrastive learning for unsupervised person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2004–2013.
[180] H. Huang, W. Yang, J. Lin, G. Huang, J. Xu, G. Wang, X. Chen, K. Huang, Improve person re-identification with part awareness learning, IEEE Trans. Image Process. 29 (2020) 7468–7481.
[181] Y. Wang, L. Wang, Y. You, X. Zou, V. Chen, S. Li, G. Huang, B. Hariharan, K.Q. Weinberger, Resource aware person re-identification across multiple resolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8042–8051.
[182] X. Li, W.-S. Zheng, X. Wang, T. Xiang, S. Gong, Multi-scale learning for low-resolution person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3765–3773.
[183] X.-Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, B. Xu, Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 695–704.
[184] X. Jing, X. Zhu, F. Wu, R. Hu, X. You, Y. Wang, H. Feng, J. Yang, Super-resolution Person re-identification with semi-coupled low-rank discriminant dictionary learning, IEEE Trans. Image Process. 26 (3) (2017) 1363–1378. A Publication of the IEEE Signal Processing Society
[185] K. Han, Y. Huang, Z. Chen, L. Wang, T. Tan, Prediction and recovery for adaptive low-resolution person re-identification, in: European Conference on Computer Vision, Springer, 2020, pp. 193–209.
[186] K. Han, Y. Huang, C. Song, L. Wang, T. Tan, Adaptive super-resolution for person re-identification with low-resolution images, Pattern Recognit. (2020) 107682.
[187] Z. Feng, J. Lai, X. Xie, Resolution-aware knowledge distillation for efficient inference, IEEE Trans. Image Process. 30 (2021) 6985–6996.
[188] Z. Cheng, Q. Dong, S. Gong, X. Zhu, Inter-task association critic for cross-resolution person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2605–2615.
[189] Y. Huang, Z.-J. Zha, X. Fu, R. Hong, L. Li, Real-world person re-identification via degradation invariance learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14084–14094.
[190] Y. Ge, D. Chen, H. Li, Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification, in: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020, OpenReview.net, 2020.
[191] Y. Dai, J. Liu, Y. Bai, Z. Tong, L.-Y. Duan, Dual-refinement: joint label and feature refinement for unsupervised domain adaptive person re-identification, IEEE Trans. Image Process. 30 (2021) 7815–7829.
[192] Y. Zheng, S. Tang, G. Teng, Y. Ge, K. Liu, J. Qin, D. Qi, D. Chen, Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 8371–8381.
[193] Y. Huang, Q. Wu, J. Xu, Y. Zhong, Z. Zhang, Unsupervised domain adaptation with background shift mitigating for person re-identification, Int. J. Comput. Vis. 129 (7) (2021) 2244–2263.
[194] S. Xuan, S. Zhang, Intra-inter camera similarity for unsupervised person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 11926–11935.
[195] K. Zheng, W. Liu, L. He, T. Mei, J. Luo, Z.-J. Zha, Group-aware label transfer for domain adaptive person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 5310–5319.
[196] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, Y. Tian, Unsupervised cross-dataset transfer learning for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1306–1315.
[197] C. Su, S. Zhang, J. Xing, W. Gao, Q. Tian, Deep attributes driven multi-camera person re-identification, in: European Conference on Computer Vision, Springer, 2016, pp. 475–491.
[198] H. Li, Z. Kuang, Z. Yu, J. Luo, Structure alignment of attributes and visual features for cross-dataset person re-identification, Pattern Recognit. 106 (2020) 107414.
[199] Z. Shi, T.M. Hospedales, T. Xiang, Transferring a semantic representation for person re-identification and search, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4184–4193.
[200] X. Zhang, Y. Ge, Y. Qiao, H. Li, Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3436–3445.
[201] Y. Bai, C. Wang, Y. Lou, J. Liu, L.-Y. Duan, Hierarchical connectivity-centered clustering for unsupervised domain adaptation on person re-identification, IEEE Trans. Image Process. 30 (2021) 6715–6729.
