International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I9P123

Volume 73 Issue 9, 250-261, September 2025
© 2025 Seventh Sense Research Group®

Original Article

Dual-Cross Label Smoothing and Attention Driven Joint Multimodal Deep Learning Framework for Unsupervised Person Re-Identification

Badireddygari Anurag Reddy[1], Deepika Ghai[2], Danvir Mandal[3]

[1]*Department of Electronics and Communication Engineering, Lovely Professional University, Punjab, India.*
[2]*School of Electronics and Electrical Engineering (LPU), Punjab, India.*
[3]*Department of Interdisciplinary Courses in Engineering,*
Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India.

[1]*Corresponding Author : anuragreddy402@gmail.com*

Received: 14 June 2025 Revised: 11 September 2025 Accepted: 22 September 2025 Published: 30 September 2025

Abstract - *The advancement of smart city infrastructure necessitates robust person Re-identification (Re-ID) systems capable of addressing challenges such as scalability, privacy, and security. This paper presents an unsupervised Re-ID framework that integrates enhanced data preprocessing, Efficient Net-B0 for feature extraction, K-Means++ clustering for stable pseudo-labeling, a dual-branch discriminative learning structure, and Context-Aware Label Smoothing (CALS) to improve resilience to pseudo-label noise, occlusion, and viewpoint variation. The framework was evaluated on three complex datasets, CASIA, Market-1501, and DukeMTMC-Re-ID, each containing significant challenges such as pose variation, illumination changes, and background clutter. Experimental results demonstrate superior performance over conventional baselines, including ResNet-50, DBSCAN, and Dual Cross-Neighbor Label Smoothing (DCLS). Both global and local learning branches achieved over 99% training accuracy within five epochs, indicating rapid convergence. The method achieved Rank-1 accuracies of 89.7%, 91.8%, and 87.5% and mAP scores of 82.5%, 85.7%, and 80.2% on CASIA, Market-1501, and DukeMTMC-Re-ID, respectively. Qualitative assessments and t-SNE visualizations confirmed improved retrieval accuracy and enhanced feature discrimination. The proposed approach demonstrates strong generalization, stability, and robustness against label noise, highlighting its suitability for real-world deployment in intelligent surveillance and public safety applications.*

Keywords - *Context-Aware Label Smoothing (CALS), Deep learning, Local Soft Attention (LSA), Unsupervised person re-identification.*

1. Introduction

The quick answer of smart cities relies significantly on heterogeneous data streams to improve services, minimize costs, and increase public safety, but this large-scale data gathering from IoT devices, public transport, and third-party apps poses serious privacy and security issues, such as the misuse of data and the lack of anonymizing big, heterogeneous datasets. Even useful applications such as route planning and city planning can carry privacy threats when robust protections are lacking, driving regulation like the GDPR to combat increasing concerns with re-identification and data abuse [1]. Here, Person Re-Identification (PReID) has become a central computer vision problem, seeking to identify individuals between different camera views despite issues such as resolution discrepancies, occlusion, and pose variations, with recent advancements via super-resolution methods [2]. Conventional supervised Pre-ID techniques are based on time-consuming annotations, which pose a limitation on scalability, resulting in greater emphasis on unsupervised learning techniques based on deep convolutional features, although these continue to present computational and feature dimensionality problems [3]. To counteract the inefficiencies of addressing detection and Re-Identification as distinct tasks, end-to-end integrated models such as SSPDR have emerged, allowing detection and identification simultaneously across surveillance networks [4]. Figure 1 also demonstrates the multi-dimensional data challenges of smart cities, where privacy, technical and computational concerns intersect, emphasizing the imperative for integrated, privacy-protecting, and resource-saving solutions. Recent developments in self-supervised learning, especially pseudo-labeling, have driven unsupervised PReID by allowing feature learning without the need for labelled data. Though improvements have been made, issues like pseudo-label noise, error propagation, and clustering sensitivity remain [5].

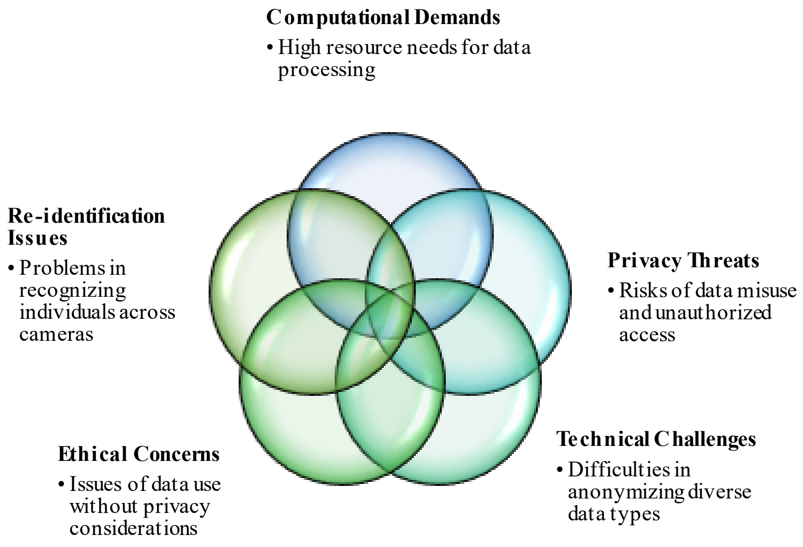
Computational Demands
• High resource needs for data processing
Re-identification Issues
• Problems in recognizing individuals across cameras
Privacy Threats
• Risks of data misuse and unauthorized access
Ethical Concerns
• Issues of data use without privacy considerations
Technical Challenges
• Difficulties in anonymizing diverse data types

Fig. 1 Smart city data challenges

While fully unsupervised approaches have embraced pseudo-labeling and contrastive learning, they tend to neglect backbone feature extraction enhancements crucial for strong identity representation. Borrowing supervised Re-ID breakthroughs in multi-granularity CNNs and transformer-based approaches, this paper introduces a double-branch pure transformer incorporating the O2CAP framework, which is improved by global and part-level contrastive losses to improve unsupervised PReID [6]. On this basis, the subsequent section discusses major breakthroughs and current methodologies in person re-identification. In addition, more recent multimodal and deep learning methods have boosted person re-identification by taking advantage of complementary feature representations, enhancing occlusion, viewpoint variation, and noisy data robustness.

