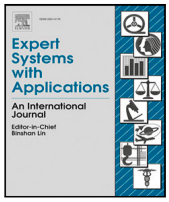
Expert Systems With Applications 239 (2024) 122419

Contents lists available at ScienceDirect

Expert Systems With Applications

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



Review

Occluded person re-identification with deep learning: A survey and perspectives

Enhao Ning [a,1], Changshuo Wang [a,b,1], Huang Zhang [c], Xin Ning [a,*], Prayag Tiwari [d]

[a] Institute of Semiconductors, Chinese Academy of Sciences, Beijing, 100083, China
[b] Center of Materials Science and Optoelectronics Engineering & School of Microelectronics, University of Chinese Academy of Sciences, Beijing, 100083, China
[c] School of Software, Xinjiang University, Xinjiang, 830000, China
[d] School of Information Technology, Halmstad University, Halmstad, 30118, Sweden

ARTICLE INFO

Keywords:
Occluded person re-identification
Literature survey and perspectives
Multimodal person re-identification
3D person re-identification

ABSTRACT

Person re-identification (Re-ID) technology plays an increasingly crucial role in intelligent surveillance systems. Widespread occlusion significantly impacts the performance of person Re-ID. Occluded person Re-ID refers to a pedestrian matching method that deals with challenges such as pedestrian information loss, noise interference, and perspective misalignment. It has garnered extensive attention from researchers. Over the past few years, several occlusion-solving person Re-ID methods have been proposed, tackling various sub-problems arising from occlusion. However, there is a lack of comprehensive studies that compare, summarize, and evaluate the potential of occluded person Re-ID methods in detail. In this review, we commence by offering a meticulous overview of the datasets and evaluation criteria utilized in the realm of occluded person Re-ID. Subsequently, we undertake a rigorous scientific classification and analysis of existing deep learning-based occluded person Re-ID methodologies, examining them from diverse perspectives and presenting concise summaries for each approach. Furthermore, we execute a systematic comparative analysis among these methods, pinpointing the state-of-the-art solutions, and provide insights into the future trajectory of occluded person Re-ID research.

1. Introduction

With the increasing integration and intelligence of surveillance equipment (Bedagkar-Gala & Shah, 2014) in recent years, person re-identification (Re-ID) technology has significantly advanced. This technology finds extensive application in various fields, such as medicine, rescue operations, criminal investigations, and surveillance. These fields often operate in complex and dynamic environments. Consequently, the rapid and accurate localization and identification of specific pedestrian targets in multi-camera occlusion scenarios hold immense practical significance.

In real-life scenes, people and objects move randomly, leading to a high likelihood of occluded individuals. Surveillance devices typically cover wide areas, further complicating the situation. Occlusion can significantly degrade visual information, making the affected features unreliable. This phenomenon may arise from various factors, including object interference, changes in pedestrian pose, clothing, and perspective. In early pedestrian representations, researchers primarily relied on basic, local visual attributes extracted from images, such as color, texture, edges, and corner points. These features capture geometric shapes and pixel distributions in images but are highly sensitive to external factors, lacking robustness and generalization. The development of deep learning has introduced high-level visual features. Compared with low-level visual features, high-level features are more adaptive to occlusions, noises and pose changes, and have stronger robustness in complex environments. Consequently, numerous researchers have developed a multitude of methods to address the prevalent occlusion problem. In general, the occlusion problem is divided into three sub-problems: (1) Noise problem. The problem of interference by multiple and mixed information from the features in the acquisition of complex scenes. (2) Missing problem. The problem of incomplete pedestrian features is due to only a part of the pedestrian being captured. (3) Alignment problem. Owing to the change in posture, perspective, and position, the features cannot correspond one-to-one, which causes distraction, shared location misalignment, and other issues. The study of occlusion involves the separation of humans from backgrounds to extract human features as the core. Simultaneously, research has focused

* Corresponding author.
E-mail addresses: ningenhao@163.com (E. Ning), wangchangshuo@semi.ac.cn (C. Wang), zhhh1998@outlook.com (H. Zhang), ningxin@semi.ac.cn (X. Ning), prayag.tiwari@ieee.org (P. Tiwari).
[1] Equal contribution.

https://doi.org/10.1016/j.eswa.2023.122419
Received 7 July 2023; Received in revised form 7 October 2023; Accepted 30 October 2023
Available online 2 November 2023

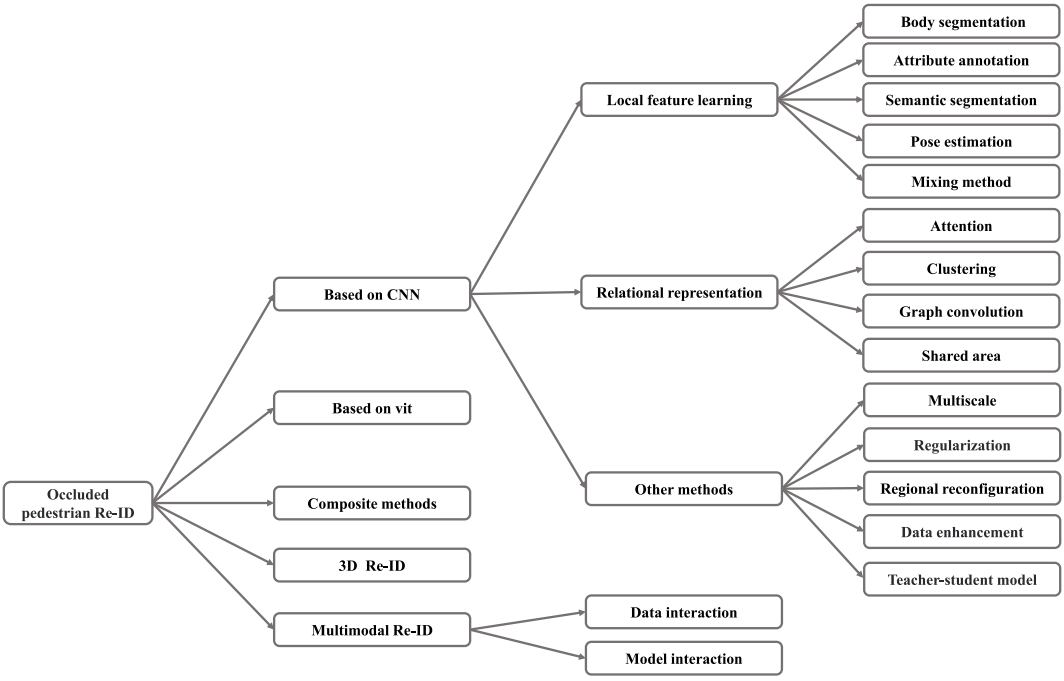


Fig. 1. Overall structure of the survey.

Table 1
The summary of person-reported surveys in recent years.

Survey	Venue
A survey of approaches and trends in person Re-ID (Bedagkar-Gala & Shah, 2014)	IVC2014
Person Re-ID Past, Present and Future (Zheng, Yang, & Hauptmann, 2016)	arXiv2016
A systematic evaluation and benchmark for person Re-ID: Features, metrics, and datasets (Gou et al., 2018)	TPAMI2019
Beyond intra-modality discrepancy: A comprehensive survey of heterogeneous person Re-ID (Wang, Wang, Wu, Wang, & Satoh, 2019)	arXiv2019
A Survey of Open-World Person ReIdentification (Leng, Ye, & Tian, 2019)	TCSVT2020
Survey on Reliable Deep Learning-Based person Re-ID Models: Are We There Yet? (Lavi, Ullah, Fatan, & Rocha, 2020)	arXiv2020
Deep Learning for Person Reidentification: A Survey and Outlook (Ye, Shen, et al., 2021)	TPAMI2021
SSS-PR: A short survey of surveys in person Re-ID (Yaghoubi, Kumar, & Proença, 2021)	PRL2021
Deep learning-based person Re-ID methods: A survey and outlook of recent works (Ming et al., 2022)	IVC2022
Deep Learning-based Occluded person Re-ID: A Survey (Peng et al., 2022)	arXiv2022

on developing methods to extract fine-grained, highly discriminative, and more essential features with reference and value.

We want to identify the current state-of-the-art and limitations of existing methods and discover unexplored areas. Specifically, we present methods for dealing with occluded person Re-ID that were submitted in top international journals or conferences before 2023. We classify deep learning-based occluded person Re-ID according to the network structure of extracted features (CNN-based, transformer-based, and hybrid structure-based), the way features are extracted (uni-modal and multi-modal), and the hierarchical structure of features (2d and 3d). (see Fig. 1). First, due to the impressive performance of convolutional neural networks (CNNs) in image matching tasks, CNN-based methods have become a mainstream approach for addressing occlusion problems in person Re-ID. Therefore, we categorize CNN-based methods as the first class of approaches for tackling occlusion issues. Secondly, following the success of transformers in the natural language processing field, Vision Transformers (ViTs) have also gained widespread use in handling occlusion problems in pedestrian re-identification, yielding promising results. Consequently, we classify transformer-based methods as the second category. The third class of methods encompasses composite approaches. One example is the exploitation of the complementary nature of CNNs and ViTs to create hybrid structures. The fourth and fifth categories of methods involve the utilization of 3D and multimodal approaches to address occlusion problems in person Re-ID. These methods are capable of handling a wider range of scenarios and represent a relatively novel approach.

In general, the contributions of this study are as follows:

(1) This study focuses on addressing the occlusion problem in person Re-ID models, which is crucial for achieving high accuracy and robustness. We present a scientific and comprehensive review of past and current state-of-the-art approaches.

(2) The current review of person Re-ID methods lacks sufficient coverage of approaches based on ViT. Given the excellent performance of ViT in occluded person Re-ID, we include a discussion of this method and its hybrid variants in our study, offering researchers new ideas and options for addressing the occlusion problem.

(3) We creatively incorporate 3D person Re-ID and multimodal person Re-ID, which have become popular in recent years. These novel methods can better solve the occlusion problem by utilizing additional depth or modal information, thus improving the performance and reliability of person Re-ID.

