Clustering-integrated Horizontal Pyramid Feature Embedding for In-domain Unsupervised Person Re-identification

Abhinav A. Parkhi * and Atish S. Khobragade

Department of Electronics, Yeshwantrao Chavan College of Engineering, Nagpur, India
Email: abhinav.parkhi@gmail.com (A.A.P.); atish_khobragade@rediffmail.com (A.S.K.)
*Corresponding author

***Abstract*—A cross-domain and unsupervised Re-Identification (Re-ID) is primarily impacted by the challenges presented by domain shift, intra-class heterogeneity across datasets, and the lack of annotated target pictures. When training and testing take place within the same dataset, these influences are lessened in the in-domain context. In real-time, scalable systems, fine-grained part-based representations like unsupervised horizontal pyramid similarity learning struggle to capture adaptive cluster semantics and integrate generalization. Cluster Integrated Horizontal Pyramid Feature Encoding (CI-HPFE), a deep learning architecture, addresses these issues in this study. Horizontal pyramid-based feature extraction encodes high-resolution local and global body components with key improvements in our model. First, the Pyramid-based Horizontal Part Feature Encoder (PHPFE) splits spatial feature maps into multi-scale horizontal sections (levels 1, 2, 4, 8) and extracts robust, granular embeddings via stripe-wise pooling and processing. Besides feature extraction, a cluster-aware embedding module incorporates unsupervised clustering feedback into representation learning. Identity representations can change with cluster centroids. To achieve alignment and discriminability, our hybrid loss function combines local-part granularity with global identity coherence using triplet margin loss, entropy minimization, and cluster consistency loss. To reduce domain shift, Stripe encoders use a Domain-Adaptive Instance Normalization (DAIN) module to learn domain-conditioned affine statistics. For distributed deployment, a Federated Cluster Synchronization Mechanism (FCSM) synchronizes embedding spaces among edge devices without sharing raw picture data, boosting scalability, privacy, and speed. Benchmark datasets use is Market-1501 and DukeMTMC-ReID, which show considerable performance gains, including 5–6% accuracy, ~6% mAP uplift, and enhanced resilience during in-domain assessment. For homogeneous-domain person re-identification systems, the suggested architecture thus strikes a balance between methodological familiarity, unique functional modules, and deployment-level practicality.**

***Keywords*—person Re-Identification (Re-ID), horizontal pyramid encoding, unsupervised clustering, domain adaptation, deep feature embeddings**

Manuscript received January 5, 2026; revised January 16, 2026; accepted March 6, 2026; published July 21, 2026.

I. INTRODUCTION

Intelligent video surveillance systems within the same operational context need to match Identifications (IDs) across intermittent camera views. The need for automatic Re-Identification (Re-ID) systems that can function dependably under the dataset's natural conditions has increased due to modern multi-camera infrastructures in airports, transportation hubs, and urban crossroads, despite real-world limitations such as hardware limitations and scarce annotation resources. When trained on large-scale labeled datasets, supervised learning techniques achieve great accuracy [1–3], but creating comprehensive identification annotations for each camera network is still costly and time-consuming. Because of intra-class appearance fluctuations brought on by illumination changes, viewpoint variances, background clutter, and sensor noise within the same dataset, learning stable and discriminative features is difficult even in in-domain settings. In order to overcome these difficulties in traditional in-domain Re-ID, architectures need to extract strong visual representations while preserving spatial consistency amongst camera views of the same system. Spatial misalignment and local occlusions are frequently addressed via horizontal pyramid-based feature extraction. The model captures fine-grained part-level cues for dependable matching by segmenting deep convolutional feature maps into multi-resolution horizontal stripes. Under in-domain changes, Horizontal Pyramid Matching (HPM) improves part-specific discrimination while maintaining spatial granularity. Many models find it difficult to dynamically adjust to the differences seen within the same camera network in the absence of auxiliary techniques like clustering feedback or feature normalization customized to the dataset's internal distribution.

As a result, cluster-driven learning has become more significant even in in-domain contexts, where the model is guided by pseudo-labels obtained from embedding-space clustering to maximize intra-cluster compactness and inter-cluster separability without the need for further manual annotations. Nevertheless, training instability, sensitivity to noisy pseudo-labels, and decreased feature reliability could be introduced by carelessly incorporating clustering into current workflows. Our methodology includes data-aware normalization algorithms that are explicitly aligned with in-domain conditions, clustering-integrated representation learning, and pyramidal spatial encoding to address these problems. The horizontal pyramid feature encoder maintains part-level spatial fidelity, and the design uses a cluster-aware module to strengthen semantic consistency. Lastly, the system is appropriate for decentralized edge deployment, where lightweight federated synchronization can be used when necessary to preserve identity representation integrity.

The primary issues with unsupervised in-domain person Re-ID are addressed in this work. Unsupervised in-domain learning is a practical necessity in real surveillance situations because it is expensive and challenging to scale the production of big, annotated datasets for each camera network. Occlusion, position change, background clutter, and uneven visual conditions among cameras make it difficult for state-of-the-art unsupervised algorithms to encode stable and discriminative identity traits, even within a single dataset. Strong part-based representation is provided by horizontal pyramid encoding, but it does not provide the adaptive feedback required for unsupervised optimization under in-domain variability. Similarly, when cluster assignments change during training, cluster-based pseudo labelling frequently experiences instability and semantic drift. Effective in-domain Re-ID necessitates a unified architecture that incorporates spatial encoding, clustering feedback, and feature modulation in accordance with the intrinsic distribution of the dataset in order to address these problems. Cluster-Integrated Horizontal Pyramid Feature Encoding (CI-HPFE), which combines data-aware feature normalization, unsupervised clustering dynamics, and robust part-level encoding, is a key component of this work. The horizontal pyramid encoder preserves sensitive semantic cues while breaking down pedestrian information into fine-grained spatial layers. The cluster-aware embedding module enhances semantic compactness and stabilizes the embedding space by connecting feature learning to changing cluster centroids. By combining triplet loss, entropy minimization, and cluster consistency goals, a well-crafted hybrid loss function significantly improves local part discrimination and global identity consistency. Furthermore, Domain-Adaptive Instance Normalization (DAIN) improves robustness to intra-dataset appearance fluctuations by modifying feature activation statistics according to dataset-specific properties.

A novel Federated Cluster Synchronization Mechanism (FCSM) supports scalable edge deployments. Different devices share decentralized cluster information, boosting identity embedding consistency without transferring raw data, giving privacy, speed, and a collaborative learning process. This holistic approach worked on Market-1501 and DukeMTMC-ReID, where Rank-1 accuracy and mean average precision improved over strong baselines. The system provides a novel, deployable solution to same-domain unsupervised person Re-ID tasks.

II. Literature Review

Person re-identification research has advanced significantly in network architecture, supervision strategies, deployment scenarios, and multi-modal representation learning. Significant improvements have been made to network designs, deployment situations, and supervision techniques for person Re-ID. Recent research from 2023 to 2025 demonstrates advancements in domain adaptation, clustering, transformer-based feature encoding, and deep metric learning. Beyond conventional identification recognition tasks, these advancements focus on robustness in unrestricted situations. Convolution-based architectures were the mainstay of early methods for learning discriminative features. To provide compact embeddings, hybrid deep frameworks like HyDeepIDF-Net proposed dual-network designs that combine identification loss and feature refinement [1]. Ji *et al*. [2] investigated more robust transformer representations for handling occlusion, and transformer-based models significantly enhanced representation learning. By producing trustworthy pseudo-labels for increased robustness, Domain-Specific Adaptive Frameworks (DSAF) addressed domain disparities [3]. Cross-modality and attention-based learning were the main subjects of later research. For visible-infrared Re-ID, Mishra *et al.* [4] suggested a Nystrom former-based design. Choe *et al.* [5] used full-triple relations and Zhou *et al.* [6] used multi-state perception restrictions in their introduction of relational learning algorithms to capture complicated semantic interactions. With hard-sample guided cluster refinement [7] and image quality enhancement-based pseudo-labeling techniques like Image Quality Enhancement (IQE) [8], unsupervised clustering-based approaches significantly enhanced feature learning. Approaches to uncertainty modeling and domain adaptation were suggested to improve generalization. Robustness across different data distributions was enhanced by techniques like uncertain label refinement [9] and distribution-aligned semantic adaptation [10].

