




Article

Dynamic Screening Strategy Based on Feature Graphs for UAV Object and Group Re-Identification

Guoqing Zhang ^{1,2,3} , Tianqi Liu ¹ and Zhonglin Ye ^{4,*}

¹ School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China; guoqingzhang@nuist.edu.cn (G.Z.); liutianqi@nuist.edu.cn (T.L.)

² Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210094, China

³ Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science & Technology, Nanjing 210044, China

⁴ The State Key Laboratory of Tibetan Intelligent Information Processing and Application, Qinghai Normal University, Xining 810008, China

* Correspondence: yezhonglin@qhnu.edu.cn

Abstract: In contemporary times, owing to the swift advancement of Unmanned Aerial Vehicles (UAVs), there is enormous potential for the use of UAVs to ensure public safety. Most research on capturing images by UAVs mainly focuses on object detection and tracking tasks, but few studies have focused on the UAV object re-identification task. In addition, in the real-world scenarios, objects frequently get together in groups. Therefore, re-identifying UAV objects and groups poses a significant challenge. In this paper, a novel dynamic screening strategy based on feature graphs framework is proposed for UAV object and group re-identification. Specifically, the graph-based feature matching module presented aims to enhance the transmission of group contextual information by using adjacent feature nodes. Additionally, a dynamic screening strategy designed attempts to prune the feature nodes that are not identified as the same group to reduce the impact of noise (other group members but not belonging to this group). Extensive experiments have been conducted on the Road Group, DukeMTMC Group and CUHK-SYSU-Group datasets to validate our framework, revealing superior performance compared to most methods. The Rank-1 on CUHK-SYSU-Group, Road Group and DukeMTMC Group datasets reaches 71.8%, 86.4% and 57.8%, respectively. Meanwhile, our method performance is explored on the UAV datasets of PRAI-1581 and Aerial Image, the infrared datasets of SYSU-MM01 and CM-Group and the NIR dataset of RBG-NIR Scene dataset; the unexpected findings demonstrate the robustness and wide applicability of our method.

Keywords: group re-identification; UAV object re-identification; graph neural networks; feature matching



Citation: Zhang, G.; Liu, T.; Ye, Z. Dynamic Screening Strategy Based on Feature Graphs for UAV Object and Group Re-Identification. *Remote Sens.* **2024**, *16*, 775. <https://doi.org/10.3390/rs16050775>

Academic Editors: Claudio Piciarelli, Pedro Melo-Pinto, Danilo Avola, Alessio Mecca and Marco Cascio

Received: 3 January 2024

Revised: 16 February 2024

Accepted: 20 February 2024

Published: 22 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Object re-identification (re-ID) is a technique used to match objects observed across different cameras without overlaps, which involves some challenges, including issues related to low resolution, occlusion and changing poses and illuminations, etc. In the last decade, many single-object re-ID approaches have been proposed and achieved great success in both public safety and video surveillance [1–7]. However, there are few works focused on object re-identification in the context of Unmanned Aerial Vehicles (UAVs). The reason for this can be attributed to the time-consuming nature of annotating datasets for object re-identification in UAVs, which necessitates annotating both the positions of bounding boxes and assigning ID numbers to each object. In addition, objects frequently get together in groups in real-world scenarios.

Group re-ID focuses on aligning groups of objects observed from various camera viewpoints. Most existing methods utilize contextual information from neighbors to assist

re-identify each group member within groups [8–11]. Utilizing group context information has demonstrated its effectiveness in mitigating ambiguity in single-object re-ID [10,12,13]. Therefore, the process of group re-ID can effectively be applied to person re-ID in groups. In comparison to single-object re-ID, group re-ID presents greater challenges because of the changes in group members and layout. The person as an object and group re-ID are shown in Figure 1a,b. The current group re-ID methods can be categorized into two types: feature aggregation-based [11,14] and crowd matching-based methods [15,16]. The first category of method aggregates the contextual information from neighbors to re-identifying each group members within groups. For example, Yan et al. [14] devised a multilevel attentional mechanism to capture intra- and inter-group contexts and suggested a novel uniform framework for group re-ID. However, these approaches neglect the significance of background features, which can be used to assist message transmission between group nodes. The second category of method usually adopts a single- or multi-object matching strategy and assigns important weights to objects to capture the dynamic changes within the group. For example, Lisanti et al. [16] presented multi-granularity representation for group images, which they combined with a dynamic weighting approach to obtain improved people matching. However, these methods cannot solve the problem of group members changes well.