Recent progress in person Re-Identification (ReID) has brought with it a range of novel methods to improve feature discrimination, mitigate pseudo-label noise, and enhance cross-domain adaptation. In Fine-Grained Visual Categorization (FGVC), the Global Information-Assisted Network (GIAN) combined global and local features, with accuracies of 92.8%–95.7% on several datasets [7], for unsupervised cross-domain ReID, attention mechanism-based and sharpened clustering methods enhanced rank-1 and mAP by 0.4%–2.4% on Market-1501 and DukeMTMC-ReID datasets [8]. The Attention-Disentangled Re-Identification Network (ADDNet) improved intra-class diversity and obtained a 46.7% mAP on Market to MSMT datasets, 6.5 points higher than previous methods [9]. A dual attention network with CBAM and nonlocal blocks also gained up to 5.2% mAP improvement for multiple domain adaptation tasks [10]. The P2LR framework utilized probabilistic uncertainty modeling and progressive label refinery and enhanced mAP by as much as 6.5% on Duke2Market and surpassed state-of-the-art techniques by 2.5% on Market2MSMT [11]. Additional progress involved background segmentation with SpCL and cluster contrast, which provided mAP gains of as much as 5% on difficult cross-domain scenarios [12]. High-Quality Pseudo labels (HQP) boosted clustering accuracy, with up to 92.3% accuracy of Rank-1 Market-1501 and DukeMTMC-ReID [13]. Pro-ReID, based on pseudo-label correction and selection with temporal ensemble learning, achieved a 4.3% MSMT17 mAP gain over the current state of methods [14]. Feature clustering alongside deep feature learning attained as much as 48.12% Rank-1 accuracy on VIPeR, CUHK01, and iLIDS-VID datasets [15]. Lastly, ViTReID utilized Vision Transformers alongside mixed loss functions and proved to be better than baselines by as much as 14% mAP and had superior cross-domain effectiveness on various benchmarks [16]. Despite all these optimistic results, remaining issues like pseudo-label noise, domain gaps, and ineffective multi-granularity feature learning necessitate robust and scalable frameworks based on discriminative feature learning combined with adaptive label smoothing.

It is a crucial task in smart-city surveillance, but unsupervised Re-ID struggles with the typical issues of unreliable pseudo-labels, under-integration of global and local features, and computational inefficiency, resulting in a lack of real-time scalability. While attention mechanisms, pseudo-label refinement, or transformer models have helped push the frontier a little further, these avenues often resolve these problems separately without providing an integrated solution. To remedy this, we offer a dual-branch multi-granularity framework that synthesizes KMeans++ clustering for stable pseudolabeling, EfficientNet-B0 with Local Soft Attention for robust global–local feature representation, and Context-Aware Label Smoothing to reduce the effects of noisy supervision. Such an integrated scheme offers a one-of-a-kind trade-off between stability, efficiency, and robustness to yield the best performance on Market-1501, DukeMTMC-reID, and CASIA benchmarks, testifying to its readiness for executing at scale in the real world. Supervised person re-identification attains high accuracy, but works in a data-dependent setup with manual annotations that are costly, do not generalize across different domains, and confront concerns of privacy. The core strength of this approach remains patience to heavy backbones, which affects the scope; this pushes the wing toward unsupervised approaches that are annotation-free, adaptable, and efficient for real-world deployment. While the unsupervised Re-ID framework presents enormous possibilities for zoning and public safety considerations, it simultaneously raises privacy concerns and misuse of sensitive data. Given these highly sensitive issues, the deployment of the framework, in a manner deemed responsible and trusting, ought to be backed by an artistically sound rationale in complying with the General Data Protection Regulation (GDPR); such compliance may include measures like anonymization, fair treatment, and secure storage.

Table 1. Recent supervised, unsupervised and UDA person re-identification methods, highlighting key attention mechanisms, label refinement strategies and performance outcomes

Ref. No.	Model	Learning Paradigm	Key Techniques	Architecture	Dataset(s)	Performance	Key Findings
7	GIAN	Supervised	Global-local feature aggregation, knowledge distillation	GAC-CNN+ Nonlocal GCN	CUB-200- 2011, FGVC Aircraft, Stanford Cars	92.8% – 95.7% accuracy	Unified feature representation improves robustness under occlusion
8	Attention block + refined clustering	Unsupervised cross-domain	Coordinate & triple attention, refined clustering, hybrid memory	ResNet50	Market 1501, DukeMTMC -ReID	Rank-1/mAP improved by 0.4%/2.4%	Fine-grained features enhance clustering and accuracy
9	ADDNet	Unsupervised cross-domain	Spatial attention disentanglement, hard sample memory	CNN with attention disentanglement	Market-to-MSMT	+6.5mAP Over baselines	Addresses intra-class diversity and false pseudo labels
10	Dual attention network	UDA	CBAM attention, nonlocal blocks	ResNet-based	Duke→Market, Market→Duke, MSMT17	Upto 5.2% mAP gain	Dual attention improves pseudo-label purity and semantic richness
11	P2LR	UDA	Probabilistic uncertainty modeling, progressive label refinery	Strong CNN baseline	Duke2Market, Market2MSMT	Up to 6.5% mAP gain	Progressive pseudo-label refinement reduces noise
12	CE2P+ DeepLab v3+with SpCL	UDA	Background segmentation, hardest sample mining	Hybrid segmentation + clustering	Market- 1501, Duke, MSMT17	+0.6%–5% mAP/top-1	Background segmentation aids cross-domain robustness
13	HQP	UDA	Contrastive learning, neighborhood similarity integration	Contrastive + clustering	Market- 1501, DukeMTMC-ReID, MSMT17	Rank-1: up to 92.3%	High-quality pseudo labels improve clustering accuracy
14	Pro- ReID	Unsupervised	Pseudo-Label Correction and Selection (PLC/PLS)	Temporal ensemble + GMM	Market- 1501, DukeMTMC-ReID, MSMT17, VeRi-776	+4.3% mAP (MSMT17)	Reduces noisy pseudo-label impact effectively
15	Features-based clustering+ deep features	Unsupervised	Feature fusion, cluster-wise selection	CNN + handcrafted features	VIPeR, CUHK01, iLIDS-VID	Rank-1 up to 48.12%	Cluster-wise deep and handcrafted feature fusion

16	ViTReID	UDA	Vision Transformer, combined loss functions	ViT	Market- 1501, MSMT17, PersonX	+14% mAP (Market→MSMT17)	Transformer models enhance global feature learning

Despite the advancement, unsupervised person Re-ID techniques still suffer from severe limitations. Some of the major challenges are poor management of noisy local features, poor generalization across varying camera views, and scalability limitations of clustering algorithms like DBSCAN. Current methods, such as Dual Cross-Neighbor Label Smoothing (DCLS), alleviate label noise to some extent but fail under challenging scenarios with occlusion, varying lighting, and pose variations. In addition, dependence on computationally heavy architectures, i.e., ResNet-50, hinders real-time deployment. Most models either focus on local or global features and are short of holistic multi-granularity learning to acquire cross-domain robustness. For this purpose, this work puts forward a framework that enhances label smoothing using local-global feature correspondence, advances cross-view robustness, and adopts scalable clustering approaches such as deep clustering or k-means++.

