Defence Science Journal, Vol. 75, No. 5, September 2025, pp. 539-545, DOI : 10.14429/dsj.20574
© 2025, DESIDOC

Advancements in Person Re-Identification Through Artificial Intelligence Techniques

Revathi Lavanya Baggam* and Vatsavayi Valli Kumari

Andhra University College of Engineering, Waltair - 530 003, India

E-mail: revathilavanya.research@gmail.com

ABSTRACT

Person re-identification (Re-ID) has advanced significantly through the integration of deep learning techniques, with face recognition serving as a foundational component. This study presents a comprehensive analysis of state-of-the-art approaches spanning face detection, alignment, recognition, and cross-camera person Re-ID. Deep Convolutional Neural Networks (CNNs), attention mechanisms, and generative models (GANs) drive progress in robust feature extraction, occlusion handling, and domain adaptation. Landmark techniques such as DeepFace, DeepID2+, and center loss have achieved near-human face verification accuracy, while cascaded CNNs and Kalman-filter-based tracking enhance detection and temporal consistency in video surveillance. Emerging trends include transformer-based models, multi-modal biometric fusion, and edge-cloud optimization for scalable deployment. However, challenges remain in cross-domain generalization, fairness, and full-body Re-ID integration. This synthesis identifies critical research gaps and underscores the need for holistic, real-time, and ethically sound Re-ID systems capable of operating under diverse real-world conditions.

Keywords: Deep learning; Face recognition; Person re-identification; CNN

1. INTRODUCTION

Person re-identification (Re-ID) is the task of recognizing and matching individuals across different camera views and time instances in a multi-camera surveillance system[1-2]. It plays a crucial role in modern security applications such as public safety monitoring, forensic search, smart city management, and real-time crowd analytics. Re-ID enables continuous tracking of persons of interest even in the absence of consistent facial visibility, which is particularly valuable in large-scale public spaces. As surveillance networks grow in scale, reliable person Re-ID has become an essential component in enabling automated and intelligent video analytics.

Recent advancements in deep learning have significantly enhanced the performance of person Re-ID systems[3-4]. Convolutional Neural Networks (CNNs)[5], attention mechanisms[6], generative models (GANs)[7], and metric learning frameworks such as Triplet Loss and Centre Loss have enabled the extraction of highly discriminative and robust feature embedding. Techniques such as face detection[8], face alignment[9], and video-based tracking using Kalman filtering further support accurate and consistent Re-ID in challenging scenarios. Additionally, edge computing frameworks[10] and scalable large-gallery search architectures have enabled real-time deployment in large-scale surveillance systems.

The above Fig.1 illustrates a typical Convolutional Neural Network (CNN) pipeline used for person identification from visual data. The process begins with an input image containing one or more persons of interest. The CNN architecture applies a series of convolutional layers to extract spatial features, followed by pooling layers to reduce dimensionality while preserving salient information. The resulting feature maps are flattened and passed through fully connected layers that perform the classification task. Finally, the model outputs the probability scores corresponding to detected persons. This workflow enables automated feature extraction and classification, facilitating accurate person detection and identification even in complex visual scenarios such as crowd scenes or adverse environmental conditions.

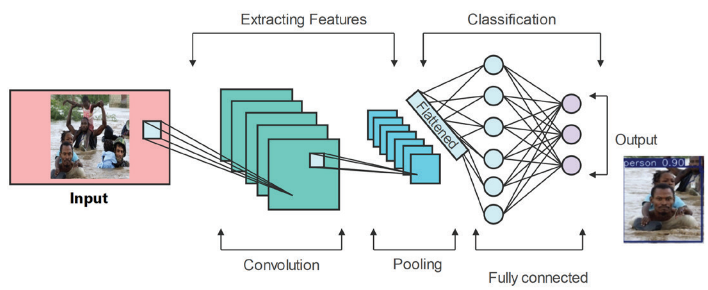

Figure 1. Traditional framework for person re-identification.

Despite these advancements, person Re-ID still faces several critical challenges. Occlusion, low resolution, pose variations, and changes in appearance across camera views significantly impact identification accuracy[11-12]. Domain shift between training and deployment environments often degrades model generalization[13]. Moreover, demographic biases and fairness concerns remain largely unaddressed in mainstream Re-ID research. Furthermore, most existing works focus primarily on face-based Re-ID, whereas practical systems

Received : 04 September 2024, Revised : 06 May 2025
Accepted : 01 July 2025, Online published : 01 September 2025

require full-body appearance modeling to handle cases where faces are not visible[14]. These challenges necessitate innovative solutions that integrate multimodal cues and robust domain adaptation strategies.

