IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 December 2024, accepted 24 January 2025, date of publication 29 January 2025, date of current version 4 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3536478

SURVEY

A Review on Person Re-Identification Techniques and Its Analysis

C. SELVAN[1], (Senior Member, IEEE), H. ANWAR BASHA[2], K. MEENAKSHI[3], AND SOUMYALATHA NAVEEN[4], (Member, IEEE)

[1]School of Computer Science and Engineering, REVA University, Bengaluru, Karnataka 560064, India
[2]Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai, Tamil Nadu 600124, India
[3]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, Tamil Nadu 600026, India
[4]Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: Soumyalatha Naveen (soumyalatha.naveen@manipal.edu)

ABSTRACT Person re-identification (Re-ID) emerges as a captivating realm within computer vision, dedicated to the task of recognizing the same individual across diverse camera angles or locations. The realm of video-based person re-identification (video re-ID) has recently captivated increasing interest, owing to its wide array of practical applications spanning surveillance, smart city solutions, and public safety measures. Nevertheless, video re-ID proves to be a formidable challenge, an ever-evolving domain fraught with a multitude of uncertainties like viewpoint variations, occlusions, pose changes, and unpredictable video sequences. Over the past few years, the realm of deep learning applied to video re-ID has consistently delivered remarkable outcomes on public datasets, showcasing a range of innovative strategies devised to tackle the array of issues encountered in video re-ID. In stark contrast to image-based re-ID, video re-ID stands out as significantly more intricate and demanding. In a bid to inspire forthcoming research endeavors and confronts emerging challenges, this paper presents a comprehensive overview of the latest advancements in deep learning methodologies tailored for video re-ID. It delves into three crucial facets; encompassing succinct explanations of video re-ID techniques along with their constraints, pivotal breakthroughs coupled with the technical hurdles faced, and the architectural framework underpinning these developments. The paper further furnishes a comparative analysis of performance across diverse datasets, offers insightful guidance on enhancing video re-ID strategies, and outlines compelling avenues for future research exploration.

INDEX TERMS Person re-identification, deep learning, vision based re-identification, review analysis.

I. INTRODUCTION

Person Re-Identification (Re-ID) [1] pertains to a specific individual retrieval issue observed across non-overlapping, disjoint camera systems. The objective of Re-ID [2] is to ascertain whether a person of interest has been present in a different location at a separate time as recorded by multi-camera setups or even within the same camera system but at a different time point.

Queries can be represented by images, video sequences, or textual descriptions. Re-ID aids law enforcement and security personnel in monitoring an individual's movements across various cameras, even if they exit a particular area. This capability is essential for conducting investigations and preventing criminal activities. Person Re-ID involves a complex mission, yet plays a vital role in enhancing the meaningful connection in analysis. Understanding Person Re-ID is akin to grasping a simple concept, something that humans effortlessly do on a regular basis. The human eyes and mind are adept at spotting, pinpointing, recognizing, and subsequently re-recognizing objects and individuals in the physical realm. Re-ID signifies the process of identifying a previously encountered individual in their subsequent appearance through a distinct characteristic of that individual. Humans possess the ability to derive such a characteristic by observing the person's facial features, stature, attire,

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

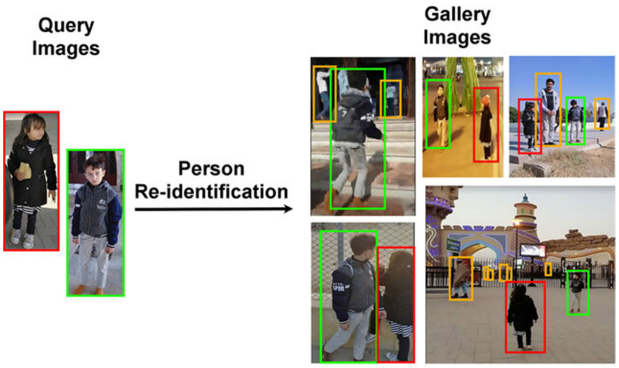

FIGURE 1. Person Re-Identification- an overview.

hair color, hairstyle, gait, and more. Among these, a person's facial features stand out as the most distinctive and dependable trait that aids in identification. Conversely, automating person Re-ID proves to be a challenging task without human involvement. Generally, automating person Re-ID poses a formidable challenge due to various factors, which will be elaborated on later in this segment. Nevertheless, the primary hurdle in Re-ID lies in the variability of a person's appearance across different surveillance cameras. As depicted in Fig. 1, images of an individual captured by diverse cameras on different occasions showcase the discrepancies in appearance. Interestingly, the visual transformation is also noticeable within the same camera perspective.

A typical Re-ID system comprises two fundamental elements: creating a unique descriptor or model for an individual and subsequently comparing these models to determine a match or mismatch. To establish a distinctive individual descriptor, the system must possess the capability to automatically detect and monitor individuals in images or videos. The typical flow process involved in the person re-identification is illustrated in the Fig. 2.

The major processes involved in the person Re-Identification application are the image or video capturing using the surveillance camera. In general, more than one number of camera were employed in the capturing of the target image/video, which assists in multi-dimensional image or video. The captured image is performed with the target person tracking or detection process, executed through the vision based identification algorithms. The target tracking process is followed by the feature extraction and the description generation process. The feature extraction [3] process shall be accomplished by employing the various and novel feature extraction algorithms which is followed by the matching process. The matching process determines the matching between the target and the test image yielding the result of person matching or person not-matching.

A common system for Re-identification might take in an image (single-shot) or a video (multi-shot) to extract features and generate descriptors. When dealing with an image input, it is crucial to accurately detect and locate the person for precise feature extraction. In cases where multiple images are present, it is vital to establish a connection between recognized individuals across frames to ensure that the extracted features pertain to the right person. This matching process, known as tracking, assigns a consistent label to each individual across different frames. Consequently, numerous instances of a person can be utilized for feature extraction and subsequent descriptor generation in Re-identification tasks. Person detection and tracking multiple individuals pose significant challenges with their unique obstacles. Extensive efforts have been dedicated to tackling the issue of person detection over the years. While Multiple Object Tracking (MOT) [4] within a single camera's field of view has been extensively studied with numerous proposed algorithms in the last twenty years, maintaining continuous tracking across diverse observation settings remains an ongoing challenge.

This manuscript reviews the existing architectures and methodologies of person re-identification process and performs the detailed literature analysis. Based on the literature analysis, this article provides a details research scope of advancements in the person re-identification applications. This manuscript is organized with a detailed introduction in section I and the various architectures along with the loss functions were presented in section II. The section III presents the results of the fruitful research performed by the existing researchers and the section IV performs the literature analysis. The review manuscript was concluded by mentioning the future scope of research in the person re-identification application.

II. ARCHITECTURES AND LOSS FUNCTIONS

This section presents the various architectures and the loss functions employed in the person re-identification process. The various methods employed for the person re-identification process are the Spatio-temporal representation [5], Deep Metric Learning [6], Gait based recognition [7], Attention mechanisms [8].

