Research and Performance Analysis of Person Re-Identification Methods in Occluded Scenarios

Yitong Yu

{13880367070@163.com}

Shude High School (Guanghua Campus), Chengdu, Sichuan 610091, China

Abstract. Pedestrian re-identification (ReID) has gradually become a core supporting technology in intelligent security systems. However, its practical deployment is still significantly hampered by occlusion issues. To tackle this key challenge of occluded person ReID, this paper reviews and categorizes existing research methods into three major types: data driven, model driven, and auxiliary information-driven approaches. Data-driven methods simulate real occlusion scenes and implement diverse data augmentation strategies to effectively compensate for the shortage of occluded samples in training data. Model driven methods focus on optimizing the feature extraction process removing interference from occluded regions via feature dropping, and enhancing effective feature representations via feature strengthening to improve the model's occlusion resistance. Auxiliary information-driven methods leverage external techniques such as pose estimation and semantic parsing to accurately locate visible pedestrian regions, thereby facilitating effective feature matching. To objectively evaluate the performance of typical algorithms, this paper summarizes and compares their experimental results on four mainstream benchmarks based on published literature. It is shown that feature-level enhancement methods achieve outstanding performance in occluded scenarios and good generalization in non-occluded scenarios, yielding the best overall effectiveness.

Keywords: Person Re-Identification; Occluded Scenarios; Methods Analysis

1 Introduction

Driven by global digital transformation and the continuous advancement of smart city and safe city strategies, distributed video surveillance networks have evolved into critical urban public infrastructure. Although the global surveillance market reached $91.66 billion in 2025, with more than 1 billion cameras deployed and 150 EB of active storage capacity, traditional manual monitoring still suffers from low efficiency, high latency, and limited coverage. Consequently, the transition toward intelligent analysis systems based on computer vision has become inevitable [1]. Manual monitoring is no longer competent for large-scale video analysis, as it is inefficient, time-consuming, and prone to missing critical events. In practice, most surveillance video data remain unprocessed and unanalyzed. Therefore, exploiting

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365361

computer vision techniques to mine effective information from massive video data represents the only feasible and scalable development direction.

Person re-identification (ReID) serves as a core technique in intelligent surveillance systems. Its objective is to retrieve the identities of the same pedestrian across multiple non-overlapping cameras. In real-world scenarios, face recognition frequently fails under long-distance imaging, low-resolution conditions, or when pedestrians are observed from the back or side views. Person ReID effectively compensates for such limitations. Reliable ReID systems provide strong technical support for criminal tracking, missing-person investigation, and intelligent security early warning.

Existing person ReID methods have achieved promising performance under ideal laboratory conditions, where pedestrians are fully visible and free from obstructions. However, in real-world deployment scenarios, occlusion remains the most critical bottleneck restricting practical performance. Occlusion occurs randomly and is inherently uncontrollable [2], and can be divided into two typical categories.

First, static occlusion occurs when pedestrians are partially blocked by stationary objects [3,4,5]. In crowded scenes such as shopping malls, transportation stations, and urban intersections, pedestrians are frequently occluded by trees, vehicles, railings, or signboards. Such occlusion disrupts the complete semantic structure of pedestrian visual features, resulting in incomplete or fragmented feature representations, which further leads to the failure of global feature matching.

Second, inter-person occlusion caused by other pedestrians is more challenging and disruptive compared with static occlusion [6,7,8]. When interfering pedestrians overlap with the target individual, pedestrian detectors cannot accurately segment the human boundary, causing extraneous visual features to be mixed into the extracted feature representations. Such interference induces severe feature distribution shifts in the embedding space, misguides model decision-making, and significantly degrades recognition accuracy.

Furthermore, a severe data distribution mismatch exists between open-source training datasets and real-world testing scenarios. For example, in the widely used DukeMTMC-ReID dataset, only 9% of training samples contain occlusion, whereas 100% of the query set is occluded. Such a large domain gap makes learned models extremely fragile when facing local feature missing, and lack effective feature reconstruction, semantic recovery, and anti-interference capabilities.

To address occluded person re-identification, existing methods can be roughly divided into three categories: data-driven, model-driven, and auxiliary information-based methods, classified by how and when occlusion is handled. Preprocessing mainly adopts data-driven strategies to add noise to samples; feature extraction relies on model-driven schemes to reduce occlusion impacts; and post-extraction classification or retrieval also uses model-driven ways to enhance features. Data-driven methods can be further divided into image-level and feature-level; model-driven methods fall into occlusion suppression and feature enhancement. Detailed classification and analysis of specific methods are provided below.

