Theory and Practice of Science and Technology
2026, VOL. 7, NO. 2, 172-179
DOI: 10.47297/taposatWSP2633-456934.20260702

MMCE-Net: A Multi-Module Collaborative Enhancement Network for Occluded Person Re-Identification

Mingjun Ma*, Jun Li

College of Control Engineering, Northeast Forestry University, Harbin, Heilongjiang, 150040, China

ABSTRACT
Person re-identification aims to match pedestrian images captured by different cameras and remains challenging due to occlusion, pose variation, and background interference in practical scenarios. This paper proposes a collaborative enhancement framework for person re-identification based on a ResNet-50 backbone, with the objective of improving robustness under partial occlusion. The main contribution of this work lies in the integration of a multi-scale attention mechanism that strengthens local feature perception and enhances sensitivity to critical body regions. In addition, an Enhanced Feature Generation Module is designed to fuse shallow and deep features through cross-layer interactions, enabling more complete and discriminative pedestrian representations. By jointly leveraging complementary information from different network stages, the proposed framework effectively mitigates the negative impact of occluded or misaligned regions. Extensive experiments conducted on multiple public datasets, including occluded and conventional person re-identification benchmarks, demonstrate that the proposed method achieves competitive and robust performance compared with representative state-of-the-art approaches, validating its effectiveness in complex real-world scenarios.

KEYWORDS
Occluded person re-identification; Multi-scale features; Feature enhancement; Attention mechanism

1 Introduction

Pedestrian re-identification (Re-ID) is a fundamental computer vision task focused on matching individuals across non-overlapping camera views under varying conditions[1]. It plays an important role in intelligent surveillance and public security systems. Occlusion significantly degrades recognition performance by corrupting discriminative appearance cues [2].

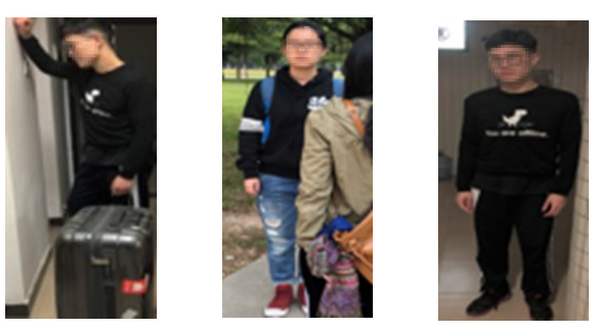


Figure 1 Occluded person Re-ID datasets images

Pose-guided methods exploit human keypoints to align visible body regions for occluded Re-ID, but they rely heavily on accurate pose estimation [3]. Graph-based methods model relationships among body parts to handle occlusion, but they often require complex graph construction and additional processing [6,8,19]. Transformer-based approaches leverage long-range dependencies and show strong performance under occlusion; however, they often rely on external pose estimators or introduce background noise through global self-attention [4,14,22-23].

The key technical contribution of MMCE-Net lies in its ability to achieve robust occluded Re-ID performance without relying on external pose estimators or segmentation models. Unlike Transformer-based methods, MMCE-Net enhances feature representation through internal multi-scale attention and feature enhancement modules integrated into the backbone network. [5,22]. This approach achieves efficient feature fusion and distinction of key regions in occluded environments while maintaining lightweight design constraints.

This study advances pedestrian re-identification by integrating a multi-scale attention mechanism into ResNet-50 [6,10,20]. This enhances local detail capture, mitigates the im-pact of occluded regions, and improves feature representation learning in non-occluded areas. We propose an Enhanced Feature Generation Module (EFGM) that fuses shallow and deep

* **Corresponding Author:** Mingjun Ma, 18336868962@163.com

features through cross-layer connection strategies, thereby enhancing the overall completeness and discriminative power of feature representation. On the occluded Re-ID dataset Partial-iLIDS, our method outperforms mainstream approaches and demonstrates strong generalization on conventional Re-ID benchmarks [21].

The main contributions of this work are summarized as follows:1. We propose a multi-scale attention-enhanced person Re-ID framework that improves robustness under occlusion without relying on external pose information.2.An Enhanced Feature Generation Module is designed to fuse shallow and deep features via cross-layer interactions. 3. Extensive experiments on occluded and conventional Re-ID benchmarks demonstrate the effectiveness and generalization ability of the proposed method.

