IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 March 2024, accepted 28 March 2024, date of publication 4 April 2024, date of current version 19 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3385342

RESEARCH ARTICLE

A Cross-Period Network for Clothing Change Person Re-Identification

SHUXIN ZHENG[1], SAI LIANG[2], CHUNYUN MENG[3], ZHONGGUO ZHANG[4], AND LI LUAN[5]

[1]School of Management, Guangdong University of Science and Technology, Dongguan 523083, China
[2]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China
[3]Faculty of Electrical, Information and Communication Engineering, Kanazawa University, Kanazawa 920-1192, Japan
[4]School of Intelligent Information Engineering, Guangdong Vocational College of Hotel Management, Dongguan 523960, China
[5]School of Public Affairs, University of Science and Technology of China, Hefei 230026, China

Corresponding author: Chunyun Meng (cymeng@stu.kanazawa-u.ac.jp)

This work was supported in part by the National Natural Science Foundation of China under Grant 10572156 and Grant 61802274.

ABSTRACT Pedestrian re-identification aims to identify the same target pedestrian among multiple non-overlapping cameras. However, in real scenarios, pedestrians often change their clothing features due to external factors such as weather and seasons, rendering traditional methods reliant on consistent clothing features ineffective. In this paper, we propose a Knowledge-Driven Cross-Period Network for Clothing Change Person Re-Identification, comprising three key components: 1) Knowledge-Driven Topology Inference Network: Leveraging knowledge graphs and graph convolution networks, this network captures spatio-temporal information between camera nodes. Knowledge embedding is introduced into the graph convolution network for effective topology inference; 2) Cross-Period Clothing Change Network: This network aggregates spatio-temporal information for clothing generation. By utilizing overall pedestrian clothing characteristics whthin logical topology cameras, it mitigates matching errors caused by external factors; and 3) Joint Optimization Mechanism: A collaborative approach involving both the topology inference network and cross-period clothing change network. Multi-camera logical topology offers auxiliary information and retrieval order for the clothing change network, while pedestrian re-identification results provide feedback to adjust the logical topology. Experimental analysis on datasets Celeb-ReID, PRCC, UJS-ReID, SLP, and DukeMTMC-ReID, demonstrates the effectiveness and robustness of our proposed model in addressing the challenges of pedestrian re-identification in scenarios involving changing clothing features.

INDEX TERMS Logical topology inference, knowledge graph, clothing change re-identification, graph convolution network, intelligent surveillance application.

I. INTRODUCTION

Recent developments in computer vision models [1], [2], especially the introduction of deep convolutional neural networks, has extensively improved the accuracy and speed of biometric identification. Pedestrian re-identification aims to locate a specific individual in a network of non-overlapping cameras. This technique is often regarded as an extension of face recognition, and numerous scholars have extensively investigated pedestrian re-identification [3], [4], achieving commendable performance in controlled environments. However, current methodologies and datasets are contingent on the assumption of clothing invariance. In real-world scenarios observed by long-term surveillance cameras, pedestrians are prone to alter their attire due to external factors such as weather and season changes. This renders methods relying on clothing consistency impractical in these dynamic situations.

The task of clothing change presents a formidable challenge in pedestrian re-identification. Traditional approaches to pedestrian re-identification primarily focus on matching individuals with consistent clothing, where clothing consistency implies that the clothing features of the target pedestrian remain unchanged across different cameras.

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

Previous methods for clothing change in pedestrian re-identification have typically employed features other than clothing as discriminative features, aiming to mitigate the effects of clothing changes on pedestrian matching. While these methods have demonstrated some success in reducing the impact of clothing changes on pedestrian matching, they often overlook the intricate interplay between external factors and clothing alterations.

Although previous researches [5], [6], [7] have made significant contributions to pedestrian re-identification in scenarios involving clothing changes, there are still several challenges that require attention: (1) Existing approaches primarily focus on pedestrian area features and overlook the intricate interplay between external factors and clothing changes, resulting in inaccuracies in the predicted outcomes. (2) They neglect the spatio-temporal relationships between cameras and the auxiliary role of logical topology, leading to generated clothing that deviates from reality. (3) Traditional topology inference methods concentrate solely on spatial information between cameras and fail to capture temporal cues between nodes, hindering the dynamic prediction of targets.

To tackle the aforementioned challenges, this paper proposes a person re-identification model for clothing changes based on a knowledge-driven cross-period network. We construct a knowledge graph encompassing various external factors, employ a knowledge embedding method to capture semantic relationships between topology information and external factors, and integrate this acquired knowledge into spatio-temporal graph convolution for camera topology inference. To enhance the precision of clothing generation, we incorporate auxiliary information from the camera's logical topology. We devise a Cross-Period Clothing Change Network (CPCCN) based on camera topology to generate features for pedestrian clothing. Our contributions are outlined below:

(i) We introduce a Knowledge-Driven Topology Inference Network (KTIN), designed to utilize a knowledge graph for capturing semantic relationships between external factors and camera topology. The acquired knowledge is then incorporated into spatio-temporal graph convolution to enhance the precision of topology inference.

(ii) A Cross-Period Clothing Change Network, based on the aggregation of spatio-temporal information, is developed for clothing generation. This network utilizes camera logical topology information to extract clothing details with robust associations and employs auxiliary information to generate clothing features for target pedestrians.

(iii) Our approach involves the joint optimization of the Topology Inference Network and Cross-Period Clothing Change Network. The camera logical topology serves as auxiliary information for CPCCN, and the recognition results from CPCCN are fed back to KTIN to refine the camera topology inference.

The remaining sections of the paper are organized as follows. Section II provided a brief history of research in person re-identification and highlights representative methods that have contributed to the advancement of the field. We delve into the details of our proposed framework, including the architecture and functionality of KTIN and CPCCN in section III. Section IV described the experimental setup and presents the results of extensive experiments conducted on various datasets. Lastly, we concluded the manuscript by summarizing the key findings and contributions of our research in section V.

II. RELATED WORK

A. PERSON RE-IDENTIFICATION

Person re-identification has evolved significantly over the years, driven by the increasing demand for robust and accurate methods [8], [9], for matching individuals across different camera views. Wu et al. [10] leveraged camera-aware self-training to improve the performance of semi-supervised person re-identification systems. Liu et al. [11] focused on adaptive transfer learning techniques to improve cross-domain person re-identification performance. Jin et al. [12] separated global distance distributions to improve the discriminative ability of unsupervised person re-identification models. Chen et al. [13] emphasized deep credible metric learning to address the challenges of unsupervised domain adaptation in person re-identification tasks. Zhang et al. [14] addressed the challenge of noise in unsupervised domain adaptation for person re-identification, aiming to enhance model robustness and generalization. Zhang and Hu [15] emphasized unified domain learning techniques for unsupervised person re-identification, aiming to achieve robust performance across diverse datasets and domains. These studies illustrate the evolution of person re-identification research and highlight the diverse approaches and techniques used to tackle this challenging problem.

B. MULTI-CAMERA LOGICAL TOPOLOGY INFERENCE

The application of multi-camera logical topology inference in pedestrian re-identification aims to enhance system performance by scrutinizing and modeling the logical topological relationships among multiple cameras. This method underscores the significance of maintaining topological consistency, thereby refining the accuracy of pedestrian re-identification through optimized matching in multi-camera systems and mitigating matching errors attributed to topological variations. Moreover, logical topological reasoning renders the system adaptable to intricate scenarios, fortifies its resilience in multi-camera environments, and strongly supports the ongoing advancement of pedestrian re-identification tasks. In prior approaches [16], the topology of multiple cameras is defined, and Loy et al. [17] introduced an unsupervised method, eliminating the need for camera calibration but inferring a logical topology with fixed positions between cameras, deviating from reality. In many

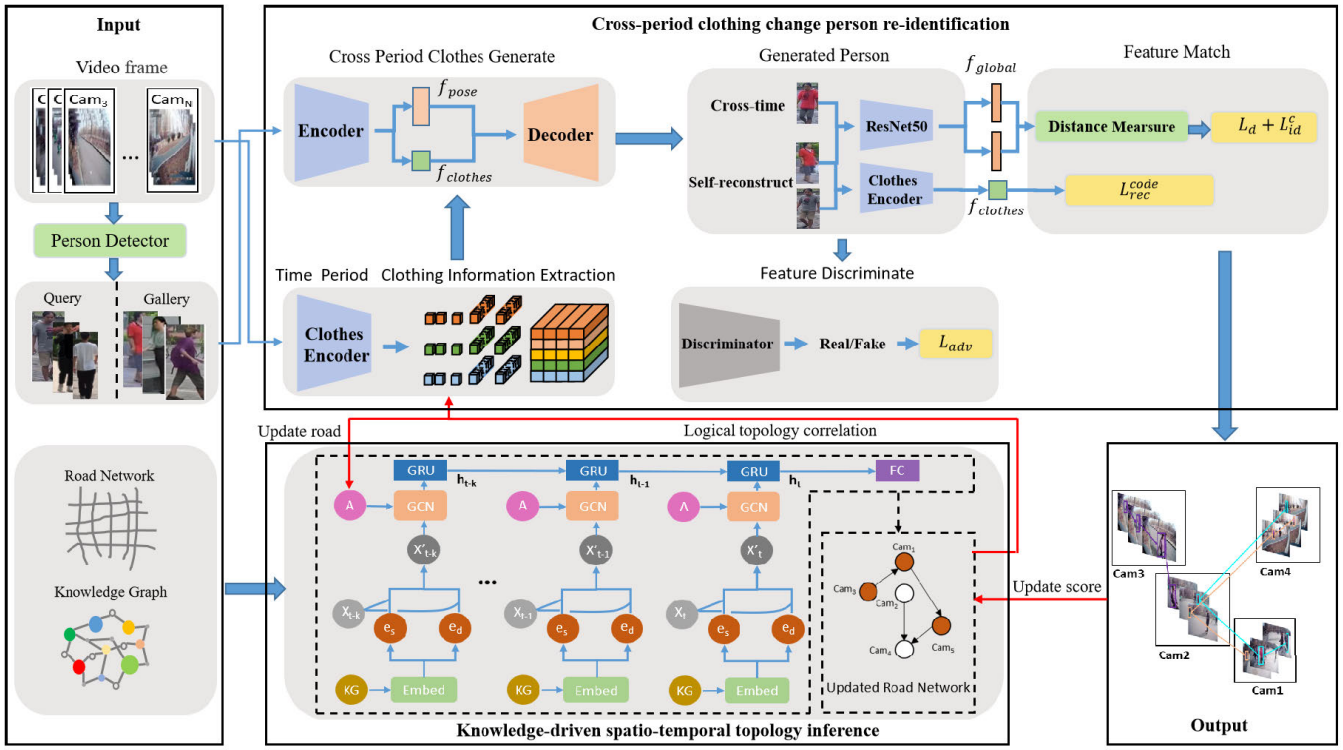

FIGURE 1. Diagram illustrating the approach for the joint optimization mechanism of the knowledge-driven cross-period network through knowledge-driven spatio-temporal topology inference and the cross-period clothing change person re-identification network.

instances, camera topology should correlate with pedestrian trajectories rather than remaining static. Javed et al. [18] proposed a method to establish the correlation between cameras based on pedestrian trajectories, yet it is susceptible to factors such as occlusion and viewpoint orientation. Cho et al. [19] concurrently train pedestrian re-identification with multi-camera logical topology inference; however, it faces limitations in dynamically updating the inferred logical topology as it may undergo dynamic changes over time.