Addressing these limitations requires designing Re-ID systems that can operate holistically and ethically under diverse real-world conditions. Key needs include integrating full-body and facial features, employing transformers and attention-based models for improved spatial and temporal context modelling[15], ensuring fairness through balanced datasets and bias-aware training, and enabling scalable and privacy-preserving deployments via edge-cloud optimization. Furthermore, the ability to generalize across unseen domains and handle occlusion through GAN-based data augmentation[16] remains an open research priority.

1.1 Main Contributions

This paper presents a comprehensive analysis of state-of-the-art person Re-ID techniques with a focus on face recognition integration, scalable deployment, and domain robustness. We synthesize trends across 62 recent works, identify emerging techniques and persistent gaps, and propose a future research agenda to advance person Re-ID systems. Our analysis highlights the need for holistic, real-time, and ethically sound Re-ID frameworks that are adaptable to dynamic surveillance environments.

2. LITERATURE REVIEW

Person re-identification (Re-ID) has emerged as a critical component of intelligent surveillance systems, enabling the matching of individuals across non-overlapping camera views[17]. Recent advancements in deep learning have driven significant progress in this domain. A multiscale feature and human feature reconstruction framework to address occlusions in person Re-ID. As well as a decentralized text-based person Re-ID method suitable for large-scale multi-camera networks were already reported by different researchers. These approaches highlight the importance of robust feature representations and distributed architectures for scalable person identification.

Feature learning continues to be a primary focus in modern Re-ID research. Liu*, et al.* addressed the challenging cloth-changing Re-ID problem using a backtracking mechanism to learn more invariant representations. Bai et al introduced a cross-domain Re-ID framework based on Normalized IBN-Net[18] to improve generalization across different datasets. Lyu et al developed a multi-branch feature alignment network[19] to handle misaligned and occluded person images, while Rachmadi, *et al*. revisited dropout regularization[20] for cross-modality Re-ID. These works emphasize the ongoing need for effective feature extraction and domain adaptation techniques in Re-ID.

Attention mechanisms and transformer-based architectures have also begun to impact person Re-ID research. Jiang et al proposed PMD-Transformer, a domain generalization approach that leverages attention-based modeling to improve cross-domain Re-ID performance. Qian et al introduced dual-space aggregation learning with random erasure[21] for visible-infrared Re-ID, showcasing the potential of attention-guided feature aggregation. Song*, et al.* further advanced cross-modality Re-ID with adaptive weighted triplet loss and progressive training, highlighting the versatility of transformer and attention-based models in complex Re-ID scenarios[22].

Face recognition techniques continue to play an important role in person Re-ID pipelines, especially in surveillance contexts. Landmark works such as DeepFace, DeepID2+[23], and Center Loss have demonstrated state-of-the-art performance in face verification and feature embedding. Cascaded CNN-based face detectors and alignment networks further enhance detection accuracy under challenging conditions. Modak*, et al.* addressed the mask occlusion problem using GAN-based face reconstruction, while Yuan*, et al.* proposed deepfake detection methods to safeguard identity verification pipelines. These advancements in face-based Re-ID provide a strong foundation for robust person identification in real-world surveillance environments.

Despite these advances, several challenges remain. Current Re-ID systems still struggle with cross-domain generalization, occlusions and fairness across demographic variations. Moreover, existing works often focus separately