(4) We anticipate ongoing advancements in occluded person Re-ID and firmly believe that continuous research and innovation will yield more effective methods and technologies for tackling the occlusion problem. These advancements are poised to serve as a source of inspiration and a catalyst for progress within the broader field of person Re-ID.

2. Literature review

In the field of person Re-ID, there is a relative scarcity of specialized reviews compared to methodological articles. The available reviews

primarily focus on specific aspects of the field, and these surveys are summarized in Table 1. For instance, Bedagkar-Gala and Shah (2014) delves into the challenges of person Re-ID, categorizing it into open-set Re-ID and closed-set Re-ID based on the fixity of the gallery. On the other hand, Zheng et al. (2016) divides person Re-ID methods into those for images and those for videos based on the matching strategy. In a more detailed examination, Gou et al. (2018) provides an in-depth study of the features, metrics, and datasets relevant to person Re-ID. Meanwhile, Wang, Wang, Wu, et al. (2019) focuses on heterogeneous person Re-ID, classifying methods into four categories based on application scenarios: low-resolution, infrared, sketch, and text. Other surveys include (Leng et al., 2019), which concentrates on open-world Re-ID tasks, and Lavi et al. (2020), which classifies Re-ID into single feature learning-based approaches and multi-feature learning-based approaches based on feature learning strategies. Additionally, Ye, Shen, et al. (2021) offers an extensive explanation of Re-ID for open and closed settings, introducing methods such as transmembrane states and unsupervised approaches. Yaghoubi et al. (2021) provides a multidimensional classification of the person Re-ID problem, while Ming et al. (2022) categorizes person Re-ID methods into four groups based on metric learning and representation learning, also incorporating the latest methodologies. Lastly, Peng et al. (2022) focuses on image-based obscured person Re-ID methods.

These investigations are inevitably constrained by certain inherent limitations. Given the pervasive challenge of occlusion in pedestrian recognition, research on occluded person Re-ID is of paramount importance. Consequently, we offer a thorough summary and comprehensive analysis of methods and future prospects in the field of occluded person Re-ID, aiming to drive forward future advancements.

3. Datasets and evaluation protocols

3.1. Datasets

Occluded person Re-ID datasets can be divided into two categories: partial person and occluded person Re-ID datasets. The pedestrian images of the occluded person Re-ID datasets have occlusion information interference and are not cropped. The pedestrian image portion of the partial person Re-ID dataset is present and artificially cropped. Examples of partial/occluded person Re-ID datasets are shown in Fig. 2.

Occluded-DukeMTMC (Miao, Wu, Liu, Ding, & Yang, 2019) was collected from DukeMTMC-reID (Zheng, Zheng, & Yang, 2017), containing 15,618 training images of 708 pedestrians, 2210 query images of 519 pedestrians, and 17,661 gallery images of 1110 pedestrians for testing. Of these images, 9% of the training set, 100% of the query set, and 10% of the gallery are occluded images. Obstacles include cars, bicycles, trees, and other pedestrians, adding complexity to the dataset.

P-ETHZ (Zheng, Li, et al., 2015) was an image-based occluded person Re-ID dataset, modified by ETHZ (Ess, Leibe, Schindler, & Van Gool, 2008). It has 3,897 images containing 85 pedestrian identities with 1 to 30 full-body and occluded pedestrian images per identity.

P-DukeMTMC-reID (Zhuo, Chen, Lai, & Wang, 2018) was modified from DukeMTMC-reID (Zheng et al., 2017), containing a total of 24,143 images of 1,299 pedestrians, and each identity has a full-body and occlusion image; the pedestrian in the image is occluded by different objects, such as other pedestrians, cars, and signage.

Occluded-REID (Zheng, Li, et al., 2015) has 2000 images of 200 pedestrians, each pedestrian corresponding to 5 occlusion and 5 whole body images, collected from Sun Yat-sen University. The dataset includes different viewpoints and types of severe occlusion, which challenges person Re-ID.

Occluded-DukeMTMC-VideoReID (Hou et al., 2021) was reorganized from the DukeMTMC-VideoReID (Wu et al., 2018) dataset. The training set contains 1,702 trajectory segments covering 702 pedestrians, the test set queries cover 661 pedestrians, and the gallery covers 1110. More than 70% of the videos are occluded, including different perspectives and a variety of obstacles, such as cars, trees, bicycles, and other pedestrians.

Fig. 2. Examples of four commonly used occluded person Re-ID datasets.

Partial-ReID (Zheng, Li, et al., 2015) has 600 images of 60 pedestrians, 5 partial and 5 full-body images for each pedestrian. Using the visible parts, they are manually cropped to form new partial images. The images are collected from different perspectives, backgrounds and occlusions in a university campus.

Partial-iLIDS (He, Liang, Li, & Sun, 2018) was derived from iLIDS (Zheng, Gong, & Xiang, 2011) and contains 238 images of 119 pedestrians. Each pedestrian corresponds to one manually cropped non-occluded partial image and one full-body image. The partial image is used as a query, and the full-body image is used as a search library. It was shot by multiple non-overlapping cameras, mostly for test sets.

Partial-CAVIAR (He et al., 2018) was derived from CAVIAR (Cheng, Cristani, Stoppa, Bazzani, & Murino, 2011) and contains 142 images of 72 pedestrians. The partial map is generated by randomly picking half of the overall image of each pedestrian.

P-CUHK03 (Kim & Yoo, 2017) was constructed based on CUHK03 (Li, Zhao, Xiao, & Wang, 2014), with a total of 1360 pedestrian images, wherein 15,080 images corresponding to 1160 pedestrians are used as a training set, and the remaining 100 pedestrians are used as a validation and test set. Two of the images are selected to generate 10 local body query images with a spatial area ratio, and the remaining three images are used as whole body gallery images.

3.2. Evaluation protocols

In the field of occluded person Re-ID, the commonly used evaluation metrics are Cumulative Matching Characteristic (CMC) curves and mean Average Precision (mAP).

CMC curves are based on the principle of ranking the similarity between the query image and the image library, and the higher the top image, the higher the similarity with the query image. Then, the top-k accuracy ACC^k of the query image is calculated based on this ranking. If the first k samples contain the query target, then ACC^k is 1, $k \in \{1, 2, 3...\}$. Otherwise, ACC^k is 0 . Finally, the ACC^kcurves for all targets are summed and divided by the total number of targets to obtain CMC-k.

MAP better reflects the extent to which all correct target images are positioned at the top of the ranked list. In comparison to the CMC curve, mAP offers a more comprehensive evaluation of the performance of Re-ID algorithms. Here, 'P' represents precision, signifying the proportion of correct samples among all samples, thus reflecting the accuracy of

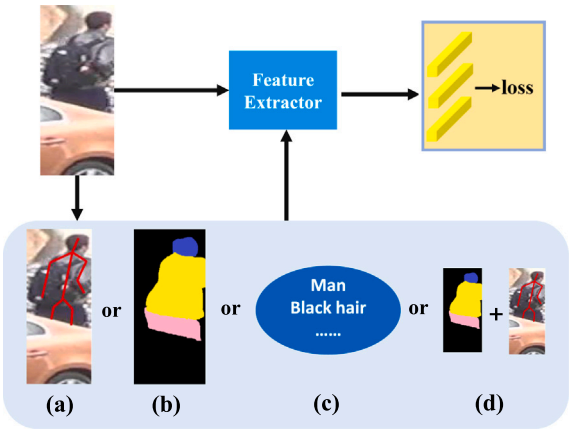


Fig. 3. Four different local feature learning methods: (a) indicates pose estimation, (b) indicates semantic segmentation, (c) indicates attribute annotations, and (d) indicates the mixing method.

the model's output with respect to correct samples. The Average Precision (AP) corresponds to the average of all correct samples predicted by the model, indicating how effectively the model operates within a specific category by averaging the accuracy of each correct prediction. Given the presence of multiple classes in recognition tasks, the average AP value is calculated across all classes. This is achieved by summing the average accuracy for each class and dividing it by the total number of classes to obtain the mAP.

The CMC curve cannot consider the hits of the samples with lower rankings, while mAP takes all samples into account. Therefore, they are important and complementary.

4. Deep learning methods

4.1. Based on CNN

Convolutional neural networks (CNNs) have emerged as one of the leading methods for learning pedestrian representations from RGB images. By using local perceptual fields and learning filters, CNNs can extract powerful features that capture regional information about local features of pedestrians. These features are then compressed and mapped to higher-level representations. Researchers have refined them to be usable for pedestrian matching tasks in complex realistic scenarios. We classify it into local feature learning, relational representation, mixing methods, and other methods.

4.1.1. *Local feature learning*

Local feature learning excels in capturing regional characteristics and offers distinct advantages in recognizing and locating occluded regions, as compared to global features. According to its implementation of different local feature methods, we divide them into human segmentation, pose estimation, human parsing, attribute annotation, and hybrid methods (see Fig. 3).

Body Segmentation. By leveraging the characteristic of pedestrians walking upright, our method extracts improved local features through the segmentation of the original image or feature map. The segmentation results can take the form of stripes, fixed regions, or small patches (see Fig. 4). However, segmentation does not have the process of identifying occlusions, so it is sensitive to noise.

Addressing these challenges, researchers have developed various approaches. For instance, CBDB-Net (Tan, Liu, Bian, Wang, & Yin, 2021) evenly divides the strips on the feature map. It then systematically discards each strip one by one to generate multiple incomplete feature maps. This approach compels the model to learn a more robust pedestrian representation even in an environment with incomplete

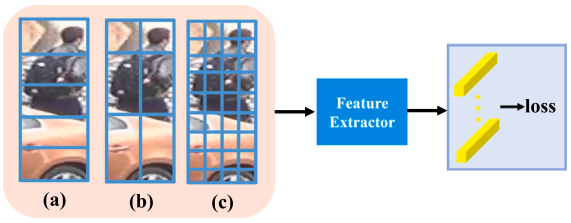


Fig. 4. Three common Body Segmentation schematics: (a) indicates stripes, (b) indicates fixed areas, and (c) indicates small blocks.

information. DPPR (Kim & Yoo, 2017) predefines thirteen bounding boxes for the whole-body image, including the whole image, half-body image, and horizontal part image, and extracts features from each part. Simultaneously, it introduces an attention-based matching mechanism, assigning higher weights to features of the same body part, thereby mitigating the impact of information loss caused by occlusion. OCNet (Kim, Cho, Lee, Cho, & Lee, 2022) uses a relationship-based approach to handle occlusion. It horizontally divides the image into three sections: top, middle, and bottom, extracting four corresponding features. It employs a relationship adaptive module with two shared layers to address alignment issues among regional features while introducing weights to reduce noise interference.