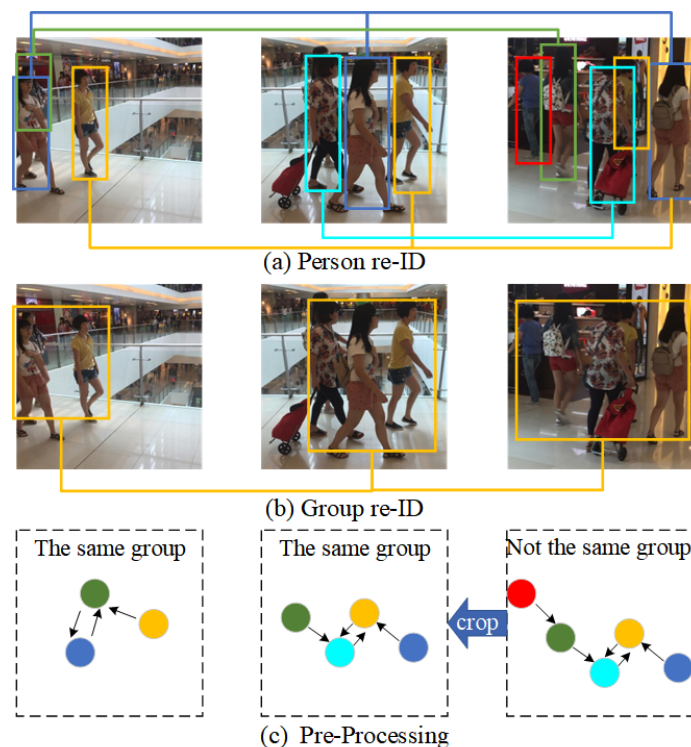


Figure 1. Demonstration of (a) single person re-ID and (b) group re-ID. (c) The Pre-Processing module of our proposed method, while each graph structure is constructed based on the corresponding pedestrians in Figure 1a. As shown in the rightmost graph of Figure 1c, it is determined whether the nodes in the graph belong to the same group. If not, the red node (meaning additional pedestrians) will be “cropped” until all pedestrians belong to the same group.

To address the aforementioned challenges, a novel dynamic screening strategy based on a feature graphs framework is proposed for UAV object and group re-ID. In particular, the Pre-Processing module (PREP) designed aims to remove objects that do not belong to the current group, as shown in Figure 1c. Specifically, the group pairs are as input and the CNN model is employed to extract person and background features. In the PREP module, the person features are regarded as nodes to construct the context graphs using the k-nearest neighbor algorithm. First, if the graph is connected, the graph nodes are

determined whether in the same group. When these nodes are not in the same group, the graph nodes and their connected edges are pruned. Otherwise, the original graph structure is retained without any processing. If the graph is not connected, the model selects a sub-graph and repeat the above steps until the nodes in the graph belong to the same group. Finally, the reconstructed graphs do not contain the persons that do not belong to the current group.

Here, the Graph-based Feature Matching (GFM) module presented attempts to enhance the transmission of group contextual information by using adjacent feature nodes. In this module, matched background features can be obtained using the method of feature matching. The background features are regarded as newly added nodes to the reconstructed graph. Inspired by the method of Multi-Attention Context Graph [14], the Multi-Objects Context Graph (MOCCG) module designed tries to capture context messages of objects by adopting multi-level attention mechanism within and between graphs. It is noteworthy that accuracy of group re-ID has seen enhancement to some extent through continuous messages transmission between person nodes and new nodes, as well as continuous updates of nodes in the context graph.