C. CLOTHING CHANGE PERSON RE-IDENTIFICATION

Traditional pedestrian re-identification methods primarily focus on matching individuals in short-term camera surveillance, often relying on clothing color as the primary feature. However, in widely used datasets like Market-1501 and DukeMTMC, individuals wear identical outfits in both QUERY and GALLERY, contrary to real-world scenarios where people change attire within days. The significant challenge in clothing change pedestrian re-recognition lies in enabling the model to recognize detached clothing or maintaining clothing consistency during the matching phase.

Clothing change pedestrian re-recognition tasks can be broadly classified into three categories. The first involves reconstructing pedestrians using depth information [20], [21], [22]. Early research aimed to address clothing dependency by leveraging non-clothing information for matching. For instance, Barbosa [20] used RGB-D sensors to acquire depth information, extracting soft biometric cues such as limb size, and matching pedestrians based on limb information. Munaro [21] expanded on this by incorporating facial information and skeleton link orientation, making the point cloud model more comprehensive. However, this sensor-based biometric approach is sensitive to viewing angle changes and highly reliant on device accuracy. Haque et al [22] applied deep learning, abandoning RGB information and recognizing pedestrians by extracting 4D spatio-temporal features, adding the time dimension to 3D point cloud features. This method requires substantial sample quality and is limited to single pedestrian recognition scenarios.

The second category involves using human silhouette information for re-recognition. To reduce reliance on clothing appearance, Zhang et al [23] emphasized the motion characteristics of pedestrians, using a Gaussian model for trajectory alignment and motion pattern recognition. While effective, this method is susceptible to clothing interference when recovering discriminative features. Yang et al. [6] proposed an SPT model to transform spatial polarity in pedestrian contour maps, reducing clothing dependency but ignoring the influence of clothing style on the silhouette.

The third category retains clothing information for re-recognition. Xu et al. [24] introduced capsule networks [25] to re-identify pedestrians changing clothes, utilizing the vectorial property to store attributes. Huang et al. [26] added training samples, removing consistent clothing from images, but their provided celebrity snapshots differed significantly from real scenes. Yu et al. [27] proposed a biometric drows network (BC-Net) with separate branches for learning

biometric and clothing features. However, the method relies on third-party data and manual template searches, and the target pedestrian's clothing is unknown in real scenarios.

III. PROPOSED METHOD

In this section, we elaborate on the proposed cross-period pedestrian re-identification model based on logical topology inference. Initially, we establish the camera logical topology using a Knowledge Graph (KG) and a Graph Convolution Network (GCN), endowing it with the capability to furnish retrieval order and clothing-assisted features. Subsequently, we introduce a joint optimization mechanism for camera topology inference and the cross-period clothing change network. The latter leverages auxiliary information from the camera logical topology to predict clothing features for target pedestrians within the corresponding time frame. The outcomes of pedestrian matching inform adjustments to the camera topology through feedback to enhance its accuracy. The detailed model structure is illustrated in Fig. 1.

A. KNOWLEDGE-DRIVEN TOPOLOGY INFERENCE NETWORK

Utilizing knowledge graphs in topology inference for person re-identification presents a novel and innovative approach. Knowledge graphs offer a structured representation of knowledge, facilitating effective reasoning and inference over complex relationships among entities. In the context of person re-identification, the application of knowledge graphs can revolutionize the way camera topology is inferred and utilized.

Knowledge graphs enable the representation of camera topology in a structured format, capturing the spatial and temporal relationships among cameras. By encoding these relationships as graph edges and cameras as graph nodes, representation learning techniques can be employed to learn informative embeddings that capture the underlying topology. And then, it can be enriched with semantic information about camera attributes, such as location. This semantic enrichment enhances the expressiveness of the topology representation, allowing for more precise inference and reasoning about camera relationships.

Moreover, knowledge graphs facilitate contextual reasoning by capturing rich contextual information about the environment in which cameras are deployed. This contextual information can include factors e.g., lighting conditions, which influence camera observations and person re-identification performance. Graph-based inference algorithms can be applied to knowledge graphs to infer camera topology relationships. Techniques such as graph neural networks (GNNs) enable message passing and aggregation over graph structures, allowing for effective propagation of information and inference of camera connections. At the same time, knowledge graphs provide a flexible framework for dynamically adapting camera topology in response to changes in the environment. By continuously updating the graph structure based on new observations and contextual cues, the re-identification system can adapt to evolving conditions and maintain robust performance over time.

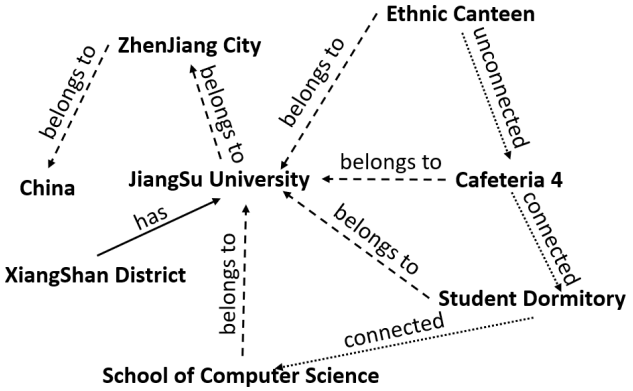

FIGURE 2. Illustrative example of a knowledge graph.

Last but not least, knowledge graphs can be seamlessly integrated with person re-identification models to enhance their performance. By incorporating graph-based features or leveraging graph-based inference for candidate matching, person re-identification models can leverage the rich contextual information encoded in the knowledge graph to improve accuracy and reliability.

1) KNOWLEDGE GRAPH

We formulate knowledge graphs to capture semantic relationships between external factors and camera topology information. These knowledge graphs serve as repositories for raw camera topology data and other data categories, structured around triplets (heads, relations, tails). In the context of a knowledge graph, the head refers to the subject entity of a triplet, which is a fundamental unit of information in the graph. It represents the entity about which information is being stated. A relation in a knowledge graph connects the head entity to the tail entity and represents the semantic link between them. It describes the nature of the connection or interaction between the entities. The tail represents the object entity of a triplet. It is the entity to which the head entity is related through the specified relation. In each triplet, there exists both semantic information and a network structure, with both the head and tail representing entities. The semantic information characterizes the relationship between the head and tail entities. An illustration of the designed knowledge graph is presented in Fig. 2.

2) KNOWLEDGE REPRESENTATION

The Knowledge Graph (KG) comprises numerous triplets, each not only expressing semantic relationships between entities but also facilitating the modeling of relationships between attributes and entities. An attribute represents a characteristic or property associated with an entity in the knowledge graph. Attributes in our model, designed as external factors, exhibit a many-to-one relationship with entities. To capture the knowledge structure and semantic relationships between the logical camera topology and external factors, we employ

KR-EAR [28], a knowledge graph representation model based on entity-attribute relationships. This model effectively distinguishes between attribute and relationship information. In this context, these terms are used to model the relationships between different cameras (entities) and infer the logical topology of the camera network. The head entities represent the cameras, the relations capture the spatial and temporal connections between cameras, and the tail entities represent the neighboring cameras. Attributes include information such as geographic location, field of view, or connectivity strength between cameras, which are used to infer the topology of the camera network. Camera nodes, attributes, and relations are represented as the triplet $KG = C, C_{att}$.

$$\begin{aligned} C &= \{(c_i, adj, c_j), i, j \in \{1, 2, \ldots, n\}\}, \\ C\_att &= \{(c_i, att_l, att_{l\_c_i}), l \in \{1, 2, \ldots, L\}\} \end{aligned} \tag{1}$$

As depicted in Equation (1), C represents a relational triplet where c_i and c_j denote non-overlapping camera nodes, adj signifies the adjacency of camera nodes, and n represents the total number of camera nodes. C_{att} forms a relation triplet between camera nodes and attributes. Each camera node encompasses multiple attribute categories, denoted as att_l, with corresponding attribute values represented as $att_{l\_c_i}$. For instance, (c_i,*weather*,*sunny*) indicates that the current weather attribute value for camera node c_i is sunny. Utilizing KR-EAR, we learn the embedding X_E of the triplet, incorporating knowledge graph information into spatio-temporal graph convolution. The objective function is formulated as follows:

$$\begin{aligned} &P(C, C_{att}|X_E) = P(C|X_E)P(C_{att}|X_E), \\ &P(C|X_E) = \prod_{(c_i,adj,c_j)\in C} P((c_i, adj, c_j)|X_E), \\ &P((c_i, adj, c_j)|X_E) = \\ &\frac{\exp\left(-||c_i M_r + adj - c_j M_r)_{L_1/L_2} + b_1||\right)}{\sum_{\widehat{c_i}\in V_c} \exp(g(\widehat{c_i}, adj, c_j))}, \\ &P(C_{att}|X_E) = \\ &\prod_{(c_i,att_l,att_{l\_c_i})\in C_{att}} P((c_i, att_l, att_{l\_c_i})|X_E), \\ &P((c_i, att_l, att_{l\_c_i})|X_E) = \\ &\frac{\exp(-||f(c_i U_{att} + b_{att}) - E_{att\_c}||_{L_1/L_2} + b_2)}{\sum_{\widehat{c_i}\in V_c} \exp(h(c_i, att_l, \widehat{att_{l\_c_i}}))} \end{aligned} \tag{2}$$

The conditional probabilities of the relation triplet, denoted as $P(C|X_E)$, and the attribute triplet, denoted as $P(C_{att}|X_E)$, are defined. Here, $V_c = c_1, c_2, \ldots, c_n$ represents the set of camera nodes, and g serves as the energy function indicating the relevance of the relationship and entity. In the equation, b_1 is a bias term, M_r is a transfer matrix, and L_1 and L_2 stand for the L_1 and L_2 norms, respectively. The function f represents a nonlinear function, E_{att_c} corresponds to the embedding vector of the attribute value att_c, and U_{att} is a linear transformation. The terms b_2 and b_{att} are bias terms. Ultimately, KR-EAR is employed to strengthen the association between entities and attributes, generating representations of relationships and attributes, respectively.