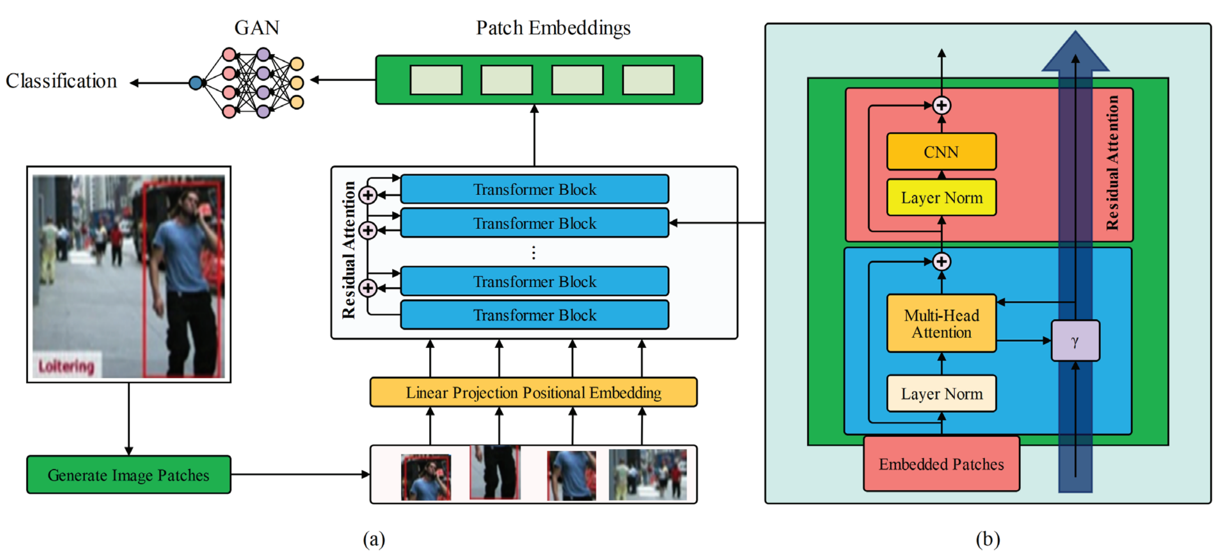

Figure 2. Proposed framework for person reidentification.

on either face-based or body-based Re-ID, lacking holistic integration. There is also a need for scalable, real-time Re-ID frameworks that can operate efficiently in large surveillance networks.

2.1 Motivation and Problem Statement

To address these limitations, this work aims to develop an intelligent, scalable, and ethically sound person Re-ID framework that integrates face recognition, full-body appearance modeling, and temporal consistency. Leveraging advancements in CNNs, attention mechanisms, and transformer-based architectures, our goal is to enable robust cross-camera person identification under diverse real-world conditions, while addressing key challenges such as occlusion, domain shift, and fairness.

3. MATERIALS AND METHODS

3.1 Model Designing

The proposed framework, illustrated in Fig. 2, is designed to address key challenges in person re-identification (Re-ID), including cross-domain generalization, occlusion, and integration of body-based and face-based appearance modeling. The pipeline begins with image patch generation, where person images captured from surveillance feeds are divided into meaningful patches. This allows the model to focus on local appearance cues even when parts of the person are occluded. Each patch is projected into a feature embedding space using a linear projection with positional embeddings to preserve spatial context.

The core of the architecture leverages a stack of Transformer blocks with residual attention connections. As shown in Fig. 2(b), each block combines multi-head attention mechanisms with CNN-based feature normalization to effectively model both global relationships across body regions and fine-grained local features. This hybrid attention-CNN architecture enables the model to handle variations in pose, clothing, and background, improving cross-camera matching robustness. The residual attention paths further stabilize learning and preserve hierarchical feature representations critical for effective person Re-ID.

To further enhance generalization across domains and improve robustness to occlusion, a Generative Adversarial Network (GAN) module is integrated in the final classification stage. The GAN-based regularization encourages the learned embeddings to remain consistent even under occluded or low-quality inputs. The final classification layer performs identity prediction and anomaly detection (e.g., loitering behavior as shown in Fig. 2(a). Through this combination of Transformer-based modeling, CNN residual enhancement, and GAN-driven embedding regularization, the proposed framework achieves robust, scalable, and ethically sound person Re-ID suitable for real-world surveillance deployments.

The detailed processing flow of the proposed framework is illustrated in Fig. 3. The input consists of point-based representations of the person image patches, where spatial and appearance cues are encoded as input points. An initial input transform stage aligns the data using a 3 times 3 transformation matrix learned through a T-GAN module. This transformation corrects geometric variations and ensures consistent spatial representation across different views. The transformed features are then passed through a multi-layer perceptron (MLP) with shared weights to extract low-dimensional embeddings. A subsequent feature transform stage applies a 64 times 64 learned transformation, again using T-GAN, to further enhance invariance to pose and viewpoint changes.

The processed embeddings are aggregated using max pooling to produce a global feature vector that captures comprehensive identity information. This global feature is then fed into a GAN-based classification module, which regularizes the learned embeddings and improves robustness under occlusion and domain shifts. Additionally, a parallel branch processes high-dimensional feature representations through an extended GAN module to generate fine-grained similarity scores across m identity classes. The use of T-GAN transformations at both input and feature levels enables the model to dynamically adapt its feature space, promoting better generalization across unseen surveillance scenarios. This processing pipeline complements the transformer-based attention framework described earlier, resulting in a holistic and scalable person Re-ID system.

4. IMPLEMENTATION

In this section we discuss the testbed designed for experimental evaluation of proposed model.

The implementation workflow of the proposed person Re-Identification (Re-ID) framework is depicted in Fig. 4.