2 The Proposed MMCE-Net Method

Existing approaches for occluded person Re-ID include pose-guided alignment, attention-based modeling, and Transformer-based architectures[21], which improve robustness by exploiting visible regions or long-range dependencies [7]. Recent studies also explore lightweight multi-scale fusion and hashing-based embeddings to improve efficiency [9,15-16]. However, such methods remain limited in handling severe occlusion and pose variation, motivating the proposed EFGM and DFEM for robust multi-scale feature enhancement.

The framework integrates three core modules: the Enhanced Feature Generation Module (EFGM), Dynamic Feature Enhancement Module (DFEM), and Part-aware Attention Module (PAM). Through synergistic multi-scale modeling, dynamic contextual enhancement, and local spatial refinement, these modules achieve robust recognition of partially occluded pedestrians. DFEM is inserted after Layer2 to enhance mid-level features via multi-branch multi-scale contextual modeling, while EFGM is deployed between Layer3 and Layer4 to refine high-level semantic representations using dynamic receptive fields and attention mechanisms. Together, these modules capture both local details and long-range contextual dependencies, improving robustness to occlusion.

Subsequently, the Enhanced Feature Generation Module (EFGM) employs the Part-aware Attention Module (PAM) to segment and extract local features, enabling precise localization of key human body regions[11-13]. The extracted features are further enhanced by a global context enhancement module to improve feature expressiveness. The back-bone-extracted features are then reduced in dimensionality via global average pooling (GAP) and mapped to the final representation space through a fully connected layer (FC), generating discriminative pedestrian identity features.

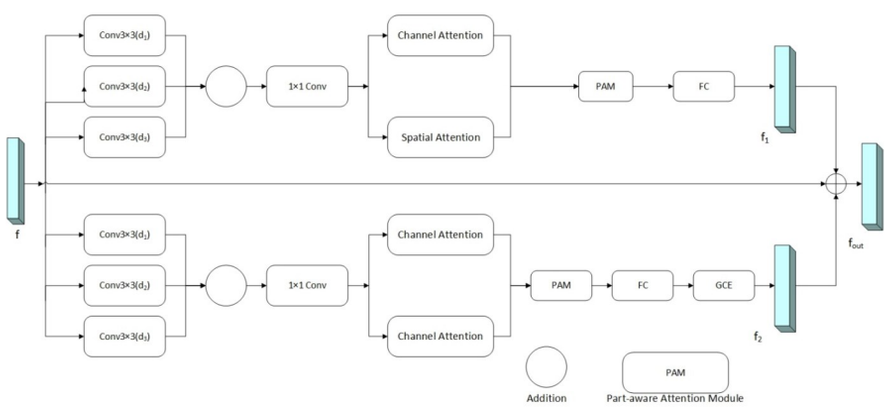


Figure 2 Enhanced feature generation module

2.1 Multi-Scale Attention Mechanism Module

To alleviate performance degradation caused by occlusion and viewpoint variations, the Enhanced Feature Generation Module (EFGM) is designed to enhance discriminative feature representation through adaptive receptive field modeling and attention-based refinement.

As illustrated in Figure 2, the EFGM processes features from the backbone to generate diverse, compact representations extracted from the backbone network, generating diverse and compact feature representations. In the initial stage, the feature map (f) undergoes three dilated convolutions with 3×3 kernels and learnable dilation rate d_1,d_2, d_3, modeling for multi-scale context extraction. The outputs are then fused and compressed to 25% of the original channel dimensionality through a 1×1 convolution, effectively reducing feature redundancy and computational overhead.

The fused features are refined using channel and spatial attention, followed by a part-aware attention mechanism to emphasize discriminative local regions while suppressing background interference[10]. A global context enhancement module is further applied to integrate long-range semantic information and improve the completeness of the final

feature representation..

The EFGM output features are obtained by concatenating the original input f with enhanced branches f_1 and f_2 as formulated:

$$f_2 = FC(ReLU(\left[CA(f_1), SA(f_1)\right])) \tag{1}$$

To enhance the discriminative power and diversity of the generated embeddings f_*, we adopt a simplified center-based regularization inspired by the center-guided pair mining idea in DEEN [21]. Unlike the original cross-modality formulation, our task involves only single-modality RGB data. Therefore, we reformulate the loss as a single-modality center loss that encourages intra-class compactness and inter-class separability.