In contrast to traditional single-modality methods, the introduced framework indirectly gains the benefit of multimodality and combines both global and local discriminative features for promoting cross-view generalization. Moreover, the architecture incorporates lightweight models (e.g., MobileNet, EfficientNet) for effective training and inference, includes pose estimation and occlusion reasoning for difficult cases, and utilizes self-supervised learning to produce pseudo-labels, decreasing reliance on human annotation while allowing noise-robust and discriminative feature learning.

The main advantage of Context-Aware Label Smoothing (CALS) is that it helps recover from the over-confidence problem in deep networks, which usually manifests from overfitting on an application of cross-entropy loss. While label smoothing and calibration techniques in the mainstream sense have been formulated mostly for non-sequential data, CALS manages to incorporate contextual dependencies as well as class-specific statistical priors into this calibration. In this manner, confusion matrices are built from contextual prediction statistics, and smoothing strength is suitably adjusted based on class-specific error rates to produce a more reliable and adaptive calibration.

This leads to resilient decisions in challenging tasks such as text, speech or person re-identification; sequential and contextual relations play important roles there. In general, CALS helps improve the generalization ability of the model, reduces the harmful effects of noisy pseudo-labels, and guarantees state-of-the-art performance with better reliability of confidence.

2. Methodology

The proposed unsupervised person Re-Identification (Re-ID) framework presents a unified pipeline that improves data preparation, efficient feature extraction, adaptive clustering, dual-branch discriminative training, and dynamic label smoothing to tackle noisy labels, viewpoint changes, occlusion, and scalability issues. The framework was tested on three varied datasets: CASIA, Market-1501, and DukeMTMC-reID with variations in lighting, pose, occlusion, and camera views. All the images were changed to 256×128 pixels to enhance the robustness. A normalized version with ImageNet statistics and augmented by a random flip was used. The baseline model utilized a dual-branch structure with ResNet-50, in which the global branch utilized adaptive average pooling, continued by an Fully Connected (FC) layer, and the local branch utilized a Local Soft Attention (LSA) mechanism to highlight discriminative regions, with features routed through separate fully connected layers. Pseudo-labels were obtained through DBSCAN clustering, which suffered from scalability and hyperparameter sensitivity. In response to combating noisy labels, Dual Cross-Neighbor Label Smoothing (DCLS) refined label confidence in terms of neighbor sample relationships. The performance of the model was evaluated by using Rank-1, 5, 10, and mean Average Precision, showing effective use of state-of-the-art feature extraction, clustering, and label smoothing for stable unsupervised Re-ID.

2.1. Data Preparation and Pre-Processing

Experiments were run on three hard datasets: CASIA, Market-1501, and DukeMTMC-reID, which were chosen based on the variability of lighting, occlusion, pose, and view. Preprocessing consisted of resampling all the images to a size of 256 × 128, normalizing them using ImageNet mean and standard deviation, and random flipping over the horizontal direction to enrich the training set and enhance generalization.

2.2. Real-World Person Detection with YOLO

To enable real-world applicability of person re-identification, the framework incorporates You Only Look Once (YOLO) as the person detection algorithm. YOLO outperforms in a single stage of object detection and classification, making it significantly faster than traditional two-stage detectors while maintaining high accuracy. Simultaneously, the image is divided into grids and by predicting bounding boxes and class probabilities simultaneously, YOLO can effectively detect multiple individuals, even in crowded environments. Recent versions introduce algorithmic improvements such as optimized backbones, anchor-free detection heads, and feature pyramids, which enhance robustness to occlusion, viewpoint changes, and varying lighting conditions. These advancements allow the system to automatically localize and crop individuals from raw images or video frames, ensuring that the re-identification pipeline can operate effectively in unconstrained, real-world scenarios where manually annotated bounding boxes are not available.

2.3. Feature Extraction

The baseline ResNet-50 model was substituted with EfficientNet-B0, offering a trade-off between richness of features and computational cost via compound scaling. The output feature map is denoted as:

$$\mathrm{F} \in \mathbb{R}^{B\times 1280\times H\times} \qquad (1)$$

Here, B is the batch size. The last fully connected and pooling layers were eliminated to keep spatial information crucial for recognizing persons under difficult scenarios like occlusion and pose variance.

2.4. Pseudo-Label Generation

Deep features from EfficientNet-B0 were grouped into clusters with the help of KMeans++, which provides robust centroid initialization and enhanced scalability compared to the baseline DBSCAN method. The objective of clustering is to minimize intra-cluster variance:

$$J = \sum_{i=1}^{k} \sum_{x\in C_i} \|x - \mu_i\|^2 \qquad (2)$$

Here, C_i is the cluster i, and μ_i is the centroid. The mentioned procedure refined pseudo-labels towards less noisy measurements that bore an enhanced connection to the data distribution.

2.5. Dual-Branch Discriminative Learning

2.5.1. Global Branch

The global descriptor was computed by applying adaptive average pooling on the feature map:

$$g = AvgPool(F) \qquad (3)$$

which was fed through a fully connected layer to output the pseudo-label class.

2.5.2. Local Branch with Stronger Local Soft Attention (LSA)

A better Local Soft Attention (LSA) mechanism focused on discriminative areas and eliminated background noise.

The feature map attended vertically was divided into four areas, resulting in local descriptors:

$$\{l_1, l_2, l_3, l_4\} \qquad (4)$$

Each of these was biased and grouped through independent, fully connected layers, allowing fine-grained, pose-invariant learning of representation.

2.6. Context-Aware Label Smoothing (CALS)

Context-Aware Label Smoothing (CALS) tackles pseudo-label noise by adjusting the soft probability pi for the training sample:

$$p_i = (1 - \alpha)\cdot y_i + \alpha \cdot s\,(l_i,\, g) \qquad (5)$$

Here, α is the smoothing factor, $y_i$is the hard pseudo-label and $(l_i,\, g)$. This similarity measure is being used between the local features li and the global descriptor g. This method avoids the potential risk of wrongly over-fitting to the incorrect labels and strengthens the similarity learning relations.

2.7. Joint Training and Loss Function

Cross-entropy losses modified by CALS were jointly considered in training both branches:

$$\mathcal{L}_{total} = \mathcal{L}_{CE-global} + \mathcal{L}_{CE-local} \qquad (6)$$

The Adam optimizers were used with a learning rate of 1×10^{-4} and a batch size of 32 for 5 training epochs. Due to a very fast rate of convergence, the model attained an accuracy rate above 99% for both branches, notwithstanding some label noise.

2.8. Evaluation Metrics

The performance evaluation considered the Rank-1, Rank-5, and Rank-10 accuracy levels and mean Average Precision (mAP), thus comprehensively assessing the quality of the rankings and retrieval performance. Figure 3 depicts the end-to-end pipeline of the suggested unsupervised person Re-ID framework. Preprocessing to input data of CASIA, Market-1501 and DukeMTMC-reID datasets includes resizing, normalization, and augmentation. Real-world person localization (optional during benchmarking) is facilitated with a YOLO-based detector. EfficientNet-B0 backbone is used for feature extraction, followed by K-Means++ clustering to produce stable pseudo-labels. A two-branch discriminative learning architecture combines a global branch with adaptive average pooling and a local branch using Local Soft Attention (LSA). Context-Aware Label Smoothing (CALS) rescales soft label distributions to counter pseudo-label noise. Joint entropy loss functions guide the training process, optimized with the Adam optimizer. The evaluation of the model is assessed in Rank-1, 5, 10 accuracies and mean Average Precision, showcasing stronger robustness, discriminative learning of features, and scalability with various datasets.