2 Method Analysis

2.1 Data-Driven Occluded Person Re-Identification

Data augmentation is an effective way to improve model robustness in computer vision tasks, solving image retrieval and recognition problems from the perspective of adversarial defense. Due to the scarcity of occluded samples in training sets, common approaches enrich data by adding simulated occlusion, including random erasing [2], random pasting, rotation, etc.

Given the limited occluded training samples in large-scale person re-identification datasets such as DukeMTMC-ReID, PGFA [3] constructs a dataset partition tailored for occluded person re-identification on the basis of DukeMTMC-ReID, facilitating training and inference (Figure 1). FED [4] enhances model robustness to static occluders by adding manually cropped occlusion patches to training samples and performing soft binarization on occlusion regions to guide targeted feature erasing.

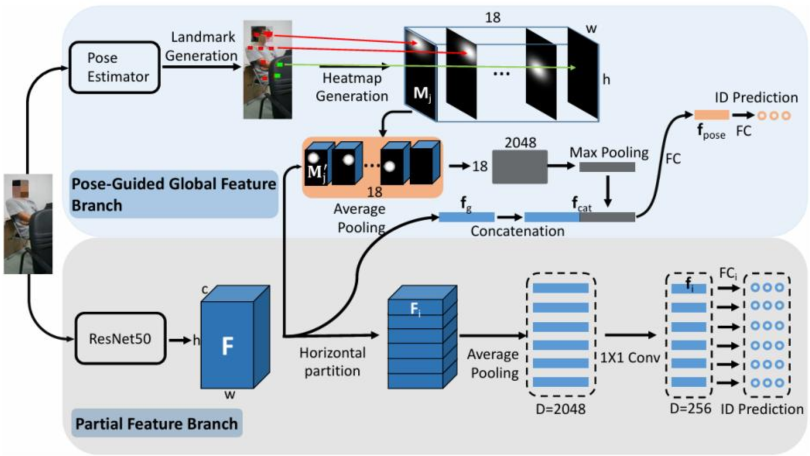

Fig. 1 Pipeline of the Pose-Guided Feature Alignment method[3]

To tackle the imbalance between heavily occluded queries and rarely occluded training samples, PADE [5] proposes a parallel augmentation mechanism that diversifies training samples with multiple augmentation strategies and extracts features via a shared backbone network for contrastive learning. LogoVit [29] converts partial image regions to grayscale through grayscale patch replacement and combines images and identities from two training samples into new samples via CutMix [6]. OSRTrans [7] aligns and mixes two samples through patch-aligned occlusion enhancement and provides binary masks to guide target pedestrian recognition. ADP [8] generates self-occlusion by pasting background patches onto auxiliary samples and produces occlusion masks for dual-path training. DPEFormer [9] leverages the Segment Anything Model (SAM) [10] to segment clear occluders from real street-view datasets and pastes them onto pedestrian edges to generate realistic occlusion. MDAE-HG [11] reconstructs occluded images via multi-scale autoencoders for noise reduction.

FED [4] specifically addresses multi-pedestrian scenes, where occlusion from non-target pedestrians is more disruptive than object occlusion. Inspired by a memory bank design [12], it diffuses pedestrian information from other images into current samples via a feature diffusion module, strengthening robustness to non-target pedestrian occlusion. SPT [13] segments pedestrian instances and background obstacles through salient patch selection and transfers instances onto different obstacles to generate realistic occluded samples (Figure 2). Inspired by generative adversarial networks (GAN), ETNDNet [14] and related works add noise to feature maps and conduct adversarial training between identity classifiers and feature extractors to force the model to focus on discriminative regions.