[202] M. Liu, L. Qu, L. Nie, M. Liu, L. Duan, B. Chen, Iterative local-global collaboration learning towards one-shot video person re-identification, IEEE Trans. Image Process. 29 (2020) 9360–9372.
[203] L. Song, C. Wang, L. Zhang, B. Du, Q. Zhang, C. Huang, X. Wang, Unsupervised domain adaptive re-identification: theory and practice, Pattern Recognit. 102 (2020) 107173.
[204] X. Jin, C. Lan, W. Zeng, Z. Chen, L. Zhang, Style normalization and restitution for generalizable person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3143–3152.
[205] W. Chen, X. Chen, J. Zhang, K. Huang, Beyond triplet loss: a deep quadruplet network for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 403–412.
[206] A. Khatun, S. Denman, S. Sridharan, C. Fookes, A deep four-stream siamese convolutional neural network with joint verification and identification loss for person re-detection, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018, pp. 1292–1301.
[207] S. Choi, T. Kim, M. Jeong, H. Park, C. Kim, Meta batch-instance normalization for generalizable person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3425–3435.
[208] H. Feng, M. Chen, J. Hu, D. Shen, H. Liu, D. Cai, Complementary pseudo labels for unsupervised domain adaptation on person re-identification, IEEE Trans. Image Process. 30 (2021) 2898–2907.
[209] G. Chen, Y. Lu, J. Lu, J. Zhou, Deep credible metric learning for unsupervised domain adaptation person re-identification, in: Proc. Eur. Conf. Comput. Vis, Springer, 2020, pp. 643–659.
[210] D. Mekhazni, A. Bhuiyan, G. Ekladious, E. Granger, Unsupervised domain adaptation in the dissimilarity space for person re-identification, in: European Conference on Computer Vision, Springer, 2020, pp. 159–174.

[211] F. Yang, Z. Zhong, Z. Luo, Y. Cai, Y. Lin, S. Li, N. Sebe, Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4855–4864.
[212] T. Xiao, H. Li, W. Ouyang, X. Wang, Learning deep feature representations with domain guided dropout for person re-identification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1249–1258.
[213] Y. Zhao, Z. Zhong, F. Yang, Z. Luo, Y. Lin, S. Li, N. Sebe, Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6277–6286.
[214] R. Quan, X. Dong, Y. Wu, L. Zhu, Y. Yang, Auto-ReID: Searching for a part-aware convnet for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 3750–3759.
[215] Q. Yang, H.-X. Yu, A. Wu, W.-S. Zheng, Patch-based discriminative feature learning for unsupervised person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3633–3642.
[216] H. Yao, S. Zhang, R. Hong, Y. Zhang, C. Xu, Q. Tian, Deep representation learning with part loss for person re-identification, IEEE Trans. Image Process. 28 (6) (2019) 2860–2871.
[217] Y. Dai, X. Li, J. Liu, Z. Tong, L.-Y. Duan, Generalizable person re-identification with relevance-aware mixture of experts, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16145–16154.
[218] A. Zhang, Y. Gao, Y. Niu, W. Liu, Y. Zhou, Coarse-to-fine person re-identification with auxiliary-domain classification and second-order information bottleneck, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 598–607.
[219] T. Zhang, L. Xie, L. Wei, Z. Zhuang, Y. Zhang, B. Li, Q. Tian, UnrealPerson: an adaptive pipeline towards costless person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 11506–11515.
[220] Z. Zhao, B. Zhao, F. Su, Person re-identification via integrating patch-based metric learning and local salience learning, Pattern Recognit. 75 (2018) 90–98.
[221] W. Song, S. Li, T. Chang, A. Hao, Q. Zhao, H. Qin, Context-interactive CNN for person re-identification, IEEE Trans. Image Process. 29 (2019) 2860–2874.
[222] Y. Zhai, Q. Ye, S. Lu, M. Jia, R. Ji, Y. Tian, Multiple expert brainstorming for domain adaptive person re-identification, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, Springer, 2020, pp. 594–611.