A. SPATIO-TEMPORAL REPRESENTATION

The Spatio-Temporal representation method of person re-identification from the video signal is introduced to overcome the drawbacks experienced by the temporal dimension images. The temporal dimension image is composed of both static and dynamic content, whereas the spatial dimension is composed of both contents along with the coarse details. The existing methods experiences a major setback of appearance of similarity, misalignment of the frames and apparently the occlusion concern. To overcome this, the Spatio-Temporal representation is introduced in the person re-identification process. To tackle these concerns, we present the innovative concept of Spatio-Temporal Representation Factorization (STRF), a versatile and efficient computational component that can seamlessly integrate within the layers of any 3D-CNN [9] framework designed for re-identification. The flexibility of STRF makes it highly attractive for real-world scenarios where tailored architectures are essential to accommodate diverse data patterns. Furthermore, the efficacy of this

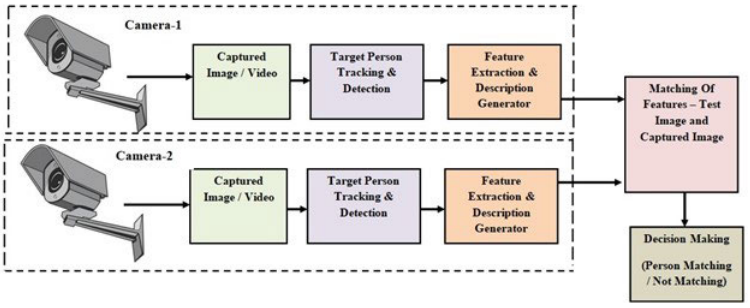

FIGURE 2. Processes involved in person Re-Identification.

module surpasses that of established specialized structures in capturing spatio-temporal re-ID features.

$$v = [v_1, v_2, \ldots v_n] \in R^{n*m*w} \tag{1}$$

where, v is the input video consists of 'n' number of frames, Let the function 'f' represents the feature encoder of the any convolutional network and the feature tensor shall be determined as in equation 2.

$$f_n \in R^{n*m*w*c} \tag{2}$$

where, c represents the number of channels, n represents the number of frames, h represents the height and w represents the width of the feature respectively. The mask for the feature shall be applied using the equation 3.

$$\bar{f}_n^{dn} = \bar{f}_n^i Md_i; dn \in \{n, s\} \tag{3}$$

where, Mdi is the factorized attention mask and the fn is the weighted volumes of the obtained features. The term 's' is the spatial index and dn is the maximum limit of frames. The spatial temporal representation is pictorially represented in Figure 3.

The spatio temporal representation depicted in the figure 3, accepts the input video sequence, from which the width, height and the time are measured as a contribution of spatio-temporal analysis. The spatio temporal representation is converted into planar representation. The planar represented video sequence is fed as the input to the neural networks to make the decision and for person re-identification process. The major setback of the spatial temporal representation is the limitation of discrete models that arises during the conversion of spatial domain to time domain.

B. DEEP METRIC LEARNING

Deep Metric Learning (DML) method of person re-identification (Re-Id) has been introduced with the influence of illumination and resolution. The DML technique of person Re-Id deals with the non-linear metric functions to overcome the pattern recognition concerns. Deep metric learning techniques embed the concept of similarity directly into the training objective. Siamese networks with contrastive and triplet loss stands out as the most notable formulations. By reducing the distance between samples of the same class, the contrastive loss enforces a gap between samples of different classes. This approach essentially drives all samples of the same class towards a singular point in the representation space while penalizing any overlap between distinct classes. On the other hand, the triplet loss offers a more flexible version of the contrastive formulation, allowing samples to move with more freedom while maintaining the margin. When presented with an anchor point, a point of the same class, and a point of a different class, the triplet loss ensures that the distance to the point of the same class is shorter than the distance to the point of the different class by a specified margin. The similarity between the input features is determined using the equation 4.

$$S = \frac{(F_1(t) * F_2(t)}{\sqrt{F_1(t)^T * F_2(t)}\sqrt{F_1(t) * F_2(t)^T}} \tag{4}$$

where, $F_1(t)$ and $F_2(t)$ are the features of the input video sequence, while S is the similarity between the two features. Siamese neural networks in their current form are bound by a unique restriction: both of their sub-networks are required to harmonize in sharing identical parameters, encompassing weights and biases. The pictorial representation of the Deep Metric Learning based person Re-Id is presented in Figure 4.

The connection function is employed to derive the relationship between the results produced by the Convolutional Neural Networks (CNN), while the cost function is employed to convert the relationship into the cost. The connection function and the cost function are derived based on the parameters like Euclidean distance [10] between the samples, vector sample, cosine functions as defined in equations 5, 6 and 7 respectively.

$$f_{Euc}(x, y) = -\sum_{n=1}^{N}(x_n - y_n)^2 \tag{5}$$

$$f_v(x, y) = \sum_{n=1}^{N} w_n |x_n : yn| \tag{6}$$

$$f_{cos}(x, y) = \frac{\sum_{n=1}^{N} x_n y_n}{(\sqrt{\sum_{n=1}^{N} x_n y_n^T})\sqrt{\sum_{n=1}^{N} y_n x_n^T}} \tag{7}$$

where, the x_n and y_n are the features extracted from the CNN models. The $f_{Euc}(x, y)$, $f_V(x, y)$ and $f_{cos}(x, y)$ are the

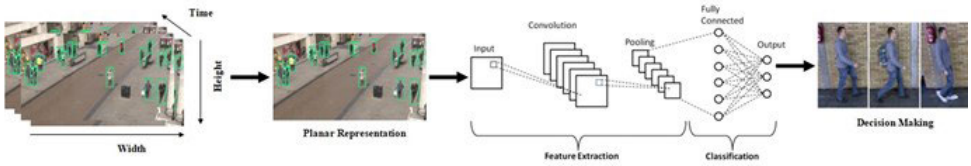

FIGURE 3. Spatio- temporal representation in person Re-ID.

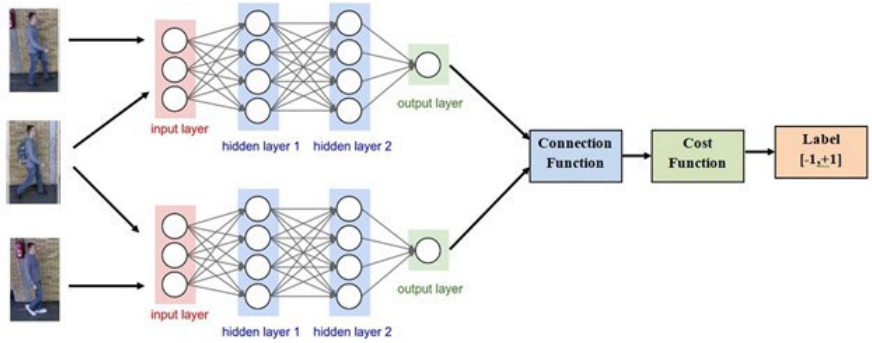

FIGURE 4. Deep metric learning based person Re-Identification.

Euclidean distance, vector sample and cost cosine functions respectively. The cosine function is employed as the cost function relating the frames and the labeling is done between the range of [−1,1].

C. GAIT RECOGNITION

Typically, appearance-dependent gait recognition techniques involve encoding pedestrian images using a sophisticated convolutional neural network and subsequently recognizing pedestrians by leveraging the acquired gait embeddings. These techniques can be categorized into three groups based on the input data: template-based methods, sequence-based methods, and silhouette-based methods. Template-dependent techniques utilize Convolutional Neural Networks (CNNs) for feature extraction from a solitary gait image, such as the Gait Energy Images (GEI) [11] or similar GEI-inspired template images. Wu et al. [12] introduced CNNs with diverse architectures aimed at enhancing the efficiency of cross-view gait recognition. In parallel, several generative models have emerged to convert gait images from one perspective to another, leveraging techniques like auto-encoders and Generative Adversarial Networks (GAN) [13].

The template-centric representation distills body movement through basic operations like the average GEI. Owing to the loss of intricate motion details in this abstraction, sequence-based or video-based methodologies have surfaced for a more thorough exploration of motion dynamics. Particularly, sequence-based strategies employ the silhouette of each frame instead of a singular template image. The prevalent models for temporal feature extraction in gait encompass Long Short Term Memory (LSTM) [14] and 3-Dimensional Convolutional Neural Network (3D-CNN) [15].

Approaches based on sets have exhibited cutting-edge performance on extensive gait datasets. It is posited that the silhouette's visual aspect serves as a substitute for temporal cues by encapsulating spatial data. The GaitSet framework employs a CNN to derive frame-level gait features from the silhouette, followed by a pooling operation to amalgamate them into a set-level feature. Analogous components can be identified in the Feature Map Pooling technique, the LSTM attention mechanism, and the Micro-motion Capture model. The gait based person Re-Id was performed based on the determination of Gait Energy Images (GEI) which is determined using equation 8.

$$G(h, w) = \frac{1}{N} \sum_{n=0}^{N-1} S_n(h, w) - B(h, w) \tag{8}$$

where, G(h,w) is the GEI of the image with dimension hxw, while Sn(h,w) is the silhouette of the input video sequence and B(h,w) is the background image to be subtracted. The Gait based person Re-Id is depicted in Figure 5.