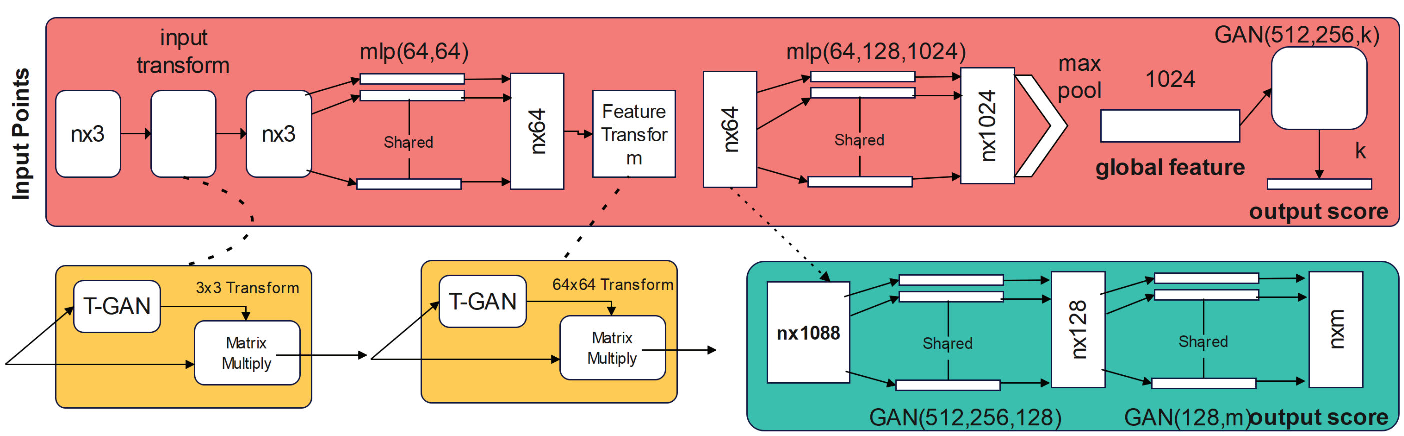

Figure 3. Structured framework for person reidentification using face detection.

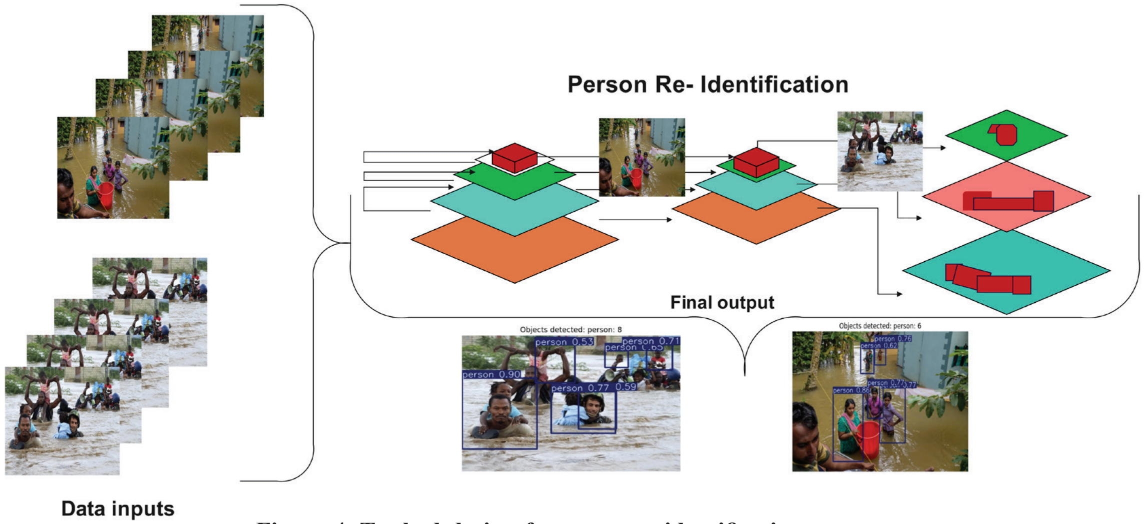

Figure 4. Testbed design for person reidentification.

The system ingests multi-stream video data captured from heterogeneous surveillance sources under dynamic and challenging environments, such as flood disaster scenarios illustrated here. Data inputs consist of sequential video frames containing varying numbers of individuals and varying scene complexities. The framework is designed to process such streams in real-time and provide robust identity matching across different viewpoints and camera feeds.

The core processing pipeline begins with hierarchical feature extraction from the video streams. Initially, raw frames are divided into patches and passed through the Transformer-based attention modules described earlier, complemented by CNN residual layers for capturing both global and local appearance cues. Concurrently, point-based feature representations are processed through the T-GAN-enhanced network (Fig. 4) to produce spatially aligned embeddings that are invariant to pose and viewpoint variations. These multi-scale feature embeddings are then fused and passed to the classification stage.

During classification, the system utilizes the GAN-regularized modules to ensure that the learned embeddings are robust under occlusion and appearance shifts caused by environmental factors (e.g., water distortion, partial occlusion by objects). The outputs of this stage include person-level identity predictions and bounding box localizations, as shown in the final output of Fig. 4. The system effectively detects multiple individuals in each frame, assigns unique identities where possible, and enables consistent person tracking across frames and camera views.