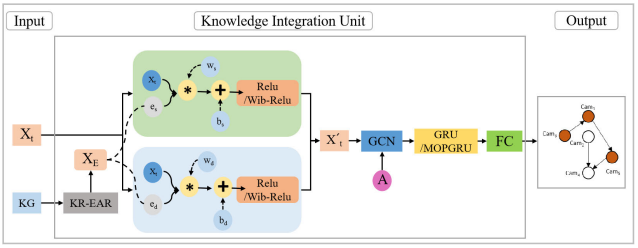

FIGURE 3. Structure of the knowledge integration unit.

3) KNOWLEDGE INTEGRATION UNITS

We have devised the Knowledge Integration Unit (KIU) to seamlessly incorporate derived knowledge into the spatio-temporal graph convolution network, enabling the capture of spatio-temporal correlation cues between external factors and camera topology. While primarily adopting the graph convolution (GC) architecture, KIU enhances this by introducing the Gated Recurrent Unit (GRU) or its enhanced version (MOPGRU [29]). This integration allows for the capture of dynamic temporal information within the logical topology, complementing the aggregation of camera node features. The specific structure of KIU is illustrated in Fig. 3. Leveraging the knowledge embedding X_E and the camera node features X_t at time t as inputs, KIU produces updated node features X_t' after passing through the knowledge fusion module. External factors, ranging from the physical location of the camera (considered as static knowledge e_s) to time-varying elements like weather (termed as dynamic knowledge e_d), are diverse. Linear transformations w_s and w_d are applied, with bias constants b_d and b_s completing the model.

We leverage Graph Convolutional Network (GCN) and GRU or MOPGRU to capture the spatio-temporal correlation among camera nodes. The adjacency matrix A and the updated camera node features X_t' serve as inputs to the GCN. The process can be formalized as follows:

$$gcn(X_t', A) = \sigma(\widetilde{D}^{-\frac{1}{2}} * \widetilde{A} * \widetilde{D}^{-\frac{1}{2}} * X_t' * W) \tag{3}$$

Here, $\sigma(*)$ represents an activation function, $\widetilde{A}$ is the adjacency matrix with self-connections, and $\widetilde{D}$ is the degree matrix of $\widetilde{A}$. The weight matrix is denoted as W. Furthermore, we employ GRU or MOPGRU to account for time dependence, comprising a reset gate and an update gate u_t. The output at time t can be expressed as follows:

$$h_t = u_t \odot h_{t-1} + (1 - u_t) \odot c_t \tag{4}$$

where c_t represents the state at the current time step.

B. CROSS-PERIOD CLOTHING CHANGE NETWORK

The model effectively handles the variability in clothing through the Cross-Period Clothing Change Network (CPCCN), a crucial component of the proposed framework. It achieves this by first extracting clothing features using a

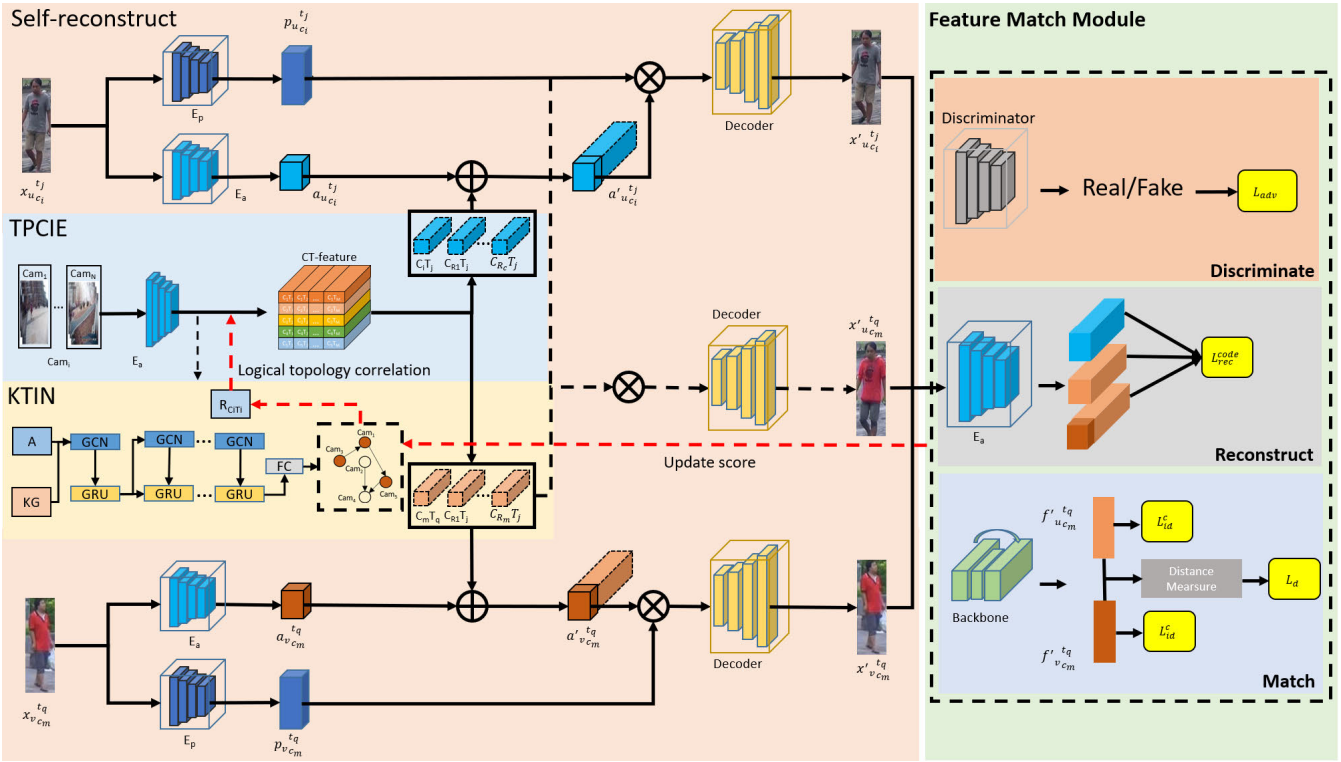

FIGURE 4. Structure of the cross-period clothing change network.

clothing appearance encoder, capturing colors, patterns, and textures characteristic of different clothing styles. Incorporating temporal modeling techniques enables the CPCCN to analyze sequences of pedestrian images, learning patterns and trends in clothing variations over time. Additionally, leveraging Generative Adversarial Networks (GANs), the model generates realistic images of pedestrians with varying clothing styles, augmenting the training data with synthetic examples of clothing variations for improved generalization. Integration of external knowledge, further refines predictions of clothing variations. The joint optimization mechanism between the Knowledge-Driven Topology Inference Network (KTIN) and CPCCN ensures that predictions align with inferred logical topology relationships between cameras, enhancing contextual relevance. Overall, by effectively leveraging clothing appearance features, temporal modeling, GANs, knowledge integration, and joint optimization, the model robustly handles clothing variability, making it highly applicable to real-world scenarios where individuals frequently change clothing.

We delve into the intricacies of the devised cross-period clothing change network. To enhance the precision of predicting clothing characteristics for target pedestrians during specific time periods, we meticulously consider the spatio-temporal relationships between cameras and the impact of external factors on clothing changes. Firstly, we categorize pedestrians' clothing features based on the distinct time periods within each group of cameras. These features encapsulate the overall clothing style under a specific camera and time segment. Subsequently, we acquire the clothing features of the source/target camera and other cameras exhibiting a strong association during a particular time period, guided by the camera logical topology relationship. Lastly, the pose features of the pedestrian are seamlessly integrated with specific clothing features using a Generative Adversarial Network (GAN). This integration aims to generate images of the target person under a specific camera and time period. The detailed structure is illustrated in Fig. 4.

1) TIME PERIOD CLOTHING INFORMATION EXTRACTION

We comprehensively analyze pedestrian clothing information from various cameras within each time period to establish clothing templates for inferring the styles worn by the target pedestrians during specific time intervals. Illustrated in Fig. 5, we employ a clothing encoder to extract clothing features, categorizing the clothing information based on distinct time periods within each camera group. All pedestrian images are segmented into M time periods, denoted as $Time = Time_1, Time_2, \ldots, Time_M$. Each time segment is defined by a time interval of T, with $Time_i = t_1, t_2, \ldots, t_T$, where $1 \leq i \leq M$. The clothing features extracted by the clothing encoder are represented as:

$$A_{T_m}^{C_i} = \left\{ \left(a_{r_{t_j}}^{C_i}, y_r \right) \mid 1 \leq r \leq N^{C_i}, 1 \leq j \leq T \right\} \quad (5)$$

where $A_{T_m}^{C_i}$ represents the set of pedestrian clothing appearance features for the current camera i at time period m, and

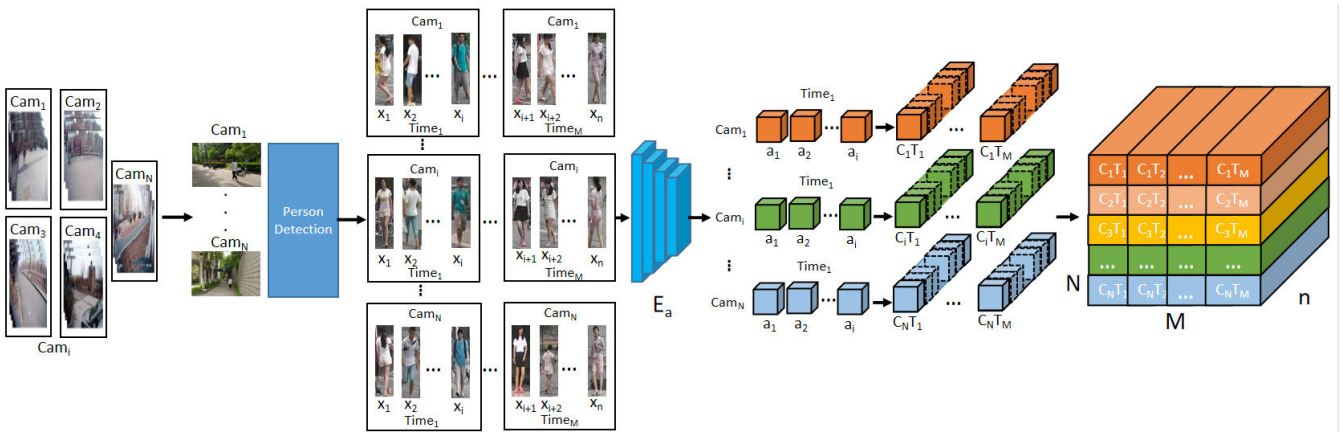

FIGURE 5. Time period clothing information extraction.