Pose Estimation. Pose estimation extracts semantic information at the image pose level by leveraging the structured human skeleton. Noise interference is mitigated through guided or fused techniques. In the case of HOReID (Wang, Yang, et al., 2020), it introduces a learnable relational matrix. The key human body points obtained from pose estimation are treated as nodes in a graph, resulting in the formation of a topology graph to further suppress noise interference. PMFB (Miao, Wu, & Yang, 2021) uses pose estimation to obtain confidence and coordinates of human keypoints. Then, a threshold is set to filter the occluded regions. Finally, the visible part is used to constrain the feature response at channel level to solve the occlusion problem. PGMANet (Zhai, Han, Ma, Gou, & Xiao, 2021) generates an attention mask using a human heat map. The interference of noise is removed jointly by the dot product of feature maps and guidance of higher-order relations.

Researchers commonly employ pose estimation in two key directions. First, it is used to extract semantic features, allowing for the identification of noisy points and improving noise interference removal. Second, pose estimation is leveraged to localize human regions, which addresses alignment issues and facilitates the extraction of local features. AACN (Xu, Zhao, Zhu, Wang, & Ouyang, 2018) uses pose points to locate pedestrian body regions and introduces a posture-guided visibility score to separate occlusions. DAReID (Xu, Zhao, & Qin, 2021) adopts a dual-branch structure. In this architecture, the mask branch is responsible for extracting highly discriminative local features using a spatial attention module guided by pose estimation. In parallel, the global branch enhances the representation of human discriminative information through feature activation.

Semantic Segmentation. By incorporating a human parsing model, noise interference is detected and eliminated through segmentation or semantic parsing. In the case of SPReID (Kalayeh, Basaran, Gökmen, Kamasak, & Shah, 2018) generates probability maps associated with five different body regions based on the trained pedestrian class semantic parsing model Inception-V3 (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016), namely, foreground, head, upper body, lower body, and shoes. These probability maps are subsequently fused with semantic region features after bilinear interpolation. This fusion process activates different parts and effectively mitigates the impact of occlusion. Co-Attention (Lin & Wang, 2021) utilizes the parsing mask of the pedestrian's local body image as a query and simultaneously builds a mapping. This process involves the introduction of a self-attentive

mechanism aimed at filtering out occlusions. MMGA (Cai, Wang, & Cheng, 2019) initially divides pedestrians within images into upper and lower body segments utilizing JPPNet (Liang, Gong, Shen, & Lin, 2018). Subsequently, the method incorporates two attention modules. The first module is employed to filter out background interference, while the second module generates spatial and channel attentions based on whole, upper, and lower body masks. Ultimately, element-level multiplication is conducted to produce the final feature. HPNet (Huang, Chen, & Huang, 2020) leverages the COCO (Lin et al., 2014) dataset to train a human body parsing model. This model provides labels for the four primary body parts, which are used for training the parsing model and the overall network concurrently in a multitasking fashion. Additionally, the process involves generating visibility scores to mitigate occlusions.

Semantic segmentation-based approaches make significant contributions to augmenting feature diversity. These approaches divide an image into distinct regions, enabling the model to scrutinize and comprehend the context of each specific part. This, in turn, facilitates a more detailed scene analysis, resulting in richer and more diverse feature representations. In the case of SORN (Zhang, Yan, Xue, Hua, & Wang, 2020), a three-branch model is employed, consisting of a global branch, a local branch, and a semantic branch. The global branch is tasked with acquiring global features through normalization and feature aggregation. Meanwhile, the local branch capitalizes on prior knowledge of pedestrian body structure to generate pedestrian body parts and extract local features through mapping, pooling, and normalization processes. The semantic branch initially employs the DANet (Fu et al., 2019) model to pre-train semantic labels for the data. Subsequently, it trains a semantic segmentation model on the DensePose-COCO dataset (Güler, Neverova, & Kokkinos, 2018), incorporating label smoothing to optimize the semantic labels. Finally, it aggregates the semantic segmentation component. This aggregation process results in the creation of a foreground pedestrian body region, effectively achieving the separation of background and pedestrians.

Attribute annotation. The occlusion problem is handled by introducing attribute annotation. ASAN (Jin, Lai, & Qian, 2021) extracts the visible part of human features by combining attribute information and weak supervision. Attribute information is a semantic level attribute annotation. Based on the visibility part determination, a region visibility matching algorithm is introduced to achieve the effect of denoising.

Mixing method. Introducing more than two kinds of external information can help the model remove the interference of noise in the form of feature interaction or co-guidance. GASM (He & Liu, 2020) introduces an architecture for learning salient information. Firstly, it separates pedestrians from the background by leveraging semantic information. Then, it addresses occlusion interference through pose estimation. Finally, it fuses these two features to guide the model's learning. SSPReID (Quispe & Pedrini, 2019) designs a joint learning method to combine salient and semantic features. Five different semantic features of the human body are obtained by human body parsing. The saliency features utilize the regions of highest attention in the graph. These features are fused with global features and finally concatenated together to form the final features. TSA (Gao, Zhang, et al., 2020) presents two types of features to address the pose change problem. One set is guided by pose keypoints, and the other is guided by partial masks from a human parsing model. These two sets of features are subsequently fused. Leveraging interaction can effectively tackle the pose change issue, and the combination of pose and segmentation helps suppress noise, thus resolving occlusion problems. FGSA (Zhou et al., 2020) proposes a pose resolution network for complex pose changes to deal with local locations and the relationships between them.

4.1.2. *Relationship representation*

By emphasizing feature relationships, occlusions can be addressed through suppression, removal, or supervision. We categorize these approaches into four groups: attention, clustering, graph convolution, and shared region (as shown in Fig. 5), based on the diverse methods and mechanisms they employ to learn these relationships.

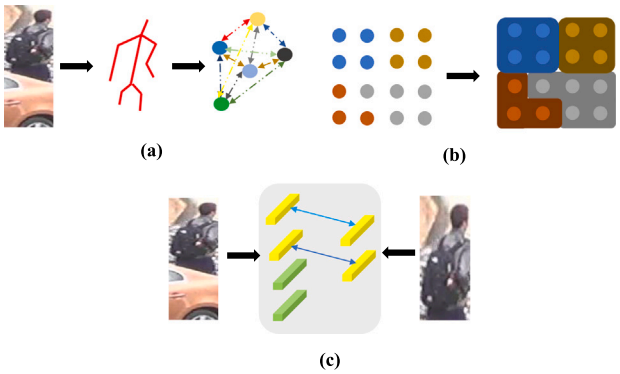


Fig. 5. (a) Schematic diagram showing figure convolution. (b) Schematic representation of clustering. (c) Schematic representation of the shared area.

Attention. By introducing the attention mechanism, the model can select the highly salient and discriminative regions to suppress the interference of noise. To address the issue of missing images in person Re-ID, DPPR (Kim & Yoo, 2017) employs an attention mechanism to emphasize the same pedestrian part across different images. Moreover, OCNet (Kim et al., 2022) mitigates the effect of noise by capturing higher-order relationships among regional features and incorporating them with weighted combinations. This method effectively suppresses the influence of noisy or irrelevant information, resulting in more robust and accurate person Re-ID outcomes. DAReID (Xu et al., 2021) introduces dual attention recognition. The local area visible to the pedestrian is obtained by gesture-guided spatial attention. Global features are extracted by feature activation and pose. Both will then be used together to guide the representation of features. MHSA-Net (Tan, Liu, Yin, & Li, 2022) multiplies attention weights with feature maps and applies a nonlinear transformation to encourage multi-headed attention mechanisms to adaptively capture key local features.

By incorporating attention mechanisms, models can allocate more attention to important features and ignore irrelevant ones. This not only increases the model's flexibility by allowing it to adapt its attention allocation according to different inputs but also enhances its interpretability as it highlights the relevant parts of the input. PAFM (Yang, Zhang, Tang, & Li, 2022) introduces an enhanced spatial attention module designed to uncover relationships among pixel points while capturing and aggregating pixel points with high semantic relevance. Subsequently, this module is multiplied with the feature map containing pose information to perform feature fusion. Co-Attention (Lin & Wang, 2021) utilizes the analytic mask of a partial pedestrian image and the whole image as targets, matching them using a self-attention mechanism (Li, Jiang, & Hwang, 2020). The result is the suppression of noise interference through a focus on pedestrian features.

Clustering. The issue of noise interference is addressed by identifying the intrinsic distribution structure of the data to categorize the pixel points. ISP (Zhu, Guo, Liu, Tang, & Wang, 2020) assigns a pseudo-label to each pixel by tandem clustering. All pixels of the human body image are firstly divided into foreground and background, based on the assumption that the foreground is more responsive than the background. Secondly, the pixels are clustered into different parts and assigned pseudo-labels. Based on the pseudo-labels, different weights are assigned to the pixels to extract local features. This not only separates occlusions from pedestrians at the pixel level, but also enables automatic alignment.

Graph Convolution. Learning high-order semantic pixel relationships suppresses noise interference through constrained information transmission. HOReID (Wang, Yang, et al., 2020) introduces a matrix describing the higher order relationships between points and later

passes information in this relationship matrix to form a topological map. With the help of the constraints of the topological map, the transfer of useless information between points is suppressed, and the purpose of noise removal is achieved.

Shared Area. The interference of noise is mitigated by sensing the same body parts of pedestrians in image pairs to extract shareable features. DPPR (Kim & Yoo, 2017) assigns greater weights to regions containing matching body parts, enhancing the model's capacity to extract essential features. VPM (Sun et al., 2019) addresses alignment and denoising challenges by perceiving the visibility of shared regions. PPCL (He, Shen, Huang, Chen, & Hua, 2021) autonomously learns component matching and ultimately computes image similarity solely based on shared semantic corresponding regions. KBFM (Han, Gao, & Sang, 2020) concentrates on extracting highly visible and shareable pose points, serving as the core areas for feature extraction, thus achieving denoising and alignment effects.