The main contributions of this paper are as follows:

- The Graph-based Feature Matching module designed attempts to enhance the transmission of group contextual information by adding matched features for the first time.
- The Pre-Processing module presented aims to solve the challenge of group member changes in the group re-ID, which can remove group members that do not belong to the current group through pruning operations.
- The Multi-Objects Context Graph module proposed tries to employ multi-level attention mechanisms within and between graphs to capture contextual information.
- Our proposed framework is examined on several datasets, and the experimental results demonstrate that the model surpasses the most advanced approaches.

2. Related Work

2.1. Single Object Re-Identification

Recently, a lot of attention has been attracted by single object re-identification from the academic and industrial perspectives in computer vision [17,18]. It faces many challenges, such as the variety of views [19,20], poor image resolutions [3,21], illumination variations [22], unconstrained postures [23–25], occlusions [26], heterogeneous modes [27,28], complex camera environments [29], backdrop clutter and unreliable boundary box generation. Early studies primarily focused on manual body structure feature construction [30–32] or distance metric learning [33–37]. Single object re-identification has shown encouraging results on commonly utilized benchmarks [38–40]. However, few works are focused on Unmanned Aerial Vehicles (UAVs) object re-ID. In this work, our model are applied to the task of UAV person re-ID for the first time, demonstrating the strong generalization ability of our proposed model.

2.2. Group Re-Identification

In comparison to single person re-identification, group re-identification aims to evaluate the group images' similarities and few works have been proposed [41–43]. For example, Cai et al. [8] earliest presented the covariance descriptor to match groups. To match the two patch sets, Zhu et al. [9] constructed group re-ID as a patch-matching task, and the patch was placed in the "Boosted Saliency Channels". To address the spatial displacement arrangement of individuals within the group, Zheng et al. [10] developed a visual descriptor and sparse feature coding approach. Then, Lisanti et al. [16] presented multi-granularity representation for group images, which they combined with a dynamic weighting approach to obtain improved people matching. With the widespread adoption of graph neural networks in recent years, Zhu et al. [11] developed a network to group analysis where context is represented by a graph. The model uses a novel technique to aggregate information from neighboring nodes using features derived from SKNNG, such as node entry degree and spa-

tial interactions between nearby nodes (i.e., relative distance and direction). Yan et al. [14] devised a novel uniformed framework to capture group-level contexts with multilevel attentional mechanism. Nevertheless, the majority of the approaches mentioned cannot solve the changes of group members well. In this work, our proposed PREP module can effectively remove the object (person) who does not belong to this group.

2.3. Graph Neural Network

Graph Neural Networks (GNNs) aim to use graphs to represent information and extract features to replace conventional handwritten features [44–46]. GNNs have been employed in a multitude of visual tasks, including object detection, semantic segmentation, and solving visual question-answering problems. The fundamental concept involves creating node embeddings using neural networks through the aggregation of local information. In recent times, there has been a growing fascination with expanding convolution operations within GNNs. Kipf et al. [44] presented the Graph Convolutional Networks (GCNs), utilizing convolutional operations to gather messages from 1-step neighbourhoods surrounding each node. Velivckovic et al. [45] presented the Graph Attention Networks (GATs), employing a self-attention mechanism for the calculation of weights to gather data from neighboring nodes. Wu et al. [46] proposed the Graph Attention Model (GAM), which addresses the graph classification problem through adaptive access to the sequence of each significant node for information processing.

3. Method

3.1. Overview

Group re-ID faces significant challenges because it not only suffers from the same problems as single object re-ID, but also faces other challenges, such as variations in the number of group members. Therefore, the key to re-identifying a group is that the model can be capable of removing group members who do not belong to this group.