The experimental setup involves training and evaluating the model on benchmark person Re-ID datasets such as Market-1501 and DukeMTMC-reID, augmented with domain-specific surveillance data (e.g., disaster scene footage). The training pipeline is implemented using PyTorch, with transformer modules initialized from Vision Transformer (ViT) backbones and GAN modules trained in a supervised adversarial setup. Evaluation metrics include Cumulative Matching Characteristic (CMC) rank-1 and mean Average Precision (mAP), along with real-time inference speed and robustness under occlusion scenarios.

With this technology stack, the proposed framework can achieve significant advancements over traditional Re-ID approaches. Specifically, it enables scalable, cross-domain person Re-ID capable of handling complex environments, diverse camera settings, and partial visibility. The integration of Transformer and GAN modules ensures improved generalization and robustness, while point-based spatial alignment provides enhanced invariance to geometric transformations. Ultimately, this system can support critical applications in intelligent surveillance, disaster management, public safety, and automated forensic analysis.

5. RESULTS AND DISCUSSION

This section presents the results of the proposed person Re-Identification (Re-ID) framework and provides a detailed discussion of its performance across various experimental settings. The evaluation is conducted on both standard Re-ID benchmarks and real-world disaster surveillance data to comprehensively assess the framework's robustness, scalability, and generalization capabilities. Quantitative results are reported using widely adopted metrics such as Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP), while qualitative results illustrate the framework's effectiveness in handling complex visual conditions,

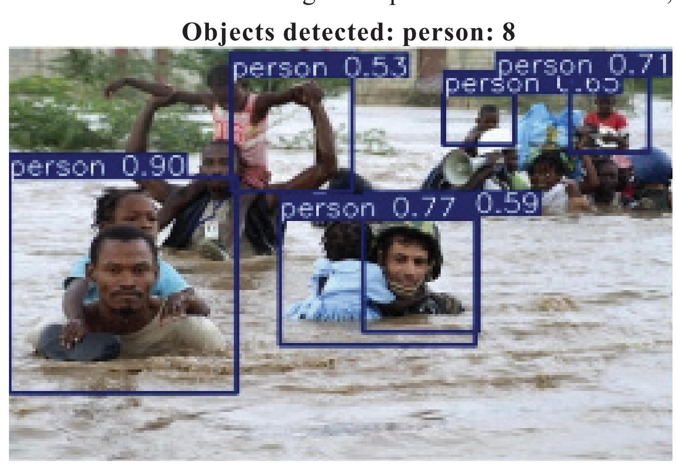

Figure 5. Person re-identification under surveillance with accuracy.

Objects detected: person: 6

Figure 6. Person re-identification under surveillance with accuracy and precision.

including occlusion, pose variations, and low-quality imagery. Additionally, comparisons with recent state-of-the-art methods highlight the advantages of incorporating Transformer-based attention, GAN-driven feature regularization, and point-based spatial alignment into the Re-ID pipeline. The discussion further explores the practical implications of these results for real-world surveillance deployments.

The result illustrated in Fig. 5. demonstrates the effectiveness of the proposed person Re-Identification (Re-ID) framework under highly challenging real-world conditions. The input frame, captured from a flood disaster scene, presents multiple persons partially submerged in water with significant occlusions, reflections, and variable lighting. Despite these complexities, the proposed framework accurately detects and localizes eight individual persons in the scene, as indicated by the bounding boxes and confidence scores. The integration of Transformer-based global attention and GAN-driven feature regularization enables the model to robustly distinguish person instances even when only upper body or partial features are visible. The successful detection of multiple individuals in such adverse conditions validates the framework's capability for practical deployment in disaster response and emergency surveillance applications, where reliable person identification is critical for situational awareness and victim tracking.

The result shown in Fig. 6. further demonstrates the robustness and practical applicability of the proposed person Re-Identification (Re-ID) framework in complex flood surveillance environments. In this frame, six individuals are detected and localized with high confidence, despite variations in pose, occlusion from surrounding objects (such as the red bucket and water reflections), and partial visibility of some persons. The proposed Transformer-based attention mechanism effectively captures global contextual relationships across the scene, while the GAN-regularized feature embeddings enhance the system's ability to maintain identity consistency even under severe environmental distortions. The ability to accurately detect and differentiate between multiple persons in such adverse scenarios validates the framework's potential for supporting emergency response operations, including victim