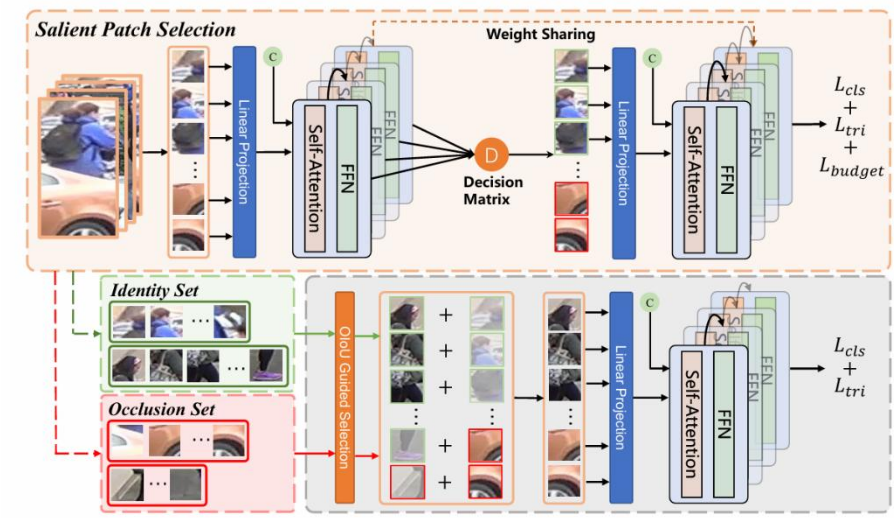

Fig. 2 The framework of Saliency-Guided Patch Transfer (SPT) [13].

2.2 Model-Driven Occluded Person Re-Identification

Model-driven methods improve occlusion robustness during training or inference by modifying model structures for occlusion-specific tasks. They generally follow two routes: reducing the influence of occluded regions via feature discarding, and enhancing extracted features to boost re-identification accuracy.

When the original visual content within occluded regions cannot be explicitly reconstructed, suppressing interference introduced by occluders constitutes a widely adopted strategy in occluded person ReID. FED [4] leverages occlusion masks generated during data augmentation to develop an occlusion erasing module, which explicitly detects and discards feature components contaminated by occlusions. OATrans employs an occlusion-aware grouping mechanism to dynamically disentangle discriminative features from non-discriminative ones, thereby down-weighting occluded regions and suppressing irrelevant noise. A series of related studies [15,16,17,18,19,20,21] focus on refining network architectures to capture multi-scale, multi-granularity local pedestrian features for enhanced robustness. Furthermore, numerous Transformer-based methods adopt token sparsification, dynamic selection, or structural reorganization to eliminate redundant image patches and reduce background computation, thereby improving training efficiency. In the context of occluded ReID, such schemes preferentially prune occluded patches with low discriminative contribution, which further alleviates the adverse effects of occlusion. DPEFormer [9], for instance, selects representative proxy tokens to discard uninformative patches and suppress occlusion-induced interference.

Other methods interact global and local features at different levels to learn deeper representations and enhance robustness. PADE [5] employs non-local computation to enhance local features with global information and then aggregates enhanced local features to refine global representations, achieving sufficient feature fusion. MVI2P [23] builds a perception and localization module using class activation maps (CAMs) [17] to selectively integrate discriminative image regions (Figure 3).

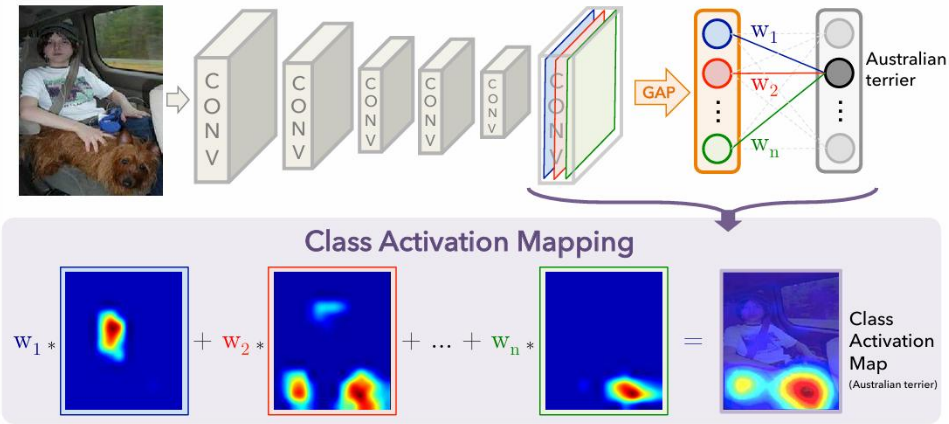

Fig.3. Class Activation Mapping [17]

Some works [18] fuse CNN local features and Transformer global features to obtain discriminative representations, while others [26] learn discriminative local features under global semantic guidance. FRT [27] divides pedestrians into four parts, ranks images via graph convolutional networks (GCN) [19], and supplements features by feeding k-nearest neighbors and the current sample into Transformer layers. LogoVit [20] modifies the JPM module of TransReID [21] with a hierarchical structure for grouped alignment to extract fine-grained local features. LGFNet [22] generates spatial attention maps with adaptive receptive fields to jointly learn global-local features, effectively alleviating weak discriminability and misalignment of body parts.