4.1.3. *Other methods*

Regional reconfiguration. This approach complements obscured or noisy areas using complete pedestrian areas. To address the issue of information loss due to occlusion, RFCNet (Hou et al., 2021) introduces an encoder–decoder architecture that leverages non-occluded remote spatial context for feature completion. The encoder is designed based on similarity region assignment, while the decoder reconstructs the occluded region by establishing correlations between the occluded and distant non-occluded regions through clustering. ACSAP (He, Yang, & Chen, 2021) combines attitude and adversarial generation networks, incorporating an attitude-guided spatial generator and spatial discriminator to eliminate noise interference.

Data enhancement. The model's sensitivity to occlusion is enhanced by incorporating data transformation techniques. IGOAS (Zhao et al., 2021) employs a progressive occlusion module, introducing small uniform occlusions on a group of images and gradually generating larger occlusions based on model learning. This approach improves occlusion recognition. OAMN (Chen, Liu, et al., 2021) uses a cropping and scaling method, predefining four corners and randomly selecting a training image to be cropped and scaled into patches at four positions. Weighted learning, combined with attention mechanisms, achieves denoising. SSGR (Yan et al., 2021) introduces a compound batch erase method, involving random erase and batch constant erase operations. It divides the image into random segments, selectively erases strips in each sub-batch, and employs a matching-based disentanglement non-local operation, enhancing feature extraction from the complete pedestrian region. ETNDNet (Dong, Zhang, Yan, Tang, & Tang, 2023) addresses the occlusion problem from an adversarial defense perspective. It tackles incomplete information, positional misalignment, and noisy data by randomly erasing feature maps, introducing random transformations, and perturbing feature maps.

Regularization. Penalties and constraints are utilized to target high-attention areas, compelling the model to prioritize the entire pedestrian region. This approach ensures the extraction of pedestrian features using comprehensive information. MHSA-Net (Tan, Liu, et al., 2022) introduces a feature regularization mechanism comprising two components: a regularization term based on attention weight embedding and a hard triplet loss based on triplet feature units. The regularization term diversifies local information representation and enhances information completeness. Meanwhile, the hard triplet loss refines feature fusion, leading to improved pedestrian matching.

Teacher-student model. Teacher models assist student models in dealing with occlusion problems. HG (Kiran et al., 2021) designs an end-to-end unsupervised teacher–student framework that lets the teacher network learn the between-class distance by inputting different combinations of images, and then the student inherits the network and learns the within-class distance by inputting more noisy images of the same class. At the same time, the attention embedding method with distance distribution matching can help the student network to remove noise interference better and extract more discriminative features. AFPB (Zhuo, Lai, & Chen, 2019) adopts a two-step approach to address occlusion challenges. Initially, it incorporates regular data and pedestrian volume data simulating occlusion into the teacher network. Subsequently, a joint training process enables the teacher network to acquire a robust model capable of handling occlusions. The student network then inherits the knowledge from the teacher network and further refines its capabilities by learning from more realistic and noisy real-world occluded pedestrian data.

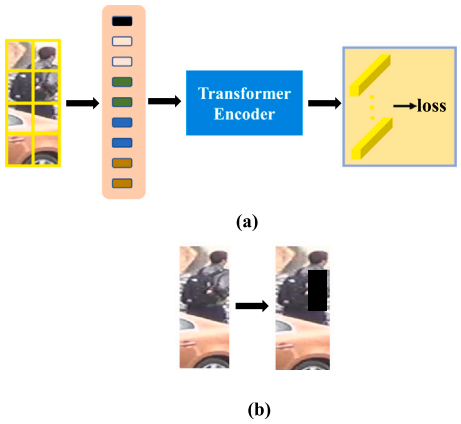


Fig. 6. (a) Schematic of the transformer-based approach. (b) Schematic representation of data enhancement.

Multiscale. The occlusion problem is effectively tackled through multi-scale feature representation methods. DSR (He et al., 2018) focuses on addressing scale differences by training a fixed-size fully convolutional network and introducing three different scales of blocks for feature extraction. Similarly, FPR (He et al., 2019) deals with scale issues using a structure composed of convolution and pooling layers. It incorporates a pyramid layer with various pooling kernels and employs an attention-based foreground probability generator to reduce background distractions.

4.2. *Based on ViT*

TransReID, proposed in 2021 (He, Luo, et al., 2021), marks a significant milestone as the first model to incorporate the ViT architecture (Dosovitskiy et al., 2020) into the realm of Re-ID (see Fig. 6). Compared to ResNet (He, Zhang, Ren, & Sun, 2016), TransReID offers two key advantages: (1) TransReID's utilization of multi-headed self-attention enables it to excel at capturing long-range relationships within the data. This feature encourages the model to focus on distinct body parts (Khan et al., 2022; Shamshad et al., 2023), enhancing its ability to distinguish between individuals effectively. (2) The Transformer architecture employed by TransReID excels in preserving intricate details even in the absence of downsampling computations. This capability results in a richer and more informative representation of the data.

Based on these characteristics, many variants of transformer have appeared in recent years, and researchers have widely used them in occlusion Re-ID. PFD (Wang, Liu, et al., 2022) uses a transformer to capture contextual relationships in image blocks and enhances body part visibility through pose-guided feature aggregation. PFT (Zhao, Zhu, Wang, & Liang, 2022) introduces a learnable enhancement patch to improve local feature extraction by transformers, focusing on both local and long-range correlations. FED (Wang, Zhu, et al., 2022) distinguishes between different types of occlusions and concentrates on non-target pedestrian occlusions. FRT (Xu, He, Liang, & Sun, 2022) classifies pedestrians into head, torso, and legs, then employs a graph-based occlusion elimination module to reduce the impact of occlusions by learning region similarities.

4.3. Composite method

By introducing more than two different networks, the occlusion problem is dealt with in an interactive or fused manner. FGMFN (Zhang, Chen, Chen, Zhang, & Zheng, 2022) employs a dual-branch network, where local features undergo an affine transformation and are processed by ResNet-50 to extract upper body features. These features are divided into three regions with the help of an attention module, while global features are extracted using a block partition scheme. The final feature is a fusion of both branches. Pirt (Ma, Zhao, & Li, 2021) pretrains an HRNet pose estimation model on the COCO dataset and then employs a two-branch structure. Intra-part features are processed by a modified ResNet-50, while inter-part relationships are guided by a transformer. The model effectively handles occlusion by establishing part-aware long-range dependencies. DRL-Net (Jia, Cheng, Lu, & Zhang, 2022) generates augmented samples with random obstacles from training images. It combines a CNN and Transformer to create a query-based semantic feature extraction layer and uses semantic bootstrapping to learn positive and negative sample comparisons, effectively eliminating interference noise.

4.4. 3D Re-ID

Compared to traditional 2D approaches, the utilization of 3D information in occlusion person Re-ID is a relatively recent development. 3D data offers robust shape and spatial depth features, which are less affected by texture information, making it effective in reducing interference from occlusions. This is achieved through techniques such as 3D feature denoising, 3D feature complementation, and multi-view construction. For instance, PersonX (Sun & Zheng, 2019) creates virtual 3D models by scanning real-world people and objects and then translating them back into 2D representations. This data manipulation enhances the representation of the data. In another approach, Wang, Liang, and Liao (2022) employs UV texture mapping to transfer clothing from real-world pedestrians to virtual 3D characters. They utilize a patch-based feature segmentation and expansion method to address occlusion challenges. A more common form of 3D information is the point cloud (Qi, Su, Mo, & Guibas, 2017), where the depth information in the point cloud can be used as an additional channel to the image. OG-Net (Zheng, Wang, Zheng, & Yang, 2022) uses Skinned Multi-Person Linear (SMPL Kanazawa, Black, Jacobs, & Malik, 2018) to generate six channels of point cloud data from 2D images, providing positional and texture information. ASSP (Chen, Jiang, et al., 2021) uses 3D body reconstruction as an auxiliary task for 2D feature extraction.

However, the research on recognition using point clouds is still limited compared to 2D images, which is an important research direction for the future.

4.5. Multi-modal Re-ID

RGB-IR multimodal Re-ID. Both day and night are important scenes of pedestrian life, and in the case of insufficient illumination, images can only be collected by infrared cameras (Nguyen, Hong, Kim, & Park, 2017; Wu, Zheng, Yu, Gong, & Lai, 2017). If there is occlusion in the scene, the infrared image still has occlusion. At the same time, the infrared image can be used as a special channel of the RGB image, which makes the representation mode of the RGB image more complete and can supplement the information well in the case of information loss caused by occlusion.

The RGB-IR multimodal Re-ID is designed to feed both infrared and conventional images into the model. Removal of interference from occlusion is achieved by the relationship of different modalities or by combining approaches, such as attention mechanisms, in dealing with single modal noise, multi-scale, etc. DDAG (Ye, Shen, J. Crandall, Shao, & Luo, 2020) presents a dynamic dual attention cross-modal graph structure. It starts by generating local attention based on feature similarity. Then, it introduces an aggregated representation for part-level relation learning. Additionally, it incorporates a graph structure to remove noise interference through relational information. For addressing intra-modal challenges, HMML (Zhang et al., 2022) introduces a pairing-based intra-modal similarity constraint to enhance features. Similarly, CMC (Wen, Feng, Li, & Chen, 2022) employs multi-scale, multi-level feature learning for refined feature extraction. To tackle image internal misalignment, DTRM (Ye, Chen, Shen, & Shao, 2021) combines attention and partial aggregation. It utilizes the contextual relationship between two modalities to enhance global features and mitigate the effects of noise.