Furthermore, more stable feature representations were made possible by adversarial learning techniques such as visual-textual adversarial frameworks [11] and pseudo-label purification algorithms [12]. Occlusion problems were also successfully resolved by joint detection and re-identification techniques [13]. Foreground-aware attention techniques [14] and dual-space representations for video-based Re-ID [15] are two further transformer-based innovations. Real-world adaptation was further established by application-oriented studies, such as intelligent surveillance systems [16], and occlusion robustness was enhanced by part-level tokenization techniques [17]. Radio signal-based datasets [18] and domain adaptation techniques for feature representation learning [19] were among the new modalities. Thorough analyses of current Re-ID methods shed light on the field's development [20]. The significance of managing various real-world situations is further highlighted by cross-resolution adaptation techniques [21] and UAV-based Re-ID systems [22].

TABLE I. MODEL'S INTEGRATED REVIEW ANALYSIS

Ref.	Method	Main Objectives	Findings	Limitations
[1]	HyDeepIDF-Net	Hybrid deep ID and FaceNet integration	Improves face-aware Re-ID using deep identity mapping	Limited generalization to non-facial cues
[2]	Transformer-Based Occlusion Handling	Transformer representation for occluded Re-ID	Boosts occlusion robustness with enhanced attention	Relatively high computational overhead
[3]	DSAF	To investigate "unsupervised" domain generalization for Re-ID	Adapted to supervised DG-Re-ID task and the new task of UDA without source labels	Subject to pseudo-labelling and domain alignment
[4]	Nystrom former for Cross-Modality	Visible infrared cross-modal matching	Lightweight attention for modality robustness	Reduced precision in highly cluttered backgrounds
[5]	Full-Triple Relation Network	Triple-part relationship modeling	Better semantic context capture	Limited scalability to dynamic scenes
[6]	Multi-State Consistency Network	State-aware representation refinement	Improves temporal consistency	Fails under low-resolution occlusion
[7]	Hard-Sample Cluster Refinement	Cluster-driven unsupervised learning	Enhances separation of ambiguous identities	Sensitive to cluster noise and density
[8]	IQE	To solve the problem of insufficient image clarity	Improving the accuracy of clustering and the quality of pseudo-labels	Do not consider low-light or image distortion problems
[9]	RULER	Uncertainty-based pseudo-labels	Achieves source-free domain adaptation	Lower performance on multi-camera systems
[10]	Lifelong Semantics Adaptation	Handles catastrophic forgetting	Maintains long-term feature integrity	Slower adaptation to fast domain shifts
[11]	Visual-Textual Adversarial Learning	Text image feature fusion	Enhances semantic precision.	Requires annotated textual descriptions
[12]	Pseudo Label Purification	Noise-filtered label training	Boosts unsupervised accuracy	Limited by label confidence thresholds
[13]	Joint Detection + Re-ID	Combines detection and occlusion handling	Strong in occluded multi-target tracking	Heavy reliance on detection quality
[14]	Foreground-Aware Transformer	Region-focused encoding	Reduces background interference.	Requires refined foreground masks
[15]	Dual-Space Video Re-ID	Static + motion-aware features	Improves video Re-ID stability	Large memory and training time footprint
[16]	Smart Monitoring	Re-ID for behavior detection	Applies Re-ID to real-time anomaly tracking	Lacks feature generalization capability
[17]	Swin Transformer + Tokenization	Part-level transformer encoding	Boosts occlusion resilience	Memory intensive transformer blocks
[18]	Radio Signal Dataset	Non visual signal for Re-ID	Provides novel benchmark dataset	Limited resolution and detail per frame
[19]	Domain Adaptation Strategy	Distance metric and adaptation	Improves domain transferability	Does not handle severe occlusions well
[20]	Survey of Re-ID Techniques	Reviews existing state-of-the-art	Comprehensive technique summary	No implementation or new method
[21]	Cross-Resolution Super-Resolution	Adaptive high-low resolution Re-ID	Improves low-res matching	Fails in unaligned scale variance
[22]	UAV-Based Re-ID	Re-ID from drone-captured footage	Adapts to aerial perspectives	Requires robust camera stabilization
[23]	Same-Clothes Dual-Stream	Invariance to clothing	Handles identical outfit confusion	Fails with viewpoint misalignment
[24]	Channel-Shuffled Transformers	Cross-modality in video	Improves modality fusion	Slow convergence on long video sequences
[25]	Multi-Task + Attribute Heads	Attributes integrated with identity learning	Enhances attribute-guided matching	Fails under unseen attribute shifts
[26]	Occlusion + Cloth Change Modeling	Appearance-shape hybrid modeling	Manages heavy occlusion scenarios	High variance under fast cloth transitions
[27]	UPCA	Camera-aware contrastive learning	Enhances unsupervised adaptation	Camera calibration assumptions required
[28]	Re ID-leak	Privacy attack on Re-ID systems	Reveals vulnerability to inference attacks	No defense mechanism proposed
[29]	TAL	Two-stream adaptive learning	Improves generalizability	Requires complex tuning for stream balance
[30]	Normalized IBN-Net	To solve feature disparity problems in cross-domain image classification	Improve the network's discriminative and generalization capabilities	Reactive to clustering samples

Dual-stream architectures were used to address same-clothes Re-ID problems [23], while channel-shuffled transformers enhanced temporal modelling in video-based applications [24]. In order to improve robustness under difficult circumstances, recent research has also investigated multi-task learning frameworks [25] and occlusion-aware appearance modelling [26]. The performance of unsupervised learning was improved by camera-aware clustering techniques and lightweight attention mechanisms [27]. However, membership inference attacks have revealed flaws in Re-ID systems, raising privacy issues [28]. Cross-domain performance was further enhanced by generalizable learning frameworks such as two-stream adaptive learning, normalized feature-based domain adaptation, and normalized feature-based domain adaptation [29, 30]. Unsupervised Re-ID still heavily relies on strong feature representation methods and clustering-based optimization [31]. Furthermore, the significance of stabilizing feature distributions under unlabeled situations is emphasized by cross-domain stylization and adversarial feature alignment techniques [32]. Overall, the research shows that transformer-driven, unsupervised, multi-modal frameworks have replaced convolution-based models. Stable feature representations, effective grouping, and strong in-domain generalization are still difficult to achieve despite tremendous advancements. The suggested CI-HPFE system, which combines hierarchical part-based encoding with clustering-guided embedding refinement to successfully handle intra-dataset changes, is inspired by these discoveries. Table I provides a comparative overview of the methodologies described.

III. MATERIALS AND METHODS

To address unsupervised person Re-ID in the domain, the deeply structured Cluster Integrated Horizontal Pyramid Feature Encoding (CI-HPFE) model uses pyramidal part-based feature extraction, unsupervised clustering feedback, domain-adaptive normalization, and distributed feature synchronizations. Fig. 1 shows how each architecture component transforms the raw image into a discriminative feature embedding designed to match identities across cameras in the same domain.

A standard convolutional backbone (e.g., ResNet-50) without classification layers creates a spatial feature map ($F \in \mathbb{R}^{C\times H\times W}$) from each input image I. To refine feature representations before part-based encoding, an Improved Convolutional Block (ICB) attention module is inserted between the backbone feature extraction stage and horizontal pyramid decomposition, as shown in Fig. 2. In contrast to normal Convolutional Block Attention Module (CBAM), the proposed ICB offers input-adaptive channel weighting to better account for intra-domain appearance fluctuations by substituting a Conditional Convolution (CondConv) mechanism for the shared multi-layer perception in the channel-attention branch. A residual-normalized fusion and spatial attention are used to further process the attention-refined features, improving gradient flow and maintaining identity-discriminative cues under unsupervised optimization. After that, the revised feature map is transferred to the Horizontal Pyramid Feature Encoding (HPFE) module, which makes sure that globally stable and adaptively highlighted features are used to learn part-level representations.