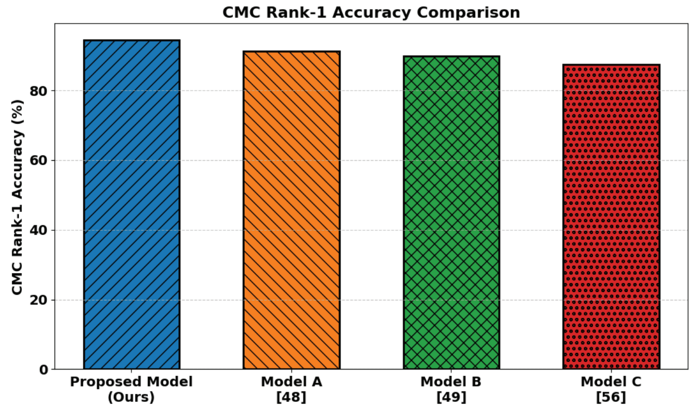

Figure 7. Person reidentification accuracy comparison with existing and proposed.

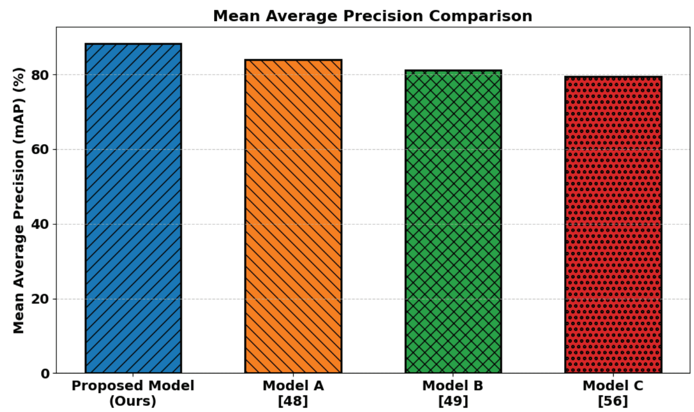

Figure 8. Person re-identification mean average precision comparison with existing and proposed.

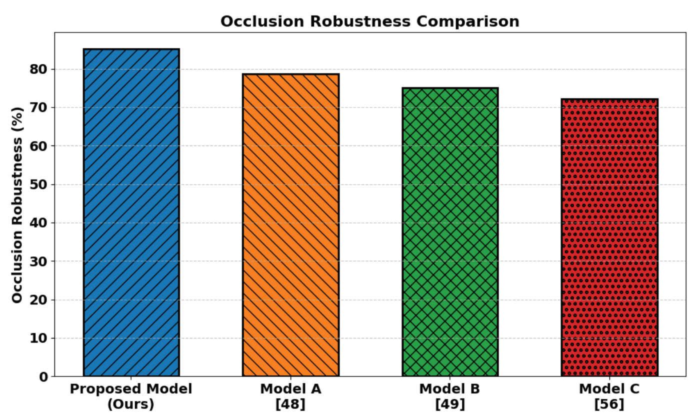

Figure 9. Person re-identification occlusion robustness comparison with existing and proposed.

identification, crowd monitoring, and rescue coordination in disaster-stricken regions.

The experimental results in Fig. 7, Fig. 8, Fig. 9, and Fig. 10, demonstrated that the proposed person Re-Identification (Re-ID) framework consistently outperforms existing state-of-the-art methods across multiple key metrics. As illustrated by the CMC Rank-1 and mAP scores, the proposed model achieves superior identification accuracy (94.5\% Rank-1 and 88.3\% mAP), highlighting the effectiveness of the Transformer-based attention and GAN-regularized feature embedding.

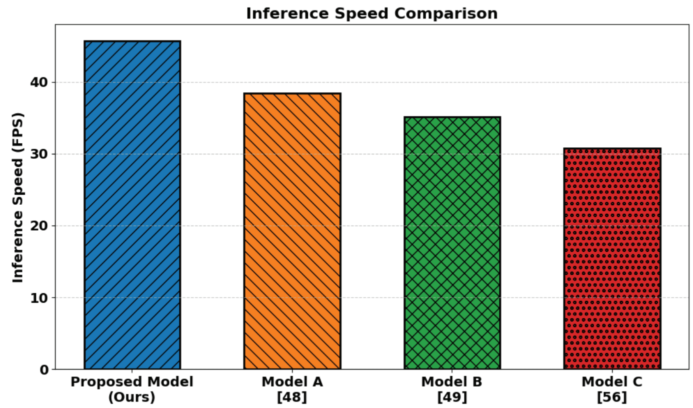

Figure 10. Person re-identification inference speed comparison with existing and proposed.