RGB-Depth multimodal Re-ID. Depth images, obtained through devices like laser radars, offer valuable body shape and skeletal information by measuring distances. In situations where information is obscured by obstacles in regular images, depth features can complement texture-based position information, providing a more comprehensive representation. Additionally, depth features can address challenges related to lighting variations and changes in clothing, which can affect pedestrian recognition. They prove particularly useful in dealing with issues like obstacle-related illumination changes and variations in clothing due to different environments. CMD (Hafner, Bhuyian, Kooij, & Granger, 2022) introduces an approach that combines embedding representation and feature distillation to tackle noise interference. This method employs a gate-controlled attention mechanism to dynamically activate more discriminative features in one modality by gating signals from another modality, effectively reducing the impact of noise.

RGB-Text multimodal Re-ID. RGB-Text multimodal Re-ID aims to introduce text data to enhance feature representation by sharing semantic information and attentional calibration to eliminate the effect of noise. In daily life, text information is one of the most frequently used types of information. When image information is missing or cannot be used owing to obstacles, it can be supplemented with text. AXM-Net (Farooq, Awais, Kittler, & Khalid, 2022) dynamically exploits multi-scale information of text and images, recalibrating each modality according to the shared semantics and adding contextual attention to the text branch to supplement the information of the convolution block. Furthermore, attention is introduced to enhance feature consistency and local information of the visual part. It can learn the alignment semantic information of different modalities and automatically remove the interference of irrelevant information.

5. Method comparison

We statistically evaluate the results of the occluded person Re-ID methods on two general datasets (Market1501 Zheng, Shen, et al., 2015, DukeMTMC-reID Ristani, Solera, Zou, Cucchiara, & Tomasi, 2016), two occluded person Re-ID datasets (Occluded-DukeMTMC Miao et al., 2019, Occluded-REID Zheng, Li, et al., 2015), and two partial person Re-ID datasets (Partial-ReID Zheng, Li, et al., 2015, Partial-iLIDS He et al., 2018). We categorize them into three groups: The experimental results based on the local feature learning method are presented in Table 2. This category includes body segmentation, pose estimation, semantic segmentation, attribute annotation, and fusion methods. The experimental results based on the relationship representation method are shown in Table 3. This group includes shared area, clustering, graph convolution, and attention-based approaches. The experimental results of other methods are presented in Table 4, which includes transformer-based methods, composite method, multiscale, data enhancement, and regularization techniques. For a detailed introduction to each method category and specifics of each study, please refer to Section 4. From the results, we can derive the following insights:

(1) From the results we can get the following information: OC-Net (Kim et al., 2022) based on local feature learning, QPM (Wang, Ding, et al., 2022) based on relational representation, DPM (Tan, Dai, et al., 2022) and PFD (Wang, Liu, et al., 2022) based on transformer perform better on the Occluded-DukeMTMC dataset. On the

Table 2
Comparison of experimental results based on local feature learning methods. The red numbers indicate the best results. (in %).

		Method	Venue	Occluded-Duke		Occluded-REID		Partial-REID		Partial-iLIDS		Market1501		DukeMTMC-reID	
				Rank-1	mAP	Rank-1	mAP	Rank-1	Rank-3	Rank-1	Rank-3	Rank-1	mAP	Rank-1	mAP
Local feature learning	Body segmentation	CBDB-Net (Tan et al., 2021)	TCSVT2021	50.09	38.9	–	–	66.7	78.3	68.4	81.5	94.4	85	87.7	74.3
		OCNet (Kim et al., 2022)	ICASSP2022	**64.30**	**54.40**	–	–	–	–	–	–	95	89.3	90.5	80.2
	Pose estimation	AACN (Xu et al., 2018)	CVPR2018	–	–	–	–	–	–	–	–	88.69	82.96	76.84	59.25
		ACSAP (He, Yang, & Chen, 2021)	ICIP2021	–	–	–	–	77	83.7	76.5	87.4	–	–	–	–
		DAReID (Xu et al., 2021)	KBS2021	63.4	–	–	–	68.1	79.5	76.7	85.3	94.6	87	88.9	78.4
		DSA-reID (Zhang, Lan, Zeng, & Chen, 2019)	CVPR2019	–	–	–	–	–	–	–	–	**95.7**	87.6	86.2	74.3
		HOReID (Wang, Yang, et al., 2020)	CVPR2020	–	–	80.3	70.2	85.3	91	72.6	86.4	94.2	84.9	86.9	75.6
		PAFM (Yang et al., 2022)	NCA2022	55.1	42.3	76.4	68	82.5	–	–	–	95.6	88.5	**91.2**	80.1
		PDC (Su et al., 2017)	ICCV2017	–	–	–	–	–	–	–	–	84.14	63.41	–	–
		PGFL-KD (Zheng et al., 2021)	MM2021	63	54.1	80.7	70.3	85.1	90.8	74	86.7	95.3	87.2	89.6	79.5
		PGMANet (Zhai et al., 2021)	IJCNN2021	–	–	–	–	82.1	85.5	68.8	78.1	–	–	–	–
		PMFB (Miao et al., 2021)	TNNLS2021	56.3	43.5	–	–	72.5	83	70.6	81.3	92.7	81.3	86.2	72.6
		PVPM (Gao, Wang, Lu, & Liu, 2020)	CVPR2020	–	–	70.4	61.2	78.3	–	–	–	–	–	–	–
	Semantic segmentation	Co-Attention (Lin & Wang, 2021)	ICIP2021	–	–	–	–	83	90.3	73.1	83.2	–	–	–	–
		HPNet (Huang et al., 2020)	ICME2020	–	–	**87.3**	**77.4**	85.7	–	68.9	80.7	–	–	–	–
		MMGA (Cai et al., 2019)	CVPR2019	–	–	–	–	–	–	–	–	95	87.2	89.5	78.1
		SGSFA (Ren, Zhang, & Bao, 2020)	PMLR2020	62.3	47.4	63.1	53.2	68.2	–	–	–	92.3	80.2	84.7	70.8
		SORN (Zhang et al., 2020)	TCSVT2020	57.6	46.3	–	–	76.7	84.3	79.8	86.6	94.8	84.5	86.9	74.1
		SPReID (Kalayeh et al., 2018)	CVPR2017	–	–	–	–	–	–	–	–	94.63	**90.96**	88.96	**84.99**
	Attribute annotation	ASAN (Jin et al., 2021)	TCSVT2021	55.40	43.80	82.50	71.80	**86.80**	93.50	**81.70**	**88.30**	94.60	85.30	87.50	76.30
	Mixing method	FGSA (Zhou et al., 2020)	TIP2020	–	–	–	–	–	–	–	–	91.50	85.40	85.90	74.10
		GASM (He & Liu, 2020)	ECCV2020	–	–	80.30	73.10	–	–	–	–	95.30	84.70	88.30	74.40
		PGFA (Miao et al., 2019)	ICCV2019	–	–	–	–	68.80	80.00	69.10	80.90	91.20	76.80	82.60	65.50
		SSPReID (Quispe & Pedrini, 2019)	IVC 2019	–	–	–	–	–	–	–	–	93.70	90.80	86.40	83.70
		LKWS (Yang et al., 2021)	ICCV2021	62.2	46.3	81	71	85.7	**93.7**	80.7	88.2	–	–	–	–
		TSA (Gao, Zhang, et al., 2020)	ACM MM2020	–	–	–	–	72.70	85.20	73.90	84.70	–	–	–	–

Table 3
Comparison of experimental results based on relationship representation methods. The red numbers indicate the best results.(in %).

		Method	Venue	Occluded-Duke		Occluded-REID		Partial-REID		Partial-iLIDS		Market1501		DukeMTMC-reID	
				Rank-1	mAP	Rank-1	mAP	Rank-1	Rank-3	Rank-1	Rank-3	Rank-1	mAP	Rank-1	mAP
Relationship representation	Shared area	KBFM (Han et al., 2020)	ICIP2020	–	–	–	–	69.7	82.2	64.1	73.9	–	–	–	–
		PPCL (He, Shen, et al., 2021)	CVPR2021	–	–	–	–	83.70	88.70	71.40	85.70	–	–	–	–
		VPM (Sun et al., 2019)	CVPR2019	–	–	–	–	67.70	81.90	65.50	74.80	90.40	75.70	–	–
	Clustering	ISP (Zhu et al., 2020)	ECCV2020	62.80	52.30	–	–	–	–	–	–	94.63	90.69	88.96	84.99
	Figure convolution	HOReID (Wang, Yang, et al., 2020)	CVPR2020	–	–	**80.3**	**70.2**	85.3	91	72.6	86.4	94.2	84.9	86.9	75.6
	Attention	AACN (Xu et al., 2018)	CVPR2018	–	–	–	–	–	–	–	–	88.69	82.96	76.84	59.25
		APN (Huo, Song, Liu, & Zhang, 2021)	ICPR2021	–	–	–	–	71.80	85.50	66.40	76.50	**96.00**	89.00	89.50	79.20
		CASN (Zheng, Karanam, Wu, & Radke, 2019)	CVPR2019	–	–	–	–	–	–	–	–	94.40	82.80	87.70	73.70
		Co-Attention (Lin & Wang, 2021)	ICIP2021	–	–	–	–	83.00	90.30	73.10	83.20	–	–	–	–
		DAReID (Xu et al., 2021)	KBS2021	63.4	–	–	–	68.1	79.5	76.7	85.3	94.6	87	88.9	78.4
		DSOP (Wang, Qi, et al., 2020)	JPCS2020	57.70	45.30	–	–	–	–	–	–	95.40	85.90	88.20	77.00
		MHSA-Net (Tan, Liu, et al., 2022)	TNNLS2022	59.70	44.80	–	–	**85.70**	**91.30**	74.90	**87.20**	95.50	**93.00**	90.70	**87.20**
		OCNet (Kim et al., 2022)	ICASSP2022	64.30	**54.40**	–	–	–	–	–	–	–	–	–	–
		PAFM (Yang et al., 2022)	NCA2022	55.1	42.3	76.4	68	82.5	–	–	–	95.6	88.5	**91.2**	80.1
		PISNet (Zhao et al., 2020)	ECCV2020	–	–	–	–	–	–	–	–	95.60	87.10	88.80	78.70
		PSE (Sarfraz, Schumann, Eberle, & Stiefelhagen, 2018)	CVPR2018	–	–	–	–	–	–	–	–	90.30	84.00	85.20	79.80
		QPM (Wang, Ding, et al., 2022)	ITM2022	**64.40**	49.70	–	–	81.70	88.00	**77.30**	85.70	–	–	–	–
		VPM (Sun et al., 2019)	CVPR2019	–	–	–	–	67.70	81.90	65.50	74.80	90.40	75.70	–	–

Occluded-REID dataset, HPNet (Huang et al., 2020) based on local feature learning, HOReID (Wang, Yang, et al., 2020) based on relational representation, FED (Wang, Zhu, et al., 2022) and PFD (Wang, Liu, et al., 2022) based on transformer perform stably. On the Partial-ReID dataset, ASAN (Jin et al., 2021) and LKWS (Yang et al., 2021) based on local feature learning, MHSA-Net (Tan, Liu, et al., 2022) based on relational representation, and FRT (Xu et al., 2022) based on transformer achieve better performance. On the Partial-iLIDS dataset, ASAN (Jin et al., 2021) based on local feature learning, MHSA-Net (Tan, Liu, et al., 2022) and QPM (Wang, Ding, et al., 2022) based on relational representation, ACSAP (He, Yang, & Chen, 2021) based on region reconstruction, and OAMN (Chen, Liu, et al., 2021) based on data augmentation achieve very stable performance. Therefore, there is no single method that achieves the best performance on all datasets.