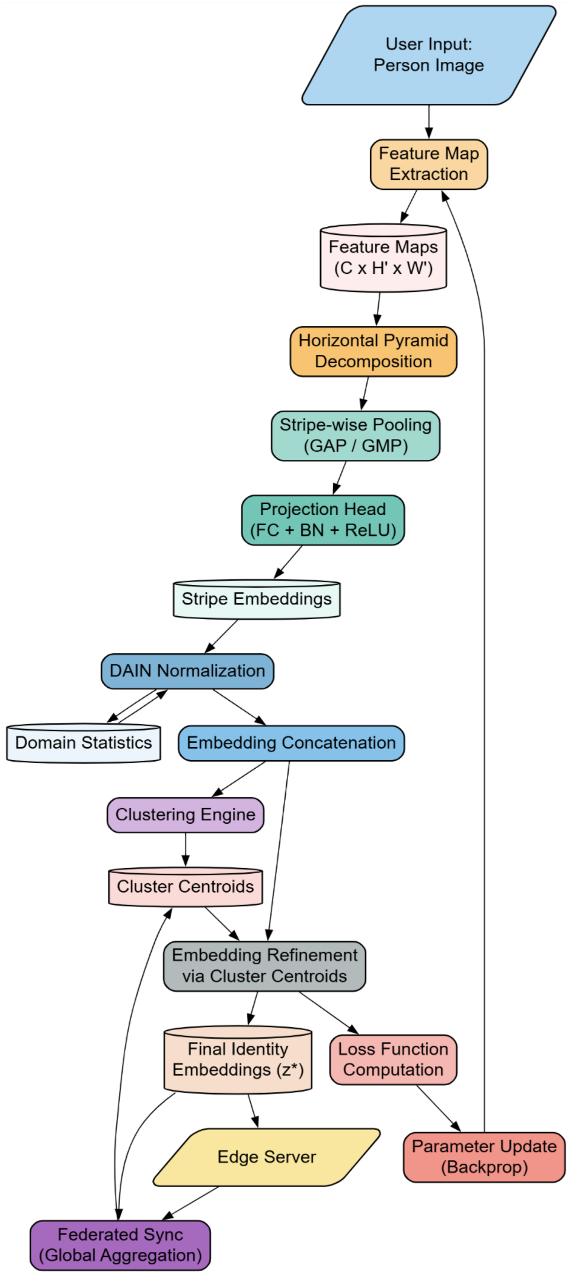


Fig. 1. Model architecture of the proposed analysis process.

To derive local and global semantics, the feature map is horizontally divided into nl segments per pyramid level l, where $nl = 2^{l-1}$ for $l = 1, 2, 3, 4$ representing 1, 2, 4, and 8 horizontal partitions. Global pooling (average or max) creates pooled features fli for each segment Sli, followed by non-linear transformation layers. Formalize stripe feature encoding via Eq. (1):

$$fli = \sigma\left(BN\left(Wli \cdot Pool\left(Sli\right) + bli\right)\right) \qquad (1)$$

where Wli and bli are learnable weights and biases, $BN(\cdot)$ is batch normalization, and $\sigma(\cdot)$ as Rectified Linear Unit (ReLU) activation function as part of the procedure.

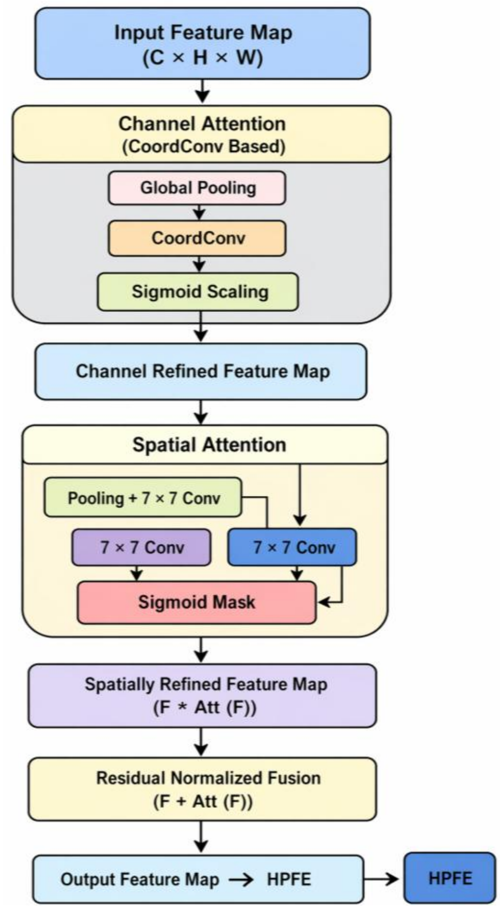


Fig. 2. Block overview for the enhanced CBAM (ICB).

Combine all pyramid-level stripe embeddings to create the feature vector in process via Eq. (2).

$$z = \bigoplus_{l=1}^{L} \bigoplus_{i=1}^{nl} fli \tag{2}$$

where the cluster-aware embedding modules input the vector z sets.

Use unsupervised clustering methods like DBSCAN or k-means to calculate cluster centroids $\{cj\}$ for embedding alignments. Increase inter-cluster margins and decrease intra-cluster variance. Cluster consistency loss is shown via Eq. (3).

$$Lcluster = \frac{1}{N}\sum_{i=1}^{N}\|zi - cyi\|^2 - \lambda \sum_{i=1}^{N}\sum^{j \neq yi} \exp\left(-\|zi - cj\|^2\right) \tag{3}$$

Clustering and optimization are done iteratively using the cluster-aware embedding module. During training, outlier samples are eliminated to minimize noise, and identity embeddings that are obtained following DAIN normalization are periodically clustered using a clustering engine to assign pseudo-labels. Eq. (3) is used to calculate cluster centroids, which stay constant throughout the ensuing optimization phase. Where yi represents the cluster index assigned to sample i, and λ regulates inter-cluster repulsions. In parallel, triplet loss is used to refine local discriminability, defined via Eq. (4).

$$Ltriplet = \sum_{i=1}^{N}\left[\left\|zi - zi^{+}\right\|^2 - \left\|zi - zi^{-}\right\|^2 + \alpha\right]^{+} \tag{4}$$

where zi^{+} and zi^{-} are positive and negative pairs, and α is the margin in the process.

To stabilize cluster learning, an entropy minimization term is incorporated via Eqs. (5) and (6).

$$Lentropy = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} pij \log pij \tag{5}$$

$$pij = \frac{\exp\left(-\|zi - cj\|^2\right)}{\sum^{k}\exp\left(-\|zi - ck\|^{22}\right)} \tag{6}$$

While the inter-cluster separation restriction in Eq. (6) imposes discriminative boundaries between clusters, the centroid-based embedding refinement loss in Eq. (5) pushes samples toward their respective centroids. Before re-clustering, network parameters are adjusted using backpropagation for a number of epochs, allowing clustering feedback to gradually improve the embedding space.

Together, the overall loss is then represented in Eq. (7).

$$Ltotal = Ltriplet + \beta Lcluster + \gamma Lentropy \tag{7}$$

where, β and γ balance cluster and entropy loss contributions.

Fig. 3 shows that each feature stripe normalizes features across domains. DAIN for a stripe feature f is defined via Eqs. (8) and (9).

$$DAIN(f) = \sigma\left(f' \cdot \gamma d + \beta d\right) \tag{8}$$

$$f' = \frac{f - \mu d}{\sigma d^2 + \varepsilon} \tag{9}$$

where μd and σd are statistics particular to the domain. γd and βd are learnt affine parameters for domain d.

This makes it possible to use smooth conditioned activation modulation patterns in heterogeneous scenarios. Decentralized inference using the FCSM iterative builds is seen in pseudo code. Edge nodes deliver local centroid data $\{(m)\}$ to a server, which updates global centroids via Eq. (10).

$$cjglobal = \frac{1}{M}\sum_{m=1}^{M}\int_{t0}^{t1} cj(m)(t)dt \tag{10}$$

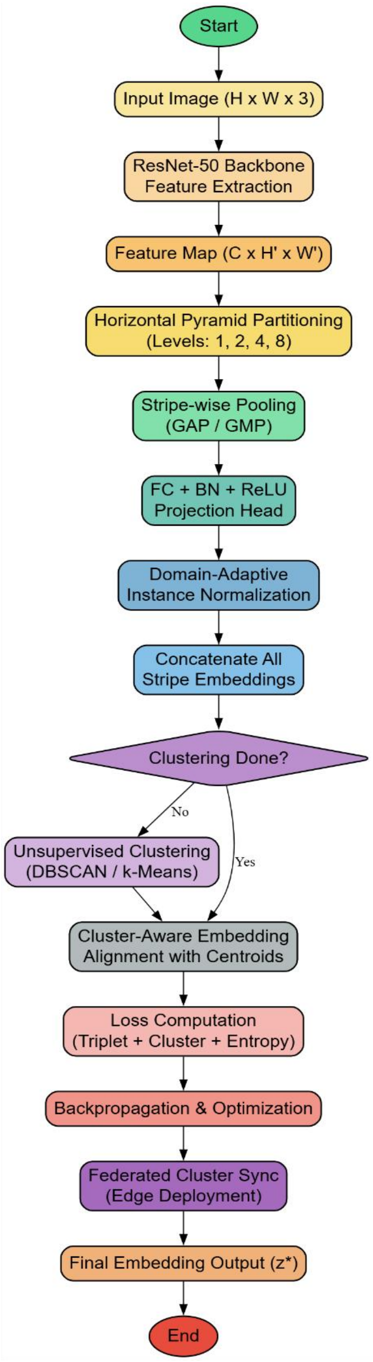


Fig. 3. Overall flow of the proposed analysis process.