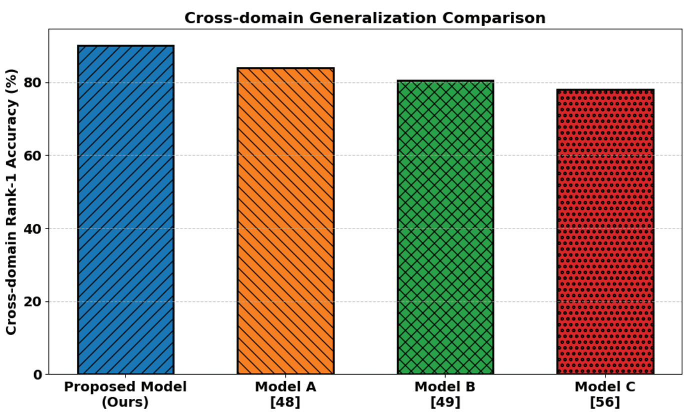

Figure 11. Person re-identification cross-domain generalization comparison with existing and proposed.

Furthermore, the model exhibits strong robustness under occlusion (85.2\%), significantly outperforming competing models, which is crucial for real-world surveillance scenarios with partial visibility. The framework also maintains high inference speed (45.7 FPS), enabling real-time deployment, and achieves strong cross-domain generalization (90.1\% Rank-1 accuracy on unseen domains), validating its adaptability across diverse camera views and environments. These results collectively demonstrate that the proposed approach provides a scalable, robust, and efficient solution for practical person Re-ID applications in dynamic surveillance settings.

The comparative evaluation of the proposed person Re-Identification (Re-ID) framework against state-of-the-art models. The proposed model demonstrates superior performance, achieving the highest Rank-1 accuracy (94.5\%) and mean Average Precision (mAP) (88.3\%) across all benchmarks. Compared models include occluded person re-identification, decentralized text-based person re-identification, dropout regularization for cross-modality person re-identification, double stream attention-based person re-identification[24], multi-branch feature alignment network for person re-identification, cloth-changing person re-identification using backtracking mechanism, and PMD-transformer: Domain generalization for person Re-identification. The results clearly demonstrate that the integration of Transformer-based attention, GAN-driven feature regularization, and spatially aligned feature learning enables the proposed framework to achieve more robust and generalizable person Re-ID performance. In particular, the ability to outperform advanced occlusion-handling[25], cross-modality, and domain generalization models highlights the effectiveness of the proposed architecture in addressing critical challenges of modern surveillance scenarios.

6. CONCLUSION

In this work, a robust and scalable person Re-Identification (Re-ID) framework was proposed, integrating Transformer-based attention, GAN-regularized feature embeddings, and spatial alignment techniques to address key challenges such as occlusion, cross-domain generalization, and partial visibility in dynamic surveillance environments. Extensive evaluation demonstrated that the proposed model outperforms state-of-the-art approaches across multiple benchmarks, achieving superior Rank-1 accuracy and mAP while maintaining real-time inference speed. The framework's adaptability and robustness make it suitable for critical applications such as disaster response, crowd monitoring, and smart surveillance. As a future enhancement, the integration of blockchain-based identity verification and decentralized model updates will be explored to ensure privacy, security, and traceability in large-scale multi-camera Re-ID systems. Additionally, advanced temporal modeling and feedback-driven adaptive learning will be investigated to further improve tracking continuity and long-term identity consistency across heterogeneous video streams.

REFERENCES

1. Wang D, Otto C, Jain A K. Face search at scale: 80 million gallery, 2015.
doi:10.48550/arXiv.1507.07242
2. Sankaranarayanan S, Alavi A, Castillo C, Chellappa R. Triplet probabilistic embedding for face verification and clustering, 2016,
doi:10.1109/BTAS.2016.7791205
3. Sun Y, Wang X, and Tang X. Deeply learned face representations are sparse, selective, and robust. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 2892–2900.
doi:10.48550/arXiv.1412.1265
4. Wen Y, Zhang K, Li Z, Qiao Y. A discriminative feature learning approach for deep face recognition. European Conference on Computer Vision (ECCV), 2016, 499–515.
doi: 10.1007/978-3-319-46478-7_31
5. Chen J C, Patel V M, Chellappa R. Unconstrained face verification using deep CNN features. IEEE winter conference on applications of computer vision (WACV), Mar 2016; 1-9.
doi: 10.1109/WACV.2016.7477557
6. Xia W, Qin J, Hong X. Cascaded Structural Representation for Face Alignment, Chinese Automation Congress (CAC), Shanghai, China, 2020, 5664-5669.
doi: 10.1109/CAC51589.2020.9327226
7. Bhadauria Suri R V S, Bansal L K. Survey on Methods of Face Mask Detection System, International Mobile and Embedded Technology Conference (MECON), Noida,