(2) In general, the performance on the occlusion dataset reflects the ability of the model to resist noise, the performance on the partial

Table 4
Comparison of experimental results with other methods. The red numbers indicate the best results.(in %).

	Method	Venue	Occluded-Duke		Occluded-REID		Partial-REID		Partial-iLIDS		Market1501		DukeMTMC-reID	
			Rank-1	mAP	Rank-1	mAP	Rank-1	Rank-3	Rank-1	Rank-3	Rank-1	mAP	Rank-1	mAP
Transformer	DPM (Tan, Dai, Ji, & Wu, 2022)	ACM MM 2022	**71.40**	**61.80**	85.50	79.70	–	–	–	–	95.50	89.70	91.00	82.60
	FED (Wang, Zhu, et al., 2022)	CVPR2022	68.10	56.40	**86.30**	79.30	83.10	–	–	–	95.00	86.30	89.40	78.00
	FRT (Xu et al., 2022)	TIP2022	70.70	61.30	80.40	71.00	**88.20**	**93.20**	73.00	87.00	95.50	88.10	90.50	81.70
	TransReID (He, Luo, et al., 2021)	ICCV2021	66.40	59.20	–	–	–	–	–	–	95.20	88.90	90.70	82.00
	PFD (Wang, Liu, et al., 2022)	AAAI2022	69.50	**61.80**	81.50	**83.00**	–	–	–	–	95.50	89.70	**91.20**	83.20
	PFT (Zhao et al., 2022)	NCA2022	69.80	60.80	83.00	78.30	81.30	79.90	68.10	81.50	95.30	88.80	90.70	82.10
Composite method	DRL-Net (Jia et al., 2022)	TMM2021	65.80	53.90	–	–	–	–	–	–	94.70	86.90	88.10	76.60
	Pirt (Ma et al., 2021)	ACMMM2021	60.00	50.90	–	–	–	–	–	–	94.10	86.30	88.90	77.60
Multiscale	DSR (He et al., 2018)	CVPR2018	40.80	30.40	72.80	62.80	50.70	70.00	58.80	67.20	83.58	64.25	–	–
	FPR (He et al., 2019)	ICCV2019	–	–	78.30	68.00	81.00	–	68.10	–	95.42	86.58	88.64	78.42
Regional reconfiguration	ACSAP (He, Yang, & Chen, 2021)	ICIP2021	–	–	–	–	77.00	83.70	76.50	**87.40**	–	–	–	–
	RFCNet (Hou et al., 2021)	TPAMI2021	63.90	54.50	–	–	–	–	–	–	95.20	89.20	90.70	80.70
Data enhancement	IGOAS (Zhao et al., 2021)	TIP2021	60.10	49.40	81.10	–	–	–	–	–	93.40	84.10	86.90	75.10
	OAMN (Chen, Liu, et al., 2021)	ICCV2021	62.60	46.10	–	–	86.00	–	**77.30**	–	93.20	79.80	86.30	72.60
	SSGR (Yan et al., 2021)	ICCV2021	65.80	57.20	78.50	72.90	–	–	–	–	**96.10**	89.30	91.10	81.30
Regularization	MHSA-Net (Tan, Liu, et al., 2022)	TNNLS2022	59.70	44.80	–	–	85.70	91.30	74.90	87.20	95.50	**93.00**	90.70	**87.20**

dataset reflects the recognition ability of the model under the condition of missing pedestrian information, and the performance on the general dataset reflects the comprehensive performance of the model. Each of these approaches addresses one or more specific problems.

(3) Attention and pose estimation are the more mainstream and typical of the many pedestrian re-identification methods for dealing with occlusion. Attribute annotation-based, clustering-based and figure convolution-based methods, on the other hand, have received less attention.

6. Future directions

6.1. Richer, higher quality datasets

While most models undergo evaluation using datasets gathered in controlled environments, it's crucial to acknowledge that real-world scenarios present uncontrollable variables that can profoundly impact model performance in such contexts. For datasets focused on occluded person Re-identification (Re-ID), it becomes imperative to incorporate one or more modal inputs. This should encompass a diverse array of data types, including images, textual information, infrared and depth maps. This multifaceted approach equips models to adeptly navigate a wider spectrum of realistic occlusion challenges. Moreover, a pressing issue in the field is the scarcity of extensive datasets that span a broader range of domains, encompass varying environmental conditions (Gou et al., 2018), and offer higher resolutions. The availability of such datasets would be instrumental in providing researchers with richer, more diverse content and superior data quality for their investigations.

6.2. More robust and varied feature extraction

3D Re-ID. Inspired by human three-dimensional cognition, some researchers advocate for a holistic pedestrian representation that combines both 3D and 2D modalities (Zheng et al., 2022). Currently, PointNet (Qi et al., 2017), a prominent deep learning method for point cloud feature extraction (C.S., H., X., S.W., & W.J., 2022; Wang, Ning, et al., 2022), has exhibited promising outcomes. Incorporating techniques like point cloud completion (Fei et al., 2022) and point cloud correction can be beneficial for addressing challenges in 3D occluded person Re-ID. Additionally, leveraging 3D pose estimation (Wang et al., 2021) and 3D semantic segmentation (Xie, Tian, & Zhu, 2020) can guide the feature extraction process in person Re-ID. Nonetheless, research in the 3D domain for pedestrian recognition (Sun et al., 2019; Zhao, Ouyang, & Wang, 2013) remains relatively limited compared to the progress made in 2D approaches. Hence, 3D occluded person Re-ID stands as a significant and promising research avenue for the future (Tirkolaee, Goli, & Weber, 2020).

Multimodal Re-ID. The information captured from different modalities demonstrates a significant diversity in content representation (Wu, Zheng, & Lai, 2017; Wu, Zheng, Yu, et al., 2017). Improving the interaction, fusion, and extraction of more comprehensive pedestrian features at both the data and feature extraction stages represents a crucial research direction for future advancements (Sekhar, Nagaraju, & Yu, 2017; Tutsoy & Tanrikulu, 2022).

Cross-resolution occluded person Re-ID. Owing to the influence of the distance and pixel size of the collection device, the resolution of the collected samples is uneven, and the feature space correspondence is also inconsistent (Li, Chen, Lin, Du, & Wang, 2019). At the same time, low resolution will lose significant spatial and detail information (Mao, Zhang, & Yang, 2019). How to extract pedestrian features at different resolutions under occlusion conditions is a problem to be solved in the future.

Unsupervised and semi-supervised occluded person Re-ID. The complex manual labeling process is omitted, and the pedestrian features are learned by using the datasets without labels (Liu, Zha, Hong, Wang, & Zhang, 2019; Zhang & Lu, 2018) or with a small number of labels (Nagaraju, Raju, Ko, & Yu, 2016; Wang, Wang, Zheng, Chuang, & Satoh, 2019; Wang, Zhang, et al., 2019). Currently, the performance of unsupervised and semi-supervised methods in the realm of occluded person Re-ID still falls short when compared to supervised methods. Supervised techniques typically rely on extensive labeled datasets for training, resulting in high performance. However, as unsupervised and semi-supervised approaches gain traction, they exhibit substantial potential for enhancing the generalization of occluded person Re-ID models.

6.3. Occluded person Re-ID system

At present, few researchers combine object detection and occluded person Re-ID together. The end-to-end person Re-ID systems are lacking, and the integrated system has more applications in real life (Martinel, Das, Micheloni, & Roy-Chowdhury, 2016). How to combine the two more effectively and rationally and design a occluded person Re-ID system that is more robust to occlusion is an important research direction.

7. Conclusion

This review presents a comprehensive and integrated analysis and discussion of deep learning methods for occluded person Re-ID, addressing both practical and research-driven requirements. Firstly, we provide an overview of occlusion problems and highlight datasets specifically designed for occluded person Re-ID. Secondly, we systematically categorize and introduce methods proposed in top international journals and conferences up to 2023 for tackling occluded person Re-ID challenges. Finally, we analyze the future prospects of occluded person Re-ID, considering data, feature, and system perspectives, respectively. In this study, we classify the most significant image feature extraction methods into five major categories: local feature learning, relational representation, transformer-based methods, mixing methods, and other approaches. This review aims to aid researchers in understanding the methodologies and goals of these methods, offering valuable references, and contributing to the research significance in advancing occluded Re-ID.

CRediT authorship contribution statement

Enhao Ning: Data curation, Literature analysis, Interpretation of results, Preparation of the manuscript. **Changshuo Wang:** Data curation, Literature analysis, Interpretation of results, Preparation of the manuscript. **Huang Zhang:** Data curation, Literature analysis, Interpretation of results, Preparation of the manuscript. **Xin Ning:** Data curation, Literature analysis, Interpretation of results, Preparation of the manuscript. **Prayag Tiwari:** Data curation, Literature analysis, Interpretation of results, Preparation of the manuscript.

Declaration of competing interest

The authors declare no conflict of interests.

Data availability

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62373343). Beijing Natural Science Foundation (No. L233036).