The figure 5 is composed of the neural networks, which has to be fed with the time series data measured from the target person's gait, based on which the test data is analyzed.

D. ATTENTION MECHANISMS

The major objective of the attention based person Re-Id is to enhance the accuracy of the identification process. The attention mechanism has proven to be a powerful tool for enhancing the performance of deep convolutional neural networks. The evolution of attention mechanism can be dissected into two main components: (1) the merging of enriched features; (2) the incorporation of channels and spatial attention. In order to enhance feature fusion effectively,

a novel approach of second-order pooling is introduced. This method involves utilizing two-dimensional convolution with a kernel size of k*k to compute spatial attention, which is then merged with channel attention. It is evident that the aforementioned techniques achieve remarkable results through the intricate design of attention modules. In contrast to traditional attention mechanisms, our focus is on acquiring efficient channel attention while simplifying the model's complexity. Typically, channel attention is attained by transforming channel features into a lower-dimensional space and then reverting them, leading to indirect relationships between channels and their weights. Research indicates that alterations in channel dimensionality and inter-channel interactions can influence the efficacy of channel attention to a certain degree. Maintaining a constant channel dimension aids in learning effective attention, while cross-channel interactions further contribute to this process. Building upon the SE module, the Efficient Channel Attention (ECA) module streamlines the model's complexity and enhances learning efficiency by promoting local cross-channel interaction and parameters sharing among channels. The expression for determining the channel weights is presented as defined in equation 9.

$$C_w(i) = \delta \Sigma_{i=1}^{N} (x_i \alpha_i) \alpha^T \tag{9}$$

where, $C_w(i)$ is the weight of the channel, δ is a constant, xi is the parameter shared and α_i is the feature representation of the input video sequence. The weight calculation can be achieved through one-dimensional convolution using a kernel size of t. The importance of t becomes evident as it dictates the extent of interaction among channels. It is commonly accepted that a larger channel dimension results in a broader interaction range, while a smaller dimension leads to a more confined interaction range. It is logical to infer that the interaction range k is directly proportional to the channel dimension c. Although a linear function provides a basic mapping, its limitations are apparent. Typically, the channel dimension C is a power of 2. Therefore, this linear correlation can be expanded into a nonlinear one, as demonstrated in equation 10.

$$C = \Theta(t) = 2^{(\mu * t - \alpha)} \tag{10}$$

where C is the channel dimension, t represents the range of the channel interaction and is defined as [−1,1]. The term μ and 'a' are the feature point extracted in random from the input video sequence. The attention based person Re-Id is depicted in Figure 6.

The enhanced network elevates the semantic details of high-level feature maps during the process of extracting pedestrian characteristics, while also suppressing less relevant feature data, thus boosting the resilience of the person ReID feature extraction network. Displayed in Figure 6 is a segment of the network structure illustration of the attention module integrated into ResNet50. Following three convolution operations of the main network, the attention network assigns weights to the feature map, which is then combined with the original feature map using the ''Eltw sum'' technique. ''Eltw sum'' denotes the addition operation of feature maps within the corresponding channel. Subsequently, the resulting new feature map is generated and transmitted to the subsequent layer through the activation function. Furthermore, this study adopts the Leaky ReLU activation function as a replacement for the original ReLU activation function utilized in the ResNet backbone network. While the ReLU activation function sets negative values to zero, potentially resulting in the loss of certain features, Leaky ReLU assigns a non-zero slope to negative values, serving as a complement to the ReLU function and effectively enhancing the extraction of pedestrian characteristics.

E. LOSS FUNCTIONS

The loss function plays a vital and pivotal role in distinguishing the acquired traits. Generally, the softmax loss segregates the acquired traits instead of distinguishing them. The primary objective of formulating a person re-ID loss function is to elevate the representation with an effective loss. This subsection sheds light on some of the most impactful loss functions for video re-ID. The different loss functions to be taken into account during the process of person re-identification include attention and CL loss, Weighted Triple Sequence Loss, Symbolic triplet loss, Weighted contractive loss, Triplet loss, and Regressive Pairwise loss.

The CL loss employs the center vectors as a label for classes representing the high power noise signals. The high power noisy signals possess high variance, in turn penalizes the re-Id model. The CL is mathematically depicted in equation 11.

$$CL = \frac{1}{N} \Sigma_{i=1}^{N} label(n) XA_{score}(n) \tag{11}$$

where A_{score} is the attention score, while the label is to be in the range of 0 to 1 for N number of frames.

The Weighted Triple Sequence Loss (WTSL) is a frame based loss which deteriorates the effects of the outline of image frame. The loss is intra class in nature and makes similar video to appear closed and pushes dissimilar videos creating reduction in the accuracy of the person Re-Id. The WTSL is mathematically represented in equation 12.

$$L_{WTS} = \Sigma_{n=1}^{N} ||f_x^n - f_y^n||^2 + ||f_x^n - f_y^n||^2 + \beta \tag{12}$$

where, f_x represents the closer feature to the centroid and f_y represents the far away feature from the centroid.

The symbolic triplet loss defines the representation concern, creates due to the distance of separation between the feature vectors as defined in equation 13.

$$L_{ST}(\omega_i, \omega_i) = \Sigma_{i=1}^{N} \Sigma_{j=1}^{M} \omega_i^{-1}(n) - \omega_j^{-1}(n) \tag{13}$$

where, w_i and w_j are the vectors of the multi-dimensional features of the input image.

The weighted contrastive loss in a loss function, which allocates necessary weight to the imaging pair. Weights are seamlessly woven into the fabric of the contrastive

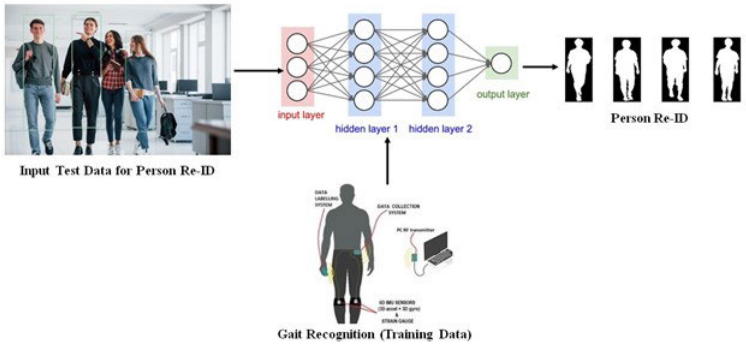

FIGURE 5. Gait based person Re-Identification process.

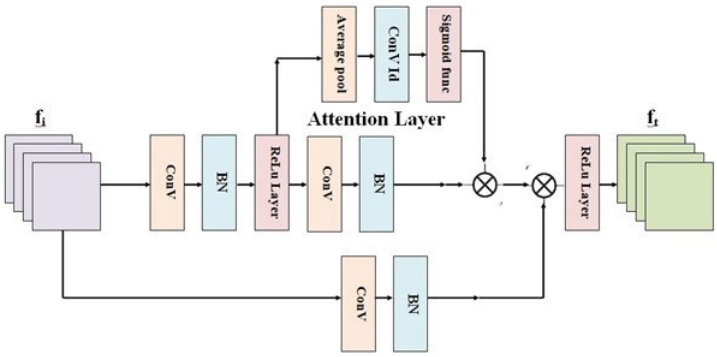

FIGURE 6. Attention based person Re-Identification process.

loss function by WCL. The significance of individual data points or comparisons is carefully taken into account by the loss function during the computation of the overall loss. The weighted contrastive loss is mathematically depicted in equation 14.

$$L_{wc}(n) = \frac{1}{N} \frac{\Sigma_{n=1}^{N} n\omega_n^T max(0, \alpha)}{\Sigma_{n=1}^{N} n\omega_n^T} \tag{14}$$

where, the w_n^T represents the contribution weightage of positive and negatives sets of the contrastive losses. The triplet loss is the another significant loss function which is used to conserve the ranking relationship among the features of the input video sequence. The area of separation between the featured pairs of the similar patterns are measured to determine the triplet losses as defined in equation 15.

$$L_T = \Sigma_{m,n=1}^{M,N} d_s(m.n) - d(m, n) + T_m \tag{15}$$

where, the T_m is the threshold margin $d_s(m, n)$ and d(m,n) are the similar features in the input video sequence.