2.3 Occluded Person Re-Identification Based on Auxiliary Information

Auxiliary information refers to results processed by models external to the main framework. Many methods solve occlusion using pose estimation or semantic parsing models. For example, PFD uses the pose estimation model HRNet [23] to generate human pose heatmaps for semantic disentanglement and selective occlusion of unoccluded parts. BPBReID [24] also uses HRNet to extract high-resolution feature maps for fine-grained attention maps(Figure 4). PGFA [3] intervenes in global features extracted by the backbone using heatmaps from a pose estimator to obtain pose-guided global representations. A series of works [25,26] adopt pose estimation models to extract human keypoints or part heatmaps, assisting feature alignment and occlusion localization to emphasize unoccluded regions.

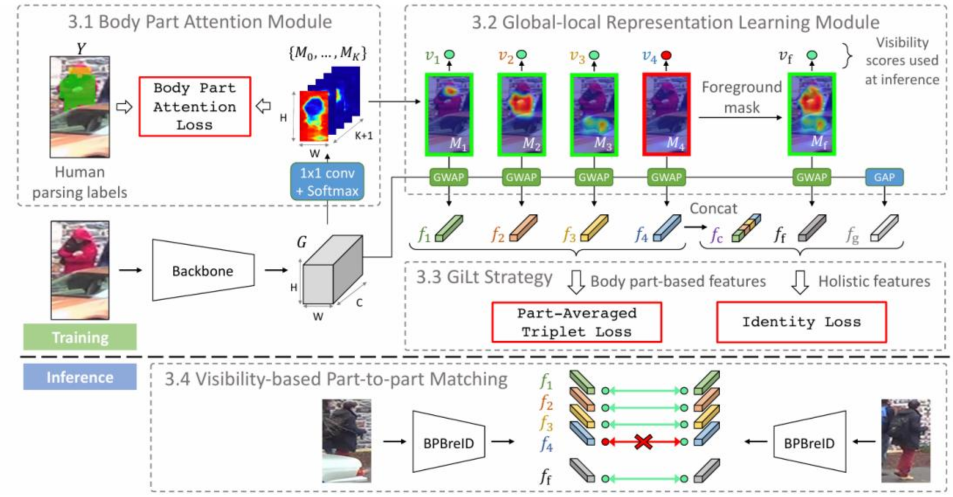

Fig. 4. Structure of BPBreID with detailed architecture [24]

However, pose-guided methods suffer from keypoint missing under heavy occlusion, leading to higher matching failure rates if over-reliance on keypoints occurs. Furthermore, some methods use human semantic parsing for fine-grained local matching. SPReID [27] employs a modified Inception-V3 [28] model as a semantic parser to explore local cues and re-identifies pedestrians via weighted feature summation. ISP proposes identity-guided human semantic parsing to locate body parts and accessories at the pixel level, explicitly identifying visible regions in occluded images without extra semantic supervision. Although auxiliary information reduces the difficulty of occluded ReID to a certain extent, it increases model size and computational cost due to external inference.

3 Experimental Comparison and Analysis

Since the Occluded-ReID dataset does not provide a dedicated training set, evaluation is performed using weights trained on Market1501. Benchmark comparison is only conducted with previous works that report performance under the same cross-dataset protocol or are reproduced in this paper. Table 1 presents a comprehensive comparison of state-of-the-art (SOTA) methods on occluded datasets.

3.1 Result Analysis on Occluded Scene Datasets

The effectiveness of external cues in occluded person ReID is evident from models like SPH [29], HOReID [30], and FAFN [31] . On Occluded-DukeMTMC, SPH [29] achieves 71.9% Rank-1 and 63.7% mAP, substantially outperforming the earlier HOReID [30] (55.1% Rank-1, 43.8% mAP). Such improvements, however, come at the cost of dependence on external models. Biased or inaccurate cues directly degrade recognition robustness, and even under optimal conditions these methods fail to surpass the best feature-level enhancement approaches. For instance, PADE [5], a feature-level method without any external dependency, attains 71.8% Rank-1 on the same dataset, closely matching SPH [29] while requiring no extra cues. ETNDNet [14] reported 68.1% Rank-1, further demonstrating the viability of feature-level enhancement.