$a_{r_{t_j}}^{C_i}$ denotes the clothing feature of pedestrian r at moment t_j, extracted from the clothing appearance encoder E_a. In other words, $a_{r_{t_j}}^{C_i} = E_a(x_{r_{t_j}}^{C_i})$. Here, $x_{r_{t_j}}^{C_i}$ represents the pedestrian at moment t_j, y_r is the label for pedestrian r, and N^{C_i} is the count of pedestrians captured by camera i. The clothing characteristics of pedestrians in time period m are denoted as:

$$C_iT_m = \frac{1}{N^{C_i}} \sum_{r=1}^{N^{C_i}} \sum_{j=1}^{T} E_a(x_{r_{c_i}}^{t_j}) \tag{6}$$

The ultimate collection of clothing features for each time period is denoted as $CT = \{C_1T_1 \dots C_iT_m \dots C_NT_M\}$, where $CT \in \mathbb{R}^{N \times M \times n}$. Here, N represents the number of cameras, and M represents the number of time periods.

2) CROSS-PERIOD CLOTHING GENERATION

Our approach involves predicting the clothing features of a target pedestrian in a specific environment based on the camera and time period of the pedestrian in the gallery. This aims to optimize the retrieval order of pedestrian images in the gallery, mitigating the adverse effects of clothing changes on pedestrian re-identification. As illustrated in Fig. 4, CPCCN primarily employs a generative adversarial architecture to forecast the clothing characteristics of the target pedestrian. The input image pair, $x_{u_{c_i}}^{t_j}$ and $x_{v_{c_m}}^{t_q}$, represents the target pedestrian captured by the source camera c_i at moment t_j and the pedestrian from the gallery of camera c_m at moment t_q. To extract the pose features $p_{u_{c_i}}^{t_j}$ and clothing features $a_{u_{c_i}}^{t_j}$ of the target pedestrian, we utilize the pose encoder E_p and clothing appearance encoder E_a, respectively. The extracted pose features typically encompass more spatial geometric location information compared to the clothing appearance features. To minimize the influence of appearance information during pose feature extraction, we convert the original image into a grayscale map, effectively separating the pose features from the appearance features. This ensures a reduction in the impact of appearance details when employing the pose encoder for feature extraction.

3) SELF-RECONSTRUCTION

A critical step in enhancing the reconstruction ability of the cross-period clothing change network involves self-reconstruction of the pedestrian $x_{u_{c_i}}^{t_j}$. Specifically, KTIN identifies the set of cameras strongly associated with c_i based on the environment C_iT_j (comprising camera c_i and time period T_j). The TPCIE module then extracts the group clothing features of the corresponding environment from the clothing feature set CT, leveraging the topology relationship among cameras. This can be represented as $C_{R_c}T_j = \{C_iT_j, C_{R_1}T_j \dots C_{R_c}T_j\}, 1 \le c \le N$, where R_c represents the cameras associated with c_i. The decoupled clothing feature $a_{u_{c_i}}^{t_j}$ is seamlessly integrated with these features to obtain the feature $a'uc_i^{t_j}$ encompassing group clothing information. Subsequently, $a'uc_i^{t_j}$ is fed into the generator along with the decoupled pose feature $p_{u_{c_i}}^{t_j}$ to predict the pedestrian image $x'uc_i^{t_j}$ within the environment C_iT_j. The same operation is iteratively applied to pedestrians in the gallery.

4) CROSS-PERIOD GENERATION

In contrast to self-reconstruction, the environmental information for this step is sourced from pedestrians in the gallery, denoted as C_mT_q. To mitigate interference from the original clothing feature, the initial clothing feature $a_{u_{c_i}}^{t_j}$ is directly substituted with the group clothing information $C_{R_c}T_q$, extracted from CT based on the camera logical topology relationship. Likewise, this information is amalgamated with the pose feature to generate images of the pedestrian $x'^{t_q}_{u_{cm}}$ at the specific time period T_q. The identity of the generated pedestrian is contingent upon the provider of the pose feature.

5) TRAINING LOSS FUNCTION

To enhance the realism of the generated images, we employ the generative adversarial loss, denoted as L_{adv}, to align the distribution of the generated images with the actual data distribution.

$$L_{\text{adv}} = \mathbb{E}\left[\log D\left(x_{u_{c_i}}^{t_j}\right) + \log\left(1 - D\left(G\left(p, a\right)\right)\right)\right] \tag{7}$$

where D and G represent the discriminator and generator, respectively, while p and a denote the pose feature and clothing appearance feature of the pedestrian. During the self-reconstruction stage, we employ the reconstruction loss, denoted as L_{rec}^{self}, to constrain the pedestrian images before and after reconstruction, as well as the associated clothing features. This constraint aims to enhance the feature extraction capability of the encoder. The loss function is formulated as follows:

$$L_{rec}^{self} = \mathbb{E}\left[\left\| x_{u_{c_i}}^{t_j} - G\left(p_{u_{c_i}}^{t_j}, a_{u'_{c_i}}^{t_j}\right)\right\|_1\right] + \mathbb{E}\left[\left\| a'^{t_j}_{u'_{c_i}} - E_a\left(x'^{t_j}_{u_{c_i}}\right)\right\|_1\right] \tag{8}$$

Moreover, directly matching pedestrians may result in larger intra-class differences, potentially disrupting ID consistency. To address this, we employ ResNet-50 to extract global features from $x'^{t_q}_{u_{cm}}$ and $x'^{t_q}_{v_{cm}}$. The similarity between images is then evaluated using the distance metric d:

$$d_{(f'^{t_q}_{u_{cm}}, f'^{t_q}_{v_{cm}})} = \frac{\exp\left(f\left\|f'^{t_q}_{u_{cm}} - f'^{t_q}_{v_{cm}}\right\|_1\right)}{1+\exp\left(f\left\|f'^{t_q}_{u_{cm}} - f'^{t_q}_{v_{cm}}\right\|_1\right)} \tag{9}$$

where f denotes the fully connected layer, $f'^{t_q}_{u_{cm}}$ and $f'^{t_q}_{v_{cm}}$ represent the global features of $x'^{t_q}_{u_{cm}}$ and $x'^{t_q}_{v_{cm}}$ respectively. The distance loss is applied to constrain $x'^{t_j}_{u_{c_i}}$ and $x_{u_{c_i}}^{t_j}$:

$$L_d = \begin{cases} y\log d + (1-y)\log(1-d) \\ y = 1; \text{ if } f'^{t_q}_{u_{cm}} = f'^{t_q}_{u_{cm}} \\ y = 0; \text{ otherwise} \end{cases} \tag{10}$$

We employ cross-entropy loss to maintain identity consistency, expressed as $L_{cls} = -\log\left(F_g\right)$.

C. JOINT ITERATIVE PROCESS

The proposed method in this paper necessitates the collaboration of the camera topology inference network and the cross-period clothing change network to derive the ultimate pedestrian re-identification results. In essence, the final recognition outcomes are influenced by both the camera topology inference network and the cross-period clothing change network. During the training process of KTIN, the initial topology weights are determined by the geographic location of the cameras. However, these initial weights lack information about the walking direction, necessitating continuous updates and convergence during the iterative process to achieve a stable topology. In the joint iteration process of the topology inference network and cross-period clothing change network, KTIN provides CPCCN with the pedestrian clothing features under the associated cameras, aiding the cross-period clothing change network in predicting the clothing features of the target pedestrians at a specific time period. Concurrently, the recognition results of CPCCN are reordered based on the provided camera logical topology relationships. Conversely, the camera logical topology is updated based on the recognition results, and the optimized iterative strategy is outlined in Algorithm 1.

With the conclusion of the iteration process, the logical topology of the camera undergoes continuous convergence, resulting in a stable camera topology relationship. Simultaneously, more accurate clothing features of the target pedestrians are predicted, thereby enhancing the precision of pedestrian re-identification. The time complexity of the algorithm we designed is $O(n^2)$, and the total loss function of the joint iteration process can be expressed as:

$$L_{total} = \lambda_1 L_{cls} + \lambda_2 L_{adv} + \lambda_3 L_{rec}^{self} + \lambda_4 L_d \tag{11}$$

In our experiment, the values of λ_1, λ_2, λ_3 and λ_4 are set to 0.5, 0.3, 0.3, and 0.2, respectively.

Algorithm 1 Process of Joint Iterative KTIN and CPCCN

Require: Video sequence data; e_{init}-the initial epochs of stable training; b-the optimal batch group in an epoch; e_{final}-the total epochs of training

Ensure: Camera logical topology and pedestrian re-identification results

```
while epochs < e_final do
    if epochs = e_init then
        Person re-identification based on CPCCN is training
    else
        while batchs%b = 0 do
            Multi-camera topology network is training
        end while
        The whole joint optimization mechanism is trained. The weight of GCN will be updated and multi-camera logical topology is inferred.
    end if
end while
Return Multi-camera logical topology and pedestrian re-identification results.
```

IV. EXPERIMENT

Our experiments are conducted on two public large-scale datasets, namely SLP and UJS-ReID, which provide multi-camera topology information and timestamps. To assess the performance of our model in scenarios where pedestrians change clothes, we conduct comparative experiments on clothing change person re-identification datasets. Additionally, we perform comparison experiments on traditional pedestrian re-identification datasets to validate the feature extraction capability of our model.