The Regressive Pairwise Loss (RPL) is the loss function which is considered to enhance the pairwise similarity among the positive sets of the image features. The RPL concentrates on the soft margin of the positive sets of features and is mathematically represented in equation 16.

$$\begin{aligned} LRP = n.max\{d(x_m, y_n) - log\alpha, 0\} \\ + (1-2)max\{\alpha - d(x_m, y_n), 0\} \end{aligned} \tag{16}$$

where $d(x_m, y_n)$ is the margin of the positive set of the feature extracted from the input sequence. These are the major loss functions considered during the person Re-Id process.

III. LITERATURE REVIEW

The person re-identification holds the significant application in identifying the person from the mass crowd through the video sequence captured through the CCTV cameras. Due to its significance, many researchers had actively involved in determining solution and had introduced numerous novel methodologies and algorithms for an enhanced level of security. Despite fruitful research results, the pitfalls still exists in the domain of person Re-Id, motivates to perform this literature review and its analysis. This review analysis examines the existing methodologies and enlists the observations in terms of technology used and the performance exhibited by the notable research results.

Qian and Tang [16] had proposed a pose attention based person Re-Id process using the guided pair of images captured from the surveillance cameras. The proposed PAPG is capable of generating the cross modality of the input image pair and to produce the consistency among the real image. In addition, the proposed work alleviated the inadequate data to reduce the risk of overfitting concern. Gwon et al. [17] proposed a new well-balanced strategy between Modality-Specific and Modality-Shared approaches, incorporating an efficient transformation method to cater to a diverse range

of instances. Initially, they suggested the Informative Weighted Gray transform (IWG), which strives to enhance the variety of instances by producing unique combinations of RGB colors. Subsequently, the introduction of the Customized Modality-Specific Enhanced Module (CMSpEM) enhances feature maps by employing an attention mechanism between pre-pooling and post-pooling features, reinforcing specific features for the modality-shared network with minimal parameters. Lastly, they presented the Pseudo Label-oriented Modality-Specific (PLMSp) Loss, which aids in explicit representation learning to diminish the modality gap by utilizing pseudo labels as reference points.

Hou et al. [18] introduced a three-phase framework for video-based VI-ReID to maximize the utilization of identity information present in various modality data, termed as the Decomposition-Mining-Aggregation framework. Our framework comprises three key stages: decomposition, mining, and aggregation. Experimental findings indicate that the amalgamation of these stages effectively enhances the capability of the feature extraction network to extract identity information from images with modality discrepancies. Bian et al. [19] put forward a fresh Occlusion-Aware Feature Recover (OAFR) model. OAFR replicates diverse occlusions to assist the model in perceiving OTP, PTP occlusions, and recovering occluded query features with unoccluded gallery features. Specifically, the Prior Knowledge-based Occlusion Simulation method is initially introduced to generate OTP, PTP occlusions, and corresponding occlusion labels, bolstering the model's ability in person perception and occlusion-awareness through self-supervised learning.

He et al. [20] introduced a Region Generation and Assessment Network (RGANet) to proficiently detect human body regions and emphasize the significant areas. In the proposed RGANet, a Region Generation Module (RGM) is devised first, leveraging pre-trained CLIP to pinpoint human body regions using semantic prototypes derived from textual descriptions. A learnable prompt is devised to bridge the domain gap between CLIP datasets and ReID datasets. Bai et al. [21] utilized the normalized IBN-Net network as the foundational structure of the ResNet50 network. Subsequently, the integration of the SimAM attention mechanism into the foundational network, an attention mechanism for inter-modal fusion primarily used in multimodal data processing tasks, aids in learning spatial attention weights in the person images to acquire person information with more distinctive features. Finally, supervised learning is conducted utilizing the cross-entropy loss function during training in the source domain.

Pang et al. [22] utilized a novel Cross-Modality Hierarchical Clustering And Refinement (CHCR) technique to enhance modality-invariant feature learning and enhance the reliability of pseudo-labels. Unlike traditional VI-ReID methods, CHCR does not depend on manual identity annotation or intra-modality initialization. Initially, the proposal involves designing a straightforward and efficient cross-modality clustering baseline that clusters across modalities.

Huang et al. [23] presented an innovative concept of visibility scores acting as the graph's focus, directing the Graph Convolutional Network (GCN) to delicately eliminate the disturbances of concealed features while transferring essential semantic details from the overall picture to the obscured one. The outcomes of our experiments on obscured benchmarks highlight the effectiveness of our approach. Zhang et al. [24] introduced a fresh framework known as Dual-Semantic Consistency Learning Network (DSCNet), attributing discrepancies in modality to the inconsistency in semantic details at the channel level. DSCNet enhances channel coherence by addressing fine-grained inter-channel semantics and comprehensive inter-modality semantics.

Zhu et al. [25] innovated a new framework named Dual Knowledge Distillation on Multiview Pseudo Labels (DKD-MPL) to tackle the challenge posed by pseudo labels. The DKD-MPL framework comprises two key modules: Global Knowledge Distillation (GKD) and Self-Knowledge Distillation (SKD). Zhang et al. [26] introduced a unique UDA technique through Dynamic Clustering and Co-segment Attentive Learning (DCCAL). DCCAL integrates a Dynamic Clustering (DC) module and a Co-segment Attentive Learning (CAL) module. The DC module dynamically clusters pedestrians during various stages to alleviate inaccurate labels, while the CAL module diminishes the domain gap via a co-segmentation-based attention mechanism. Moreover, we incorporate Kullback-Leibler (KL) divergence loss to enhance performance by minimizing feature distribution disparities between the two domains.

Yang et al. [27] devised a dual consistency-constrained learning framework (DCCL) that simultaneously encompasses off-line refinement of cross-modality labels and on-line interaction learning of features. The core concept revolves around maintaining the consistency between cross-modality instance-instance and instance-identity relationships. DCCL establishes an instance memory, an identity memory, and a domain memory for each modality. At the start of each training epoch, DCCL explores the off-line coherence of cross-modality instance-instance and instance-identity similarities to enhance the credibility of cross-modality identities. Liu et al. [28] developed a weakly supervised tracklet association learning (WS-TAL) model solely based on video labels. Initially, we introduce an intra-bag tracklet discrimination learning (ITDL) term.

Lu et al. [29] introduced an Illumination Distillation Framework (IDF) that employs illumination enhancement and distillation strategies to facilitate the training of Re-ID models. IDF includes a primary branch, an illumination enhancement branch, and an illumination distillation module. Wei et al. [30] proposed a novel Dual-Adversarial Representation Disentanglement (DARD) model to segregate modality-specific attributes from intertwined pedestrian features and effectively acquire resilient modality-invariant representations. Our approach leverages dual-adversarial learning, integrating image-level channel exchange and

feature-level magnitude adjustment to introduce diversity in modality-specific representations.

Chen et al. [31] introduced a Hierarchical Attention-aware Spatio-temporal Interaction (HASI) network, featuring an Attention-aware Temporal Interaction (ATI) module and a Hierarchical Local-spatial Enhancement (HLE) module designed for person re-identification in video sequences. Zheng et al. [32] introduced a Visionary Time-Traveling Network for Garment Transformation Person Identification, consisting of three pivotal elements. 1) Insightful Topology Discovery Network: By merging knowledge graphs and convolution networks, this network captures the essence of space and time among camera nodes. Infusing knowledge embedding into the convolution network enhances effective topology discovery; 2) Garment Transformation Network Across Time Periods: This network amalgamates spatial-temporal details for attire creation. It leverages overall pedestrian clothing traits within camera topologies to reduce matching errors from external factors; and 3) Collaborative Enhancement Mechanism: A unified strategy involving both the topology discovery network and garment transformation network.

Yang et al. [33] proposed a Spatial-Temporal Feature Enrichment (STFE) network from a comprehensive space-time perspective, merging the strengths of the aforementioned methods to extract more holistic insights from video fragments. STFE comprises two primary modules: Space Feature Projection Module (SFPM) and Global Low-frequency Amplification Module (GLAM).