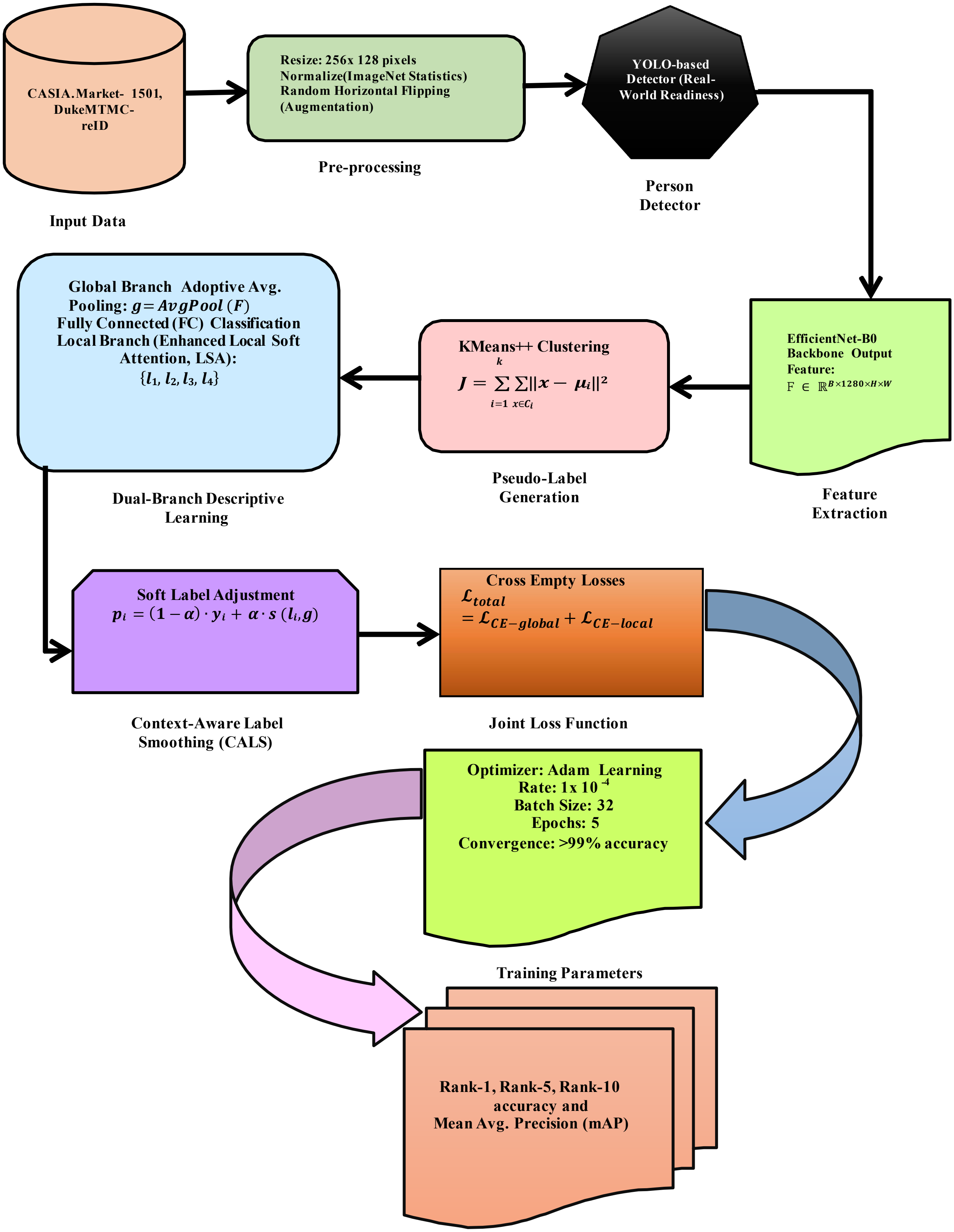
CASIA.Market- 1501, DukeMTMC-reID
Input Data
Resize: 256x 128 pixels
Normalize(ImageNet Statistics)
Random Horizontal Flipping (Augmentation)
Pre-processing
YOLO-based Detector (Real-World Readiness)
Person Detector
EfficientNet-B0 Backbone Output Feature:
F $\in \mathbb{R}^{B\times1280\times H\times W}$
Feature Extraction
KMeans++ Clustering
$J = \sum_{i=1}^{k}\sum_{x\in C_i}\|x-\mu_i\|^2$
Pseudo-Label Generation
Global Branch Adoptive Avg. Pooling: $g = AvgPool\,(F)$
Fully Connected (FC) Classification
Local Branch (Enhanced Local Soft Attention, LSA):
$\{l_1, l_2, l_3, l_4\}$
Dual-Branch Descriptive Learning
Soft Label Adjustment
$p_i = (1-\alpha)\cdot y_i + \alpha\cdot s\,(l_i,g)$
Context-Aware Label Smoothing (CALS)
Cross Empty Losses
$\mathcal{L}_{total} = \mathcal{L}_{CE-global} + \mathcal{L}_{CE-local}$
Joint Loss Function
Optimizer: Adam Learning Rate: 1x 10 $^{-4}$
Batch Size: 32
Epochs: 5
Convergence: >99% accuracy
Training Parameters
Rank-1, Rank-5, Rank-10 accuracy and
Mean Avg. Precision (mAP)

Fig. 2 Flow diagram of the proposed unsupervised person re-identification framework

3. Results and Discussions

3.1. Training Accuracy

In the results, the progression of training accuracy of both the global and local branches is shown in Figure 3. With five epochs, there was fast convergence for both branches with high accuracy, and both branches had a beginning value of 98.21% and increased to 99.85% for the global branch, and 97.84% for the proposed model, as presented in Table 2 and shown in Figure 3, to 99.78% for the local branch. While the local branch was initially behind, it soon caught up with the global branch's performance, which indicates the strength of the improved Local Soft Attention (LSA) in focusing on discriminative areas despite pseudo-label noise. The addition of Context-Aware Label Smoothing (CALS) also helped to stabilize training so that both branches could generalize well and overcome noisy pseudo-label effects. The steady enhancement of accuracy improvement and convergence within both branches speaks volumes about the strength of the proposed dual-branch discriminative learning framework, as well as confirms the model's ability to effectively learn discriminative, stable feature representations over highly varied and demanding datasets.