A. DATASETS

The selection of datasets for testing encompasses five diverse datasets, each offering unique challenges and representing various real-world scenarios. The Celeb-ReID dataset provides a large-scale collection of pedestrian images, enabling comprehensive evaluation of re-identification algorithms. PRCC dataset, known for its long-term person

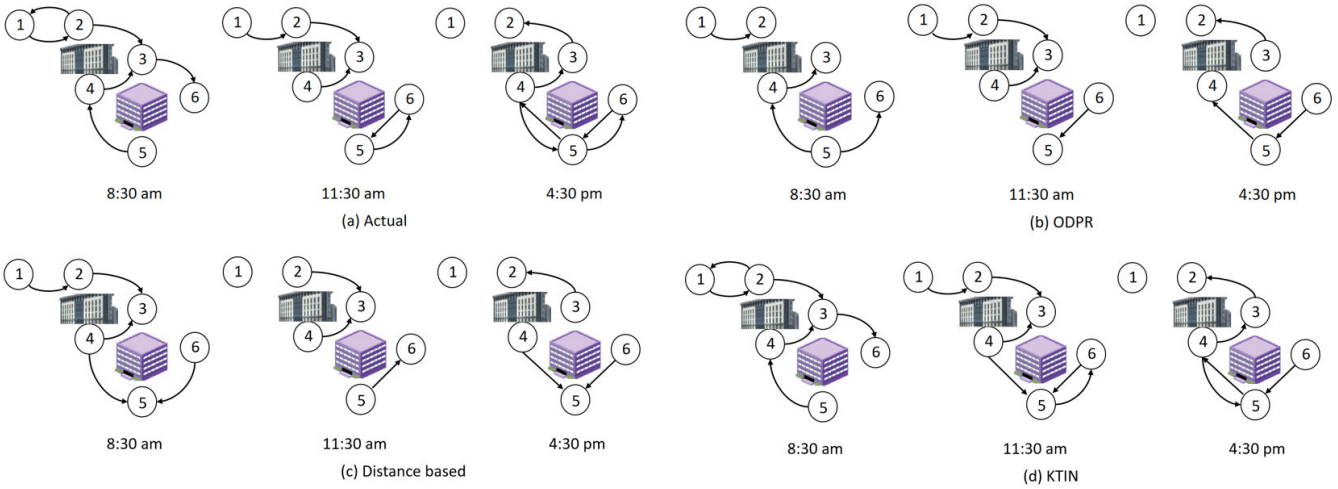

FIGURE 6. Comparison of dynamic logical topological structures on UJS-ReID inferred from different models.

re-identification scenarios, features individuals wearing different clothing across cameras, simulating real-world scenarios. UJS-ReID dataset captures pedestrians' diverse walking trajectories in different scenes, offering realistic representations of pedestrian movement. The SLP dataset offers synchronized large-scale person re-identification data with comprehensive annotations and camera synchronization, providing valuable insights into real-world surveillance scenarios. Finally, the DukeMTMC-ReID dataset, captured by multiple cameras at Duke University, presents diverse pedestrian appearances and environmental conditions, contributing to the evaluation of algorithm robustness across varied settings.

Celeb-ReID: The dataset comprises 34,186 pedestrian images, each with dimensions of 256 × 128, featuring a total of 1052 unique IDs. It is segmented into three subsets: training, query, and gallery. The training set consists of 20,208 pedestrian images representing 632 IDs. The testing set includes 420 IDs, with 2,972 images in the query subset and 11,006 pedestrian images in the gallery subset.

PRCC: This dataset is a large-scale, long-term person re-identification dataset featuring 221 unique IDs. Pedestrians are captured by three cameras, wearing identical clothing in cameras A and B, and different attire in camera C. Each individual has approximately 152 images, totaling 33,698 pedestrian images. The training set includes 150 IDs, and the test set comprises 71 IDs.

UJS-ReID: In this dataset, we strategically deploy several non-overlapping field-of-view cameras across the campus to capture the diverse walking trajectories of pedestrians during different time periods. The real-scene data used in our experiments is recorded at a frame rate of 15 FPS, providing a realistic portrayal of pedestrians with dynamically changing walking patterns in various scenes.

SLP: The SLP dataset stands out as a synchronized large-scale person re-identification dataset, offering not only a substantial number of person IDs and cameras but also supplementary information like comprehensive annotations, camera synchronization, and camera parameters.

DukeMTMC-ReID: The dataset, captured by 8 cameras at Duke University, boasts a training set comprising 16,522 pedestrian images and a test set containing 17,661 images, with 2,228 images designated for query purposes. Both the test set and the training set encompass 702 unique IDs. We will evaluate the model's effectiveness on this dataset for traditional pedestrian re-identification.

B. EVALUATION METRICS

The choice of evaluation metrics is crucial for assessing the performance of the proposed model accurately. In person re-identification tasks, both mean average precision (mAP) and Rank-k accuracy are commonly used metrics, each providing valuable insights into the model's performance from different perspectives.

1) MEAN AVERAGE PRECISION (MAP)

mAP is a widely used metric in object detection and re-identification tasks. It takes into account the precision-recall curve across all possible thresholds, providing a comprehensive measure of the model's performance across different levels of recall. In the context of pedestrian re-identification, mAP is particularly suitable as it accounts for the varying levels of difficulty in matching pedestrians across different camera views and conditions.

2) RANK-K ACCURACY

Rank-k accuracy measures the percentage of correct matches within the top-k retrieved results. This metric is essential as it reflects the practical utility of the system in real-world scenarios, where users are typically interested in retrieving the correct match within the top few results. Rank-k accuracy provides insights into the effectiveness of the model in retrieving relevant pedestrian matches, especially considering

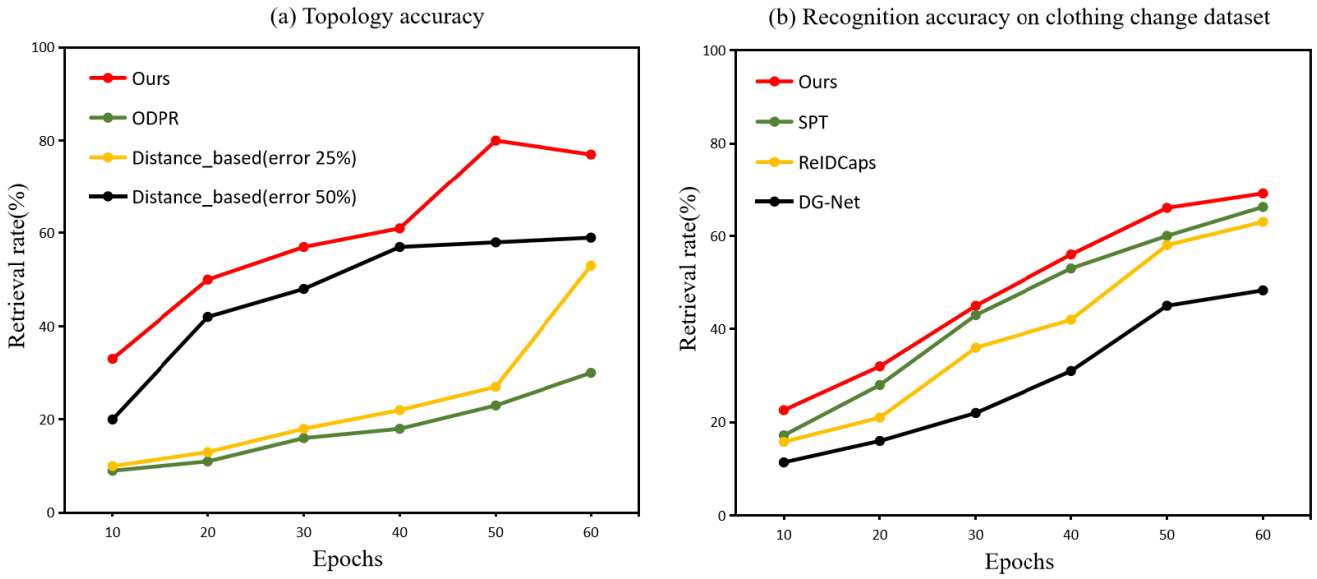

FIGURE 7. Topology and re-identification accuracy based on average search range.

the challenges posed by changing clothes and variations in camera viewpoints.

By using mAP and Rank-k accuracy as evaluation metrics, the effectiveness of the proposed person re-identification system can be comprehensively assessed, capturing both the overall performance across all matches (mAP) and the practical utility of the system in retrieving relevant matches within the top-k results (Rank-k accuracy).

C. IMPLEMENTATION DETAILS

We employed ResNet-50 as the backbone network, removing the fully connected layer and global average pooling layer. The clothing appearance encoder (E_a) underwent pre-training on ImageNet using ResNet-50, incorporating an adaptive max pooling layer to produce appearance features in the dimensions of $2048\times4\times1$. The structural characteristics of the E_p output are 128 × 64 x 32, primarily comprising four convolutional layers of residual blocks. Since the existing methods all use ResNet-50 as the backbone network in the comparison experiments, using ResNet-50 allows for a consistent benchmark across different methods, and to highlight the contribution of our model in the clothing change. While ResNet-50 may serve as a backbone network for fair comparison, we also integrated a better backbone CMSFL [30] to further improve the performance of our framework. The knowledge embedding configuration used in our model is based on the pretraining module described in KST-GCN [31]. We leveraged the findings from KST-GCN to inform our hyperparameter choices and training configurations, ensuring robustness and effectiveness in our person re-identification framework. Our training batch size was set to 64, activation function was ReLU or WIB-ReLU [32] (clothing change Person Re-ID on Celeb-ReID and PRCC), and the training epoch spanned 60. Adam optimizer was utilized, with an exponential decay rate of 0.5 for first-order moment estimation, 0.99 for second-order moment estimation, and an initial learning rate of 0.1. Our model is implemented using the PyTorch framework (PyTorch 1.1) based on the Python language (Python 3.6). The deep learning setup utilizes a server configuration with CentOS 7 operating system, Intel(R) Xeon(R) E5-2630 v4 CPU, 256GB of memory, and TITAN RTX×4 GPUs, with each GPU having 24GB of graphics memory.

TABLE 1. Comparative Re-identification test results on datasets SLP and UJS-ReID.

METHODS	SLP		UJS-ReID	
	mAP(%)	R1(%)	mAP(%)	R1(%)
without topology	56.3	65.7	68.7	76.3
Distance-based [33]	58.9	65.8	75.9	82.3
ODPR [34]	43.5	49.6	56.8	63.8
PGFA [35]	46.3	60.6	58.2	67.3
HOReID [36]	-	-	71.2	76.4
KTINet(ours)	**63.4**	**68.5**	**78.0**	**85.1**

TABLE 2. Multi-dimensional comparative analysis results.