Wu et al. [34] crafted a versatile Adaptive Style Harmonization (ASH) module to eliminate intrinsic style distractions while preserving distinctive content through dual-level adaptive distribution adjustment and discriminative enhancement. Furthermore, acknowledging the significance of a suitable modality mediator in conveying pertinent information on inter-modality distribution shift, we introduce Mutual Modality Connection Learning (MMCL) to refine the modality connection process.

Xuan and Zhang [35] propose the Exemplar and Camera Tone Balancing (ECTB) to reinforce resilience against domain variations. ECTB alleviates intra-camera fluctuations by dynamically learning a blend of exemplar and batch balancing. ECTB also enhances resistance to inter-camera variations through TNorm, transforming the original feature style into target styles. Liu et al. [36] presented a discerning identity-feature exploration and differential-conscious learning (IDEAL) framework to cultivate more distinctive intra-identity portrayals. Particularly, IDEAL extracts noteworthy intra-instance features through synthetic complementary attention, and further delves into the distinctive identity features by mapping the correlation among these noteworthy features using graph neural networks.

Xiahou et al. [37] introduced a multi-perspective fusion model, which seeks to expand the Re-ID task query approach from single to multi-viewpoints to tackle diverse perspectives and delve into inherent gallery insights for query refinement. Nguyen et al. [38] unveiled AG-ReID v2, a dataset tailored for person Re-ID in mixed aerial and ground contexts. This dataset encompasses 100,502 images of 1,615 distinct individuals, each tagged with matching IDs and 15 soft attribute descriptors. Data were gathered from varied viewpoints using a UAV, stationary CCTV, and smart glasses-integrated camera, offering a diverse array of intra-identity nuances.

Han et al. [39] proposed an innovative concept, one-shot self-supervised cross-domain for person ReID and investigated the adaptability using minimal images in the target domain during training. Primarily, the author introduced a pioneering Collective Normalization (CN) based domain-flexible ReID model. Cui et al. [40] suggested a unique Dual Modality-conscious Alignment (DMA) model for VI-ReID, capable of upholding distinctive identity details and suppressing misleading information within a unified framework.

Tao et al. [41] unveiled an innovative Visionary Enriched Collaborative Tutoring system (VECT) to tackle the aforementioned issue. Harnessing the core mutual-mean-teaching concept, VECT introduces two groundbreaking strategies: the utilization of GAN for source domain enhancement (GSDA) and cross-branch collaborative oversight (CBCO) for paired networks. Tan et al. [42] introduced a cutting-edge identity recognition framework, dubbed as multihead self-focus network (MHSFN), aimed at refining essential details while capturing crucial local insights from individual visuals. MHSFN incorporates two pivotal components: multihead self-focus segment (MHSFS) and attention rivalry mechanism (ARM). Ma et al. [43] devised a dual-branch multi-tier feature integration system, enhancing the interpretative capacity of pedestrian attributes amidst partial obstructions by discerning distinctive traits. By integrating a streamlined attention module into Residual neural network-50 (Resnet-50), the system extracts image sequence attributes from the dimensional spectrum while suppressing cluttered background interference.

Zhao et al. [44] introduced an ingenious Content-Adaptive Auto-Concealment System (CAACS), capable of dynamically selecting optimal concealment zones within an image based on content and ongoing training phase. CAACS comprises two key elements: the Recognition Identification (RI) system and Concealment Adjustment Controller (CAC) unit. Li et al. [45] proposed an innovative Dual-Focus Biometric-Garment Fusion System (DFBGFS) for CTCC-ReID, proficient in extracting biometric traits from primary images and garment features from templates, merging them via a Dual-Focus Fusion Module. Yu et al. [46] introduced a novel dual-phase UDA strategy named Guiding And Refining (GAR) comprising a preliminary guidance phase and a subsequent refinement phase. Initially, GAR adapts the model from source to target domain by acquiring prompts for both domains, followed by fine-tuning the complete framework to enhance accuracy.

Zheng et al. [47] proposed a straightforward yet potent method to combat this challenge. Given an image, our approach adeptly identifies possible proxies, introducing credibility to gauge the significance of treating each proxy centroid as a positive cue rather than purely negative. Extensive trials on three prevalent person re-ID datasets affirm the efficacy of our devised strategy. Mao et al. [48] devised an inventive attention map-driven (AMD) transformer refinement technique, eliminating surplus tokens and heads with attention map guidance in a hardware-friendly manner. Peng et al. [49] introduced an Adaptive Memorization Strategy with Clustered tags (AMSC) framework for unsupervised person re-ID, resistant to erratic labels, leveraging sample diversity via a multi-faceted structure with adaptive memorization. Qian et al. [50] explored a dual-space amalgamation learning (DSAL) technique amalgamating instance-batch normalization (IBN) and residual contraction (RC) into a foundational model for feature enhancement and channel-level compression. The abstracted difference observed from the performed literature review is tabulated in Table 1.

IV. LITERATURE ANALYSIS

Based on the above performed literature review, the following challenges have been identified in the person Re-Identification process. Current Re-identification techniques, particularly those utilizing supervised learning, demand a substantial volume of labeled data to undergo training. It is crucial that this data is meticulously annotated with individuals identities across various camera perspectives.

The performance of the existing state of the art methodologies have been analyzed in terms of accuracy, precision, recall and F score. Accuracy in person identification (Re-ID) pertains to the system's ability to accurately recognize the same individual across various cameras or images. It quantifies how frequently the system retrieves the accurate matching image for a specific query image. Rank-1 Accuracy serves as a yardstick showing the proportion of query images where the system places the correct matching image as the top result (rank 1) in the retrieval list. Mean Average Precision (mAP) is a measure that takes into consideration the entire retrieval list and assesses how many relevant images are retrieved at each rank position, offering a more thorough gauge of precision. It is common for researchers to present both rank-1 accuracy and mAP in order to assess the effectiveness of Re-ID models. The mathematical relationship among the true and false prediction to determine the accuracy is defined in equation 17.

$$Accuracy(Accu_y) = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

where TP is the True Positive, TN is the True Negative, FP is the False Positive and FN is the False Negative. In the realm of individual re-identification (Re-ID), precision plays a specific role within the broader framework of Mean Average Precision (mAP). This element does not operate independently to gauge Re-ID accuracy. Precision, nestled in mAP, delves into the significance of retrieved images at a particular rank in the sequence. It unveils the ratio of relevant images retrieved at a specific rank that genuinely correspond (i.e., portray the same individual as the query image).

Equation 18 defines the mathematical relation for determining the precision during the person re-identification process.

$$Precision(Prec_n) = \frac{TP}{TP + FP} \tag{18}$$

In the domain of individual re-identification (Re-ID), recall stands as a yardstick utilized to evaluate the efficacy of a system in fetching all pertinent images of an individual from various cameras or snapshots. It spotlights how many actual manifestations of a specific individual the system successfully identifies, regardless of their positioning in the retrieval lineup. This yardstick scrutinizes the complete retrieval lineup. It discloses the percentage of all pertinent images (images of the same individual) the system retrieves, irrespective of their order. The mathematical equation for determining the recall is presented in equation 19.

$$Recall(Reca_l) = \frac{TP}{TP + FN} \tag{19}$$

F1 Score emerges as a metric that serves as a harmonious blend of precision and recall, striking a balance between the number of relevant items retrieved (recall) and the number of retrieved items that truly hold significance (precision). This metric frequently finds application in classification assignments where distinct positive and negative classes are discernible. Equation 20 presents the mathematical relationship among true and false prediction in determining F score.

$$F_{Score} = \frac{2X(Prec_n XReca_l)}{Prec_n + Reca_l} \tag{20}$$

The comparison of the performance exhibited by the state of the art (SOTA) methodologies has been compared and is presented in the Figure 7.

The Market-1501 dataset is widely used for training and testing the system for person Re-Id process. Is it composed of 32,668 images and most systems are employing this dataset. The average accuracy, precision, recall and F score achieved by the person Re-Id system using this market-1501 dataset are, 77.50%, 75.06%, 73.53% and 74.29% respectively. In continuation of this, the state of art methods were analyzed using CUHK-03 dataset and the observed values are shown in Figure 8.