In response to the aforementioned challenge, a novel dynamic screening strategy based on feature graphs framework (DSFG) is proposed for group re-ID and the pipeline is shown in Figure 2. Overall, the DSFG framework can be viewed as a uniform approach with the capacity to (1) extracting the features of single person and background by CNN respectively, (2) removing members that do not belong to the group and (3) using background features to enhance the transmission of group contextual information to update individual features. Specifically, the group pairs are used as input to extract person and background features from each group. The person features are regarded as nodes, with the contextual graphs constructed using the k-nearest neighbor algorithm. The Pre-Processing module proposed attempts to remove the members who do not belong to current group. At the same time, the Graph-based Feature Matching module designed aims to enhance the transmission of group contextual information by using adjacent background semantic information. Subsequently, the background features are added as new nodes to the reconstructed graph. Finally, the Multi-Objects Context Graph (MOCG) module updates the node features in the contextual graph with both intra-group and inter-group information. The architectural details are clarified in the following subsections.

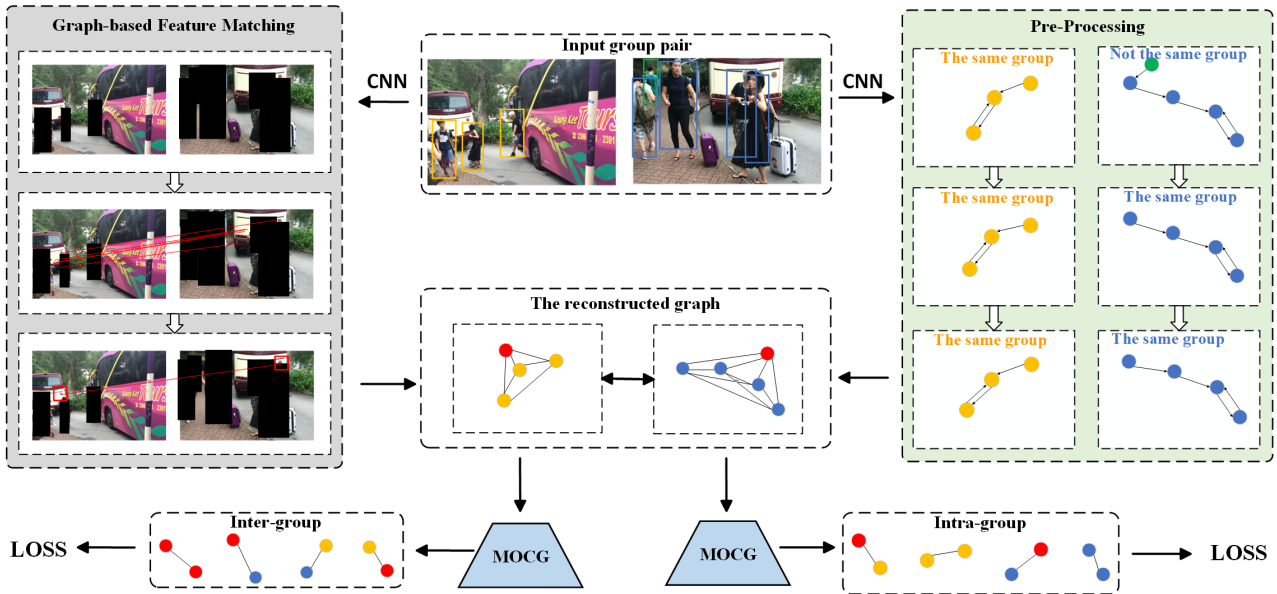


Figure 2. Illustration of our proposed framework. First, group pairs are used as input and the model extracts features of single persons and areas that do not contain a person (background information) by CNN respectively. Then, a contextual graph with complete connectivity is constructed using person features as nodes by k-nearest algorithm and removes pedestrians (marked by green box and green node) that do not belong to the same group through the Pre-Processing module. Meanwhile, the Graph-based Feature Matching module aims to obtain matched background features (marked with red boxes) that do not contain pedestrians. Subsequently, the background features as nodes are added to the reconstructed graph. Finally, the node features within the contextual graph are transferred while considering both intra- and inter-group information in the MOCG module.