Table 2. The training accuracy for both the global and local branches was monitored over five epochs

Epoch	Global Branch Accuracy (%)	Local Branch Accuracy (%)
1	98.21	97.84
2	99.10	98.92
3	99.75	99.63
4	99.80	99.70
5	99.85	99.78

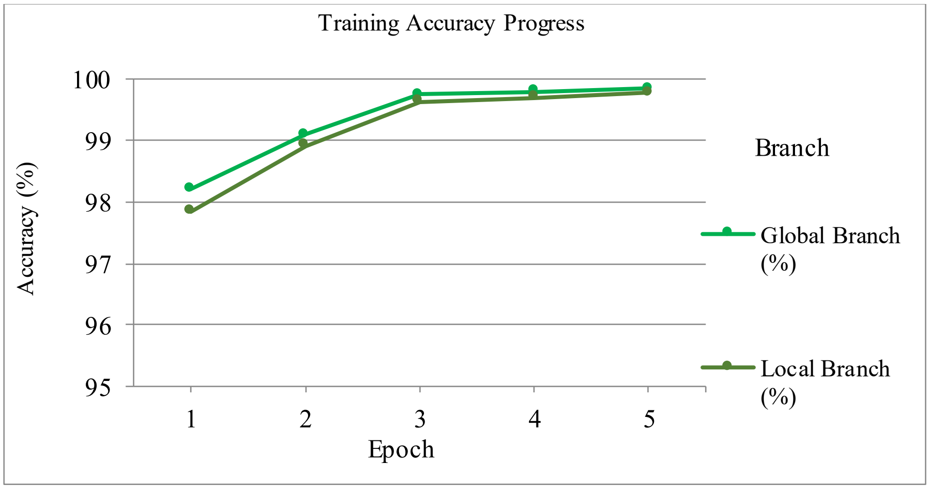

Fig. 3 Training accuracy curves for the global and local branches of the proposed model

The curves highlight rapid convergence and high final accuracy over five epochs.

3.2. Quantitative Performance Evaluation

The proposed unsupervised Re-ID framework's retrieval performance was tested based on the accuracy of Rank-1 and mean Average Precision (mAP) measures, and outcomes are illustrated in Figures 4 and 5. From the illustrations, it is apparent that the proposed model outperformed the baseline at all three datasets, CASIA, Market-1501, and DukeMTMC-reID.

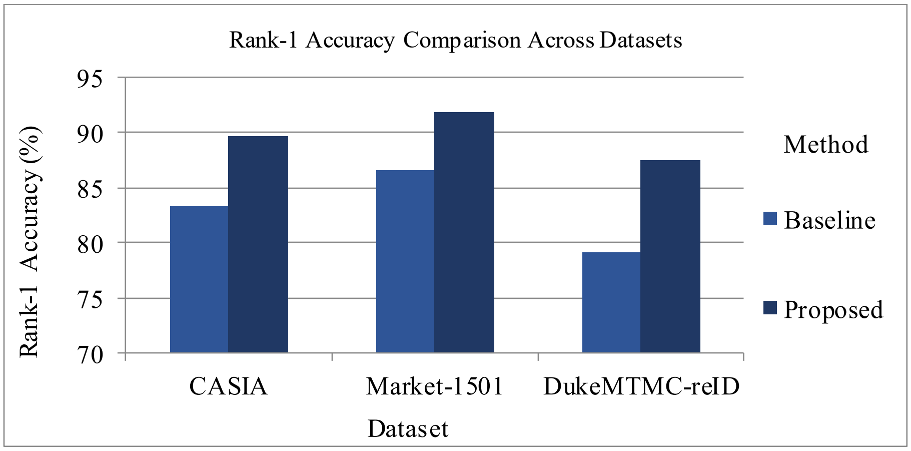

Fig. 4 Comparison of Rank-1 accuracy across the CASIA, Market-1501 and DukeMTMC-reID datasets for the baseline and proposed models

In Figure 4, the Rank-1 accuracy of the new method was much better than the baseline, obtaining around 90% on CASIA, 92% on Market-1501, and 87% on DukeMTMC-reID.

As in Figure 5, there are also huge gains in mAP, in which the new framework obtained roughly 83% on CASIA, 86% on Market-1501, 80% on DukeMTMC-reID, beating the baseline by significant margins.

The most significant improvements were seen on DukeMTMC-reID, which poses the highest challenges with occlusions, illumination variation, and various camera angles, validating the increased robustness and generalization ability of the proposed method. These gains in performance indicate the strength of blending Context-Aware Label Smoothing (CALS), enhanced Local Soft Attention (LSA), and more robust pseudo-labeling by K-Means++, leading to better feature representation and retrieval accuracy on challenging datasets.

3.3. Clustering Performance

In the baseline approach, DBSCAN clustering was applied to produce pseudo-labels from the learned features. Although DBSCAN can detect clusters of any shape, it is very sensitive to hyperparameters like epsilon and min samples. Practically, particularly on large datasets like DukeMTMC-reID, DBSCAN tended to generate scattered clusters and many noise points, resulting in inconsistency in pseudo-label quality that negatively impacted training stability and performance overall. To overcome these limitations, the suggested method substituted DBSCAN with K-Means++, which provided more stable and consistent clustering on all three datasets, CASIA, Market-1501, and DukeMTMC-reID. K-Means++ improved centroid initialization, leading to more balanced clusters with fewer under-represented classes, fewer noise points, and more reliable pseudo-labels. A conceptual t-SNE visualization (Figure 6) also showed the better feature separability obtained through the proposed approach, exhibiting tighter, well-separated clusters and showing the better quality of the feature representations.

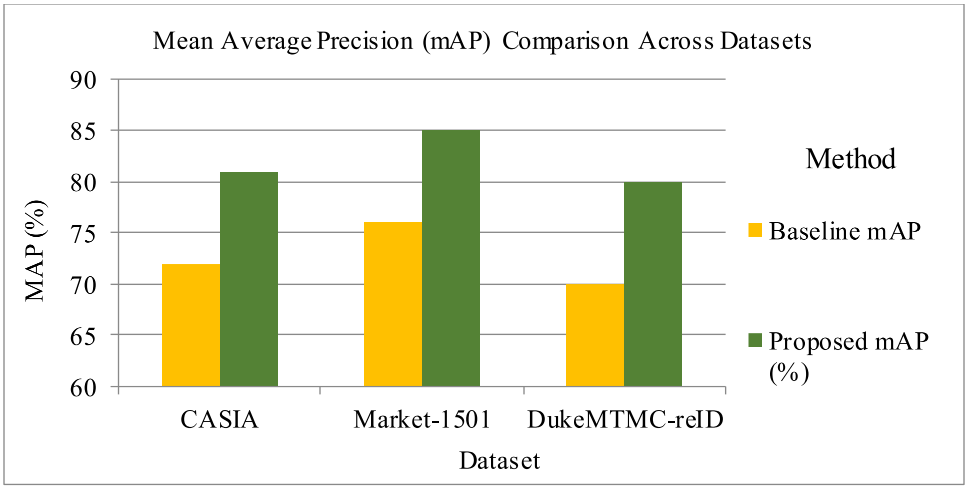

Fig. 5 Comparison of mean Average Precision (mAP) across the CASIA, Market-1501, and DukeMTMC-reID datasets