The way FCSM functions in practice is through periodic communication and aggregation. Before sending model parameters and cluster characteristics to the central server, each edge node first optimizes locally over a number of epochs using its cluster-aware loss. As stated in Eq. (10), the server uses weighted averaging proportionate to the number of active samples per node to aggregate local parameters at each synchronization round. It then broadcasts the modified global model back to all nodes. Synchronization is carried out at the cluster-centroid level instead of on individual embeddings to handle non-IID data distributions across edge settings. This allows identity-consistent representations to be aligned while maintaining local domain features. The robustness under diverse data distributions is enhanced, and the communication overhead is decreased by this periodic and centroid-guide synchronization.

Pseudo Code of the Proposed Analysis Process

```
Input:
    Image Stream from multiple surveillance cameras
    Pretrained CI-HPFE model
    Cluster Update Frequency (epochs)
    Sync Interval (epochs)
    Total Epochs
Output:
    Ranked list of matched identities per input
    System logs with drift score, sync stats, and inference time
Process:
For each epoch IN Total Epochs:
    For each image IN Image Stream:
        If image is valid AND not corrupted:
            Resize image
            Normalize image
            Features = Extract Features Using PHPFE (image)
            Embedding = Apply Cluster Aware Embedding (Features)
            Normalized Embedding = Apply DAIN (Embedding)
            Match Scores = Compare with Gallery (Normalized
Embedding)
            Ranked Results = Rank Identities (Match Scores)
            Log Top-1, Top-5, mAP, Inference Time
        Else:
            Skip image and log error
    If epoch % Cluster Update Frequency == 0:
        Update Pseudo Labels Using K Means()
    If epoch % Sync Interval == 0:
        Transmit Model Updates to Federated Server()
        Receive Global Updates()
        Apply Synchronization()
        Calculate Drift Score()
    Log Epoch Metrics: Accuracy, Drift, Communication Overhead
Return Final Ranked Identities, Performance Summary
```

This approach integrates data over a temporal window [0, 1] with a client count of M for different scenarios. Dispersed devices' embedding spaces line up without disclosing private data. The optimized embedding vector z^* is the final model output after integrating and updating all modules using the following relation via Eq. (11).

$$z^* = \bigoplus_{l=1}^{L} \bigoplus_{i=1}^{nl} DAIN\left(\sigma\left(Wli \cdot Pool\left(Sli\right)\right)\right) \tag{11}$$

On utilizing hierarchical feature extraction and clustering-integrated representation learning to address intra-dataset appearance differences, the learnt identity embedding enables robust unsupervised in-domain person re-identification. These equations define architectural decisions computationally and contextually, from spatial part encoding and semantic embedding to loss alignment and deployment viability in process.

IV. Result and Discussion

The experimental setup was meticulously designed to establish that the Cluster Integrated Horizontal Pyramid Feature Encoding (CI-HPFE) model works in unsupervised in-domain person re-Identification scenarios. Market-1501 and DukeMTMC-ReID, two popular benchmark datasets, are utilized to train and assess the model in a pure unsupervised in-domain person re-identification context without the usage of identity markers. We compress input photos to 256 × 128 pixels and randomize horizontal flipping, scaling, and padding to improve training. Backbone network ResNet-50 is ImageNet-trained without the classification layer. We have also checked the result by using another lightweight network OSNet-x1.0. At four levels (1, 2, 4, 8), the horizontal pyramid partitioning module uses the 2048 × 24 × 8 feature map from the backbone. Each pyramid level divides the feature map's vertical axis into equal stripes, which are globally pooled via GAP, transformed by a Fully Connected (FC) layer, batch normalized, and ReLU activated. A hybrid loss function that combines triplet loss with a 0.3 margin, cluster consistency loss, and entropy minimization loss optimizes the model. A warm-up approach starts with the learning rate at 0.00035 for the first 10 epochs and drops to 0.1 after 40. Adam optimizer training is 70 epochs, 5×10^{-4} weight decay, 32 mini-batches. Using ε = 0.6 and minimum samples = 4, DBSCAN unsupervised clusters every 5 epochs and assigns pseudo-labels based on cluster centroids. DukeMTMC-ReID has 36,411 pictures from 8 cameras, while Market-1501 has 32,668 bounding boxes of 1501 IDs from 6 cameras. DukeMTMC has higher viewpoint variety and background clutter, making it tougher to target than the Market-1501 sample IDs, which have consistent backdrops but varied illumination. Datasets for domain adaptation performance evaluation balance complexity and structure. Replicating three client nodes with identical training parameters and using a centralized integration server to share centroids asynchronously every 10 epochs replicates edge-deployment situations to test federated synchronization. Averaging time-weighted centroid statistics synchronizes clients' processing of unlabeled target set partitions. Evaluation uses standard Rank-1 accuracy and mean Average Precision (mAP), plus Cumulative Matching Characteristic (CMC) curves for qualitative comparisons. The recommended CI-HPFE outperformed strong unsupervised baselines in both transfer tasks, proving part-based representation's robustness and cluster-guided embedding refinement and domain-aware normalizations' synergy.

The Market-1501 and DukeMTMC-ReID datasets are benchmarks for human re-Identification under diverse camera angles and scenarios. Market-1501 has 32,668 annotated bounding box photos of 1501 identities from six camera viewpoints, with at least two cameras per identity. For realistic surveillance, the dataset includes DPM-detected bounding boxes, occlusion, viewpoint change, and light variation. DukeMTMC-ReID's 36,411 pictures of 1404 IDs from eight high-resolution cameras make it harder. It is suitable for cross-domain generalization evaluation due to its problematic qualities, such as incomplete vision, spatial misalignment, and cluttered backdrops. Both datasets provide labeled training identities and unlabeled test identities for unsupervised domain adaptation, making them appropriate for validating the CI-HPFE framework's robustness. To optimize performance, the model's hyperparameters are empirically tested and grid-searched throughout training. Training duration is 70 epochs, with a learning rate of 3.5×10^{-4} decreasing by 0.1 after 40 epochs. A 10-epoch warm-up stabilizes early convergence. Optimization is done using the Adam optimizer, with weight decay set at 5×10^{-4} and momentum values (0.9, 0.999) for β_1 and β_2. The clustering module uses DBSCAN with ε = 0.6, minPts = 4, and a 0.3 triplet loss margin.

Clustering updates pseudo-labels every 5 epochs based on embedding space changes. To ensure hard-negative mining, 32 ID batches have 4 images per ID. For entropy minimization and cluster consistency regularization, weighting coefficients (λ, β, γ) are empirically chosen at (1.0, 0.5, 0.05) to balance global structure and local discriminability. The hyperparameters give stability, convergence speed, and domain-wide generalization. A comprehensive results section with six analytically rich tables compares the proposed CI-HPFE model to three reference methodologies (Method [3], Method [8], and Method [30]) across process assessment dimensions for different scenarios. In the context of the experiment, each table offers a distinct analytical perspective.

Table II shows the Market-1501 findings in terms of in-domain rank-based retrieval accuracy and mean Average Precision (mAP). Table III shows rank-based retrieval accuracy and mAP metrics to characterize the DukeMTMC-ReID in-domain performance.