Let c_k denote the learnable center of class k, and $f_i$be the embedding of the i-th sample with label y_i. The center-based loss is defined as

$$L_{conter} = \frac{1}{2}\sum_i ||f_i - c_{y_i}||_2^2 \tag{2}$$

where $||\cdot||_2$denotes the Euclidean norm. The class centers are updated dynamically during training via backpropagation, ensuring that the feature distribution becomes more compact within classes and more separable across classes.

2.2 Feature Fusion Module

The proposed network enhances the ResNet-50 backbone through hierarchical feature fusion at multiple stages.DFEM is employed in the lower and middle layers to strengthen multi-scale contextual representation, while EFGM is introduced in higher layers to refine semantic features via attention-guided enhancement.

This coordinated fusion strategy effectively improves feature interaction and robustness under occlusion and viewpoint variations.

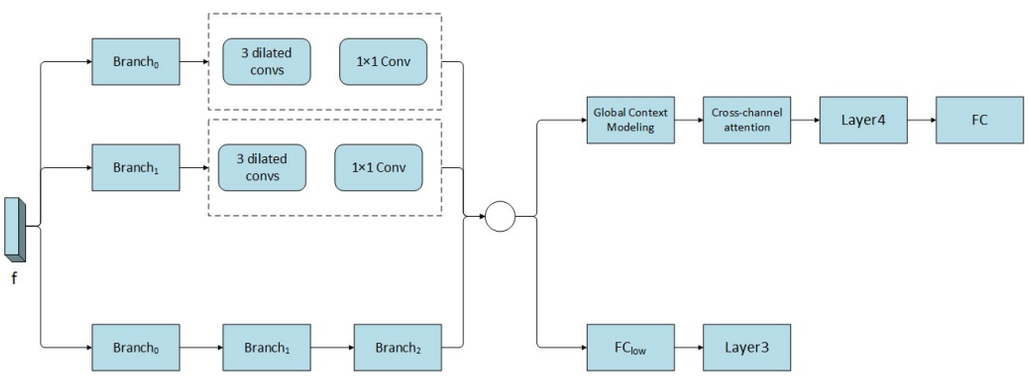


Figure 3 Part-aware Attention Module

Overall, the model achieves horizontal fusion between multi-scale branches through the DFEM at all levels and deep fusion through the EFGM at the middle and upper levels under attention enhancement. Compared to traditional shallow feature concatenation methods, the fusion strategy adopted in this paper enhances information interaction capabilities while maintaining feature diversity, enabling the final feature representation to exhibit stronger discriminative power and robustness in complex scenarios such as occlusion and viewpoint changes.

2.3 Feature Enhancement Module

Although RFB-s effectively enlarges receptive fields, its fixed dilation and limited inter-branch interaction restrict its adaptability to complex backgrounds and scale variations in person re-identification.

To address these limitations, the proposed Dynamic Feature Enhancement Module (DFEM) introduces learnable dilation rates and cross-branch interaction to enable dynamic multi-scale feature modeling.

By integrating global context enhancement, DFEM improves semantic representation and robustness in challenging occluded scenarios..

This work focuses on enhancing the model's dynamic feature modeling, structural diversity, cross-branch interaction, and context-aware enhancement. These aspects are achieved through learnable dilation rates, multibranch architectures, inter-branch information fusion, and global attention mechanisms, respectively. Together, they contribute to a more robust and adaptive feature representation framework.

To overcome the limitations of the original module, the Dynamic Feature Enhancement Module (DFEM) comprehensively improves dynamic feature adaptation, multi-branch structural diversity, cross-branch interaction, global context modeling, and sparse activation. Specifically, DFEM introduces three parallel branches to construct multi-scale receptive fields: Branch 1 employs standard 3 × 3 convolutions for basic spatial feature extraction; Branch 2 combines 1 × 3 and 3 × 1 separable convolutions to form an asymmetric receptive field, followed by dilated convolutions to expand the

receptive field size; and Branch 3 mirrors Branch 2 but swaps the convolution order (3 × 1 first, then 1 × 3), enabling a complementary spatial perspective. Unlike conventional RFB modules, DFEM introduces learnable dilation rates by defining a trainable vector self-dilation = [d_1, d_2, d_3]. This allows each branch's receptive field to adapt dynamically during training, enabling precise multi-scale semantic modeling of input features. The dilation rates d_1, d_2, and d_3 are updated via backpropagation, allowing the model to flexibly adjust receptive field sizes for varying target scales and complex scenes. These improvements collectively enhance the model's feature representation capabilities. These three outputs are then concatenated along the channel dimension. Finally, the enhanced feature map is combined with global context information and passed through ReLU activation to pro-duce the final output Y_{out} (Formula (5)).