3.2. Graph-Based Feature Matching Module

Inspired by the SuperGlue [47] method, to extract the features that do not contain a person, the areas within the person bounding boxes are first removed in each image, and then the background semantic information is obtained through the method of feature matching. Given two images A and B , each contains keypoint positions denoted as \mathbf{p} and their corresponding description visuals denoted as \mathbf{d} , which are combined into local features (\mathbf{p}, \mathbf{d}) . The i -th keypoint positions are composed of the coordinates (x, y) within images, along with a detection confidence score c , $\mathbf{p}_i = (x, y, c)_i$.

To incorporate keypoint i to high-dimensional vector, the original keypoint representation $x_i^{(0)}$ is mixed its description visual with its position through Multilayer Perceptron (MLP), as follows:

$$x_i^{(0)} = \mathbf{d}_i + MLP_{enc}(\mathbf{p}_i). \quad (1)$$

The keypoints features are constructed to unified complete graph as the nodes. To facilitate information aggregation along edges within graphs, the message passing [48] combined with GNN can be calculated to aggregate representation at every level from other nodes. The intermediary representation for the keypoint i of image A in the ℓ -th layer presents $x_i^{(\ell)}$. And final message passing with all keypoints in A is updated:

$$x_i^{(\ell+1)} = x_i^{(\ell)} + MLP([\cdot || x_i^{(\ell)} || \mathbf{m}_{\varepsilon \rightarrow i}]), \quad (2)$$

where $[\cdot || \cdot]$ indicates concatenation and $\mathbf{m}_{\varepsilon \rightarrow i}$ means the results of aggregating information from all other keypoints to keypoint i . The ultimate matching descriptors are obtained through linear projection:

$$\mathbf{f}_i^A = \mathbf{W} \cdot x_i^{(L)} + \mathbf{b}, \forall i \in \mathcal{A}, \quad (3)$$

and the keypoints in image B are in the same manner. The matching descriptors f_i^A and f_j^B are combined as the final matching descriptor to calculate the pairwise score:

$$S_{i,j} = \langle f_i^A, f_j^B \rangle \quad \forall (i,j) \in \mathcal{A} \times \mathcal{B}, \quad (4)$$

where $\langle \cdot, \cdot \rangle$ is the inner product. The module has been trained using a supervised methodology and verified with the matches of ground truth $\mathcal{M} = \{(i,j) \in \mathcal{A} \times \mathcal{B}\}$ where $I \subseteq \mathcal{A}$ and $J \subseteq \mathcal{B}$ as not matched when they are not near any reprojection. The background matching loss function L^{match} is calculated as:

$$L^{match} = - \sum_{(i,j) \in \mathcal{M}} \log \bar{P}_{i,j} - \sum_{i \in I} \log \bar{P}_{i,N+1} - \sum_{j \in J} \log \bar{P}_{M+1,j}. \quad (5)$$

Through the above steps, some rough background matching points are obtained in the two images. To obtain the unique background matching point, the following calculation strategy is designed by calculating smallest difference between the distance of all keypoints to a person in image A and the distance of matched keypoints to a person in image B:

$$s = \min \left(\left| \frac{\sum_{i=1}^H \sqrt{(x_b^A - x_i^A)^2 + (y_b^A - y_i^A)^2}}{H} - \frac{\sum_{j=1}^K \sqrt{(x_b^B - x_j^B)^2 + (y_b^B - y_j^B)^2}}{K} \right| \right), \quad (6)$$

where H and K represent the count of persons in images A and B . (x_b^A, y_b^A) and (x_b^B, y_b^B) represent the coordinates of background matching point b in images A and B , respectively. (x_i^A, y_i^A) and (x_j^B, y_j^B) represent the center coordinates of i -th and j -th bounding boxes of person in images A and B , respectively.