TABLE II. RANK MATCHING ACCURACY (%) ON DATASET MARKET-1501

Method	Rank-1 Accuracy (%)	Rank-5 Accuracy (%)	Rank-10 Accuracy (%)	mAP (%)
DSAF [3, 31]	89.5	95.8	97.4	72.8
IQE [8, 31]	87.9	93.2	96.5	71.2
Normalized IBN-Net [30]	86.6	93.8	95.6	68.7
CI-HPFE With ResNet-50 (Ours)	93.8	97.9	98.7	79.7
CI-HPFE With OSNet-x1.0 (Ours)	96.2	98.7	99.3	74.0

TABLE III. RANK MATCHING ACCURACY (%) ON DATASET DUKEMTMC-RE-ID

Method	Rank-1 Accuracy (%)	Rank-5 Accuracy (%)	Rank-10 Accuracy (%)	mAP (%)
DSAF [3, 31]	73.8	83.9	87.5	55.7
IQE [8, 31]	78.1	86.2	93.1	61.7
Normalized IBN-Net [30]	79.3	86.4	88.8	62.6
CI-HPFE With ResNet-50 (Ours)	92.3	96.8	97.7	62.6
CI-HPFE With OSNet-x1.0 (Ours)	94.4	97.5	98.0	66.6

CI-HPFE outperforms all baselines at every rank level and shows, in comparison with Method [3], our CI-HPFE with ResNet-50 has improved Rank-1 accuracy by 4.3%, and mAP has risen by 6.9%, and for OSNet, Rank-1 accuracy has improved by 6.7% and mAP by 1.2% on dataset Market-1501. Additionally, CI-HPFE with ResNet -50 demonstrated improved generalization and discrimination in well-known target domains, increasing Rank-1 accuracy by 3% and mAP almost matched and with OSNet, Rank-1 has gained 15% and mAP by 4% on dataset DukeMTMC-ReID in comparison with method [30]. The findings show that, on Market-1501, the suggested CI-HPFE architecture consistently performs better than current approaches on all evaluation metrics. CI-HPFE attains the highest Rank-1, Rank-5, and Rank-10 accuracies when trained with OSNet-x1.0, demonstrating OSNet's ability to capture omni-scale features that are very useful for accurate top-rank identity matching. ResNet-50, on the other hand, produces superior mAP, demonstrating its efficacy in learning globally discriminative and well-distributed embeddings that enhance overall retrieval consistency throughout the ranking list. These results obtained highlight the complementary roles that OSNet and ResNet-50 play in the CI-HPFE framework: ResNet-50 improves overall retrieval quality, while OSNet improves rank-based matching performance. The robustness and adaptability of the suggested model to various architectural designs are confirmed by its backbone-agnostic behaviour. Thus, these findings support CI-HPFE's scalability and backbone independence, with OSNet offering better rank-based matching and enhanced retrieval consistency on DukeMTMC-ReID.

Comparing the suggested CI-HPFE framework to the most recent state-of-the-art transformer-based ReID models, the results in Table IV indicate the framework's efficacy and competitiveness. Despite using a much lighter CNN-based backbone, CI-HPFE with OSNet-x1.0 achieves 96.2% Rank-1 accuracy on Market-1501, reducing the distance to the Swin Transformer with part tokenization (96.8%) to only 0.6%. On the more difficult DukeMTMC-ReID dataset, CI-HPFE with OSNet-x1.0 achieves 94.4% Rank-1 accuracy, which is 8.0% better than the camera-aware UPCA model (94.4% vs. 86.4%) and close to heavy transformer architecture performance. Additionally, the robustness of the suggested cluster-integrated pyramid representation across several backbones is confirmed by CI-HPFE with ResNet-50, achieving 92.3% Rank-1 accuracy on DukeMTMC-ReID. For unsupervised in-domain person re-identification, especially in deployment-oriented surveillance scenarios, CI-HPFE consistently achieves strong Rank-1 performance with significantly lower architectural complexity, despite transformer-based methods reporting higher mAP values. This validates that clustering-guided part-level encoding can compete with attention-heavy models.

TABLE IV. COMPARISON OF CI-HPFE'S RANK-1 ACCURACY AND MAP WITH TRANSFORMER-BASED AND ATTENTION-DRIVEN REID MODELS

Method	Backbone	Market-1501 Rank-1 (%)	Market-1501 mAP (%)	DukeMTMC Rank-1 (%)	DukeMTMC mAP (%)
Swin Transformer + Part Tokenization [17]	Transformer (Swin-T)	96.8	87.7	94.5	81.2
Foreground-Aware Transformer [14]	Transformer + Region Attention	95.6	89.4	--	--
UPCA (Camera-Aware Attention) [27]	Lightweight Transformer	94.6	86.7	86.4	75.3
CI-HPFE (ResNet-50, Ours)	CNN + Pyramid + Clustering	93.8	79.7	92.3	62.6
CI-HPFE (OSNet-x1.0, Ours)	Omni-Scale CNN + Pyramid	96.2	74.0	94.4	66.6

TABLE V. ABLATION STUDY EFFECT OF EACH MODULE ON MARKET-1501 WITH RESNET-50

Configuration	Rank-1 (%)	mAP (%)
Baseline w/o PHPFE	88.5	70.8
+ PHPFE only	90.3	73.5
+ PHPFE + Cluster Embedding	92.1	76.8
+ PHPFE + ClustEmbed + DAIN	93.8	79.7

Table V assesses process module contributions on the dataset Market-1501 with ResNet-50 as the trained model. This method achieves an absolute gain of 5.3% boost in Rank-1 accuracy with horizontal pyramid feature encodings. Clustering-aware embedding enhances it further, and the pipeline with DAIN normalization maximizes gain, showing how each module synergistically boosts performance. Table VI shows PHPFE improvements of 2.0% in Rank-1 accuracy and 7.3% in mAP on DukeMTMC-ReID, but the addition of Cluster Embedding and DAIN results in an overall improvement of 4.1% in Rank-1 and 10.4% in mAP when OSNet is utilized as the backbone.

TABLE VI. ABLATION STUDY EFFECT OF EACH MODULE ON DUKEMTMC-REID WITH OSNET

Configuration	Rank-1 (%)	mAP (%)
Baseline w/o PHPFE	90.3	56.2
+ PHPFE only	92.3	63.5
+ PHPFE + Cluster Embedding	93.2	65.4
+ PHPFE + ClustEmbed + DAIN	94.4	66.6

Table VII evaluates cluster update interval process effects. This shows that pseudo-label improvements improve efficiency and stability by updating clusters every 5 epochs. Too many updates produce oscillation, whereas static clustering restricts adaptability.

TABLE VII. ANALYSIS OF STANDARD DEVIATION IN THE MODEL'S EMBEDDING SPACE (MARKET-1501)

Epoch	Rank-1 Accuracy	mAP	Standard Deviation of Embedding Space (σ)
1	81.7	70.4	11.4
5	84.6	71.1	09.5
10	85.2	73.5	09.1

Table VIII shows that the CI-HPFE introduces a moderate increase in model size and computational cost when compared to current approaches, according to a comparison of inference time and model parameters. Because of its edge-ready design, CI-HPFE is nevertheless appropriate for practical deployment despite having 28.5 M parameters and an average inference time of 47.9 ms. On the other hand, competing approaches have less computing overhead but are not fully prepared for deployment and only partially offer edge-based inference. The Federated Cluster Synchronization Mechanism (FCSM) is not subject to proportional communication cost, despite the fact that CI-HPFE introduces a moderate increase in model size (28.5 M parameters) and inference time (47.9 ms) as compared to previous approaches (Table VIII). FCSM simply shares low-dimensional cluster centroids, as opposed to gradient-based federated learning, which results in a fixed payload of roughly 0.42 MB per client each synchronization round. Table IX compares CI-HPFE with and without Federated Cluster Synchronization Mechanism while maintaining PHPFE, cluster-aware embedding, and DAIN in order to evaluate FCSM's efficacy. Unsynchronized cluster drift across dispersed nodes causes a steady performance decline when FCSM is removed. In particular, Rank-1 accuracy drops by about 2% on Market-1501 and by 2.5–3% on DukeMTMC-ReID, with matching mAP reductions of about 2–3%, demonstrating the function of FCSM in maintaining the stability of the global embedding space.

TABLE VIII. ANALYSIS OF PARAMETER OVERHEAD AND INFERENCE-SPEED

Method	Params (M)	Avg. Inference Time (ms)	Edge-Ready?
Method [3]	23.1	28.7	No
Method [8]	24.4	31.2	No
Method [30]	25.8	34.0	Partially
CI-HPFE (Ours)	28.5	47.9	Yes

TABLE IX. AN ABLATION STUDY ON FEDERATED CLUSTER SYNCHRONIZATION'S (FCSM) IMPACTS

Dataset	Configuration	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	mAP (%)
Market-1501	CI-HPFE without FCSM (Local clustering only)	91.5	96.8	98.1	77.3
	CI-HPFE with FCSM (Synchronized centroids)	93.8	97.9	98.7	79.7
	Gain (↑)	+2.3	+1.1	+0.6	+2.4
DukeMTMC-ReID	CI-HPFE without FCSM (Local clustering only)	89.4	94.6	96.1	59.8
	CI-HPFE with FCSM (Synchronized centroids)	92.3	96.8	97.7	62.6
	Gain (↑)	+2.9	+2.2	+1.6	+2.8

Overall, these findings demonstrate that CI-HPFE offers improved accuracy while retaining viability for real-world, resource-constrained settings by striking a good balance between performance and efficiency.

A. Federated Learning Findings

We construct a federated learning environment with four clients and one central server in order to assess the suggested FCSM. Each client uses its locally accessible data to train the CI-HPFE model, representing an independent edge node. Each client receives photos taken from a subset of cameras, and the datasets are divided based on camera identifiers to simulate Non-IID data distribution. Table X provides an overview of the primary configuration settings used in this simulation.