The multi-branch structure extends the original design via a third branch. This branch employs cross-channel spatially separable convolutions (vertical and horizontal kernels) followed by Sigmoid activation to generate an adaptive weight matrix for the concatenated features Y yielding adaptively weighted output $Y_{weighted}$. This design significantly enhances multi-scale feature extraction. To minimize inter-branch feature redundancy, a cross-branch attention mechanism is proposed, fusing features through concatenation and applying adaptive weighting:

$$Y = Cat\left(W_1, W_2, W_3\right) \tag{3}$$

$$Y_{weighted} = Y \cdot \sigma(f_{1\times 1_{conv}}(Y)) \tag{4}$$

The sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ applies channel-wise weighting, assigning higher weights to important features to generate the weight matrix. Here, Y denotes the concatenated multi-branch feature map. To enhance global semantic representation, a global context enhancement module (ContextEnhance) and the ReLU activation function are introduced after feature fusion.

ContextEnhance combines Global Average Pooling (GAP) with an attention mechanism to focus on critical features, while ReLU further refines features and suppresses noise in the final output stage

$$Y_{out} = ReLU(Y_{weighted} + ContextEnhance\left(Y_{weighted}\right)) \tag{5}$$

where $Y_{weighted}$ denotes the feature map after cross-branch weighting, and ReLU serves as a sparse activation function to enhance key features and suppress redundant activations.

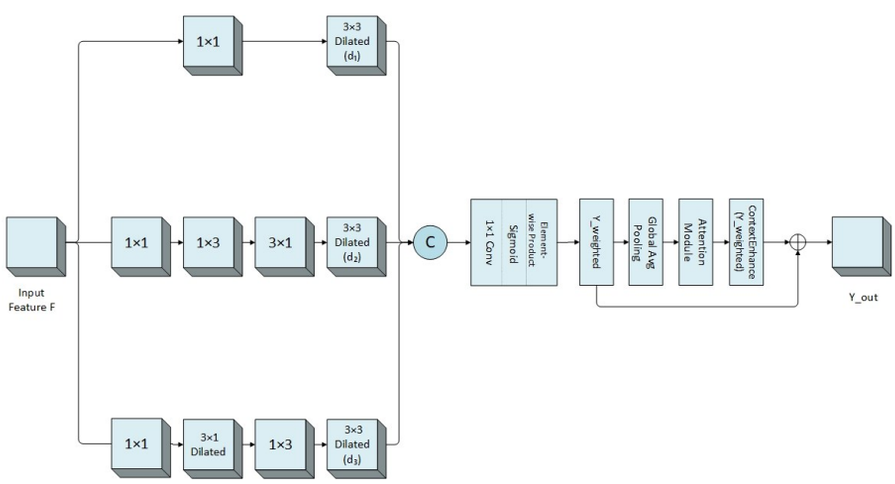


Figure 4 Dynamic Feature Enhancement Module

3 Experiments and Analyses

3.1 Experimental Setup

The experiments were conducted on Windows 10, equipped with an NVIDIA GeForce RTX 3060 GPU (8 GB VRAM) (Nvidia Corporation, Santa Clara, CA, USA) and an Intel i5-12490F CPU (Intel Corporation, Santa Clara, CA, USA). The PyTorch framework (Py-thon 3.7.11) was used, with ResNet-50 as the baseline network. During data preprocessing, all images were resized to a fixed resolution of 256 × 128. The model was trained for 360 epochs using the Adam optimizer with an initial learning rate of 2.0×10^{-4} and a weight decay of 1.0×10^{-3}. A warmup strategy was applied for the first 10 epochs with a warmup factor of 0.1 to stabilize training. The learning rate was multiplied by 0.4 at epochs 40, 80, 120, 160, 180, and 240.