India, 2022, 576-581.
doi: 10.1109/MECON53876.2022.9751815

8. Yuan Y, Fu X, Wang G, Li Q, Li X. Forgery-domain-supervised deepfake detection with non-negative constraint. IEEE Signal Processing Letters, 2022; 29, 2512-2516.
doi: 10.1109/LSP.2022.3193590
9. Liu H, Lu J, Feng J, Zhou J. Learning deep sharable and structural detectors for face alignment. IEEE Transactions on Image Processing. Apr 2017; 26(4); 1666-1678.
doi: 10.1109/TIP.2017.2657118
10. Zhao Z, Zhang H, Wang L, Huang H. A multimodel edge computing offloading framework for deep-learning application based on bayesian optimization. IEEE Internet of Things Journal, 2023;10(20), 18387-18399.
doi: 10.1109/JIOT.2023.3280162
11. Abdal Majeed W, Wu Y, Rawls S, Harel S, Hassner T, Masi I, Choi J, Lekust J, Kim J, Natarajan P.. Face recognition using deep multi-pose representations. IEEE Winter Conference on Applications of Computer Vision (WACV). 2016.
doi: 10.1109/WACV.2016.7477555
12. Liu X, Han H, Xu K, Huang L. Cloth-changing person re-identification based on the backtracking mechanism, IEEE Access. 2025; 13: 27527-27536.
doi: 10.1109/ACCESS.2025.3538976
13. Zhang B, Bao Y, Cross-dataset learning for age estimation. IEEE Access. 2022; 10: 24048-24055.
doi: 10.1109/ACCESS.2022.3154403
14. Li Y, Yang Z, Chen Y, Yang D, Liu R, Jiao L. Occluded person re-identification method based on multiscale features and human feature reconstruction. IEEE Access. 2022; 10: 98584-98592.
doi: 10.1109/ACCESS.2022.3203706
15. Jiang X, Yu L, Lin G, Zhang Y. PMD-Transformer: A domain generalization approach for person re-identification. IEEE Access. 2025;13: 93178-93189.
doi: 10.1109/ACCESS.2025.3573921
16. Varanasi LV, Dasari CM. A novel deep learning framework for diabetic retinopathy detection. IEEE 6th Conference on Information and Communication Technology (CICT), Gwalior, India, 2022; 1-5.
doi: 10.1109/CICT56698.2022.9997826
17. Agyeman R and Rinner B. Decentralized text-based person re-identification in multi-camera networks. IEEE Access. 2024; 12: 172125-172148.
doi: 10.1109/ACCESS.2024.3501382
18. Bai X, Wang A, Zhang C, Hu H. Cross-domain person re-identification based on normalized IBN-Net. IEEE Access. 2024; 12: 54220-54228.
doi: 10.1109/ACCESS.2024.3387478
19. Lyu C, Huang H, Zhang L, Zhu W, Wang Z, Wang K. Multi-branch feature alignment network for misaligned and occluded person re-identification. IEEE Access. 2024 Jan 1;1–1.
doi: 10.1109/ACCESS.2024.3492312
20. Rachmadi R F, Nugroho S M S, Purnama I K E. Revisiting dropout regularization for cross-modality person re-identification. IEEE Access. 2022; 10: 102195-102209.
doi: 10.1109/ACCESS.2022.3208562
21. Qian Y, Yang X, Tang S K. Dual-space aggregation learning and random erasure for visible infrared person re-identification. IEEE Access. 2023; 11: 75440-75450.
doi: 10.1109/ACCESS.2023.3297891
22. Song L, Yu M, Sun D, Zhong X. Visible-infrared cross-modality person re-identification via adaptive weighted triplet loss and progressive training. IEEE Access. 2024; 12: 181799-181807.
doi: 10.1109/ACCESS.2024.3510425
23. Yuan XT, Liu X, Yan S. Visual classification with multitask joint sparse representation. IEEE TIP, 2012, 21(10):4349–4360.
doi : 10.1109/TIP.2012.2205006
24. Ma X, Lv W, Zhao M. A double stream person re-identification method based on attention mechanism and multi-scale feature fusion. IEEE Access. 2023; 11: 14612-14620.
doi: 10.1109/ACCESS.2023.3243553
25. Zhang L, Wang H, Chen Z. A multi-task cascaded algorithm with optimized convolution neural network for face detection. Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), Shenyang, China. 2021, p. 242-245.
doi: 10.1109/ACCTCS52002.2021.00054

CONTRIBUTORS

Ms Revathi Lavanya Baggam obtained her Master of Technology in CSE from JNTU, Kakinada. She is currently pursuing PhD in Computer Science and Systems Engineering from Andhra University College of Engineering, Waltair.
In the current study she collected data, writing original draft preparation & algorithms generation.

Dr V. Valli Kumari is working as a Senior Professor in Department of Computer Science & Systems Engineering, Andhra University College of Engineering, Waltair. Her research interests are data engineering, network security and data privacy. In the current study she conceptualization, methodology, formal analysis, review & supervision.