Methods	layers	parameters	mAP(%)	time(s)
BUFF [37]	100+	10M	44.4	22
TCTS [38]	100+	10M	46.8	23.3
BINet [39]	100+	8M	45.3	16.3
NAE+ [40]	50+	5M	44.0	13.7
CPCCN(ours)	**50+**	**3M**	**46.2**	**11.5**

D. COMPARATIVE EXPERIMENTS

1) COMPARISON OF LOGICAL TOPOLOGY INFERENCE AND PERSON RE-IDENTIFICATION ACCURACY

We conducted comparative experiments on the SLP and UJS-ReID datasets, presenting the results in Table 1. Our method, along with Distance-based [33], ODPR [34], PGFA [35] and HOReID [36], participated in the experiments on UJS-ReID. The comparison table provides a comprehensive view of the

TABLE 3. Performance comparison on clothing change person Re-ID.

Methods	Clothing change Person Re-ID					
	Celeb-ReID			PRCC(Cross-clothes)		
	mAP	rank-1	rank-5	rank-1	rank-10	rank-20
HACNN [41]	9.5	47.6	63.3	21.8	59.4	67.4
PCB [42]	8.2	37.1	57.0	22.8	61.2	78.2
DG-Net [43]	11.4	48.3	65.2	24.3	64.5	79.6
ReIDCaps [24]	15.8	63.0	76.3	-	-	-
SPT [6]	17.2	66.2	79.0	34.3	77.3	88.5
CASE-Net [44]	18.2	66.4	78.1	-	-	-
IRANet [45]	19.0	64.1	78.7	-	-	-
CPCCN (ResNet-50 + ReLU + GRU)	**22.6**	**69.1**	**82.4**	**36.2**	**78.5**	**89.1**
CPCCN (CMSFL + Wib-ReLU + MOPGRU)	**22.8**	**69.2**	**83.0**	**36.4**	**78.9**	**89.8**

TABLE 4. Performance comparison on traditional person Re-ID.

Methods	Traditional Person Re-ID				
	DukeMTMC-ReID		PRCC(Same-clothes)		
	mAP(%)	rank-1(%)	rank-1(%)	rank-10(%)	rank-20(%)
HACNN [41]	63.8	80.5	82.4	98.1	99.0
PCB [42]	65.3	81.9	86.8	98.7	99.6
PGFA [35]	65.5	82.6	-	-	-
ReIDCaps [24]	67.8	83.8	-	-	-
SPT [6]	-	-	64.2	92.6	96.6
CPCCN(ours)	**69.3**	**85.1**	**68.3**	**95.2**	**98.4**

performance of various methods, and it is evident that our method stands out among the competitors. The subsequent discussions will delve into specific aspects of the results.

For KTIN, the rationale behind topology inference can be made interpretable by visualizing the inferred logical topology relationships between cameras. This visualization can include graphical representations of the camera network, showing the connections and relationships between different cameras over time. The inferred logical topology is visualized in Fig. 6. The outcomes illustrate that our method's inferred topology aligns well with the actual dynamic topology across different time periods. The visual representation of the inferred topology is crucial in understanding the model's ability to adapt to dynamic changes. The alignment observed in the figure supports the effectiveness of our approach in capturing the evolving relationships between cameras. Overall, by incorporating interpretability into the knowledge-driven components of the model, users can gain a deeper understanding of how the model operates and can trust the decisions it makes in the context of pedestrian re-identification.

For a more intuitive understanding of the performance of knowledge-driven logical topology inference and the overall person re-identification network, we present the retrieval rate curve in Fig. 7(a). The retrieval rate signifies the accuracy of matching candidates derived from the camera network topology. The retrieval rate curve provides insights into how well the network is leveraging the inferred topology for accurate person re-identification. Analyzing the curve helps in understanding the network's effectiveness in leveraging topological information. Additionally, we assessed the accuracy of our person re-identification framework based on the cross-period clothing change network, depicted in Fig. 7(b). Evaluating the accuracy of person re-identification in the context of clothing changes is a crucial aspect. The results depicted in the figure will be further discussed to understand how well our model handles challenges related to changing clothing appearances over time.

2) COMPARISON OF BACKBONE NETWORKS AT THE EFFICIENCY LEVEL

The proposed model is compared with methods such as BUFF [37], TCTS [38], BINet [39], and NAE+ [40] in terms of network layers, model parameters, identification accuracy, and test time dimensions. Table 2 shows that while our method may not achieve the highest recognition accuracy, it delivers comparable accuracy with significantly fewer model parameters. The comparative analysis in Table 2 highlights a crucial trade-off between model complexity and recognition accuracy. Our model demonstrates efficiency by achieving competitive accuracy with a reduced number of parameters, indicating its potential for resource-constrained scenarios. Additionally, the proposed model exhibits the shortest time cost for recognition, making it more suitable for deployment on terminal devices. The efficiency of our proposed model in terms of test time is a critical advantage, especially for real-world applications where rapid and resource-efficient recognition is essential. This is particularly relevant for deployment on devices with limited computational capabilities. The primary contributing factor is our method's utilization of logical topology for searching

pedestrian targets in video feeds, improving the efficiency of the recognition process. Leveraging logical topology for target search demonstrates a novel and effective approach to enhance the efficiency of the recognition process. This aligns with the practical considerations of real-world deployments, emphasizing the importance of not only achieving high accuracy but also optimizing resource utilization.

3) ROBUSTNESS COMPARISON OF THE OVERALL MODEL

The proposed model is rigorously compared with other clothing change re-identification methods, such as HACNN [41], PCB [42], DG-Net [43], ReIDCaps [24], SPT [6], CASE-Net [44] and IRANet [45], on Celeb-ReID and PRCC (Cross-clothes). The results presented in Table 3 affirm the effectiveness of our model in addressing clothing changes. This is particularly crucial in scenarios where individuals may undergo variations in clothing over time, as demonstrated by its superior performance compared to specialized clothing change re-identification methods. Additionally, traditional re-identification methods are compared with our model on DukeMTMC-ReID and PRCC (Same clothes), as depicted in Table 4, showcasing the efficacy of our model when contrasted with traditional re-identification approaches. The comparison against traditional re-identification methods emphasizes the versatility of our model, excelling not only in scenarios with clothing changes but also in traditional re-identification contexts. This versatility positions our proposed model as a robust solution applicable across various real-world scenarios. Table 3 indicates a notable disparity in the results between traditional re-identification methods [41], [43], [46] and those designed for changing clothes re-identification. This observation underscores the inadequacy of relying solely on clothing appearance for identifying individuals over time. It underscores the need for a comprehensive approach, as adopted by our model, that considers multiple factors for robust and accurate re-identification. Our CPCCN achieves an accuracy of 69.1% at rank-1 and a mAP of 22.6%, demonstrating superior performance compared to traditional re-identification methods in scenarios involving clothing changes. These metrics underscore the practical viability of our proposed model in real-world applications where individuals may exhibit variations in clothing over time. The proposed model excels in addressing the traditional re-identification problem with changes in clothing, thanks to the incorporation of classification and reconstruction losses that effectively constrain features. This dual-focus approach enhances feature robustness by considering both global and local features. Our model may also generate erroneous classification results under various circumstances. Firstly, when individuals undergo clothing changes over time intervals, the model may misclassify them as different entities. Additionally, if multiple pedestrians wear similar clothing during the same time period, the model may erroneously identify them as the same individual. Variations in pose, occlusions, and changes in illumination can further contribute to misclassification. Moreover, inadequate or non-representative training data may limit the model's ability to capture diverse variations, leading to erroneous classifications.

TABLE 5. Ablation study results on Celeb-ReID and PRCC (Cross-clothes).

Dataset	Model	mAP	rank-1	rank-5
Celeb-ReID	baseline	5.8	43.3	54.6
	baseline+KTIN	8.1	45.2	56.8
	baseline+CPCCN	17.6	62.1	76.2
	Ours	**22.6**	**69.1**	**82.4**
Dataset	Model	rank-1	rank-10	rank-20
PRCC (Cross-clothes)	baseline	19.4	52.3	66.4
	baseline+KTIN	27.6	63.5	75.3
	baseline+CPCCN	32.1	75.2	84.4
	Ours	**36.2**	**78.5**	**89.1**

E. ABLATION STUDY

To showcase the effectiveness of each proposed component, we conducted ablation experiments. Table 5 presents the results of our proposed method using DG-Net as the baseline, validating the effectiveness of KTIN and CPCCN on Celeb-ReID and PRCC (Cross-clothes), respectively. The results underscore the significance of each component in enhancing the overall model performance, highlighting their complementary roles in addressing the challenges of re-identification under various conditions. The symbol "+" indicates the inclusion of the following components. The experiments demonstrate that all our proposed components contribute significantly to the overall model performance.

After comparing with other existing pedestrian re-identification models, we also explored the various components of the model in ablation experiments. In the pose encoder module, we utilize generative adversarial networks to swap the poses of the target pedestrian and candidate pedestrians, ensuring consistency in poses between them during the pedestrian matching stage, thereby achieving pose normalization. This approach effectively reduces intra-class differences and the influence of pose variations. Unlike previous pose-based pedestrian re-identification methods, our module does not require additional pose feature extraction tools but directly uses a pose encoder for pose extraction. To prevent interference from other features, we grayscale the images before pose extraction to ensure the robustness of pose features. We conducted visual analysis of the pose encoder module. As shown in Fig. 8, we use a pose encoder to extract pose features from the original image of the target pedestrian and different pose providers. Then, using generative adversarial networks, we perform pose normalization based on the provided pose features. The results demonstrate that our pose encoder module accurately swaps the poses of the target pedestrian based on different pose features while ensuring that the clothing features of the target pedestrian are not affected by pose features. This is achieved by incorporating reconstruction loss into the pose encoder module to enhance feature decoupling capability,

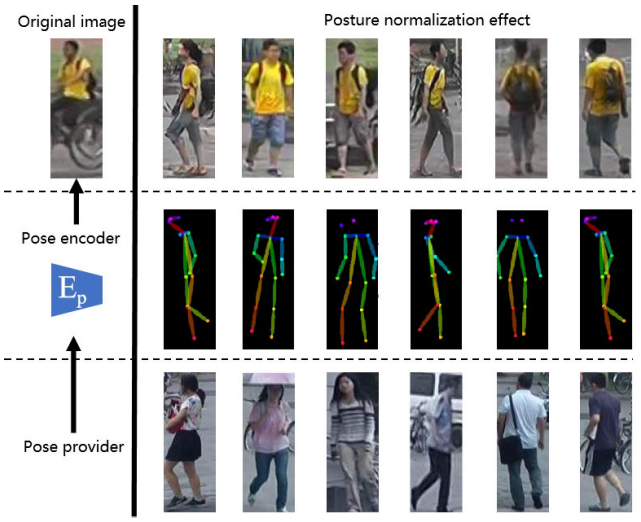

FIGURE 8. Posture normalization effect.

ultimately maintaining the same pose as the pose provider during the pedestrian matching stage to avoid matching errors caused by misalignment of features.