CUHK-03 is a dataset similar to the Market-1501 dataset, composed of 14,097 images for person Re-Id process. The performance exhibited by the existing systems in performing person Re-Id using CUHK-03 is shown in Figure 8 and proves that the performance is slightly lower than that of exhibited by the Market-1501 dataset. The systems provided an average accuracy of 70.13%,

TABLE 1. Observation on person re-identification models and its performance levels.

Ref.No.	Author and Year	Technology Used	Dataset	Performance
[16]	Y.Qian (2024)	Attention based method	Market-1501	Accuracy: 77.87%
[17]	S. Gwon (2024)	Modality-Shared Representations	Market-1501	Accuracy: 81.21%
[18]	W. Hou (2024)	Decomposition-Mining- Aggregation framework	CUHK03	Enhancement: +3.14%
[19]	Y.Bian (2024)	Occlusion-Aware Feature Recover Model	MS COCO	MAP score: 2.2%
[20]	S.He (2024)	RGANet Model	Market-1501	Error Rate: 11.91%
[21]	X.Bai (2024)	IBM-Net	PRID 2011	Accuracy: 86.6% Precision: 68.7%
[22]	Z.Pang (2024)	Cross-Modality Hierarchical Clustering And Refinement	Market-1501	Error Rate: 12.35%
[23]	M.Huang (2023)	Reasoning and Tuning Graph Attention Network	Market-1501	Accuracy: 79.61%
[24]	Y.Zhang (2023)	Dual-Semantic Consistency Learning Network	Market-1501	Accuracy: 73.89%
[25]	W.Zhu (2024)	Global Knowledge Distillation	Market-1501	Accuracy: 72.19%
[26]	F.Zhang (2024)	Unsupervised Domain Adaptation	Re-ID dataset	Enhancement: 3.1%
[27]	B.Yang (2024)	dual consistency-constrained learning framework	Market-1501	Accuracy: 73.48%
[28]	M.Liu (2024)	Intra-bag tracklet discrimination learning	CUHK03	Accuracy: 88.1%
[29]	A.Lu (2024)	Illumination Distillation Framework	Night600	Accuracy: 73.48%
[30]	Z.Wei (2024)	Dual-Adversarial Representation Disentanglement	Self dataset	Accuracy: 72.62%
[31]	S.Chen (2024)	Hierarchical Attention-aware Spatio-temporal Interaction	MARS, iLIDS-VID, and PRID-2011	Accuracy: 79.13%
[32]	S.Zheng (2024)	Knowledge-Driven Cross-Period Network	Celeb-ReID, PRCC, UJS-ReID, SLP, and DukeMTMC-ReID	Accuracy: 79.23%
[33]	X.Yang (2024)	Spatio-Temporal Feature Enhancement	MARS	Accuracy: 95.5%
[34]	J.Wu (2024)	Cross Modality Blending	STAR	Accuracy: 73.03%
[35]	S.Xuan (2024)	Instance and Camera Style Normalization (ICSN)	MSMT17	Accuracy: 64.4%
[36]	Y.Liu (2024)	Discriminative Identity-Feature Exploring And Differential Aware Learning (DiDAL)	MARS	Accuracy: 74.21%
[37]	Y.Xiahuo (2024)	Identity consistency pose transfer framework	Market-1501 DukeMTMC-reid CUHK03-labeled	Accuracy: 87.97%
[38]	H.NguYen (2024)	AG-ReID.v2	MSMT17	Accuracy: 73.49%
[39]	G.Han (2024)	Unsupervised Cross domain	Market-1501	Accuracy: 71.41%
[40]	Z.Cui (2024)	Dual Modality-conscious Alignment (DMA) model	VI-Re-Id	Accuracy: 76.07%
[41]	Y.Tao (2023)	DiveRsityEnlArged Mutual Teaching framework (DREAMT)	Market-1501	Accuracy: 78.17%
[42]	H.Tan (2023)	Multihead self-attention network (MHSA-Net)	Own Dataset	Accuracy: 81.06%
[43]	X.Ma (2023)	Double stream method	Re-ID dataset	Accuracy: 79.41%
[44]	C.Zhao (2023)	Content-Adaptive Auto-Occlusion Network (CAAO)	Re-ID dataset	Accuracy: 80.46%
[45]	S.Li (2023)	DualBCT-Net	Market-1501	Accuracy: 76.24%
[46]	S.Yu (2023)	Two stage UDA	Re-ID dataset	Accuracy: 78.26%
[47]	D.Zheng (2023)	DBSCAN	Re-Id dataset	Accuracy: 71.25%
[48]	J.Mao (2023)	Attention mapped method	ViT-Base	Accuracy: 29.4%
[49]	J.Peng (2023)	Adaptive Memorization model	Re-Id dataset	Accuracy: 74.48%
[50]	Y.Qian (2023)	Dual space aggregate learning	Re-Id dataset	Rank 1: 16.47%

precision of 68.27%, 64.51% of recall and 66.33% of F Score. This reduction is performance percentage is due to the volume of the Market-1501 is huge when compared with the CUHK dataset. In alteration to these universal datasets, most systems employed their own datasets for training and testing the designed model. In this way, the performance exhibited by the systems using local dataset were listed and depicted in the Figure 9.

The local dataset is composed by the researchers based on their necessity and with restricted volume of images. The analysis of the models using local dataset, yields an average accuracy of 78.07%, 76.34% of precision, 74.09% of recall and a F score with 75.19%. This motivates most researchers to create and employ own dataset for their models, in a wish to achieve enhanced level of performance.

The process of labeling data is both costly and time-intensive, thereby constraining the scalability of these techniques in real-world scenarios featuring diverse camera configurations. Challenges arise for existing techniques when

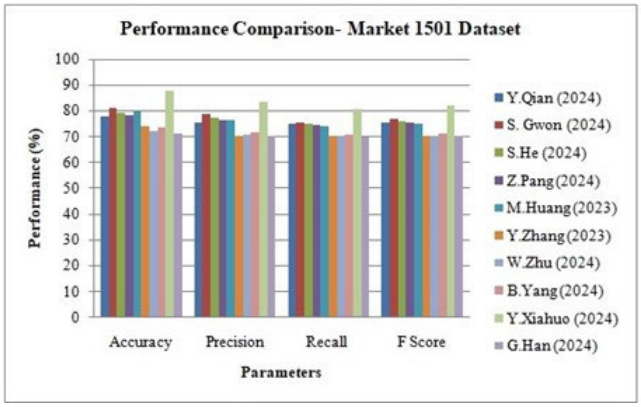

FIGURE 7. Performance comparison of existing works with Market-1501 dataset.

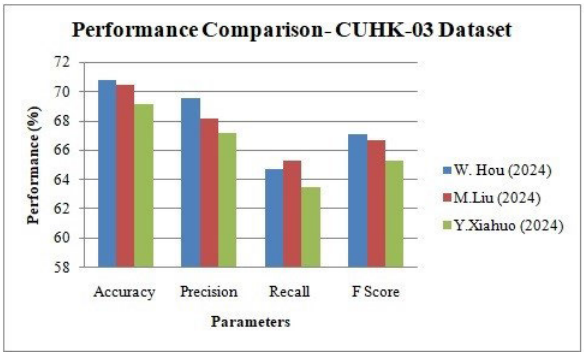

FIGURE 8. Performance comparison of existing works with CUHK-03 dataset.