3.3. Pre-Processing Module

In this module, the group pairs are used as input. The CNN model is employed to extract person features, which can be regarded as nodes to construct a graph through k-nearest neighbor algorithm. In general, pedestrians are often divided into groups of three-persons, four-persons, five-persons or even more. Ukita et al. [49] demonstrated that all three-persons groups were judged as the same group. For other groups, there exist additional pedestrians that do not belong to this group, thereby introducing the noise for group re-ID. In order to remove additional pedestrians, a threshold f is designed to determine whether the current group members are in the same group.

$$f = \frac{N_p}{N_p + N_N}, \quad (7)$$

where N_p represents the count of persons and edges, N_N represents the count of node pairs that are not in the same group. Since all three-persons groups are all judged to be the same group, $f = \frac{2}{2+1} \approx 66.6\%$, where N_p denotes the edges of node 1 \rightarrow node 2 and node 2 \rightarrow node 3, N_N denotes the nodes pairs of node 1 and node 3, as shown in Figure 3b. Therefore, when $f < 66.6\%$ they are judged as different groups, and $f \geq 66.6\%$, they are judged as the same group. If group members are not in the same group, a person is (or persons are) removed who do not belong to the group.

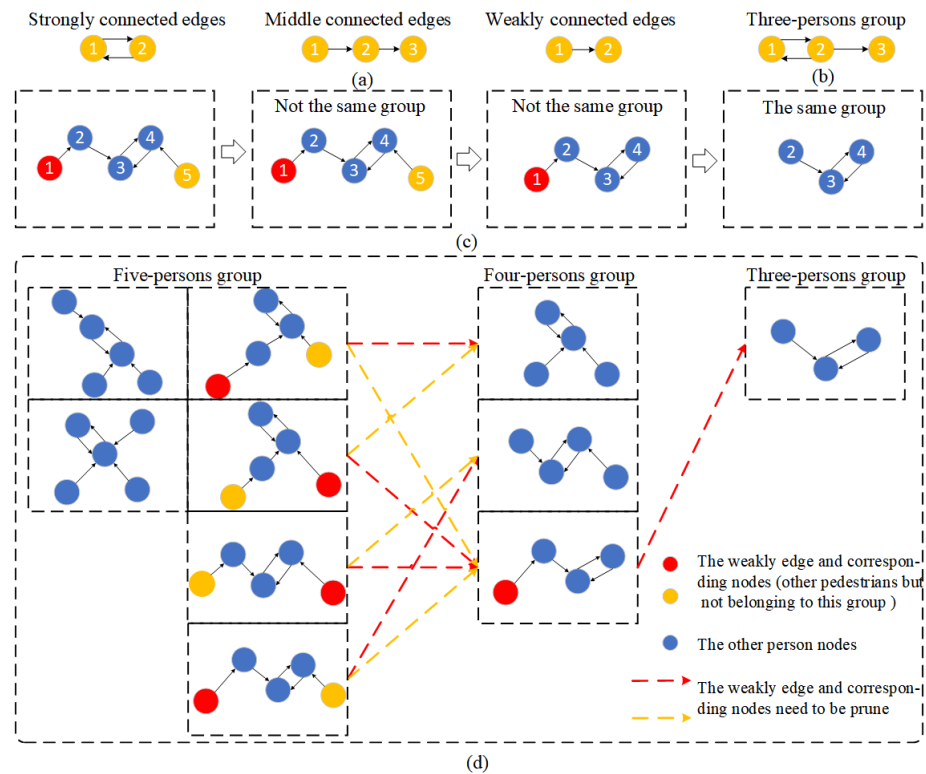


Figure 3. Demonstration of the Pre-Processing module. (a) Strongly, moderately and weakly connected edges. (b) The structure of three-persons group. (c) The specific pruning operation of five-persons group. (d) All pruning process of five-persons, four-persons and three-persons groups.