To further demonstrate the effectiveness of our approach, we conducted visual analysis by Style Erasure Method [47]. As shown in Fig. 9 (a), the input pedestrian image is first subjected to style erasure using instance normalization to obtain features ($\overline{F}$) that exclude style features. It can be clearly observed from the figure that some discriminative features of the pedestrian are lost during the style erasure process. To address these lost features (D_i), we use content attention mechanism to separately extract discriminative features (D_i^+) and task-irrelevant features (D_i^-), and finally integrate them with the style-erased features. From the visual results, it can be seen that if the erased features are integrated with task-irrelevant features, it cleverly avoids the discriminative features of the pedestrian, resulting in noise unrelated to the pedestrian. Conversely, if integrated with the lost discriminative features, it produces images focusing on the discriminative features of the pedestrian. On the other hand, we also conducted visual ablation experiments on our approach, as shown in Fig. 9 (b). We visualized the attention regions of pedestrians in environments of original, contrast ratio change, and illumination change, and compared them with the baseline. In the original environment, there is little difference in performance between our approach and the baseline, as both can focus on the discriminative features of the pedestrian. In the contrast ratio change environment, our approach can focus on more discriminative features compared to the baseline, which is affected by background noise. In the illumination change environment, the performance difference between our approach and the baseline is more significant. This is because our model can effectively erase style differences in the images, while the baseline is heavily affected by background noise, leading to its inability to focus on the discriminative features of the pedestrian and thus affecting the recognition performance of the model.

As depicted in Fig. 10, ablation experiments were conducted on the CPCCN to assess the network's performance in recognizing pedestrians undergoing clothing changes, offering a comprehensive understanding of how each component influences the model's ability to handle clothing changes. Visualizations of the generated images depicting clothing changes and an example of cross-period clothing change generation are presented in Fig. 11. Examining specific examples of cross-period clothing change generation enhances our understanding of the model's performance in scenarios where clothing characteristics evolve over time. These visual results complement the quantitative evaluation, providing a more holistic assessment of the proposed approach.

F. DISCUSSION OF SCALABILITY OF CPCCN

Scalability is a crucial aspect to consider when evaluating the effectiveness of a person re-identification model, especially in environments with diverse and large-scale camera networks. The scalability of CPCCN can be enhanced by leveraging graph-based representations of camera networks. Graph structures provide a flexible and scalable framework for representing complex relationships among cameras in diverse environments. As the number of cameras increases, graph-based representations can efficiently capture and model the topology of large-scale camera networks. To handle large-scale camera networks, the model can employ distributed processing techniques. By distributing computation across multiple nodes or processing units, the model can effectively scale to accommodate the computational demands of analyzing data from diverse cameras distributed across a wide geographic area.

CPCCN can support incremental learning techniques to adapt and scale to changing environments. By continuously updating its parameters based on new data from additional cameras or evolving conditions, the model can maintain its performance and adaptability in dynamic environments with diverse camera networks. Scalability also involves efficient utilization of computational resources. The model can be optimized to minimize resource consumption while maximizing performance, enabling it to scale effectively in resource-constrained environments or scenarios with limited computational resources. A modular architecture can enhance scalability by allowing for the integration of additional components or modules as needed. The model can be designed in a modular fashion, with well-defined interfaces between components, making it easier to scale and adapt to different environments or requirements.

The model can leverage parallel processing techniques to expedite computation and analysis across multiple cameras simultaneously. By parallelizing tasks such as feature extraction, inference, or graph traversal, the model can efficiently handle the computational load associated with large-scale camera networks. In environments where real-time operation

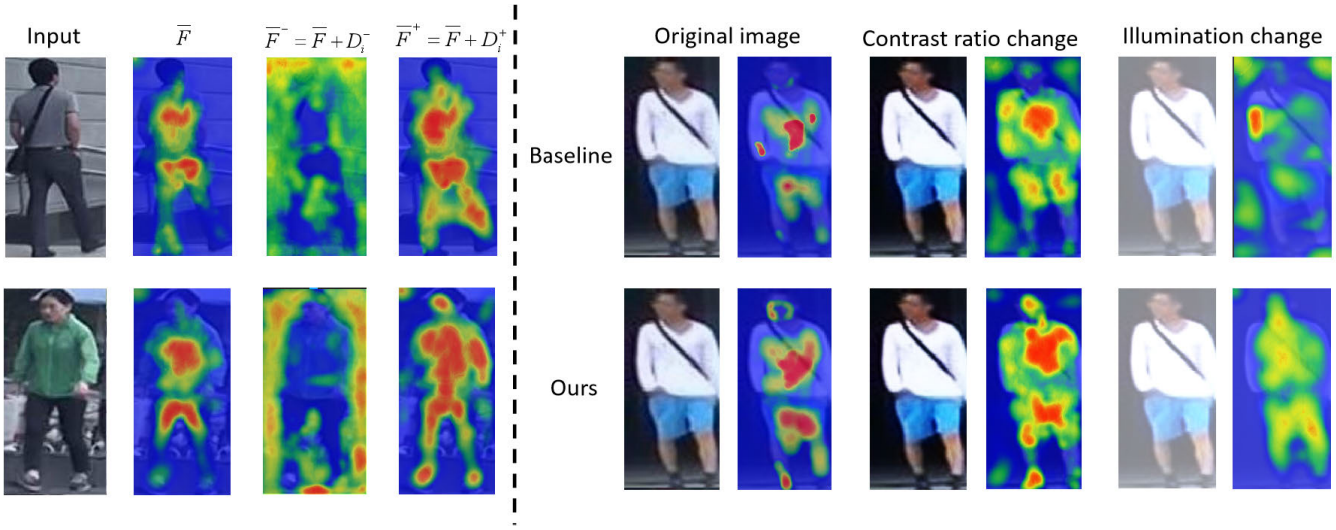

FIGURE 9. Ablation experiments via style erasure in different environments.

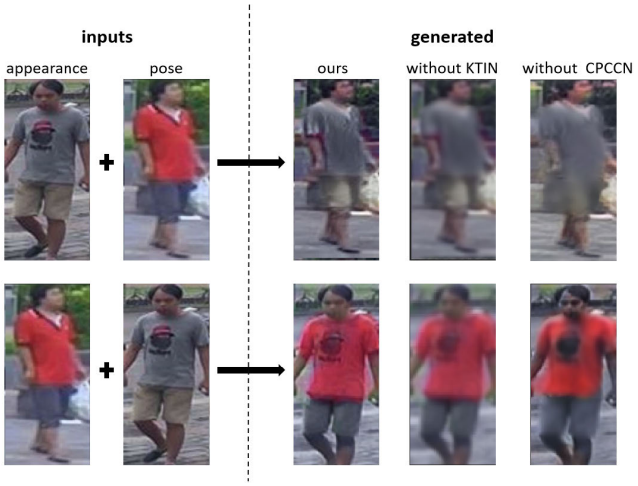

FIGURE 10. Results of clothing change generation.

FIGURE 11. Example of cross-period clothing change generation.

is critical, the model can be optimized for low-latency processing. By prioritizing efficiency and minimizing processing delays, the model can effectively handle the demands of real-time re-identification in diverse and large-scale camera networks.

Overall, the scalability of the model is essential for its practical applicability in environments with diverse and large-scale camera networks. By adopting strategies such as graph-based representation, distributed processing, incremental learning, resource efficiency, modular architecture, parallel processing, and optimization for real-time operation, the model can effectively scale to meet the challenges of the person re-identification in complex and dynamic environments.

V. CONCLUSION

This paper introduces a comprehensive framework that effectively addresses the intricate challenges associated with changing clothes in the context of pedestrian re-identification. The variability introduced by clothing changes can significantly disrupt the consistency of identity representation. In response to this, our proposed framework comprises two key components: the Knowledge-Driven Topology Inference Network (KTIN) and the Cross-Period Clothing Change Network (CPCCN).

The Knowledge-Driven Topology Inference Network (KTIN) is specifically designed to counteract the influence of external factors and clothing appearances. By capturing logical topology relationships among cameras, KTIN provides a robust foundation for understanding the spatial and temporal connections between different camera nodes. This is a crucial aspect in maintaining accurate identity consistency amidst changing clothing. The Cross-Period Clothing Change Network (CPCCN) serves as a pivotal element in predicting the clothing characteristics of target pedestrians at distinct time intervals. This network is tailored to accommodate the challenges posed by temporal variations in clothing, offering a nuanced understanding of how clothing styles evolve over time. To synergize the capabilities of KTIN and CPCCN, a joint optimization mechanism is introduced. This strategic combination aims to boost the overall performance of pedestrian re-identification. By jointly refining the logical topology relationships and enhancing clothing change predictions, the framework achieves a comprehensive and integrated solution. Extensive experiments were conducted on diverse datasets, encompassing both traditional pedestrian

re-identification scenarios and datasets specifically tailored to clothing change scenarios. The results of these experiments provide compelling evidence for the effectiveness of the proposed framework. The framework not only demonstrates proficiency in handling conventional re-identification challenges but also excels in scenarios where clothing changes pose additional complexities.

In real-world deployment, the use of person re-identification technology raises significant ethical concerns regarding privacy, surveillance, and potential misuse of personal data. The proposed model should prioritize the protection of individuals' privacy by ensuring that personal data, such as images and biometric features, are handled securely and anonymously. It should implement techniques such as anonymization, data encryption, and access control mechanisms to safeguard sensitive information. To minimize the risk of privacy breaches, the model should only collect and retain data necessary for its intended purpose. It should avoid unnecessary data collection and implement data minimization techniques to reduce the potential impact on individuals' privacy. The proposed model should encourage public engagement and dialogue on the ethical implications of Re-ID technology. It should facilitate discussions among stakeholders, including policymakers, researchers, industry experts, and civil society organizations, to develop ethical guidelines, regulations, and best practices for the responsible deployment of Re-ID technology. By addressing these ethical considerations, the proposed model can help ensure that Re-ID technology is deployed in a manner that respects individuals' privacy, rights, and dignity, while also promoting public safety and security.