3.2 Datasets and Evaluation Indicators

This paper evaluated the proposed method by training on the publicly available Market-1501 dataset and testing on

three widely used occluded/partial ReID datasets, Occluded-ReID, Partial-iLIDS, and Partial-ReID, achieving competitive accuracy. The Mar-ket-1501 dataset contains 1501 identities captured by six cameras, partitioned into 12,936 training images, 3368 query images, and 15,913 gallery images. For performance evaluation, we adopt three standardized metrics with precise definitions: Rank-n accuracy = percentage of queries where correct identity appears within top n results (e.g., Rank-1 = 82.5% means 82.5% of queries have correct match at position 1), CMC curve = plot of Rank-n accuracy for n = 1, 5, 10 showing cumulative matching probability, and mAP = average precision across all ranks. In occluded scenarios, Rank-1 often fails when top-1 matches are blocked by occlusions, while mAP quantifies performance when correct matches appear at lower ranks (e.g., Rank-3, Rank-7). All results follow a single-query protocol.

3.3 Experimentation and Evaluation

3.3.1 Visualization of Retrieval Results

To provide more intuitive evidence of the effectiveness of MMCE-Net under occlusion, we present qualitative retrieval results in Figure 5. Each row corresponds to one query image. The baseline ResNet-50 often mismatches identities when the discriminative re-gions are occluded (red boxes), whereas MMCE-Net correctly focuses on visible parts (e.g., clothing details, legs) and retrieves the correct identities (green boxes). These visual com-parisons demonstrate that MMCE-Net achieves stronger robustness to partial occlusion by effectively exploiting both local and global cues.

3.3.2 Ablation of Different Modules

To comprehensively evaluate the effectiveness of the MMCE-Net, experiments were conducted on three representative occlusion and partial visibility pedestrian re-identification (ReID) datasets: Occluded-ReID, Partial-ReID, and Partial-iLIDS. The experimental results include comparative experiments with existing mainstream methods (Table 1) and module-level ablation studies (Tables 2–4). Table 1 presents a performance comparison between the MMCE-Net and several state-of-the-art ReID methods across these three datasets, where each result is shown in the format Rank-1 (mAP) within a sin-gle cell. Specifically, on the Occluded-ReID dataset, MMCE-Net achieves a Rank-1 accuracy of 78.2%, which improves by 6.1% compared.

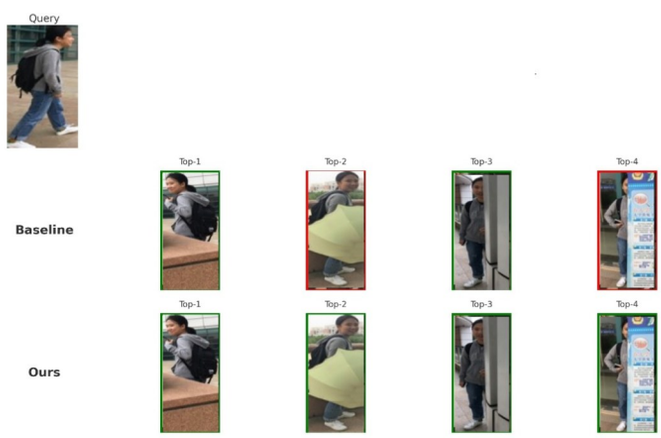


Figure 5 Qualitative Results

To OONet (72.1%). It also surpasses HO-Net (77.3%) and FPR (76.3%), demonstrating strong robustness under heavy occlusion. On the Partial-iLIDS dataset, MMCE-Net achieves 73.1% Rank-1 accuracy, outperforming OONet (71.0%) and HONet (70.6%) by 2.1% and 2.5%, respectively. On the Partial-ReID dataset, MMCE-Net achieves a Rank-1 accuracy of 80.6%, which is slightly lower than HONet (81.3%). However, our model attains a competitive mAP of 86.6%, comparable to OONet (86.3%) and close to HONet (87.8%), demonstrating strong robustness in partially visible scenarios. The results in Table 1 indicate that our approach is comparable or superior in handling occluded and partially visible targets, particularly showing a notable improvement of +6.1% on Occluded-ReID, and consistent gains of about +2–2.5% on Partial-iLIDS compared with strong baselines.

To further validate the contributions of each proposed module to overall performance enhancement, Tables 2–4 detail systematic ablation experiments conducted on the Occluded-ReID, Partial-ReID, and Partial-iLIDS datasets, evaluating the individual and combined effects of the following three modules.

As shown in Tables 2 and 4, the EFGM improves Rank-1 accuracy from 70.5% to 73.9% (+3.4%) and mAP from 77.8% to 81.9% (+4.1%) on the Partial-iLIDS dataset, and from 73.0% to 73.8% (+0.8%) on Occluded-ReID. This indicates that the cross-scale feature perception mechanism (EFGM) effectively extracts discriminative granular features by enhancing multi-semantic, cross-scale pedestrian representations.