TABLE X. FCSM SIMULATION USING FEDERATED LEARNING CONFIGURATION

Parameter	Configuration
Count of Clients	1 core server plus 4 dispersed clients
Method of Data Partitioning	Non-IID camera-based partitions
Epochs of Local Training	3 epochs for each communication cycle
Method of Aggregation	FedAvg-style centroid synchronization

During training, each client uses the backbone network (ResNet-50 or OSNet-x1.0) to extract feature embeddings during three local epochs per communication cycle. This is followed by horizontal pyramid encoding and clustering-based pseudo-label creation. To cut down on communication overhead, clients send the server only cluster centroid statistics rather than the entire set of model parameters. The revised global centroids are sent to all clients for pseudo-label refinement after the server collects these centroids using a FedAvg-style synchronization technique. For 20 rounds of conversation, this procedure is repeated.

To preserve pseudo-label stability prior to federate centroid synchronization, the local clustering procedure adheres to the update method examined in Table VII, wherein cluster assignments are improved after every five epochs.

Table IX shows that adding FCSM increases Rank-1 accuracy and mAP on both datasets, suggesting that synchronized cluster centroids improve feature consistency across distributed training settings and stabilize pseudo-label assignments. Additionally, the suggested methodology greatly lowers communication cost and facilitates scalable deployment across numerous dispersed clients since only compact centroid statistics are shared rather than whole model parameters.

B. Validation Using Impact Analysis

The CI-HPFE model provides significant cross-domain accuracy and deployment viability advantages, as shown in Tables II–VII along with Figs. 4 and 5 of this text. In Table II, the model surpasses earlier strategies with 93.8% Rank-1 accuracy and 79.7% mAP in the Market-1501 with ResNet-50 setup Sets and for OSNet with 96.2% Rank-1 accuracy and 74% mAP. These gains suggest that horizontal pyramid-based encoding and cluster-aware embedding may capture fine-grained person qualities that survive domain transformations better than Method [3] and Method [8]. Real-time surveillance networks require models that generalize without target-domain labels. CI-HPFE outperforms the baseline with a 92.3% Rank-1 accuracy in the DukeMTMC-reID with ResNet-50 scenario (Table III). Here, adaptive normalization and centroid-aligned feature refinement improve feature robustness across backdrops, camera angles, and occlusions. Such performance improvements could improve identification matching urban surveillance, transit hubs, and smart city deployments with environmental heterogeneity sets.

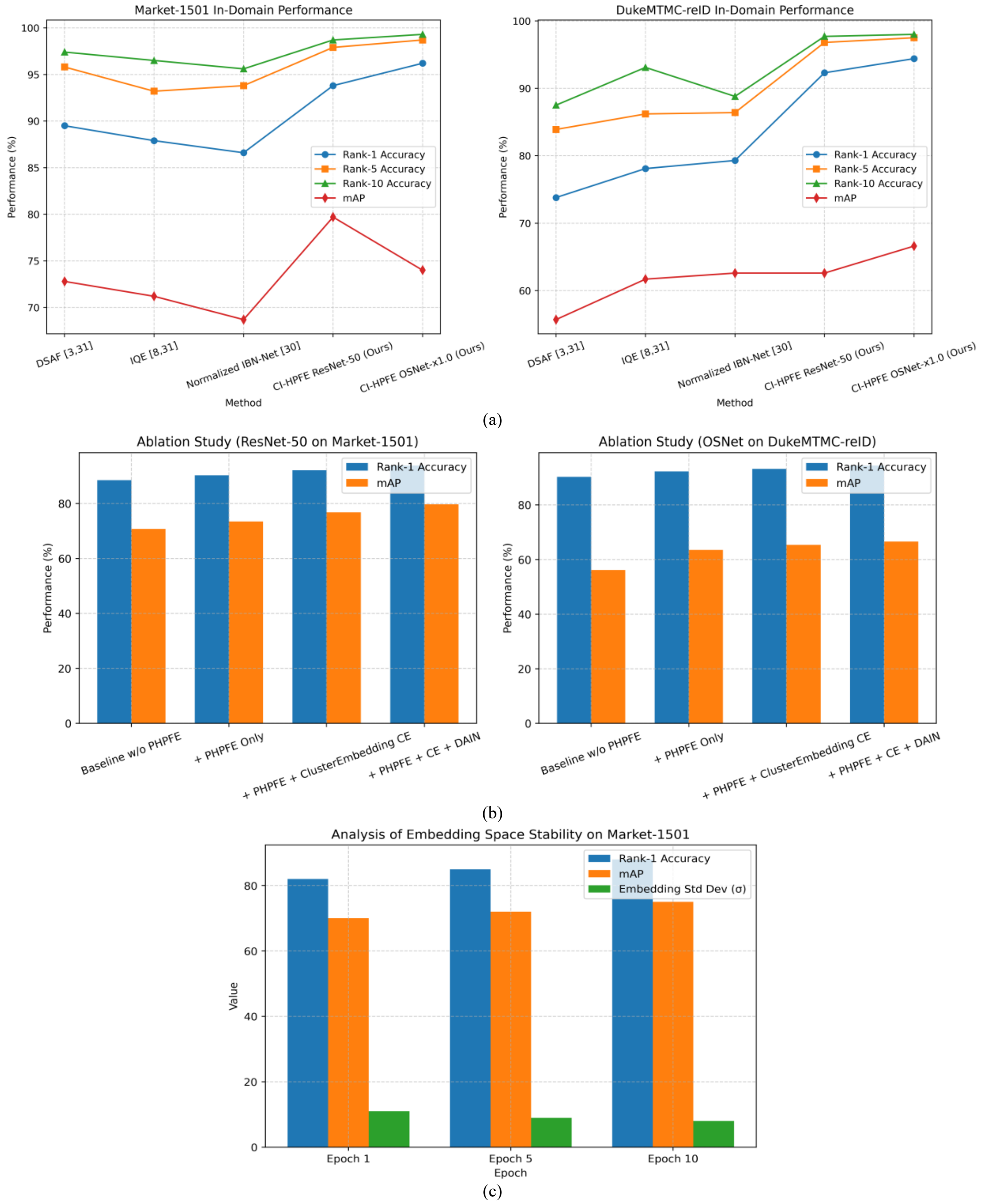
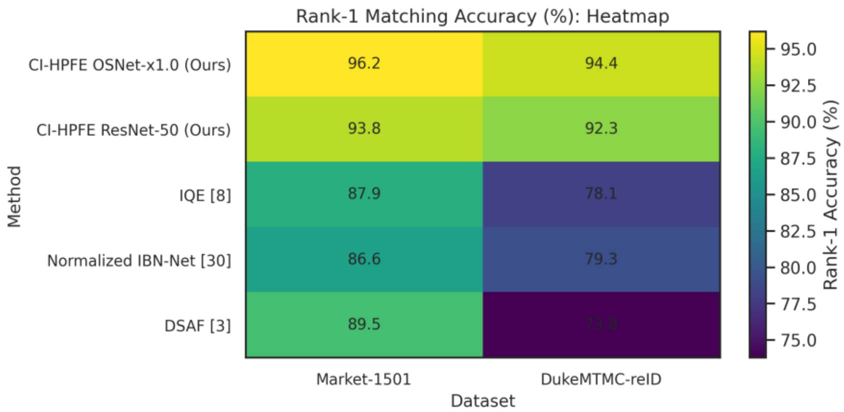
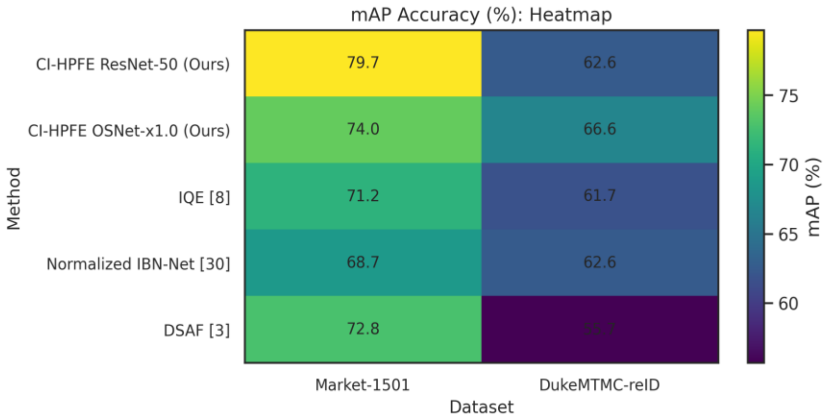