In this module, the edges of the graph are divided into strongly, moderately and weakly connected edges, as shown in Figure 3a. When two nodes choose each other as the nearest node, the edges between them are denoted as strongly connected edges. When a node is selected as the nearest node by some other nodes and also selects other nodes as the nearest node, the edge connecting the node to its nearest neighbor is denoted as a moderately connected edge, such as node 2 → node 3, where → denotes the edges of node 2 choosing node 3 as the nearest node through k-nearest algorithm. When a node selects another node as its nearest node but is not selected by other nodes, the connection between this node and its nearest node is called a weakly connected edge, such as node 1 → node 2 (the rightmost image in Figure 3a). The specific pruning process of the five-persons group is shown in Figure 3c. Specifically, for the five-persons group, $f = \frac{4}{4+3} \approx 57.1\% < 66.6\%$ which is judged as different group, where $N_P = 4$ represents the edges of node 1 → node 2, node 2 → node 3, node 3 → node 4, node 4 → node 5 and $N_N = 3$ represents the node pairs of node 1 and node 2, node 1 and node 3, node 1 and node 4, which are not in the same group. Therefore, the weakly connected edges (node 1 → node 2 or node 4 → node 5) and their nodes need to be pruned to obtain the four-persons group. If the edge of node 4 → node 5 is longer than node 1 → node 2, the longer weakly connected edges and its node (yellow node 5) should be removed. Subsequently, the four-persons group continue to determine whether the group is the same group, $f = \frac{3}{3+2} \approx 60\% < 66.6\%$ which is judged as different group, where $N_P = 3$ represents the edges of node 1 → node 2, node 2 → node 3, node 3 → node 4 and $N_N = 2$ represents the node pairs of node 1 and node 2, node 1 and node 3, which are not in the same group. Finally, the weakly connected edges and their node (red node 1) are pruned to obtain the three-persons group. In addition, the pruning process of all five-persons, four-persons, and three-persons groups are simulated respectively, as shown in Figure 3d.

3.4. Multi-Object Context Graph Module

After obtaining the reconstructed graph, the background features are added as nodes to the graph. In this module, a multi-level attention mechanism is adopted to capture intra-graph and inter-graphs contextual information between background and person features.

Each person feature h_{si} and background feature h_{rb} by CNN model can be divided into P parts as $h_{si} = [h_{si1}, \dots, h_{siP}]$ and $h_{rb} = [h_{rb1}, \dots, h_{rbP}]$, respectively. The descriptions of symbols are provided in Table 1.

Table 1. Descriptions of main symbols mentioned in this paper.

Symbols	Description
I_s	s -th input group image
N_s	number of person in I_s
G_s	group context graph in I_s
t	the t -th layer
$h_{sbp}^{(t)}$	the feature vector from the p -th part of background information b in G_s
$h_{sbq}^{(t)}$	the feature vector from the q -th part of background information b in G_s
$m_{sip}^{(t)}$	the message from the p -th intra-part in G_s
$n_{sip}^{(t)}$	the message from the p -th inter-parts in G_s
$\mu_{sip}^{(t)}$	the message from the p -th inter-graphs in G_s
h_s	graph representation in G_s
h_r	graph representation in G_r

3.4.1. The Relationship of Intra-Graph

The person and background features are divided into four parts. The correlation calculations of intra-graph can be divided into the same (intra-part) and different parts (inter-part) of the background features and person features respectively.