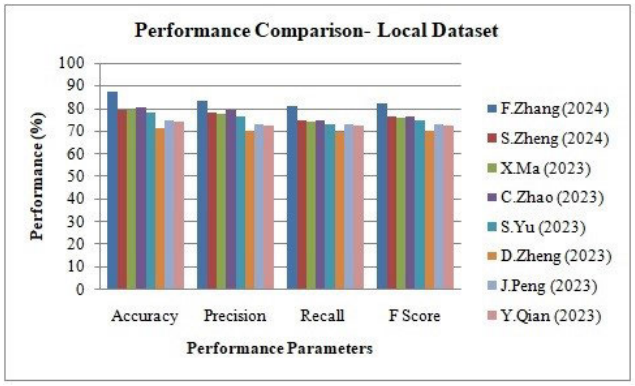

FIGURE 9. Performance comparison of existing works with local dataset.

confronted with factors commonly found in real-world settings that can greatly influence image quality. Variations in lighting, posture, viewpoint, and background distractions can pose difficulties for algorithms to effectively match characteristics across different cameras. Instances where a person's body parts are concealed by items like bags, umbrellas, or other objects further complicate the extraction of dependable features for identification. Although prevailing Re-identification models tend to excel on the datasets they were trained on, transferring this performance to unfamiliar scenarios can prove to be daunting. This is primarily due to datasets potentially failing to encompass the full spectrum of variations observed in real-world settings. The utilization of Re-identification technology gives rise to privacy apprehensions, particularly when implemented in public areas. There exists a necessity for approaches that strike a balance between the security advantages and the safeguarding of individuals' identities.

The drawbacks identified in terms of performance and security in the person Re-Id motivates the following scope of research works.

- **Unsupervised and Weakly Supervised Learning:** The current focus revolves around methodologies that demand reduced amounts of annotated data. This encompasses harnessing unlabeled data or data with fainter annotations (such as bounding boxes around individuals) to train Re-ID models.
- **Domain Adaptation and Generalizability:** This pertains to crafting algorithms capable of adjusting to novel camera perspectives and settings not encountered during the training phase. Such advancements will enhance the resilience and practicality of Re-ID systems in real-world scenarios.
- **Handling Variations:** Scholars are devising approaches to fortify Re-ID models against fluctuations in posture, lighting, viewpoint, and background commotion. This could entail methods for standardization, attention mechanisms, or integrating supplementary details like temporal hints from videos.
- **Occlusion Handling:** Novel techniques are striving to extract dependable features even when segments of a person are concealed. This approach might entail employing generative models to fill in obscured regions or concentrating on features less susceptible to obstruction such as walking patterns.
- **Low-Resolution Re-ID:** The objective is to enhance Re-ID accuracy on images taken by low-quality cameras. This could involve utilizing super-resolution methods to refine image clarity or formulating models tailored specifically for low-resolution data.
- **Privacy-by-Design Techniques:** This pertains to creating Re-ID models that inherently provide heightened privacy assurances. This could encompass anonymization methods or mastering representations that safeguard a person's identity.
- **Explainable AI for Re-ID:** Constructing Re-ID models capable of elucidating their decision-making process can foster trust and transparency in their utilization.

V. CONCLUSION

Person Re-identification (Re-Id) holds the significant position as it plays a major role in criminal investigation, identifying missing persons, crowd and traffic management etc. In this way, technology plays a vital role in performing the person Re-Id using neural networks. Various methodologies have been introduced in performing the person Re-Id process

more effectively. Accuracy is the major parameter considered in the person re-Id process and this manuscript performs the comparative analysis of the existing methodologies and algorithms of the person Re-Id process. In addition, this article performs the comparison of various architectures and loss functions involved in this re-Id domain. The manuscript is concluded with the existing challenges in this research domain and the scope of research to enhance the performance level along with the security in the person Re-Id process.

REFERENCES

[1] M. Ye and P. C. Yuen, "PurifyNet: A robust person re-identification model with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2655–2666, 2020.

[2] K. Eguchi, T. Inoue, H. Zhu, S. Terada, and F. Ueno, "Design of a charge-pump type AC-DC converter for RE-ID tags," in *Proc. Int. Symp. Commun. Inf. Technol.*, Bangkok, Thailand, Oct. 2006, pp. 1203–1206.

[3] P. Zhu, L. Tian, and Y. Cheng, "Improvement of defect feature extraction in eddy current pulsed thermography," *IEEE Access*, vol. 7, pp. 48288–48294, 2019.

[4] S. Khan, S. A. AlQahtani, S. Noor, and N. Ahmad, "PSSM-Sumo: Deep learning based intelligent model for prediction of sumoylation sites using discriminative features," *BMC Bioinf.*, vol. 25, p. 284, Aug. 2024.

[5] D. H. Kim, W. J. Baddar, J. Jang, and Y. M. Ro, "Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 2, pp. 223–236, Apr. 2019.

[6] M. R. Desai, S. A. Patel, M. Peerzade, and G. Chawhan, "Person re-identification via deep metric learning," in *Proc. 3rd Int. Conf. Adv. Electron., Comput. Commun. (ICAECC)*, Bengaluru, India, Dec. 2020, pp. 1–9.

[7] C. Benedek, B. Gálai, B. Nagy, and Z. Jankó, "LiDAR-based gait analysis and activity recognition in a 4D surveillance system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 101–113, Jan. 2018.

[8] J. Wang, K. Jiang, Z. Lu, X. Lu, and Z. She, "Cross-modality person re-identification method using cross-dimensional interactive attention mechanism," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Rome, Rome, Italy, Sep. 2023, pp. 95–99.

[9] D. Chen, S. Zhang, W. Ouyang, J. Yang, and Y. Tai, "Person search by separated modeling and a mask-guided two-stream CNN model," *IEEE Trans. Image Process.*, vol. 29, pp. 4669–4682, 2020.

[10] F. Huang, Y. Zhu, and Z. Zhou, "Irregular Euclidean distance constellation design for quadrature index modulation," *IEEE Commun. Lett.*, vol. 27, no. 11, pp. 2928–2932, Nov. 2023.

[11] S. Zhang, Y. Wang, and A. Li, "Gait energy image-based human attribute recognition using two-branch deep convolutional neural network," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 5, no. 1, pp. 53–63, Jan. 2023.

[12] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Feb. 2017.

[13] A. Dash, J. Ye, and G. Wang, "A review of generative adversarial networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing," *IEEE Access*, vol. 12, pp. 18330–18357, 2024.

[14] X. Zheng, Y. Zhao, B. Peng, M. Ge, Y. Kong, and S. Zheng, "Information filtering unit-based long short-term memory network for industrial soft sensor modeling," *IEEE Sensors J.*, vol. 24, no. 8, pp. 13530–13544, Apr. 2024.

[15] H. Lee, Y.-S. Kim, M. Kim, and Y. Lee, "Low-cost network scheduling of 3D-CNN processing for embedded action recognition," *IEEE Access*, vol. 9, pp. 83901–83912, 2021.

[16] Y. Qian and S.-K. Tang, "Pose attention-guided paired-images generation for visible-infrared person re-identification," *IEEE Signal Process. Lett.*, vol. 31, pp. 346–350, 2024.

[17] S. Gwon, S. Kim, and K. Seo, "Balanced and essential modality-specific and modality-shared representations for visible-infrared person re-identification," *IEEE Signal Process. Lett.*, vol. 31, pp. 491–495, 2024.

[18] W. Hou, W. Wang, Y. Yan, D. Wu, and Q. Xia, "A three-stage framework for video-based visible-infrared person re-identification," *IEEE Signal Process. Lett.*, vol. 31, pp. 1254–1258, 2024.

[19] Y. Bian, M. Liu, X. Wang, Y. Tang, and Y. Wang, "Occlusion-aware feature recover model for occluded person re-identification," *IEEE Trans. Multimedia*, vol. 26, pp. 5284–5295, 2024.

[20] S. He, W. Chen, K. Wang, H. Luo, F. Wang, W. Jiang, and H. Ding, "Region generation and assessment network for occluded person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 120–132, 2024.

[21] 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.

[22] Z. Pang, C. Wang, L. Zhao, Y. Liu, and G. Sharma, "Cross-modality hierarchical clustering and refinement for unsupervised visible-infrared person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2706–2718, Apr. 2024.

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

[24] Y. Zhang, Y. Kang, S. Zhao, and J. Shen, "Dual-semantic consistency learning for visible-infrared person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 1554–1565, 2023.

[25] W. Zhu, B. Peng, and W. Q. Yan, "Dual knowledge distillation on multiview pseudo labels for unsupervised person re-identification," *IEEE Trans. Multimedia*, vol. 26, pp. 7359–7371, 2024.

[26] F. Zhang, F. Chen, Z. Su, and J. Wei, "Unsupervised domain adaptation via dynamic clustering and co-segment attentive learning for video-based person re-identification," *IEEE Access*, vol. 12, pp. 29583–29595, 2024.