What distinguishes the more successful feature-level methods is their use of cross-identity augmentation and missing information recovery, which simulate realistic occlusion patterns more effectively than simple instance-internal noise injection. This leads to more robust global feature representations and explains why these methods generally outperform external-cue-based ones. Nevertheless, not all feature-level methods achieve the same level of robustness. Transformer-based or pre-trained models[31,32,33,34], such as PRE-Net [34], TransReID [21] and CLIP-ReID [35], exhibit competitive performance on occluded scenes, with TransReID [21] obtaining 66.4% Rank-1 and 59.2% mAP on Occluded-DukeMTMC. The global context modeling capability of Transformers provides an advantage, but the loss of local features due to occlusion remains a limiting factor, preventing such models from exceeding top feature-level methods. Conversely, methods dedicated to fine-grained local feature extraction, including PGMA-Net [33] (62.2% Rank-1) and PRE-Net [34] (68.3% Rank-1), confirm the importance of local feature mining for occluded ReID.

Taken together, the evidence suggests that feature-level enhancement without external dependencies strikes the best balance between robustness and generalization, while external cues and Transformer-based global modeling each face distinct limitations in handling occlusion.

Table 1 Comparison with the state-of-the-art methods on the Occluded-DukeMTMC and Occluded-ReID datasets

Method	Occluded-DukeMTMC Rank-1(%)	Occluded-DukeMTMC mAP(%)	Occluded-ReID Rank-1(%)	Occluded-ReID mAP(%)
HOReID	55.1	43.8	80.3	70.2
FAFN	55.9	44.1	79.1	75.1
RTGAT	61.0	50.1	71.8	51.0
PGMA-Net	62.2	52.2	-	-
PRE-Net	68.3	55.2	-	-
SAP	70.0	62.2	83.0	76.8
FRT	70.7	61.3	80.4	71.0
MSPL	71.2	63.2	84.2	78.5
SPH	71.9	63.7	80.1	78.5
TransReID	66.4	59.2	70.2	67.3
MTIPE	66.4	57.8	-	-
CLIP-ReID	67.2	60.3	-	-
ETNDNet	68.1	57.6	-	-

FED	68.1	56.4	86.3	79.3
PVT	69.0	61.2	83.3	77.5
PFT	69.8	60.8	83.0	78.3
SCLFT	69.8	60.8	83.0	78.3
DPEFormer	69.9	58.9	-	-
SSPEM	70.2	62.8	82.8	78.5
DCG	70.5	61.4	-	-
PADE	71.8	62.3	83.1	78.8
OAT	71.8	62.2	82.6	78.2
OSRTrans	72.9	61.5	-	-
HEFormer	73.2	61.4	-	-
CrossViT-ReID	73.7	63.6	-	-
TRGA	73.8	63.9	84.4	79.7

3.2 Result Analysis on Non-Occluded Scene Datasets

To verify the generalization capability of the proposed method, we further conduct evaluations on two standard holistic person re-identification datasets: Market-1501 and DukeMTMC-ReID [1]. As presented in Table 2, occlusion-specialized models exhibit divergent generalization performance in occlusion-free scenes, which is highly correlated with their core design paradigms.

Methods that focus on robust feature purification and enhancement, such as TRGA and PADE, demonstrate strong capability in resisting occlusion interference. Meanwhile, approaches that achieve effective local - global feature balance, including HEFormer and DGSN, also exhibit favorable performance. Representative methods such as DGSN and TransReID achieve high recognition accuracy, with Rank-1 exceeding 95% on Market-1501 and 90.7% on DukeMTMC-ReID. This indicates that feature purification mechanisms not only alleviate occlusion interference but also suppress background clutter and intra-class appearance variations, thereby facilitating the learning of discriminative identity - preserving representations.

Transformer-based methods (TransReID [21], HEFormer [36]) also achieve competitive performance on holistic datasets. HEFormer [36] obtains 95.0% Rank-1 and 86.2% mAP on Market-1501, as well as 89.4% Rank-1 on DukeMTMC-ReID [1], benefiting from its strong ability to capture global context and aggregate identity consistent features. Local feature-oriented methods such as PGMA-Net [33] and PRE-Net [34] also deliver satisfactory performance (95.0% Rank-1 and 88.4% mAP for PGMA-Net [33] on Market-1501), even without reaching state-of-the-art levels. This further validates the effectiveness of fine-grained local feature exploitation in holistic person re-identification.