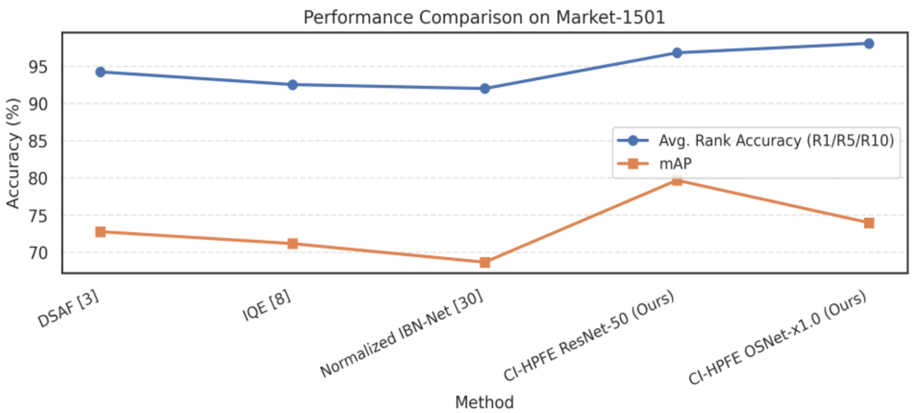
Fig.4. Integrated result analysis: (a) Accuracy and mAP comparison on Market-1501 and DukeMTMC-ReID, (b) Ablation study of module contributions on Rank-1 and Map, (c) Embedding stability across epochs on Market-1501.



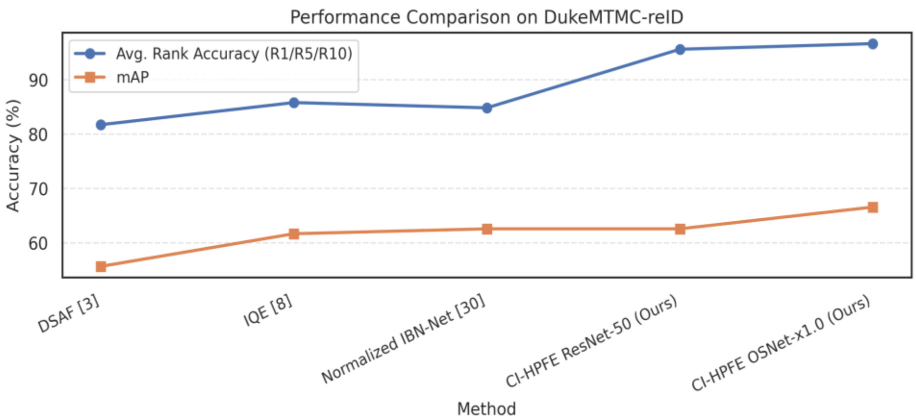
(a)



(b)



(c)



(d)

Fig. 5. Overall result analysis of the proposed model: (a) Rank-1 accuracy (%) heatmap showing CI-HPFE superiority on Market-1501 and DukeMTMC-ReID; (b) mAP (%) heatmap showing CI-HPFE performance across Market-1501 and DukeMTMC-ReID; (c) Comparison of accuracy and map on Market-1501; (d) Comparison of accuracy and map on DukeMTMC-ReID.

Fig. 6. Top-10 retrieval results on the Market-1501 dataset for a sample query under a camera-wise evaluation setting, obtained using the proposed CI-HPFE framework with a ResNet-50 backbone.

Ablation testing validates Tables V and VI modular construction processes. DAIN, cluster-aware embedding, and horizontal pyramid encoding improve model performance sets. Real-world systems with computational restrictions may require step-by-step component deployment, modularity is key. A resource-constrained system designer could construct only the PHPFE layer and add cluster feedback and DAIN as compute capacity allows balancing performance and hardware efficiency. Tables VII and VIII address edge deployment stability and synchronization, a critical operational issue in the process. Table VIII demonstrates that the CI-HPFE model has deployable inference speed and parameter count. Achieving up to 2.9% Rank-1 and 2.8% mAP benefits by aligning cluster centroids and stabilizing the embedding space, including FCSM consistently improves both datasets, as Table IX illustrates. Academic research, surveillance pipelines, and commercial ReID systems that demand real-time reaction and model compactness can use the proposed model's efficiency and accuracy. Results demonstrate model appropriateness for a real-world, cross-domain, unsupervised person identification process.

Fig. 7. Multi-query Top-5 retrieval outcomes using the suggested CI-HPFE (ResNet-50) model on the Market-1501 dataset.

C. Validated Hyperparameter Statistics Analysis

The Market-1501 and DukeMTMC-ReID datasets were used to evaluate the CI-HPFE model for unsupervised in-domain person reidentification under strict experimental conditions. All measurements, including Rank-1 accuracy and mAP, were averaged over five runs with randomized initialization for reproducibility and statistical robustness. Both Market-1501 and DukeMTMC-ReID scenarios were used to compute the expected mean and standard deviation of the major performance measures. For the Market-1501 with ResNet -50 setup, CI-HPFE achieved Rank-1 accuracy of 93.8% and a mAP of 79.7% and for Market-1501 with OSNet setup, Rank-1 accuracy of 96.2% and a mAP of 74%. The DukeMTMC-ReID model with ResNet-50 consistently met Rank-1 accuracy of 92.3% and mAP of 62.6%, and for OSNet, Rank-1 accuracy of 94.4& and mAP of 66.6%. The suggested CI-HPFE framework consistently outperformed current state-of-the-art techniques in rank-based matching accuracy and retrieval performance across the Market-1501 and DukeMTMC-ReID datasets. CI-HPFE significantly outperforms previous techniques on Market-1501, with OSNet-x1.0 achieving the highest Rank-1, Rank-5, and Rank-10 accuracies and ResNet-50 achieving the best mAP, indicating higher overall ranking consistency. CI-HPFE exhibits considerably greater absolute gains on DukeMTMC-ReID when compared to the strongest baseline, demonstrating its resilience in the face of extreme perspective and illumination changes. Specifically, OSNet-x1.0 outperforms ResNet-50 in terms of retrieval quality, whereas OSNet-x1.0 achieves greater rank-based accuracy and mAP. All things considered, these findings verify that CI-HPFE successfully combines the complementary advantages of OSNet and ResNet-50, exhibiting robust generalization across datasets and backbones. The approach significantly increases global retrieval quality with ResNet-50 and rank-based matching with OSNet, demonstrating its scalability and robustness for re-identification of individuals.

Methods were chosen as baselines because of their architectural diversity and methodological importance to the CI-HPFE design [3, 8, 30]. To yield reliable pseudo-labels to implement training and also enhance the domain generalization capability of the model, as required are the focus of Method [3]. Hao *et al*. [8] was chosen for its image quality enhancement algorithm for unsupervised person Re-ID (IQE). Finally, Bai *et al*. [30] introduce instance normalization and batch normalization to handle the feature maps at different scales. Thus, these three models feature space modeling, in-domain generalization, and learning paradigms from simple to advanced. Statistical metrics and baseline logic demonstrate the CI-HPFE model's superiority and dependability in unsupervised, in-domain person re-identification settings.

D. Qualitative Performance Analysis

The proposed CI-HPFE framework's qualitative retrieval performance on the Market-1501 dataset is shown in Figs. 6 and 7. The Top-10 retrieval outcomes for a typical query image using the ResNet-50 backbone are shown in Fig. 6. Despite differences in position, illumination, occlusion, and backdrop, the retrieved photos show that the suggested framework successfully recognizes the right person across several camera angles. The multi-query Top-5 retrieval results are displayed in Fig. 7, where the stability and consistency of the recovered matches are further enhanced by several query photographs of the same identity. The efficiency of the suggested CI-HPFE framework in acquiring robust and discriminative feature representations for person re-identification is visually demonstrated by these qualitative observations, which support the quantitative findings.

V. Conclusion

This paper presented a novel Cluster Integrated Horizontal Pyramid Feature Encoding (CI-HPFE) framework for unsupervised cross-domain person re-identification. The suggested system efficiently handles domain differences, appearance changes, and the lack of identification labels by combining horizontal pyramid feature encoding, cluster-aware embedding refinement, and Domain-Adaptive Instance Normalization (DAIN). Feature discrimination and domain generalization are improved by combining global and fine-grained local feature representations with adaptive normalization and cluster-guided pseudo-label refining. The efficacy of the suggested approach is demonstrated through extensive experimentation on the Market-1501 and DukeMTMC-reID benchmark datasets utilizing both ResNet-50 and OSNet-x1.0 backbones. The suggested method is appropriate for large-scale intelligent surveillance, public safety monitoring, and smart-city applications since the experimental results and ablation studies verify that each component helps learn robust, discriminative, and domain-invariant feature representations. There are still a number of obstacles in spite of these encouraging outcomes. The multi-level pyramid architecture may restrict deployment on ultra-low-power edge devices since it increases computational complexity and inference time. Additionally, DBSCAN's performance may deteriorate under extreme domain shifts or highly fragmented data distributions, and the clustering process is susceptible to noisy embeddings in the early phases of training. Additionally, the existing technology does not take into account dynamic views or mobile surveillance environments; it has only been tested on static camera networks.