The DFEM increases Rank-1 from 73.0% to 78.2% (+5.2%) and mAP from 64.4% to 68.7% (+4.3%) on Occluded-ReID,

and from 85.6% to 86.6% (+1.0%) on Partial-ReID. This validates the effectiveness of the pedestrian semantic enhancement module (DFEM) in generating diverse embeddings, particularly for occluded person re-identification.

Similarly, the PAM boosts Rank-1 from 73.0% to 76.5% (+3.5%) and mAP from 64.4% to 68.0% (+3.6%) on Occluded-ReID, and from 85.6% to 86.3% (+0.7%) on Partial-ReID, demonstrating that the region attention mechanism (PAM) can effectively focus on salient regions through a multi-granularity non-local fusion framework.

Table 1 Comparison of the results of the method described in this paper and existing methods on three datasets.

Methods	Backbone	Occluded-ReID		Partial-ReID		Partial-iLIDS	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
MMCE-Net	ResNet-50	78.2	69.0	80.6	86.6	73.1	82.1
OONET	ResNet-50	72.1	64.0	79.3	86.3	71.0	78.0
HONet	ResNet-50	77.3	67.2	81.3	87.8	70.6	81.4
TCSDO	ResNet-50	73.7	67.9	82.7	—	—	—
FPR	ResNet-50	76.3	68.0	80.0	—	68.1	—
PGFA	ResNet-50	—	—	68.0	80.0	69.1	80.9
VPM	ResNet-50	—	—	67.7	81.9	65.5	74.8
DSR	ResNet-50	72.8	62.8	50.7	70.0	58.8	67.2

On Partial-iLIDS, EFGM alone delivers the largest Rank-1 gain (+3.4% to 73.9%), as its cross-scale perception best suits the dataset's partial occlusions. DFEM alone (Rank-1: 66.3%, mAP: 75.7%) underperforms the baseline due to over-reliance on local semantics, which disrupts global representations. However, when combined with EFGM/PAM, DFEM's embeddings benefit from cross-scale context (EFGM) or region prioritization (PAM), improving mAP to 82.1% (vs. 77.8% baseline) despite no further Rank-1 boost. This indicates that modules can exhibit selective synergy: EFGM dominates precision tasks, while DFEM/PAM augment recall via complementary mechanisms. Notably, on Occluded-ReID, combining EFGM with DFEM yields slightly lower Rank-1 and mAP than DFEM alone (Table 2). This likely reflects redundancy in feature modeling and interference be-tween local and global cues, which may complicate optimization. Future work could explore adaptive fusion strategies or regularization to better coordinate their interaction.

Table 2 Validation results of different modules on the Occluded-ReID dataset

Module			Results			
EFGM	DFEM	PAM	Rank-1	Rank-5	Rank-10	mAP
×	×	×	73.0	85.8	90.7	64.4
√	×	×	73.8	86.9	90.6	65.1
×	√	×	78.0	88.9	92.5	68.7
√	√	×	76.8	89.8	93.4	67.8
√	×	√	76.5	89.6	93.3	68.0
√	√	√	78.2	91.1	94.6	69.0

Table 3 Validation results of different modules on the Partial-ReID dataset

Module			Results			
EFGM	DFEM	PAM	Rank-1	Rank-5	Rank-10	mAP
×	×	×	85.6	93.3	96.3	78.6
√	×	×	84.6	92.3	95.3	78.1
×	√	×	86.6	94.3	97.0	79.7
√	√	×	86.3	94.0	96.6	80.7
√	×	√	83.6	92.3	95.3	80.1
√	√	√	86.6	93.6	96.3	80.6

Table 4 Validation results of different modules on the Partial-iLIDS dataset

Module			Results			
EFGM	DFEM	PAM	Rank-1	Rank-5	Rank-10	mAP
×	×	×	70.5	88.2	94.9	77.8
√	×	×	73.9	93.2	95.7	81.9
×	√	×	66.3	88.2	92.4	75.7
√	√	×	73.1	88.2	95.7	79.5
√	×	√	73.1	90.0	95.7	79.7
√	√	√	73.1	92.4	95.7	82.1

3.3.3 A Study of EFGM Embedding Locations

To analyze the impact of inserting the EFGM at different stages of the ResNet-50 backbone network, experiments were conducted on the Partial-iLIDS dataset, where the EFGM demonstrates the most significant performance gains. The experimental setup is summarized in Table 5. As shown in the table, inserting the EFGM after Layer3 achieves a Rank-1

accuracy of 73.9%, which improves by 1.7% compared to insertion after Layer2 (72.2%) and by 5.0% compared to insertion after Layer1 (68.9%). The mAP also increases from 77.1% to 81.9% (+4.8%). In contrast, inserting the module after Layer4 leads to a performance drop to 68.7% in Rank-1 accuracy, indicating that high-level feature fusion may interfere with semantic information preservation. Therefore, inserting the EFGM after Layer3 yields the optimal performance.