Some early methods that rely on external auxiliary cues, such as FAFN, exhibit relatively limited performance, achieving 92.3% Rank-1 and 81.6% mAP on holistic datasets. This is mainly attributed to their over-specialized occlusion handling mechanisms, which neglect the generalization capacity required for occlusion-free scenarios. In contrast, methods such as

OA-ReID and PFD achieve more competitive performance, with 95.8% Rank-1 and 89.1% mAP on Market-1501, owing to their effective integration of local detailed information and global semantic context. Such designs enhance the discriminative capacity and robustness of identity representations.

These observations demonstrate that the joint utilization of complementary local and global features significantly improves the stability and reliability of person re-identification models.

The generalization performance on holistic datasets fundamentally depends on the universality of feature extraction mechanisms. Methods that can effectively capture representative identity-consistent features while suppressing background and occlusion interference perform favorably in both occluded and non-occluded scenarios. In comparison, over-specialized occlusion-specific models tend to suffer from restricted generalization ability.

Table 2 Comparison with the state-of-the-art methods on the Market1501 and DukeMTMC-ReID datasets

Method	Market1501 Rank-1(%)	Market1501 mAP(%)	DukeMTMC-ReID Rank-1(%)	DukeMTMC-ReID mAP(%)
FAFN	92.3	81.6	83.4	73.4
MHSA-Net	94.6	84.0	87.3	73.1
DRL-Net	94.7	86.9	88.1	76.6
MSAMGO	94.7	87.2	86.0	76.9
HEFormer	95.0	86.2	89.4	79.2
CENet	95.0	87.3	-	-
AMG	95.0	88.9	-	-
DGSN	95.0	87.9	90.7	80.9
PGMA-Net	95.0	88.4	90.1	79.0
Lu et al.	95.1	88.8	90.6	82.0
HCGA	95.2	88.4	-	-
TransReID	95.2	88.9	90.7	82.0
RFCNet	95.2	89.2	90.7	80.7
PRE-Net	95.3	86.5	89.3	77.8
AAFormer	95.4	87.7	90.1	80.0
PAT	95.4	88.0	88.8	78.2
NetVLAD-M	95.5	88.9	90.7	81.3
FRT	95.5	88.1	90.5	81.7
PFD	95.5	89.6	90.6	82.2
HAT	95.6	89.5	90.4	81.4

OA-ReID	95.8	89.1	90.4	81.2
TRGA	96.1	89.9	91.0	82.2

4 Conclusion

To address the occlusion problem in person re-identification, this paper systematically reviews existing approaches and categorizes them into three paradigms: data-driven, model-driven, and auxiliary information-driven methods. Data-driven methods simulate real-world occlusion scenarios and adopt diverse data augmentation strategies, thereby alleviating the scarcity of occluded samples in training data. Model-driven methods concentrate on optimizing feature extraction pipelines: one branch directly discards features from occluded regions to suppress interference, while another enhances feature representations extracted from visible regions to strengthen occlusion robustness. Auxiliary information-driven methods depend on external techniques such as pose estimation and semantic parsing. These methods first localize visible body parts, and then conduct feature matching based on the located visible regions. Furthermore, this paper summarizes and compares the performance of representative algorithms within the scope of the reviewed methods on four mainstream benchmarks. Among them, Occluded-DukeMTMC and Occluded-ReID are two typical occlusion-specific datasets, while Market-1501 and DukeMTMC-ReID are standard holistic (non-occluded) datasets. The comparison covers most state-of-the-art occluded person ReID algorithms reported in recent literature. According to the evaluated results under the selected datasets and metrics, feature-level enhancement methods that do not require external auxiliary cues achieve superior overall performance on occlusion-specific benchmarks and maintain satisfactory generalization ability on non-occluded holistic datasets. Auxiliary information-based methods can further improve recognition accuracy under heavy occlusion, but they inevitably introduce additional computational overhead and model complexity.

For future prospects, the combination of Transformer-based global modeling and fine-grained local feature extraction is regarded as a promising research direction. Transformers are capable of capturing long-range semantic dependencies in images, enabling the model to infer features of occluded regions by leveraging contextual information from visible parts under occlusion scenarios. Meanwhile, learning more discriminative and precise fine-grained local features remains critical. The integration of global semantic modeling and fine-grained local representation is therefore worthy of in-depth exploration in future research.