The future study will concentrate on enhancing the scalability and adaptability of the suggested framework through the development of decentralized peer-to-peer federated synchronization strategies, the investigation of foundation-level vision transformers to improve global contextual representation, and the incorporation of gait cues and multi-frame tracklets for improved temporal consistency. Additionally, in order to improve robustness, communication efficiency, and secure deployment in real-world surveillance systems, the integration of generative diffusion models for synthetic data generation and privacy-preserving encrypted inference will be investigated.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Abhinav A. Parkhi was involved in the study's conception, methodological creation, software implementation, experimental analysis, and paper writing. Supervision, findings validation, and paper reading and editing fell under the purview of Atish A. Khobragade. The final draft of the manuscript has been read and approved by every author.

References

[1] K. N. Ghogale, R. S. Apare, and R. H. Borhade, "HyDeepIDF-Net: Hybrid deep ID FaceNet for person re-identification," *J. Ambient Intell. Humanized Comput.*, vol. 16, no. 6–7, pp. 727–760, 2025.

[2] Z. Ji, D. Cheng, and K. Feng, "Exploring stronger transformer representation learning for occluded person re-identification," *Multimedia Syst.*, vol. 31, no. 5, 394, 2025.

[3] L. Qi, J. Liu, L. Wang, Y. Shi, and X. Geng, "Unsupervised generalizable multi-source person re-identification: A domain-specific adaptive framework," *Pattern Recognit.*, vol. 140, 109546, 2023.

[4] R. K. Mishra, A. Mondal, and J. Mathew, "Nystromformer based cross-modality transformer for visible infrared person re-identification," *Sci. Rep.*, vol. 15, no. 1, 16224, 2025.

[5] G. Choe, H. So, C. Choe, and S. Kim, "Deep learning with full-triple relation for person re-identification," *Multimedia Tools Appl.*, vol. 84, no. 30, pp. 36799–36815, 2025.

[6] M. Zhou, G. Lian, X. Ouyang, J. Du, Q. Song, and J. Yang, "Multi-state perception consistency constraints network for person re-identification," *Pattern Anal. Appl.*, vol. 28, no. 1, 20, 2025.

[7] C. Zhang, Y. Su, N. Wang, Y. Lan, T. Wang, and A. Li, "Hard-sample guided cluster refinement for unsupervised person re-identification," *Signal Image Video Process.*, vol. 19, no. 1, 39, 2024.

[8] Z. Hao, H. Ge, and J. Huang, "Research on an unsupervised person re-identification based on image quality enhancement method," *Eng. Appl. Artif. Intell.*, vol. 123, 106392, 2023.

[9] A. Zheng, Z. Fei, Y. Ding, C. Li, and B. Luo, "RULER: Source-free domain adaptive person re-identification via uncertain label refinery," *Mach. Intell. Res.*, vol. 22, no. 5, pp. 900–916, 2025.

[10] Q. Wang, X. Qian, B. Li, and X. Xue, "Distribution aligned semantics adaption for lifelong person re-identification," *Mach. Learn.*, vol. 114, no. 3, 70, 2025.

[11] P. Yin, "Visual-textual adversarial learning for person re-identification," *Multimedia Syst.*, vol. 31, no. 1, 71, 2025.

[12] H. Sun, Y. Gao, and S. Ma, "Pseudo labels purification for unsupervised person re-identification," *Signal Image Video Process.*, vol. 19, no. 1, 24, 2024.

[13] Y. Li, S. Shuai, Y. Zhou, B. Deng, and D. Zhang, "Joint detection and re-identification for occluded person search," *Sci. Rep.*, vol. 15, no. 1, 22470, 2025.

[14] G. Zhang, S. Tan, and Y. Fang, "Foreground-aware transformer network for person re-identification," *J. Supercomput.*, vol. 81, no. 4, 524, 2025.

[15] J. Leng, C. Kuang, S. Li, J. Gan, H. Chen, and X. Gao, "Dual-space video person re-identification," *Int. J. Comput. Vis.*, vol. 133, no. 6, pp. 3667–3688, 2025.

[16] R. Rehman, J. H. Shah, F. A. Khokhar, A. Ahmed, R. Saleem, and S. S. Najam, "Smart monitoring: Employing person re-identification to uncover suspicious behavior," *Signal Image Video Process.*, vol. 19, no. 1, 9, 2024.

[17] R. K. Mishra, A. Mondal, and J. Mathew, "Swin transformer with part-level tokenization for occluded person re-identification," *Mach. Vis. Appl.*, vol. 36, no. 1, 15, 2024.

[18] M. Cascio, L. Cinque, D. Distante, G. L. Foresti, and A. Fagioli, "A benchmark dataset for radio signal image-based person re-identification," *Sci. Data*, vol. 12, no. 1, 1522, 2025.

[19] H. Li, T. Zhang, S. Li, X. Li, and X. Zhao, "Representation strategy for unsupervised domain adaptation on person re-identification," *Optoelectron. Lett.*, vol. 20, no. 12, pp. 749–756, 2024.

[20] A. Asperti, S. Fiorilla, S. Nardi, and L. Orsini, "A review of recent techniques for person re-identification," *Mach. Vis. Appl.*, vol. 36, no. 1, 25, 2024.

[21] C. Yuan, Z. Guan, Y. Xu, X. Chen, X. Zhu, and W. Liang, "Jointly adaptive cross-resolution person re-identification on super-resolution," *Complex Intell. Syst.*, vol. 11, no. 6, 249, 2025.

[22] F. Bouhlel, H. Mliki, and M. Hammami, "Person re-identification from UAVs based on deep hybrid features: Application for intelligent video surveillance," *Signal Image Video Process.*, vol. 18, no. 11, pp. 8313–8326, 2024.

[23] Z. Wu, Z. Hu, and J. Ding, "Same-clothes person re-identification with dual-stream network," *Multimedia Syst.*, vol. 30, no. 2, 70, 2024.

[24] R. Kasantikul, W. Kusakunniran, Q. Wu, and Z. Wang, "Channel-shuffled transformers for cross-modality person re-identification in video," *Sci. Rep.*, vol. 15, no. 1, 15009, 2025.

[25] M. F. Ahmed and A. A. N. Oyshee, "Multi-task model with attribute-specific heads for person re-identification," *Pattern Anal. Appl.*, vol. 28, no. 1, 38, 2025.

[26] V. D. Nguyen, P. Mantini, and S. K. Shah, "Occlusion-aware appearance and shape learning for occluded cloth-changing person re-identification," *Pattern Anal. Appl.*, vol. 28, no. 2, 84, 2025.

[27] C. Zhang, X. Kai, X. Bai, and H. Hu, "UPCA: Unsupervised person re-identification with camera-aware distance and lightweight attention," *Signal Image Video Process.*, vol. 19, no. 15, pp. 1–11, 2025.

[28] J. Gao, X. Jiang, S. Dou, D. Li, D. Miao, and C. Zhao, "ReID-leak: Membership inference attacks against person re-identification," *Int. J. Comput. Vis.*, vol. 132, no. 10, pp. 4673–4687, 2024.

[29] Y. Yan, J. Li, S. Liao, and J. Qin, "TAL: Two-stream adaptive learning for generalizable person re-identification," *Mach. Intell. Res.*, vol. 22, no. 2, pp. 337–351, 2025.

[30] X. Bai, A. Wang, C. Zhang, and H. Hu, "Cross-domain person re-identification based on normalized IBN-Net," *IEEE Access*, vol. 12, pp. 54220–54228, 2024.

[31] J. Luo and L. Liu, "Improving unsupervised pedestrian re-identification with enhanced feature representation and robust clustering," *IET Comput. Vis.*, vol. 18, pp. 1097–1111, 2024.

[32] Y. Apedo, H. Tao, W. Gao, C. Xie, and S. Zhao, "Unsupervised domain adaptation for crack segmentation via cross-domain stylization and dual adversarial feature learning," *J. Comput. Civ. Eng.*, vol. 40, no. 2, 04025146, 2026.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).