[223] Z. Zhuang, L. Wei, L. Xie, T. Zhang, H. Zhang, H. Wu, H. Ai, Q. Tian, Rethinking the distribution gap of person re-identification with camera-based batch normalization, in: European Conference on Computer Vision, Springer, 2020, pp. 140–157.
[224] C. Luo, C. Song, Z. Zhang, Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup, in: European Conference on Computer Vision, Vol. 2, Springer, 2020, p. 7.
[225] Y. Zou, X. Yang, Z. Yu, B.V. Kumar, J. Kautz, Joint disentangling and adaptation for cross-domain person re-identification, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, Springer, 2020, pp. 87–104.
[226] Y. Bai, J. Jiao, W. Ce, J. Liu, Y. Lou, X. Feng, L.-Y. Duan, Person30K: a dual-meta generalization network for person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2123–2132.
[227] P. Chen, P. Dai, J. Liu, F. Zheng, Q. Tian, R. Ji, Dual distribution alignment network for generalizable person re-identification, in: Proceedings of AAAI Conference on Artificial Intelligence, Vol. 6, 2021.
[228] X. Liu, S. Zhang, Graph consistency based mean-teaching for unsupervised domain adaptive person re-identification, IJCAI, 2021.
[229] Z. Zhong, L. Zheng, Z. Luo, S. Li, Y. Yang, Invariance matters: exemplar memory for domain adaptive person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 598–607.
[230] S. Lin, C.-T. Li, A.C. Kot, Multi-domain adversarial feature generalization for person re-identification, IEEE Trans. Image Process. 30 (2020) 1596–1607.
[231] H. Zhang, Y. Li, Z. Zhuang, L. Xie, Q. Tian, 3D-GAT: 3D-guided adversarial transform network for person re-identification in unseen domains, Pattern Recognit. 112 (2021) 107799.
[232] X. Jin, C. Lan, W. Zeng, Z. Chen, Global distance-distributions separation for unsupervised person re-identification, in: European Conference on Computer Vision, Springer, 2020, pp. 735–751.
[233] Z. Bai, Z. Wang, J. Wang, D. Hu, E. Ding, Unsupervised multi-source domain adaptation for person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12914–12923.
[234] D. Fu, D. Chen, J. Bao, H. Yang, L. Yuan, L. Zhang, H. Li, D. Chen, Unsupervised pre-training for person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14750–14759.
[235] T. Isobe, D. Li, L. Tian, W. Chen, Y. Shan, S. Wang, Towards discriminative representation learning for unsupervised person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 8526–8536.
[236] Y. Zhao, Q. Shu, K. Fu, P. Wei, J. Zhan, Joint patch and instance discrimination learning for unsupervised person re-identification, Image Vis. Comput. 103 (2020) 104000.
[237] K. Jiang, T. Zhang, Y. Zhang, F. Wu, Y. Rui, Self-supervised agent learning for unsupervised cross-domain person re-identification, IEEE Trans. Image Process. 29 (2020) 8549–8560.
[238] H. Zhang, H. Cao, X. Yang, C. Deng, D. Tao, Self-training with progressive representation enhancement for unsupervised cross-domain person re-identification, IEEE Trans. Image Process. 30 (2021) 5287–5298.
[239] J. Sun, Y. Li, H. Chen, Y. Peng, J. Zhu, Unsupervised cross domain person re-identification by multi-loss optimization learning, IEEE Trans. Image Process. 30 (2021) 2935–2946.
[240] M. Li, X. Zhu, S. Gong, Unsupervised tracklet person re-identification, IEEE Trans. Pattern Anal. Mach. Intell. 42 (7) (2020) 1770–1782.
[241] H. Li, S. Yan, Z. Yu, D. Tao, Attribute-identity embedding and self-supervised learning for scalable person re-identification, IEEE Trans. Circuits Syst. Video Technol. 30 (10) (2019) 3472–3485.
[242] J. Meng, S. Wu, W.-S. Zheng, Weakly supervised person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 760–769.
[243] Y. Zhai, S. Lu, Q. Ye, X. Shan, J. Chen, R. Ji, Y. Tian, Ad-cluster: augmented discriminative clustering for domain adaptive person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9021–9030.