References

Bedagkar-Gala, A., & Shah, S. K. (2014). A survey of approaches and trends in person re-identification. *Image and Vision Computing*, *32*(4), 270–286.

Cai, H., Wang, Z., & Cheng, J. (2019). Multi-scale body-part mask guided attention for person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*.

Chen, J., Jiang, X., Wang, F., Zhang, J., Zheng, F., Sun, X., et al. (2021). Learning 3D shape feature for texture-insensitive person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8146–8155).

Chen, P., Liu, W., Dai, P., Liu, J., Ye, Q., Xu, M., et al. (2021). Occlude them all: Occlusion-aware attention network for occluded person re-id. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 11833–11842).

Cheng, D. S., Cristani, M., Stoppa, M., Bazzani, L., & Murino, V. (2011). Custom pictorial structures for re-identification. In *Bmvc, vol. 1, no. 2* (p. 6). Citeseer.

C.S., W., H., W., X., N., S.W., T., & W.J., L. (2022). 3D point cloud classification method based on dynamic coverage of local area. *Journal of Software*.

Dong, N., Zhang, L., Yan, S., Tang, H., & Tang, J. (2023). Erasing, transforming, and noising defense network for occluded person re-identification. arXiv preprint arXiv:2307.07187.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Ess, A., Leibe, B., Schindler, K., & Van Gool, L. (2008). A mobile vision system for robust multi-person tracking. In *2008 IEEE conference on computer vision and pattern recognition* (pp. 1–8). IEEE.

Farooq, A., Awais, M., Kittler, J., & Khalid, S. S. (2022). AXM-net: Implicit cross-modal feature alignment for person re-identification. In *Proceedings of the AAAI conference on artificial intelligence, vol. 36, no. 4* (pp. 4477–4485).

Fei, B., Yang, W., Chen, W.-M., Li, Z., Li, Y., Ma, T., et al. (2022). Comprehensive review of deep learning-based 3D point cloud completion processing and analysis. *IEEE Transactions on Intelligent Transportation Systems*.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., et al. (2019). Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3146–3154).

Gao, S., Wang, J., Lu, H., & Liu, Z. (2020). Pose-guided visible part matching for occluded person reid. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11744–11752).

Gao, L., Zhang, H., Gao, Z., Guan, W., Cheng, Z., & Wang, M. (2020). Texture semantically aligned with visibility-aware for partial person re-identification. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 3771–3779).

Gou, M., Wu, Z., Rates-Borras, A., Camps, O., Radke, R. J., et al. (2018). A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(3), 523–536.

Güler, R. A., Neverova, N., & Kokkinos, I. (2018). Densepose: Dense human pose estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7297–7306).

Hafner, F. M., Bhuyian, A., Kooij, J. F., & Granger, E. (2022). Cross-modal distillation for RGB-depth person re-identification. *Computer Vision and Image Understanding*, *216*, Article 103352.

Han, C., Gao, C., & Sang, N. (2020). Keypoint-based feature matching for partial person re-identification. In *2020 IEEE international conference on image processing* (pp. 226–230). IEEE.

He, L., Liang, J., Li, H., & Sun, Z. (2018). Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7073–7082).

He, L., & Liu, W. (2020). Guided saliency feature learning for person re-identification in crowded scenes. In *Computer Vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part XXVIII 16* (pp. 357–373). Springer.

He, S., Luo, H., Wang, P., Wang, F., Li, H., & Jiang, W. (2021). Transreid: Transformer-based object re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 15013–15022).

He, T., Shen, X., Huang, J., Chen, Z., & Hua, X.-S. (2021). Partial person re-identification with part-part correspondence learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9105–9115).

He, L., Wang, Y., Liu, W., Zhao, H., Sun, Z., & Feng, J. (2019). Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8450–8459).

He, Y., Yang, H., & Chen, L. (2021). Adversarial cross-scale alignment pursuit for seriously misaligned person re-identification. In *2021 IEEE international conference on image processing* (pp. 2373–2377). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hou, R., Ma, B., Chang, H., Gu, X., Shan, S., & Chen, X. (2021). Feature completion for occluded person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(9), 4894–4912.

Huang, H., Chen, X., & Huang, K. (2020). Human parsing based alignment with multi-task learning for occluded person re-identification. In *2020 IEEE international conference on multimedia and expo* (pp. 1–6). IEEE.

Huo, L., Song, C., Liu, Z., & Zhang, Z. (2021). Attentive part-aware networks for partial person re-identification. In *2020 25th international conference on pattern recognition* (pp. 3652–3659). IEEE.

Jia, M., Cheng, X., Lu, S., & Zhang, J. (2022). Learning disentangled representation implicitly via transformer for occluded person re-identification. *IEEE Transactions on Multimedia*.

Jin, H., Lai, S., & Qian, X. (2021). Occlusion-sensitive person re-identification via attribute-based shift attention. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(4), 2170–2185.

Kalayeh, M. M., Basaran, E., Gökmen, M., Kamasak, M. E., & Shah, M. (2018). Human semantic parsing for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1062–1071).

Kanazawa, A., Black, M. J., Jacobs, D. W., & Malik, J. (2018). End-to-end recovery of human shape and pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7122–7131).

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys (CSUR)*, *54*(10s), 1–41.

Kim, M., Cho, M., Lee, H., Cho, S., & Lee, S. (2022). Occluded person re-identification via relational adaptive feature correction learning. In *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing* (pp. 2719–2723). IEEE.

Kim, J., & Yoo, C. D. (2017). Deep partial person re-identification via attention model. In *2017 IEEE international conference on image processing* (pp. 3425–3429). IEEE.

Kiran, M., Praveen, R. G., Nguyen-Meidine, L. T., Belharbi, S., Blais-Morin, L.-A., & Granger, E. (2021). Holistic guidance for occluded person re-identification. arXiv preprint arXiv:2104.06524.

Lavi, B., Ullah, I., Fatan, M., & Rocha, A. (2020). Survey on reliable deep learning-based person re-identification models: Are we there yet? arXiv preprint arXiv:2005.00355.

Leng, Q., Ye, M., & Tian, Q. (2019). A survey of open-world person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(4), 1092–1108.

Li, Y.-J., Chen, Y.-C., Lin, Y.-Y., Du, X., & Wang, Y.-C. F. (2019). Recover and identify: A generative dual model for cross-resolution person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8090–8099).

Li, Y., Jiang, X., & Hwang, J.-N. (2020). Effective person re-identification by self-attention model guided feature learning. *Knowledge-Based Systems*, *187*, Article 104832.

Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 152–159).

Liang, X., Gong, K., Shen, X., & Lin, L. (2018). Look into person: Joint body parsing & pose estimation network and a new benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(4), 871–885.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft coco: Common objects in context. In *Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, September 6-12, 2014, proceedings, Part V 13* (pp. 740–755). Springer.

Lin, C.-S., & Wang, Y.-C. F. (2021). Self-supervised bodymap-to-appearance co-attention for partial person re-identification. In *2021 IEEE international conference on image processing* (pp. 2299–2303). IEEE.

Liu, J., Zha, Z.-J., Hong, R., Wang, M., & Zhang, Y. (2019). Deep adversarial graph attention convolution network for text-based person search. In *Proceedings of the 27th ACM international conference on multimedia* (pp. 665–673).

Ma, Z., Zhao, Y., & Li, J. (2021). Pose-guided inter-and intra-part relational transformer for occluded person re-identification. In *Proceedings of the 29th ACM international conference on multimedia* (pp. 1487–1496).

Mao, S., Zhang, S., & Yang, M. (2019). Resolution-invariant person re-identification. arXiv preprint arXiv:1906.09748.

Martinel, N., Das, A., Micheloni, C., & Roy-Chowdhury, A. K. (2016). Temporal model adaptation for person re-identification. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part IV 14* (pp. 858–877). Springer.

Miao, J., Wu, Y., Liu, P., Ding, Y., & Yang, Y. (2019). Pose-guided feature alignment for occluded person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 542–551).

Miao, J., Wu, Y., & Yang, Y. (2021). Identifying visible parts via pose estimation for occluded person re-identification. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(9), 4624–4634.

Ming, Z., Zhu, M., Wang, X., Zhu, J., Cheng, J., Gao, C., et al. (2022). Deep learning-based person re-identification methods: A survey and outlook of recent works. *Image and Vision Computing*, *119*, Article 104394.

Nagaraju, G., Raju, G. S. R., Ko, Y. H., & Yu, J. S. (2016). Hierarchical Ni–Co layered double hydroxide nanosheets entrapped on conductive textile fibers: A cost-effective and flexible electrode for high-performance pseudocapacitors. *Nanoscale*, *8*(2), 812–825.

Nguyen, D. T., Hong, H. G., Kim, K. W., & Park, K. R. (2017). Person recognition system based on a combination of body images from visible light and thermal cameras. *Sensors*, *17*(3), 605.

Peng, Y., Hou, S., Cao, C., Liu, X., Huang, Y., & He, Z. (2022). Deep learning-based occluded person re-identification: A survey. arXiv preprint arXiv:2207.14452.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652–660).

Quispe, R., & Pedrini, H. (2019). Improved person re-identification based on saliency and semantic parsing with deep neural network models. *Image and Vision Computing*, *92*, Article 103809.

Ren, X., Zhang, D., & Bao, X. (2020). Semantic-guided shared feature alignment for occluded person re-identification. In *Asian conference on machine learning* (pp. 17–32). PMLR.

Ristani, E., Solera, F., Zou, R., Cucchiara, R., & Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. In *Computer vision–ECCV 2016 workshops: Amsterdam, the Netherlands, October 8-10 and 15-16, 2016, proceedings, Part II* (pp. 17–35). Springer.

Sarfraz, M. S., Schumann, A., Eberle, A., & Stiefelhagen, R. (2018). 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* (pp. 420–429).

Sekhar, S. C., Nagaraju, G., & Yu, J. S. (2017). Conductive silver nanowires-fenced carbon cloth fibers-supported layered double hydroxide nanosheets as a flexible and binder-free electrode for high-performance asymmetric supercapacitors. *Nano Energy*, *36*, 58–67.

Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., et al. (2023). Transformers in medical imaging: A survey. *Medical Image Analysis*, Article 102802.

Su, C., Li, J., Zhang, S., Xing, J., Gao, W., & Tian, Q. (2017). Pose-driven deep convolutional model for person re-identification. In *Proceedings of the IEEE international conference on computer vision* (pp. 3960–3969).

Sun, Y., Xu, Q., Li, Y., Zhang, C., Li, Y., Wang, S., et al. (2019). Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 393–402).

Sun, X., & Zheng, L. (2019). Dissecting person re-identification from the viewpoint of viewpoint. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 608–617).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).

Tan, L., Dai, P., Ji, R., & Wu, Y. (2022). Dynamic prototype mask for occluded person re-identification. In *Proceedings of the 30th ACM international conference on multimedia* (pp. 531–540).

Tan, H., Liu, X., Bian, Y., Wang, H., & Yin, B. (2021). Incomplete descriptor mining with elastic loss for person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(1), 160–171.

Tan, H., Liu, X., Yin, B., & Li, X. (2022). MHSA-Net: Multihead self-attention network for occluded person re-identification. *IEEE Transactions on Neural Networks and Learning Systems*.

Tirkolaee, E. B., Goli, A., & Weber, G.-W. (2020). Fuzzy mathematical programming and self-adaptive artificial fish swarm algorithm for just-in-time energy-aware flow shop scheduling problem with outsourcing option. *IEEE Transactions on Fuzzy Systems*, *28*(11), 2772–2783.

Tutsoy, O., & Tanrikulu, M. Y. (2022). Priority and age specific vaccination algorithm for the pandemic diseases: A comprehensive parametric prediction model. *BMC Medical Informatics and Decision Making*, *22*(1), 4.

Wang, P., Ding, C., Shao, Z., Hong, Z., Zhang, S., & Tao, D. (2022). Quality-aware part models for occluded person re-identification. *IEEE Transactions on Multimedia*.

Wang, Y., Liang, X., & Liao, S. (2022). Cloning outfits from real-world images to 3D characters for generalizable person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4900–4909).

Wang, T., Liu, H., Song, P., Guo, T., & Shi, W. (2022). Pose-guided feature disentangling for occluded person re-identification based on transformer. In *Proceedings of the AAAI conference on artificial intelligence, vol. 36, no. 3* (pp. 2540–2549).

Wang, C., Ning, X., Sun, L., Zhang, L., Li, W., & Bai, X. (2022). Learning discriminative features by covering local geometric space for point cloud analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–15.

Wang, Q., Qi, M., Jin, K., & Jiang, J. (2020). Deep-shallow occlusion parallelism network for person re-identification. *Journal of Physics: Conference Series*, *1518*(1), Article 012026.

Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., et al. (2021). Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding*, *210*, Article 103225.

Wang, Z., Wang, Z., Wu, Y., Wang, J., & Satoh, S. (2019). Beyond intra-modality discrepancy: A comprehensive survey of heterogeneous person re-identification. *4*, arXiv preprint arXiv:1905.10048.

Wang, Z., Wang, Z., Zheng, Y., Chuang, Y.-Y., & Satoh, S. (2019). Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 618–626).

Wang, G., Yang, S., Liu, H., Wang, Z., Yang, Y., Wang, S., et al. (2020). 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* (pp. 6449–6458).

Wang, G., Zhang, T., Cheng, J., Liu, S., Yang, Y., & Hou, Z. (2019). RGB-infrared cross-modality person re-identification via joint pixel and feature alignment. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3623–3632).

Wang, Z., Zhu, F., Tang, S., Zhao, R., He, L., & Song, J. (2022). Feature erasing and diffusion network for occluded person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4754–4763).

Wen, X., Feng, X., Li, P., & Chen, W. (2022). Cross-modality collaborative learning identified pedestrian. *The Visual Computer*, 1–16.

Wu, Y., Lin, Y., Dong, X., Yan, Y., Ouyang, W., & Yang, Y. (2018). Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5177–5186).

Wu, A., Zheng, W.-S., & Lai, J.-H. (2017). Robust depth-based person re-identification. *IEEE Transactions on Image Processing*, *26*(6), 2588–2603.

Wu, A., Zheng, W.-S., Yu, H.-X., Gong, S., & Lai, J. (2017). RGB-infrared cross-modality person re-identification. In *Proceedings of the IEEE international conference on computer vision* (pp. 5380–5389).

Xie, Y., Tian, J., & Zhu, X. X. (2020). Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geoscience and Remote Sensing Magazine*, *8*(4), 38–59.

Xu, B., He, L., Liang, J., & Sun, Z. (2022). Learning feature recovery transformer for occluded person re-identification. *IEEE Transactions on Image Processing*, *31*, 4651–4662.

Xu, Y., Zhao, L., & Qin, F. (2021). Dual attention-based method for occluded person re-identification. *Knowledge-Based Systems*, *212*, Article 106554.

Xu, J., Zhao, R., Zhu, F., Wang, H., & Ouyang, W. (2018). Attention-aware compositional network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2119–2128).

Yaghoubi, E., Kumar, A., & Proença, H. (2021). Sss-pr: A short survey of surveys in person re-identification. *Pattern Recognition Letters*, *143*, 50–57.

Yan, C., Pang, G., Jiao, J., Bai, X., Feng, X., & Shen, C. (2021). Occluded person re-identification with single-scale global representations. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 11875–11884).

Yang, J., Zhang, C., Tang, Y., & Li, Z. (2022). PAFM: Pose-drive attention fusion mechanism for occluded person re-identification. *Neural Computing and Applications*, *34*(10), 8241–8252.

Yang, J., Zhang, J., Yu, F., Jiang, X., Zhang, M., Sun, X., et al. (2021). 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* (pp. 11885–11894).

Ye, M., Chen, C., Shen, J., & Shao, L. (2021). Dynamic tri-level relation mining with attentive graph for visible infrared re-identification. *IEEE Transactions on Information Forensics and Security*, *17*, 386–398.

Ye, M., Shen, J., J. Crandall, D., Shao, L., & Luo, J. (2020). Dynamic dual-attentive aggregation learning for visible-infrared person re-identification. In *Computer Vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part XVII 16* (pp. 229–247). Springer.

Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. (2021). Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(6), 2872–2893.

Zhai, Y., Han, X., Ma, W., Gou, X., & Xiao, G. (2021). PGMANet: Pose-guided mixed attention network for occluded person re-identification. In *2021 international joint conference on neural networks* (pp. 1–8). IEEE.

Zhang, G., Chen, C., Chen, Y., Zhang, H., & Zheng, Y. (2022). Fine-grained-based multi-feature fusion for occluded person re-identification. *Journal of Visual Communication and Image Representation*, *87*, Article 103581.

Zhang, L., Guo, H., Zhu, K., Qiao, H., Huang, G., Zhang, S., et al. (2022). Hybrid modality metric learning for visible-infrared person re-identification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*(1s), 1–15.

Zhang, Z., Lan, C., Zeng, W., & Chen, Z. (2019). Densely semantically aligned person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 667–676).

Zhang, Y., & Lu, H. (2018). Deep cross-modal projection learning for image-text matching. In *Proceedings of the European conference on computer vision* (pp. 686–701).

Zhang, X., Yan, Y., Xue, J.-H., Hua, Y., & Wang, H. (2020). Semantic-aware occlusion-robust network for occluded person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(7), 2764–2778.

Zhao, S., Gao, C., Zhang, J., Cheng, H., Han, C., Jiang, X., et al. (2020). Do not disturb me: Person re-identification under the interference of other pedestrians. In *Computer Vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part VI 16* (pp. 647–663). Springer.

Zhao, C., Lv, X., Dou, S., Zhang, S., Wu, J., & Wang, L. (2021). Incremental generative occlusion adversarial suppression network for person reid. *IEEE Transactions on Image Processing*, *30*, 4212–4224.

Zhao, R., Ouyang, W., & Wang, X. (2013). Unsupervised salience learning for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3586–3593).

Zhao, Y., Zhu, S., Wang, D., & Liang, Z. (2022). Short range correlation transformer for occluded person re-identification. *Neural Computing and Applications*, *34*(20), 17633–17645.

Zheng, W.-S., Gong, S., & Xiang, T. (2011). Person re-identification by probabilistic relative distance comparison. In *CVPR 2011* (pp. 649–656). IEEE.

Zheng, M., Karanam, S., Wu, Z., & Radke, R. J. (2019). Re-identification with consistent attentive siamese networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5735–5744).

Zheng, K., Lan, C., Zeng, W., Liu, J., Zhang, Z., & Zha, Z.-J. (2021). Pose-guided feature learning with knowledge distillation for occluded person re-identification. In *Proceedings of the 29th ACM international conference on multimedia* (pp. 4537–4545).

Zheng, W.-S., Li, X., Xiang, T., Liao, S., Lai, J., & Gong, S. (2015). Partial person re-identification. In *Proceedings of the IEEE international conference on computer vision* (pp. 4678–4686).

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision* (pp. 1116–1124).

Zheng, Z., Wang, X., Zheng, N., & Yang, Y. (2022). Parameter-efficient person re-identification in the 3D space. *IEEE Transactions on Neural Networks and Learning Systems*.

Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). Person re-identification: Past, present and future. arXiv preprint arXiv:1610.02984.

Zheng, Z., Zheng, L., & Yang, Y. (2017). Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE international conference on computer vision* (pp. 3754–3762).

Zhou, Q., Zhong, B., Lan, X., Sun, G., Zhang, Y., Zhang, B., et al. (2020). Fine-grained spatial alignment model for person re-identification with focal triplet loss. *IEEE Transactions on Image Processing*, *29*, 7578–7589.

Zhu, K., Guo, H., Liu, Z., Tang, M., & Wang, J. (2020). 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* (pp. 346–363). Springer.

Zhuo, J., Chen, Z., Lai, J., & Wang, G. (2018). Occluded person re-identification. In *2018 IEEE international conference on multimedia and expo* (pp. 1–6). IEEE.

Zhuo, J., Lai, J., & Chen, P. (2019). A novel teacher-student learning framework for occluded person re-identification. arXiv preprint arXiv:1907.03253.