In conclusion, our proposed framework presents a holistic and effective approach to tackle the multifaceted challenges associated with changing clothes in pedestrian re-identification. The integration of KTIN and CPCCN, coupled with joint optimization, showcases the adaptability and robustness of the framework across various scenarios, making it a valuable contribution to the field of person re-identification research.

ACKNOWLEDGMENT

The authors would like to express their sincere thanks to the editors and the anonymous reviewers for their valuable comments and contributions.

(Shuxin Zheng, Sai Liang, and Chunyun Meng contributed equally to this work.)

REFERENCES

[1] B. A. U. Olimov, K. C. Veluvolu, A. Paul, and J. Kim, "UzADL: Anomaly detection and localization using graph Laplacian matrix-based unsupervised learning method," *Comput. Ind. Eng.*, vol. 171, Sep. 2022, Art. no. 108313.

[2] A. A. U. Rakhmonov, B. Subramanian, B. Olimov, and J. Kim, "Extensive knowledge distillation model: An end-to-end effective anomaly detection model for real-time industrial applications," *IEEE Access*, vol. 11, pp. 69750–69761, 2023.

[3] K. Jiang, T. Zhang, X. Liu, B. Qian, Y. Zhang, and F. Wu, "Cross-modality transformer for visible-infrared person re-identification," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2022, pp. 480–496.

[4] Z. Zhang, H. Zhang, S. Liu, Y. Xie, and T. S. Durrani, "Part-guided graph convolution networks for person re-identification," *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 108155.

[5] F. Wan, Y. Wu, X. Qian, Y. Chen, and Y. Fu, "When person re-identification meets changing clothes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 3620–3628.

[6] Q. Yang, A. Wu, and W.-S. Zheng, "Person re-identification by contour sketch under moderate clothing change," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2029–2046, Jun. 2021.

[7] J. Zheng, X. Hu, T. Xiang, and P. P. K. Chan, "Dual-path model for person re-identification under cloth changing," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Dec. 2020, pp. 291–297.

[8] Y. Dong, H. Che, M.-F. Leung, C. Liu, and Z. Yan, "Centric graph regularized log-norm sparse non-negative matrix factorization for multi-view clustering," *Signal Process.*, vol. 217, Apr. 2024, Art. no. 109341.

[9] Y. Cai, H. Che, B. Pan, M.-F. Leung, C. Liu, and S. Wen, "Projected cross-view learning for unbalanced incomplete multi-view clustering," *Inf. Fusion*, vol. 105, May 2024, Art. no. 102245.

[10] A. Wu, W.-S. Zheng, and J.-H. Lai, "Distilled camera-aware self training for semi-supervised person re-identification," *IEEE Access*, vol. 7, pp. 156752–156763, 2019.

[11] J. Liu, Z.-J. Zha, D. Chen, R. Hong, and M. Wang, "Adaptive transfer network for cross-domain person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7195–7204.

[12] X. Jin, C. Lan, W. Zeng, and Z. Chen, "Global distance-distributions separation for unsupervised person re-identification," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 735–751.

[13] G. Chen, Y. Lu, J. Lu, and J. Zhou, "Deep credible metric learning for unsupervised domain adaptation person re-identification," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*. Glasgow, U.K.: Springer, Aug. 2020, pp. 643–659.

[14] S. Zhang, Y. Zeng, H. Hu, and S. Liu, "Noise resistible network for unsupervised domain adaptation on person re-identification," *IEEE Access*, vol. 9, pp. 60740–60752, 2021.

[15] S. Zhang and H. Hu, "Unsupervised person re-identification using unified domanial learning," *Neural Process. Lett.*, vol. 55, no. 6, pp. 6887–6905, Dec. 2023.

[16] T. Ellis, D. Makris, and J. Black, "Learning a multi-camera topology," in *Proc. Joint IEEE Workshop Vis. Surveill. Perform. Eval. Tracking Surveill. (VS-PETS)*, 2003, pp. 165–171.

[17] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 106–129, Oct. 2010.

[18] O. Javed and M. Shah, *Tracking in Multiple Cameras With Disjoint Views*. Cham, Switzerland: Springer, 2008, pp. 1–26.

[19] Y.-J. Cho, S.-A. Kim, J.-H. Park, K. Lee, and K.-J. Yoon, "Joint person re-identification and camera network topology inference in multiple cameras," *Comput. Vis. Image Understand*, vol. 180, pp. 34–46, Mar. 2019.

[20] I. B. Barbosa, M. Cristani, A. D. Bue, L. Bazzani, and V. Murino, "Re-identification with RGB-D sensors," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 433–442.

[21] M. Munaro, A. Basso, A. Fossati, L. Van Gool, and E. Menegatti, "3D reconstruction of freely moving persons for re-identification with a depth sensor," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 4512–4519.

[22] A. Haque, A. Alahi, and L. Fei-Fei, "Recurrent attention models for depth-based person identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1229–1238.

[23] P. Zhang, Q. Wu, J. Xu, and J. Zhang, "Long-term person re-identification using true motion from videos," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 494–502.

[24] Y. Huang, J. Xu, Q. Wu, Y. Zhong, P. Zhang, and Z. Zhang, "Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3459–3471, Oct. 2020.

[25] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–17.

[26] Y. Huang, Q. Wu, J. Xu, and Y. Zhong, "Celebrities-ReID: A benchmark for clothes variation in long-term person re-identification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[27] S. Yu, S. Li, D. Chen, R. Zhao, J. Yan, and Y. Qiao, "COCAS: A large-scale clothes changing person dataset for re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3397–3406.

[28] M. Sun, Z. Liu, and Y. Lin, "Knowledge representation learning with entities, attributes and relations," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 41–52, Jan. 2016.
[29] B. Subramanian, B. Olimov, S. M. Naik, S. Kim, K.-H. Park, and J. Kim, "An integrated mediapipe-optimized GRU model for Indian sign language recognition," *Sci. Rep.*, vol. 12, no. 1, p. 11964, Jul. 2022.
[30] B. Olimov, B. Subramanian, R. A. A. Ugli, J.-S. Kim, and J. Kim, "Consecutive multiscale feature learning-based image classification model," *Sci. Rep.*, vol. 13, no. 1, p. 3595, Mar. 2023.
[31] J. Zhu, X. Han, H. Deng, C. Tao, L. Zhao, P. Wang, T. Lin, and H. Li, "KST-GCN: A knowledge-driven spatial–temporal graph convolutional network for traffic forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15055–15065, Sep. 2022.
[32] B. Olimov, S. Karshiev, E. Jang, S. Din, A. Paul, and J. Kim, "Weight initialization based-rectified linear unit activation function to improve the performance of a convolutional neural network model," *Concurrency Computation, Pract. Exper.*, vol. 33, no. 22, p. e6143, Nov. 2021.
[33] Y.-J. Cho and K.-J. Yoon, "Distance-based camera network topology inference for person re-identification," *Pattern Recognit. Lett.*, vol. 125, pp. 220–227, Jul. 2019.
[34] N. Jiang, S. Bai, Y. Xu, C. Xing, Z. Zhou, and W. Wu, "Online inter-camera trajectory association exploiting person re-identification and camera topology," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1457–1465.
[35] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, "Pose-guided feature alignment for occluded person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 542–551.
[36] G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou, and J. Sun, "High-order information matters: Learning relation and topology for occluded person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6448–6457.
[37] W. Yang, H. Huang, X. Chen, and K. Huang, "Bottom-up foreground-aware feature fusion for practical person search," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 262–274, Jan. 2022.
[38] C. Wang, B. Ma, H. Chang, S. Shan, and X. Chen, "TCTS: A task-consistent two-stage framework for person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11949–11958.
[39] M. Liu, Y. Zhang, J. Xu, and Y. Chen, "Deep bi-directional interaction network for sentence matching," *Int. J. Speech Technol.*, vol. 51, no. 7, pp. 4305–4329, Jul. 2021.
[40] D. Chen, S. Zhang, J. Yang, and B. Schiele, "Norm-aware embedding for efficient person search and tracking," *Int. J. Comput. Vis.*, vol. 129, no. 11, pp. 3154–3168, Nov. 2021.
[41] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2285–2294.
[42] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling and a strong convolutional baseline," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 480–496.
[43] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2133–2142.
[44] Y.-J. Li, X. Weng, and K. M. Kitani, "Learning shape representations for person re-identification under clothing change," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2431–2440.
[45] W. Shi, H. Liu, and M. Liu, "IRANet: Identity-relevance aware representation for cloth-changing person re-identification," *Image Vis. Comput.*, vol. 117, Jan. 2022, Art. no. 104335.
[46] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling and a strong convolutional Baseline," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2018, pp. 501–518.
[47] X. Jin, C. Lan, W. Zeng, and Z. Chen, "Style normalization and restitution for domain generalization and adaptation," *IEEE Trans. Multimedia*, vol. 24, pp. 3636–3651, 2022.

SHUXIN ZHENG received the Ph.D. degree from the University of Toyama, Japan, in 2020. She is currently with the School of Management, Guangdong University of Science and Technology, China. Her current research interests include tourism, service management, human resource management, and artificial neural network application in tourism decision-making and service management.

SAI LIANG received the B.E. degree from Qingdao Institute of Technology, Qingdao, China. He is currently pursuing the M.Eng. degree with the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China. His research interests include computer vision and pattern recognition.

CHUNYUN MENG received the M.Eng. degree from the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China. He is currently pursuing the Ph.D. degree with the Faculty of Electrical, Information and Communication Engineering, Kanazawa University, Kanazawa, Japan. His research interests include computer vision, natural language processing, and artificial neural network application in information management.

ZHONGGUO ZHANG received the Ph.D. degree in engineering mechanics from Beijing Institute of Technology, Beijing, China, in 2004. He is currently a Professor with the School of Intelligent Information Engineering, Guangdong Vocational College of Hotel Management, Dongguan, China. His current research interests include the dynamic response of the structure and the artificial neural network application in engineering.

LI LUAN received the B.A. degree from Beijing Language and Culture University, Beijing, China. She is currently pursuing the J.M. degree with the School of Public Affairs, University of Science and Technology of China, Hefei, China. Her research interests include intellectual property and the artificial neural network application in intellectual property.