References

[1] Ristani E., Solera F., Zou R., Feris R.S., Metaxas D.N., Tomasi C.: Performance measures and a data set for multi-target, multi-camera tracking. In: Lect. Notes Comput. Sci., vol. 9914, pp. 17–35 (2016). doi:10.1007/978-3-319-48881-3_2

[2] Zhang N., Cheng D.Q., Kou Q.Q., Qian J.S.: Person re-identification based on random occlusion and multi-granularity feature fusion. J. Beijing Univ. Aeronaut. Astronaut. 49(12), 3511–3519 (2023). doi:10.13700/j.bh.1001-5965.2022.0091

[3] Miao J., Wu Y., Liu P., Ding Y., Yang Y.: Pose-guided feature alignment for occluded person re-identification. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 542–551 (2019).
[4] Wang Z., Zhu F., Tang S., Zhao R., He L., Song J.: Feature erasing and diffusion network for occluded person re-identification. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 4754–4763 (2022).
[5] Wang Z., Huang H., Zheng A., He R.: Parallel augmentation and dual enhancement for occluded person re-identification. In: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 3590–3594 (2024). doi:10.1109/ICASSP48485.2024.10446014
[6] Yun S., Han D., Oh S.J., Chun S., Choe J., Yoo Y.: CutMix: Regularization strategy to train strong classifiers with localizable features. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 6023–6032 (2019).
[7] Zhang Z., Han S., Liu D., Ming D.: Focus and imagine: Occlusion suppression and repairing transformer for occluded person re-identification. Neurocomputing 578, 127442 (2024). doi:10.1016/j.neucom.2024.127442
[8] Xia J., Tan L., Dai P., Zhao M., Wu Y., Cao L.: Attention disturbance and dual-path constraint network for occluded person re-identification. In: Proc. AAAI Conf. Artif. Intell. 38(6), 6198–6206 (2024). doi:10.1609/aaai.v38i6.28437
[9] Zhang X., Fu K., Zhao Q.: Dynamic patch-aware enrichment transformer for occluded person re-identification. Knowl.-Based Syst. 327, 114193 (2025). doi:10.1016/j.knosys.2025.114193
[10] Kirillov A., Mintun E., Ravi N., Mao H., Rolland C., Gustafson L., Xiao T., Whitehead S., Berg A.C., Lo W.-Y., Dollár P., Girshick R.: Segment anything. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 3992–4003 (2023). doi:10.1109/ICCV51070.2023.00371
[11] Zhu Y., Duan S.Z., Guo Y.F., Chen X.: Research and application of occluded person re-identification based on multi-scale denoising autoencoder. J. Appl. Res. Comput. 42(7), 2220–2226 (2025).
[12] Ge Y., Zhu F., Chen D., Zhao R., Li H.: Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In: Adv. Neural Inf. Process. Syst. 33, 11309–11321 (2020).
[13] Tan L., Xia J., Liu W., Dai P., Wu Y., Cao L.: Occluded person re-identification via saliency-guided patch transfer. In: Proc. AAAI Conf. Artif. Intell. 38(5), 5070–5078 (2024). doi:10.1609/aaai.v38i5.28312
[14] Dong N., Zhang L., Yan S., Tang J.: Erasing, transforming, and noising defense network for occluded person re-identification. IEEE Trans. Circuits Syst. Video Technol. 34(6), 4458–4472 (2023).DOI: 10.1109/TCSVT.2023.3339167
[15] Yuan C.S., Liang Y.C., Hu X.Y., Li Y.: Occluded person re-identification algorithm based on occlusion-aware transformer. J. China Acad. Electron. Inf. Technol. 20(5), 475–489 (2025).
[16] Ye Y.T., Zhou H., Cai J., Gao C., Zhang Y., Wang J., Hu Q., Yu J., Yang W.: Dynamic feature pruning and consolidation for occluded person re-identification. In: Proc. AAAI Conf. Artif. Intell. 38(7), 6684–6692 (2024). doi:10.1609/aaai.v38i7.28632
[17] Zhou B., Khosla A., Lapedriza A., Oliva A., Torralba A.: Learning deep features for discriminative localization. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2921–2929 (2016). doi:10.1109/CVPR.2016.319
[18] Wang X., Hu X.G., Fu Z.Y., Yu Y.: Transformer-based global-local fusion feature for occluded person re-identification. J. Front. Comput. Sci. Technol. 19(7), 1832–1850 (2025).
[19] Jiang B., Zhang Z., Lin D., Tang J., Luo B.: Semi-supervised learning with graph learning-convolutional networks. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 11313–11320 (2019). doi:10.1109/CVPR.2019.01157