Table 5 Table EFGM embedding location study

Module Insertion Position	Single-Query			
	Rank-1	Rank-5	Rank-10	mAP
After Layer1	68.9	88.2	94.9	77.1
After Layer2	72.2	92.4	94.9	81.0
After Layer3	73.9	93.2	95.7	81.9
After Layer4	68.7	87.3	94.3	77.0

3.3.4 Study of DFEM Embedding Positions

To investigate the impact of inserting the DFEM at different stages of the ResNet-50 backbone network, experiments were conducted on the Occluded-ReID dataset. The per-formance of DFEM on this dataset is detailed in Table 6.

Results indicate that inserting DFEM after Layer2 yields optimal performance, achieving a Rank-1 accuracy of 78.2% (an increase of 6.0% compared to insertion after Layer1) and an mAP of 68.7% (+5.4% relative to Layer1). In contrast, insertion after Layer3 or Layer4 results in degraded performance, with Rank-1 accuracy declining to 75.2% and 72.9%, respectively. Therefore, Layer2 is identified as the optimal insertion position for the DFEM (providing a Rank-1 improvement ⩾6.0%), effectivel enhancing cross-scale feature perceptioncapabilities.

Table 6 Study on the embedding position of the DFEM

Module Insertion Position	Single-Query			
	Rank-1	Rank-5	Rank-10	mAP
After Layer1	73.8	85.7	90.4	65.2
After Layer2	78.2	88.9	92.5	68.7
After Layer3	75.2	87.6	91.1	66.4
After Layer4	72.9	86.5	91.0	65.0

3.3.5 Computational Complexity Analysis

To further validate the practicality of our method, we conduct a comprehensive com-putational complexity analysis, including parameter size, FLOPs, GPU memory usage, and inference time. The results are summarized in Table 7. Our model contains 87.16 M parameters and requires 14.62 G FLOPs for an input of 256 × 128. The peak memory usage is 614.6 MB when processing a single image, and the average inference latency is 68.7 ms per image on an NVIDIA GPU, corresponding to approximately 14 FPS.

Compared to the ResNet50 baseline (23.5 M parameters, 4.1 G FLOPs, and 12 ms per image), our model introduces additional computational overhead due to the integration of multiple enhancement modules. However, this increase remains within a feasible range, and the improved recognition accuracy under challenging occlusion scenarios justifies the additional cost. Importantly, the efficiency results demonstrate that our method achieves a balanced trade-off between lightweight design constraints and practical deployability.

Table 7 Computational complexity and inference performance analysis

Model	Params (M)	FLOPs (G)	Memory (MB)	Time (ms/img)
ResNet50	23.5	4.1	210.0	12.0
Baseline	87.1	14.1	614.6	72.2
MMCE-Net	87.2	14.6	615.0	68.7

4 Conclusions

This paper presents a collaborative enhancement framework for person re-identification aimed at addressing practical challenges such as occlusion, pose variation, and background interference. By integrating multi-scale attention into a ResNet-based architecture, the proposed approach effectively enhances local feature representation and improves robustness under partial occlusion.A key contribution of this work lies in the design of effective feature enhancement and fusion strategies, which enable more complete and discriminative pedestrian representations by jointly leveraging information from different network stages. Extensive experiments conducted on multiple public datasets, including occluded and conventional person re-identification benchmarks, demonstrate that the proposed framework achieves competitive and robust performance compared with representative state-of-the-art methods. Despite these promising results, further improvements are still possible. Future work will focus on developing more lightweight attention

mechanisms and exploring modality-adaptive strategies to extend the proposed framework to cross-domain scenarios and resource-constrained deployment environments. Nevertheless, the proposed framework introduces additional computational overhead compared to the baseline model, which may limit its applicability in real-time or resource-constrained scenarios.