[244] S. Liao, L. Shao, Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16, Springer, 2020, pp. 456–474.
[245] J. Liu, Z.-J. Zha, D. Chen, R. Hong, M. Wang, Adaptive transfer network for cross-domain person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7202–7211.
[246] S. Zhou, J. Wang, J. Wang, Y. Gong, N. Zheng, Point to set similarity based deep feature learning for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3741–3750.
[247] Y. Tang, X. Yang, N. Wang, B. Song, X. Gao, CGAN-TM: a novel domain-to–domain transferring method for person re-identification, IEEE Trans. Image Process. 29 (2020) 5641–5651.
[248] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, J. Jiao, Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 994–1003.
[249] N. Pu, W. Chen, Y. Liu, E.M. Bakker, M.S. Lew, Lifelong person re-identification via adaptive knowledge accumulation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7901–7910.
[250] G. Chen, C. Lin, L. Ren, J. Lu, J. Zhou, Self-critical attention learning for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 9637–9646.
[251] B. Chen, W. Deng, J. Hu, Mixed high-order attention network for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 371–381.
[252] C.-P. Tay, S. Roy, K.-H. Yap, AANet: attribute attention network for person re-identifications, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7134–7143.
[253] J. Si, H. Zhang, C.-G. Li, J. Kuen, X. Kong, A.C. Kot, G. Wang, Dual attention matching network for context-aware feature sequence based person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5363–5372.
[254] J. Zhou, S.K. Roy, P. Fang, M. Harandi, L. Petersson, Cross-correlated attention networks for person re-identification, Image Vis. Comput. 100 (2020) 103931.
[255] W. Li, X. Zhu, S. Gong, Scalable person re-identification by harmonious attention, Int. J. Comput. Vis. 128 (6) (2020) 1635–1653.
[256] D. Fu, B. Xin, J. Wang, D. Chen, J. Bao, G. Hua, H. Li, Improving person re-identification with iterative impression aggregation, IEEE Trans. Image Process. 29 (2020) 9559–9571.
[257] S. Xu, Y. Cheng, K. Gu, Y. Yang, S. Chang, P. Zhou, Jointly attentive spatial-temporal pooling networks for video-based person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4733–4742.
[258] A. Khatun, S. Denman, S. Sridharan, C. Fookes, Joint identification-verification for person re-identification: a four stream deep learning approach with improved quartet loss function, Comput. Vision Image Understanding 197 (2020) 102989.
[259] S. Dou, X. Jiang, Q. Zhao, D. Li, C. Zhao, Towards privacy-preserving person re-identification via person identify shift, 2022, arXiv preprint arXiv:2207.07311
[260] B. RichardWebster, B. Hu, K. Fieldhouse, A. Hoogs, Doppelganger saliency: towards more ethical person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2847–2857.
[261] X. Chen, X. Liu, W. Liu, X.-P. Zhang, Y. Zhang, T. Mei, Explainable person re-identification with attribute-guided metric distillation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11813–11822.
[262] D. Goyal, N. Patel, T. Truong, S. Yanushkevich, Towards explainable person re-identification, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2021, pp. 1–8.

[263] A. Specker, A step towards explainable person re-identification rankings, in: Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory, 2021, p. 107.
[264] Y. Zhao, S. Luo, Y. Yang, M. Song, DeepSSH: deep semantic structured hashing for explainable person re-identification, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 1653–1657.
[265] E. Bekele, W.E. Lawson, Z. Horne, S. Khemlani, Implementing a robust explanatory bias in a person re-identification network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2165–2172.
[266] J. Dietlmeier, J. Antony, K. McGuinness, N.E. O'Connor, How important are faces for person re-identification? in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 6912–6919.
[267] S. Ahmad, G. Scarpellini, P. Morerio, A. Del Bue, Event-driven re-id: a new benchmark and method towards privacy-preserving person re-identification, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 459–468.