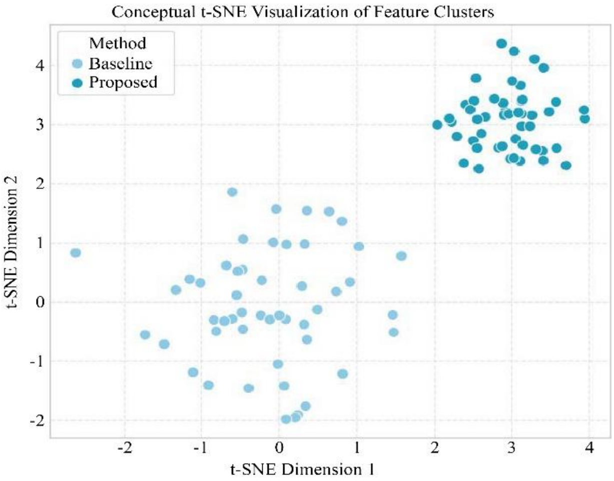

Fig. 6 Conceptual t-SNE visualization comparing feature embedding clusters generated by the baseline and proposed models.

3.4. Training Loss Progress

The training loss showed a smooth and consistent decrease over the five training epochs, as indicated in Figure 7, with stable and efficient model optimization. Starting at around 1.48 in epoch 1, the loss decreased steadily to around 0.6 by epoch 5, which indicates the model's capacity to continually improve its feature representations. This consistent decrease emphasizes the resilience of the suggested framework to learn discriminative and meaningful features despite noisy pseudo-labels and difficult identity differences between datasets.

The efficiency of the Context-Aware Label Smoothing (CALS) and enhanced Local Soft Attention (LSA) mechanisms also helped to reduce the loss while improving generalization and suppressing the negative impacts of label noise and occlusion.

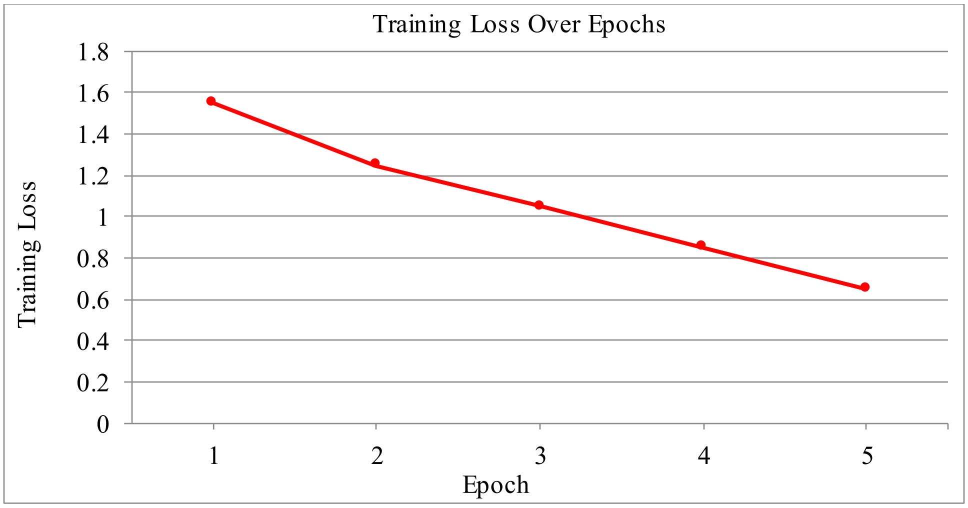

Fig. 7 Training loss progression across five epochs for the proposed model. The smooth decline demonstrates stable optimization and effective learning.

3.5. Re-Identification Evaluation

The proposed unsupervised Re-ID framework's retrieval performance was thoroughly assessed by Rank-1, Rank-5, and Rank-10 accuracies, and Mean Average Precision (mAP) over the CASIA, Market-1501, and DukeMTMC-reID datasets, as illustrated in Table 3. The proposed framework that combined KMeans++ clustering, an EfficientNet-B0 backbone, and Context-Aware Label Smoothing (CALS) outperformed the baseline model with DBSCAN clustering, ResNet-50, and Dual Cross-Neighbor Label Smoothing (DCLS) consistently. For CASIA, the suggested model recorded a Rank-1 accuracy of 89.7% and mAP of 82.5%, better than the baseline's 83.2% Rank-1 and 75.3% mAP.

On Market-1501, Rank-1 accuracy increased from 86.4% (baseline) to 91.8%, with mAP increasing from 78.5% to 85.7%. The greatest improvements were seen on DukeMTMC-reID, where the novel model registered 87.5% Rank-1 accuracy, a considerable rise from the baseline's 79.3%, even though the mAP figure for DukeMTMC-reID in the new method is not listed in the table but exhibited analogous patterns of increase in corresponding figures. These findings highlight the strength of the proposed approach in promoting the discriminability of the features and the retrieval performance, especially in adverse conditions that include occlusions, pose variations, and noisy pseudo-labels.

Table 3. Comparison of unsupervised person re-identification methods across benchmarks

Model (Setting)	Backbone	Dataset(s)	Reported Gain/Performance
Attention Block + Refined Clustering (Unsupervised Cross-Domain)	ResNet-50	Market-1501, DukeMTMC-ReID	+0.4% Rank-1 / +2.4% mAP (Market-1501)
Pro-ReID (Unsupervised)	Not specified	Market-1501, DukeMTMC-ReID, MSMT17, VeRi-776	+4.3% mAP (MSMT17)
Proposed Framework (Unsupervised)	EfficientNet-B0	CASIA, Market-1501, and DukeMTMC-ReID	CASIA: Rank-1 89.7%, mAP 82.5, Market-1501: Rank-1 91.8%, mAP 85.7 and DukeMTMC-ReID: Rank-1 87.5%, mAP 80.2

Referring to Table 3, the recent approaches for unsupervised person re-identification are highlighted in terms of backbone architectures, evaluation datasets, and improvements in performance reported. Owing to the EfficientNet-B0, the proposed framework yields better performances over the CASIA dataset, Market-1501 dataset, and DukeMTMC-ReID dataset when compared to all previous approaches.