[20] Phan N., Huy T.D., Duong S.T.M., Hoang N.T., Tran S., Hung D.H., Nguyen C.D.T., Bui T., Truong S.Q.H.: LogoViT: Local-global vision transformer for object re-identification. In: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 1-5 (2023). doi:10.1109/ICASSP49357.2023.10096126

[21] He S., Luo H., Wang P., Wang F., Li H., Jiang W.: TransReID: Transformer-based object re-identification. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 15013–15022 (2021). doi:10.1109/ICCV48922.2021.01474

[22] Wang X.Y., Ren H.E., Tao R.: Occluded person re-identification fusing self-attention and local features. J. Harbin Univ. Sci. Technol. 30(3), 24–31 (2025).

[23] Wang J., Sun K., Cheng T., Jiang B., Deng C., Zhao Y., Liu D., Mu Y., Tan M., Wang X., Liu W., Xiao B.: Deep high-resolution representation learning for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 43(10), 3349–3364 (2021). doi:10.1109/TPAMI.2020.2983686

[24] Somers V., De Vleeschouwer C., Alahi A.: Body part-based representation learning for occluded person re-identification. In: Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), pp. 1613–1623 (2023). doi:10.48550/arXiv.2211.03679

[25] Peng X.C., Zhou W., Duan Z.: Improved occluded person re-identification algorithm based on HRNet. Mod. Comput. 30(24), 67–72 (2024).https://doi.org/10.48550/arXiv.1804.02792

[26] Liu T., Luo Q., Zhang H.Y.: Occluded person re-identification fusing pose guidance and multi-scale features. J. Image Graph. 29(8), 2364–2376 (2024).

[27] Kalayeh M.M., Basaran E., Gökmen M., Kamasak M.E., Shah M.: Human semantic parsing for person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1062–1071 (2018). https://doi.org/10.48550/arXiv.1804.00216

[28] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z.: Rethinking the inception architecture for computer vision. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2818–2826 (2016). doi:10.1109/CVPR.2016.308

[29] Gao Z., Chen P., Zhuo T., Liu M., Zhu L., Wang M., Chen S.: A semantic perception and CNN-transformer hybrid network for occluded person re-identification. IEEE Trans. Circuits Syst. Video Technol. 34(4), 2010–2025 (2023). doi:10.1109/TCSVT.2023.3296680

[30] Wang G., Yang S., Liu H., Wang Z., Yang Y., Wang S., Yu G., Zhou E., Sun J.: High-order information matters: Learning relation and topology for occluded person re-identification. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 6449–6458 (2020). (DOI not found)

[31] Zhuang X., Wei D., Liang D.: Feature attention fusion network for occluded person re-identification. Image Vis. Comput. 143, 104921 (2024). doi:10.1016/j.imavis.2024.104921

[32] Huang M., Hou C., Yang Q., Wang Z.: Reasoning and tuning: Graph attention network for occluded person re-identification. IEEE Trans. Image Process. 32, 1568–1582 (2023).

[33] Chen Y., Yang Y., Liu W., Huang Y., Li J.: Pose-guided counterfactual inference for occluded person re-identification. Image Vis. Comput. 128, 104587 (2022). doi:10.1016/j.imavis.2022.104587

[34] Yan G., Wang Z., Geng S., Yu Y., Guo Y.: Part-based representation enhancement for occluded person re-identification. IEEE Trans. Circuits Syst. Video Technol. 33(8), 4217–4231 (2023). doi:10.1109/TCSVT.2023.3241764

[35] Li S., Sun L., Li Q.: CLIP-ReID: Exploiting vision-language model for image re-identification without concrete text labels. In: Proc. AAAI Conf. Artif. Intell. 37(1), 1405–1413 (2023). doi:10.1609/aaai.v37i1.25225

[36] Han S., Zhang Z., Yuan X., Ming D.: Enhance heads in vision transformer for occluded person re-identification. IEEE Sensors J. 25(4), 6894–6904 (2025). DOI: 10.1109/JSEN.2024.3523475