[268] K. Wang, Z. Ma, S. Chen, J. Yang, K. Zhou, T. Li, A benchmark for clothes variation in person re-identification, Int. J. Intell. Syst. 35 (12) (2020) 1881–1898.
[269] X. Qian, W. Wang, L. Zhang, F. Zhu, Y. Fu, T. Xiang, Y.-G. Jiang, X. Xue, Long-term cloth-changing person re-identification, in: Proceedings of the Asian Conference on Computer Vision, 2020.
[270] X. Gu, H. Chang, B. Ma, S. Bai, S. Shan, X. Chen, Clothes-changing person re-identification with RGB modality only, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1060–1069.
[271] P. Zhang, J. Xu, Q. Wu, Y. Huang, X. Ben, Learning spatial-temporal representations over walking tracklet for long-term person re-identification in the wild, IEEE Trans. Multimedia 23 (2020) 3562–3576.
[272] X. Lu, X. Li, W. Sheng, S.S. Ge, Long-term person re-identification based on appearance and gait feature fusion under covariate changes, Processes 10 (4) (2022) 770.
[273] Y. Huang, Q. Wu, J. Xu, Y. Zhong, Z. Zhang, Clothing status awareness for long-term person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11895–11904.
[274] Y. Huang, J. Xu, Q. Wu, Y. Zhong, P. Zhang, Z. Zhang, Beyond scalar neuron: adopting vector-neuron capsules for long-term person re-identification, IEEE Trans. Circuits Syst. Video Technol. 30 (10) (2019) 3459–3471.
[275] R. Zhang, Y. Fang, H. Song, F. Wan, Y. Fu, H. Kato, Y. Wu, Specialized re-ranking: a novel retrieval-verification framework for cloth changing person re-identification, Pattern Recognit. 134 (2023) 109070.
[276] E. Yaghoubi, A. Kumar, H. Proença, SSS-PR: a short survey of surveys in person re-identification, Pattern Recognit. Lett. 143 (2021) 50–57.
[277] B. Heo, S. Yun, D. Han, S. Chun, J. Choe, S.J. Oh, Rethinking spatial dimensions of vision transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11936–11945.
[278] M. Usman, T. Zia, A. Tariq, Analyzing transfer learning of vision transformers for interpreting chest radiography, J. Digit Imaging 35 (6) (2022) 1445–1462.
[279] W. Wang, R. Jiang, N. Cui, Q. Li, F. Yuan, Z. Xiao, Semi-supervised vision transformer with adaptive token sampling for breast cancer classification, Front. Pharmacol. 13 (2022) 929755.
[280] O. Uparkar, J. Bharti, R. Pateriya, R.K. Gupta, A. Sharma, Vision transformer outperforms deep convolutional neural network-based model in classifying X-ray images, Procedia Comput. Sci. 218 (2023) 2338–2349.

Asmat Zahra has obtained her MS degree in Computer Science from the University of Punjab, Lahore, Pakistan. At present she is PhD scholar at the National University of Sciences and Technology, Islamabad, Pakistan

Nazia Pwevaiz has obtained her PhD degree in Computer Science from the School of Electrical Engineering and Computer science (SEECS), National University of Sciences and Technology (NUST), Islamabad, Pakistan and serving as Assistant Professor. He research interests include Person Reidentification and Biometrics.

Muhammad Shahzad has obtained his MS degree and PhD from Technical University of Munich Germany. Later he joined the Department of Informatics at TUM as and served as Associate Professor (W2). At present he is Senior Lecturer at the Department of Computer Science, University of Reading, United Kingdom. His area of expertise is 3D Computer Vision, Point Cloud processing and Human Activity Recognition

Muhammad Moazam Fraz has obtained PhD from Kingston University London, UK. After that he had been Post Doc research fellow at the Kingston University London and the University of Warwick, UK. At present, besides working as an Associate Professor and HoD Data Science and AI at the NUSTSEECS, Islamabad, he is Rutherford Visiting Fellow at The Alan Turing Institute, London, United Kingdom. His research interests include automated visual surveillance, visual recognition, biometrics and medical image analysis.