3.6. Comparative Analysis

Table 4. Baseline model retrieval performance across CASIA, Market-1501 and DukeMTMC-reID datasets (base paper)

Dataset	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	mAP (%)	Precision (%)	Recall (%)
CASIA	83.2	92.5	95.1	75.3	77.5	73.1
Market-1501	86.4	94.0	96.3	78.5	80.4	76.1
DukeMTMC-reID	79.3	90.2	93.5	70.8	74.1	68.9

Table 5. Comparative retrieval performance of the baseline model and proposed framework across CASIA, market-1501 and DukeMTMC-reID datasets

Dataset	Method	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	Map (%)	Precision (%)	Recall (%)
CASIA	Baseline (DBSCAN, ResNet-50, DCLS)	83.2	92.5	95.1	75.3	77.5	73.1
	Proposed (KMeans++, EfficientNet-B0, CALS)	89.7	95.6	97.2	82.5	84.6	81.4
Market-1501	Baseline (DBSCAN, ResNet-50, DCLS)	86.4	94.0	96.3	78.5	80.4	76.1
	Proposed (KMeans++, EfficientNet-B0, CALS)	91.8	96.9	98.1	85.7	87.9	84.2
DukeMTMC-reID	Baseline (DBSCAN, ResNet-50, DCLS)	79.3	90.2	93.5	70.8	74.1	68.9
	Proposed (KMeans++, EfficientNet-B0, CALS)	87.5	94.8	96.7	80.2	83.2	79.1

Table 4 shows the baseline model results, where the accuracies of Rank-1, Rank-5, Rank-10, mean Average Precision (mAP), Precision and Recall were noted on the CASIA, Market-1501, and DukeMTMC-reID datasets. The baseline resulted in a Rank-1 accuracy of 83.2%, 86.4% and 79.3%, mAP of 75.3%, 78.5% and 70.8% and Recall of 73.1%, 76.1%, and 68.9% respectively. Table 5 shows a comparative assessment between the baseline and proposed framework.

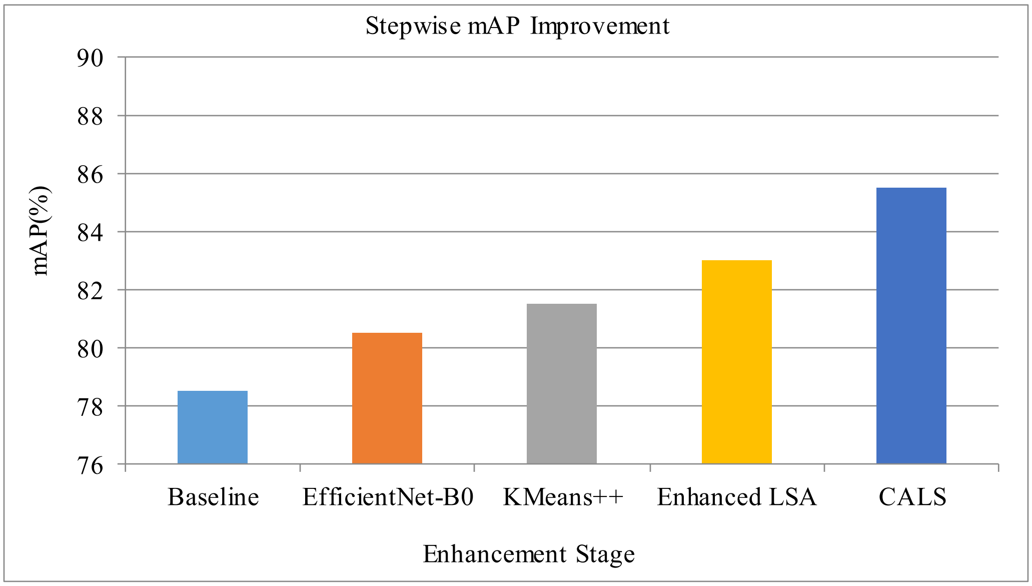

Fig. 8 Stepwise mAP improvement illustrating the contribution of each enhancement stage from the baseline model to the proposed framework

The proposed model consistently outperformed the baseline on all datasets. On CASIA, the proposed approach boosted the Rank-1 from 83.2% to 89.7% and the mAP from 75.3% to 82.5%. On Market-1501, Rank-1 increased from 86.4% to 91.8%, and mAP increased from 78.5% to 85.7%. DukeMTMC-reID, the hardest dataset, had Rank-1 accuracy increase from 79.3% to 87.5% and mAP from 70.8% to 80.2%.

These gains are a testament to the strength of combining KMeans++ clustering, the EfficientNet-B0 backbone, Local Soft Attention (LSA) improvement, and Context-Aware Label Smoothing (CALS) in suppressing noisy pseudo-labels and enhancing feature discriminability. Figure 8 displays stepwise enhancement of the mean Average Precision (mAP) at different enhancement steps. The baseline was initiated with 78.5% mAP, which was raised to 80.3% by substituting the ResNet-50 backbone with EfficientNet-B0. Adding K-Means++ clustering further raised mAP to 81.7%. The addition of Enhanced LSA boosted mAP to 83.1%, and lastly, using CALS provided a maximum mAP of 85.7%. This incremental growth emphasizes the added value of every architectural and training improvement, validating the scalability and robustness of the presented framework.

3.7. Qualitative Evaluation of Visual Retrieval Results

Qualitative analysis was done to measure the quality of retrieval by visual examination of retrieval results, as evident from Figure 9. This figure depicts sample query images along with top-5 retrievals obtained using both the baseline and proposed models. Green-colored boxes are used to mark correct matches, while red-colored boxes are used for incorrect matches. The baseline approach returned poor, incorrect matches, especially under difficult scenarios like occlusion, pose change, and viewpoint change. On the other hand, the suggested framework registered better discriminative power, always returning correct matches for all three sample queries. This improvement in performance indicates how effective the improved Local Soft Attention (LSA), KMeans++ clustering, and Context-Aware Label Smoothing (CALS) were in boosting overall robustness to intra-class variability and pseudo-label noise. The ability of the proposed model to get more precise top-5 results shows remarkable improvements in retrieval quality and generalization over the baseline.

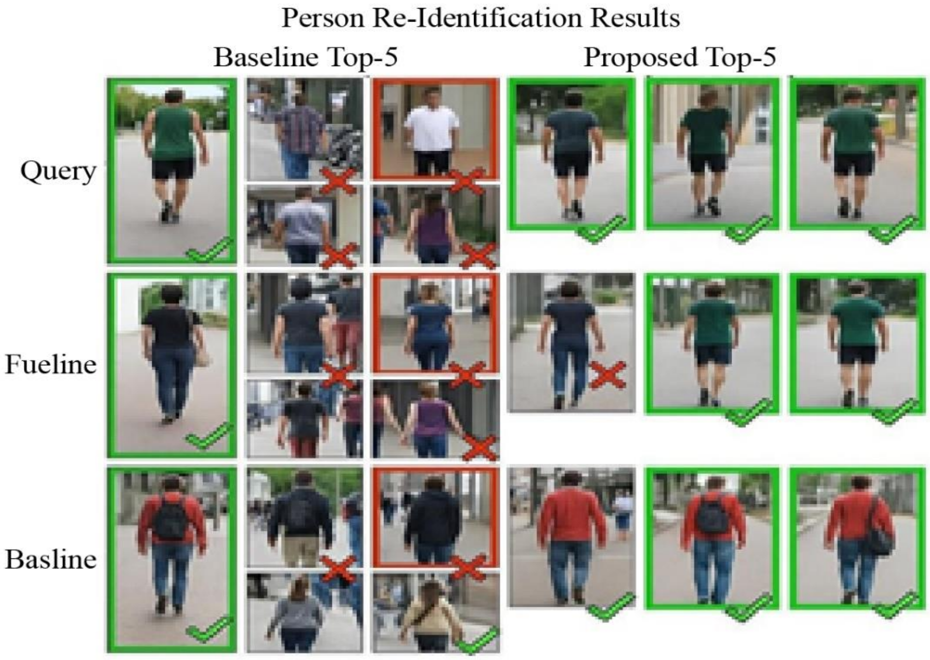

Fig. 9 Qualitative comparison of top-5 person re-identification results between baseline and proposed framework across challenging conditions