[27] B. Yang, J. Chen, C. Chen, and M. Ye, "Dual consistency-constrained learning for unsupervised visible-infrared person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 1767–1779, 2024.

[28] M. Liu, Y. Bian, Q. Liu, X. Wang, and Y. Wang, "Weakly supervised tracklet association learning with video labels for person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 5, pp. 3595–3607, May 2024.

[29] A. Lu, Z. Zhang, Y. Huang, Y. Zhang, C. Li, J. Tang, and L. Wang, "Illumination distillation framework for nighttime person re-identification and a new benchmark," *IEEE Trans. Multimedia*, vol. 26, pp. 406–419, 2024.

[30] Z. Wei, X. Yang, N. Wang, and X. Gao, "Dual-adversarial representation disentanglement for visible infrared person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2186–2200, 2024.

[31] S. Chen, H. Da, D.-H. Wang, X.-Y. Zhang, Y. Yan, and S. Zhu, "HASI: Hierarchical attention-aware spatio-temporal interaction for video-based person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 6, pp. 4973–4988, Jun. 2024.

[32] S. Zheng, S. Liang, C. Meng, Z. Zhang, and L. Luan, "A cross-period network for clothing change person re-identification," *IEEE Access*, vol. 12, pp. 53517–53532, 2024.

[33] X. Yang, X. Wang, L. Liu, N. Wang, and X. Gao, "STFE: A comprehensive video-based person re-identification network based on spatio-temporal feature enhancement," *IEEE Trans. Multimedia*, vol. 26, pp. 7237–7249, 2024.

[34] J. Wu, H. Liu, W. Shi, M. Liu, and W. Li, "Style-agnostic representation learning for visible-infrared person re-identification," *IEEE Trans. Multimedia*, vol. 26, pp. 2263–2275, 2024.

[35] S. Xuan and S. Zhang, "Intra-inter domain similarity for unsupervised person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1711–1726, Mar. 2024.

[36] Y. Liu, H. Ge, Z. Wang, Y. Hou, and M. Zhao, "Discriminative identity-feature exploring and differential aware learning for unsupervised person re-identification," *IEEE Trans. Multimedia*, vol. 26, pp. 623–636, 2024.

[37] Y. Xiahou, N. Li, and X. Li, "Identity consistency multi-viewpoint generative aggregation for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 3, pp. 1441–1455, Mar. 2024.

[38] H. Nguyen, K. Nguyen, S. Sridharan, and C. Fookes, "AG-ReID.v2: Bridging aerial and ground views for person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2896–2908, 2024.

[39] G. Han, X. Zhang, and C. Li, "One-shot unsupervised cross-domain person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 3, pp. 1339–1351, Mar. 2024.

[40] Z. Cui, J. Zhou, and Y. Peng, "DMA: Dual modality-aware alignment for visible-infrared person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2696–2708, 2024.

[41] Y. Tao, J. Zhang, J. Hong, and Y. Zhu, "DREAMT: Diversity enlarged mutual teaching for unsupervised domain adaptive person re-identification," *IEEE Trans. Multimedia*, vol. 25, pp. 4586–4597, 2023.
[42] H. Tan, X. Liu, B. Yin, and X. Li, "MHSA-net: Multihead self-attention network for occluded person re-identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8210–8224, Nov. 2023.
[43] X. Ma, W. Lv, and M. Zhao, "A double stream person re-identification method based on attention mechanism and multi-scale feature fusion," *IEEE Access*, vol. 11, pp. 14612–14620, 2023.
[44] C. Zhao, Z. Qu, X. Jiang, Y. Tu, and X. Bai, "Content-adaptive auto-occlusion network for occluded person re-identification," *IEEE Trans. Image Process.*, vol. 32, pp. 4223–4236, 2023.
[45] S. Li, H. Chen, S. Yu, Z. He, F. Zhu, R. Zhao, J. Chen, and Y. Qiao, "COCAS+: Large-scale clothes-changing person re-identification with clothes templates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1839–1853, Apr. 2023.
[46] S. Yu, Z. Dou, and S. Wang, "Prompting and tuning: A two-stage unsupervised domain adaptive person re-identification method on vision transformer backbone," *Tsinghua Sci. Technol.*, vol. 28, no. 4, pp. 799–810, Aug. 2023.
[47] D. Zheng, J. Xiao, M. Sun, H. Bai, and J. Hou, "Plausible proxy mining with credibility for unsupervised person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 7, pp. 3308–3318, Jul. 2023.
[48] J. Mao, Y. Yao, Z. Sun, X. Huang, F. Shen, and H.-T. Shen, "Attention map guided transformer pruning for occluded person re-identification on edge device," *IEEE Trans. Multimedia*, vol. 25, pp. 1592–1599, 2023.
[49] J. Peng, G. Jiang, and H. Wang, "Adaptive memorization with group labels for unsupervised person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 10, pp. 5802–5813, Oct. 2023.
[50] Y. Qian, X. Yang, and S.-K. Tang, "Dual-space aggregation learning and random erasure for visible infrared person re-identification," *IEEE Access*, vol. 11, pp. 75440–75450, 2023.

C. SELVAN (Senior Member, IEEE) received the B.E. degree in CSE from Manonmaniam Sundaranar University, India, in 2002, and the M.E. and Ph.D. degrees in CSE from Anna University, Chennai, India, in 2007 and 2013, respectively. During the Ph.D. degree, he was a JRF and a SRF with the Government College of Technology, Coimbatore, under University Grant Commission (UGC), New Delhi, India. He was a PDF with NIT, Tiruchirappalli, under UGC, from June 2017 to June 2022. He has been a Professor with the School of CSE, REVA University, Bengaluru, India, since February 2023.

H. ANWAR BASHA received the B.E. degree from Anna University, Chennai, the M.Tech. degree from the Dr. M. G. R. Educational and Research Institute University, Chennai, and the Ph.D. degree in CSE from the Saveetha Institute of Medical and Technical Sciences, Chennai. He is currently an Associate Professor with the Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai, Tamil Nadu, India. He is also having many certifications like Microsoft Certified Azure Fundamentals, AWS Certified Cloud Practitioner, and IBM Certified Data Science Foundation. He has more than 16 years of teaching experience. He has around two years of industrial work experience. He has published papers in various international conferences and peer-reviewed international journals. He has authored a text book *Cloud based Security Management*. Currently, he is working on multi-cloud storage, big data analytics, and cyber security.

K. MEENAKSHI received the bachelor's degree (Hons.) in computer science and engineering from the Raja College of Engineering and Technology, Madurai Kamaraj University, in April 2003, the master's degree (Hons.) in computer science and engineering from the Dr. M. G. R. Educational and Research Institute, Chennai, in May 2014, and the Ph.D. degree from the Saveetha Institute of Medical and Technical Sciences (Saveetha Deemed University), Chennai, in November 2022. She is currently an Assistant Professor with the CSE Department, SRM Institute of Science and Technology, Vadapalani Campus, Chennai. With a wealth of experience spanning 20 years, she is not only a Seasoned Researcher but also a Proficient Educator. Her commitment to excellence in teaching has positively impacted countless students over the years. Throughout her academic career, she has showcased her dedication to scholarly research, with a prolific publication record of 23 articles in esteemed international and national journals, alongside contributing to four patents. Her research interests include the intersection of cutting-edge technologies, particularly in cloud computing, blockchain technology, and network security.

SOUMYALATHA NAVEEN (Member, IEEE) received the bachelor's degree in computer science and engineering and the master's degree from Visvesvaraya Technological University, Belgaum, India, and the Ph.D. degree in edge intelligence from REVA University, Bengaluru. She is currently an Assistant Professor-Senior Scale with the Department of Computer Science and Engineering, Manipal Institute of Technology, Bengaluru, Manipal Academy of Higher Education. Prior to this, she was associated with REVA University, Brindavan College of Engineering, and was a Software Engineer. She has more than 12 years of teaching, research, and administrative experience. She has published over more than eight indexed journal articles, more than 25 international conference papers, and three book chapters. Her research interests include artificial intelligence, deep learning, edge computing, the Internet of Things, and the intelligent IoT systems.