About the Author

Ming;jun Ma, M.S., Northeast Forestry University, Harbin 150040, China. His research interests include person re-identification and computer vision.

Jun Li, Professor and Ph.D. supervisor at Northeast Forestry University. She received the Ph.D. degree in Computer Application Technology from Harbin Institute of Technology and completed postdoctoral research in Mechanical Engineering. Her research interests include intelligent materials and devices, multi-agent systems, and cooperative control methods. She has published more than 30 academic papers, including SCI and EI indexed papers, and has presided over several national and provincial research projects.

References

[1] Zheng, L.; Yang, Y.; Hauptmann, A.G. Person re-identification: Past, present and future. Int. J. Comput. Vis. 2016, 127, 211–238.

[2] Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z. Person re-identification by local maximal occurrence representation. Proc. CVPR 2015, 2197–2206.

[3] Sun, Y.; Zheng, L.; Deng, W.; Wang, S. SVDNet for pedestrian retrieval. Proc. ICCV 2017, 3820–3828.

[4] 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. Proc. CVPR 2020, 6449–6458.

[5] Miao, J.; Wu, Y.; Liu, P.; Ding, Y.; Yang, Y. Pose-guided feature alignment for occluded person re-identification. Proc. ICCV 2019, 542–551.

[6] Gao, S.; Wang, J.; Lu, H.; Liu, Z. Pose-guided visible part matching for occluded person re-identification. Proc. CVPR 2020, 11744–11752.

[7] Fan, X.; Luo, H.; Zhang, X.; He, L.; Zhang, C.; Jiang, W. SCPNet: Spatial-channel parallelism network for joint holistic and partial person re-identification. ACCV 2018.

[8] Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; Wang, S. Beyond part models: Person retrieval with refined part pooling. Proc. ECCV 2018.

[9] Fu, Y.; Wei, Y.; Wang, G.; Zhou, Y.; Shi, H.; Huang, T. Horizontal pyramid matching for person re-identification. Proc. AAAI 2019.

[10] Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. Proc. CVPR 2018.

[11] Sun, Y.; Xu, Q.; Li, Y.; Zhang, C.; Li, Y.; Wang, S.; Sun, J. Perceive where to focus: Visibility-aware part-level features for partial person re-identification. Proc. CVPR 2019.

[12] He, L.; Liang, J.; Li, H.; Sun, Z. Partial person re-identification with aligned spatial features. Proc. ICCV 2018

[13] Liu, J.; Zha, Z.J.; Wu, F. Partial re-identification via probabilistic alignment. IEEE Trans. Image Process. 2020.

[14] Jia, M.; Cheng, X.; Lu, S.; Zhang, J. Learning disentangled representation implicitly via transformer for occluded person re-identification. IEEE Trans. Multimed. 2023.

[15] Mikriukov, G.; Ravanbakhsh, M.; Demir, B. Deep Unsupervised Contrastive Hashing for Large-Scale Cross-Modal Text-Image Retrieval in Remote Sensing. arXiv preprint, 2022, arXiv:2201.08125. https://arxiv.org/abs/2201.08125

[16] Cui, X.; Liang, Y.; Zhang, W. Dual-branch person re-identification algorithm based on multi-feature representation. Electronics 2023.

[17] Bai, X.; Wang, A.; Zhang, C.; Hu, H. Cross-domain person re-identification based on normalized IBN-Net. IEEE Access 2024.

[18] Zhang, S.; Yang, J.; Schiele, B.Occluded person re-identification with occlusion-aware adversarial feature learning.Proc. ICCV, 2021.

[19] Li, S.; Sun, S.; Li, Y.; et al.Learning discriminative representation via self-paced adversarial learning for occluded person re-identification.IEEE Trans. Image Process., 2020.

[20] Wang, G.; Yuan, Y.; Chen, X.; et al. Learning discriminative features with multiple granularities for person re-identification. Proc. ACM MM 2018.

[21] Yang, Y.; Yu, S.; Li, X.; et al. HOReID: High-order relation-aware network for occluded person re-identification. Proc. CVPR 2020.

[22] Wang, T.; Liu, H.; Song, P.; et al. Pose-guided feature disentangling for occluded person re-identification based on Transformer. Proc. AAAI 2022.

[23] Xu, B.; He, L.; Liang, J.; Sun, Z. Learning Feature Recovery Transformer for Occluded Person Re-Identification. IEEE Transactions on Image Processing, 2022, 31, 4651–4662.