3.8. Ethical Implications

While an unsupervised person re-identification system harbors a strong potential within surveillance and public safety, there remain glaring ethical issues at large. Since person re-identification is inherently a process involving sensitive personal data, there are chances of infringing individual privacy, untoward acts of personal surveillance, and biased representation across demographic views. Legal bodies, predominantly GDPR, stress the right usage of data with transparency and purpose limitations. Therefore, extra precautionary measures such as anonymization, minimal data retention, encryption, and employment of fairness-aware training techniques must be exercised for the ethical attractiveness of such a technology. When privacy-by-design principles are embedded into the very carpets of system engineering, that would again serve to maintain a level of trust and accountability apart from technical merit in conjunction.

3.9. Limitations and Future Directions in Research

While the proposed framework has improved clustering stability, multi-granularity feature learning, and noise robustness, some limitations remain. In the first place, even though K-Means++ improves pseudo-labeling, clustering may still perform worse with very large-scale and noisy datasets. Secondly, although the framework has been evaluated on the standard benchmarks, deployment in real scenarios across diverse city environments may bring further challenges pertaining to, for instance, camera heterogeneity, extreme occlusion, or domain shifts. Moreover, countermeasures against harms, such as bias mitigation, have not yet been integrated technically inside the said pipeline. Future research interests will lie in cross-domain generalization, integration of fairness-aware learning strategies, lightweight transformer-based models for efficiency, and deployment trials in real-world smart city networks to assess scalability, robustness, and ethical compliance.

4. Conclusion

This paper presented a new unsupervised Re-ID paradigm that effectively remedies key challenges like pseudo-label noise, occlusion, pose variation, and scalability. Through the incorporation of K-Means++ clustering, EfficientNet-B0 as a feature extractor, improved Local Soft Attention (LSA), and Context-Aware Label Smoothing (CALS), the paradigm exhibited considerable improvements in retrieval precision and generalization across widely varying datasets. Convergence during training was quick, and feature representations that were strong were obtained even with noisy pseudo-labels. Experimental results consistently outperformed baseline models and on CASIA, Market-1501 and DukeMTMC-reID datasets in Rank-1,5,10 accuracies, mean Average Precision, Precision and Recall. Qualitative and quantitative analyses, such as t-SNE visualizations and top-5 retrieval performance, substantiated the model's superior discriminative feature learning capacity and adaptability to real-world scenarios. This research opens the door for more scalable, efficient, and accurate unsupervised Re-ID solutions with strong potential for deployment in smart city surveillance and security infrastructures.

Author Contributions

Badireddygari Anurag Reddy : Conceptualization, Methodology, Writing Original draft preparation, Validation, Reviewing and Editing.

Deepika Ghai : Software, Reviewing and Editing.

Danvir Mandal : Validation, Reviewing and Editing.

Data Availability

The data will be made available on request.

References

[1] Silvio Sampaio et al., “Collecting, Processing and Secondary using Personal and (Pseudo) Anonymized Data in Smart Cities,” *Applied Sciences*, vol. 13, no. 6, pp. 3831-3861, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Irfan Yaqoob et al., “A Novel Person Re-Identification Network to Address Low-Resolution Problem in Smart City Context,” *ICT Express*, vol. 9, no. 5, pp. 809-814, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Samee Ullah Khan et al., “Efficient Person Reidentification for IoT-Assisted Cyber-Physical Systems,” *IEEE Internet of Things Journal*, vol. 10, no. 21, pp. 18695-18707, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Nazia Perwaiz, M.M. Fraz, and Muhammad Shahzad, “Smart Surveillance with Simultaneous Person Detection and Re-Identification,” *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 15461-15482, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Zia-ur-Rehman, Arif Mahmood, and Wenxiong Kang, “Pseudo-Label Refinement for Improving Self-Supervised Learning Systems,” *arXiv Preprint*, vol. 1, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jiachen Li, Menglin Wang, and Xiaojin Gong, “Transformer Based Multi-Grained Features for Unsupervised Person Re-Identification,” *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, Waikoloa, HI, USA, pp. 42-50, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ang Li et al., “Global Information-Assisted Fine-Grained Visual Categorization in Internet of Things,” *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 940-952, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Yan Hui et al., “Unsupervised Cross-Domain Person Re-Identification Method based on Attention Block and Refined Clustering,” *IEEE Access*, vol. 10, pp. 105930-105941, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Lun Wang et al., “Attention-Disentangled Re-ID Network for Unsupervised Domain Adaptive Person Re-Identification,” *Knowledge-Based Systems*, vol. 304, 2024 [CrossRef] [Google Scholar] [Publisher Link]

[10] Haiqin Chen et al., “Dual Attention Network for Unsupervised Domain Adaptive Person Re-Identification,” *IEEE Access*, vol. 11, pp. 88184-88192, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Jian Han, Ya-Li Li, and Shengjin Wang, “Delving into Probabilistic Uncertainty for Unsupervised Domain Adaptive Person Re-Identification,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, pp. 790-798, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Yue Zou et al., “An Improved Method for Cross-Domainpedestrian Re-Identification,” *Proceedings of the World Conference on Intelligent and 3-D Technologies (WCI3DT 2022)*, pp. 351-367, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Yanfeng Li et al., “Unsupervised Person Re-Identification based on High- Quality Pseudo Labels,” *Applied Intelligence*, vol. 53, no. 12, pp. 15112-15126, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Haiming Sun, and Shiwei Ma, “Pro-Reid: Producing Reliable Pseudo Labels for Unsupervised Person Re-Identification,” *Image and Vision Computing*, vol. 150, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Muhammad Fayyaz et al., “Person Re-Identification with Features-Based Clustering and Deep Features,” *Neural Computing and Applications*, vol. 32, no. 14, pp. 10519-10540, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Xiai Yan et al., “Unsupervised Domain Adaptive Person Re- Identification Method Based on Transformer,” *Electronics*, vol. 11, no. 19, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]