The intra-part correlation between $h_{sip}^{(t-1)}$ and $h_{sbp}^{(t-1)}$ can be calculated in image I_s . These features consist of features extracted from person node i and background node b , which correspond to the same part p . It is able to determine the important weights of intra-part messages transmitted from b to i :

$$e_{sibp} = \varphi(\mathbf{W}_e^{(t-1)} h_{sip}^{(t-1)}, \mathbf{W}_e^{(t-1)} h_{sbp}^{(t-1)}), \quad (8)$$

where φ represents a correlation measurement function and $\mathbf{W}_e^{(t-1)}$ represents a weight matrix. All important weights e_{sibp} are normalized using the softmax function to calculate the attention weights:

$$a_{sibp} = \text{soft max}(e_{sibp}). \quad (9)$$

Finally, the update attention weights $o_{sibp}^{(t)}$ of the intra-part are obtained by merging the background nodes b to the person nodes i :

$$o_{sibp}^{(t)} = a_{sibp} \mathbf{W}_e^{(t-1)} h_{sbp}^{(t-1)}. \quad (10)$$

The inter-part correlation calculation are defined in a similar manner. The inter-part messages between $h_{sip}^{(t-1)}$ and $h_{sbq}^{(t-1)}$ can be transferred from person node i and the background node b relating to distinct parts (p, q) . The final update attention weights $r_{sib}^{(t)}$ of the inter-part are obtained by merging the background nodes b to the person nodes i :

$$e_{sib}^{pq} = \varphi(\mathbf{W}_e^{(t-1)} h_{sip}^{(t-1)}, \mathbf{W}_e^{(t-1)} h_{sbq}^{(t-1)}), \quad (11)$$

$$a_{sib}^{pq} = \text{soft max}(e_{sib}^{pq}) = \frac{\exp(e_{sib}^{pq})}{\sum_{k \in E_s} \sum_{l \neq p} \exp(e_{sib}^{pl})}, \quad (12)$$

$$r_{sib}^{(t)} = \sum_i \sum_{q \neq p} a_{sib}^{pq} \mathbf{W}_e^{(t-1)} \mathbf{h}_{sbq}^{(t-1)}. \quad (13)$$

After acquiring both intra-part message $\mathbf{o}_{sibp}^{(t)}$ and inter-part message $\mathbf{r}_{sib}^{(t)}$, both of them belong to the intra-graph messages. The attention mechanism within the graph is depicted in Figure 4, showing how it focuses on both the intra-part and inter-part.

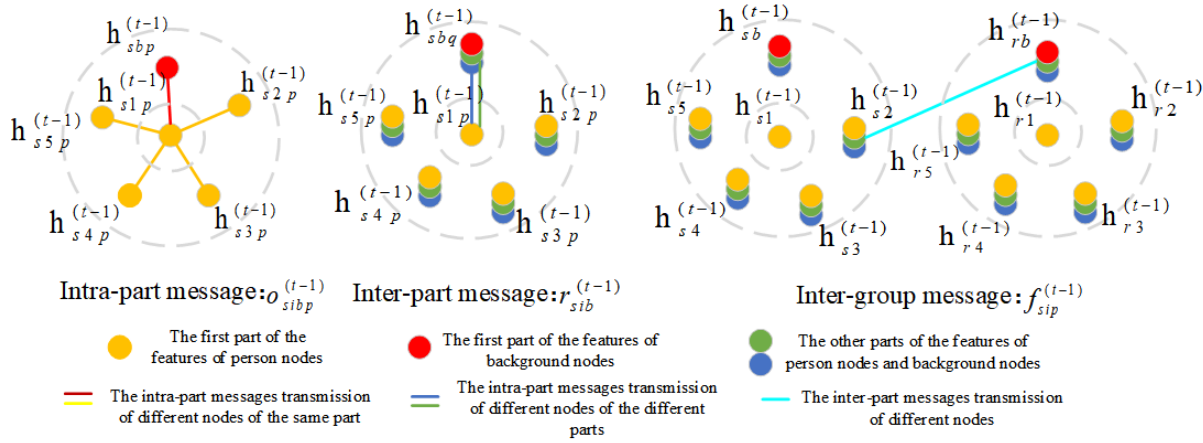


Figure 4. The visualization of intra-graph (intra- and inter-part) and inter-graph attentions.

3.4.2. The Relationship of Inter-Graph

The intra-graph context can be effectively modeled by intra-graph message passing, as mentioned above. However, it is essential to study the correlation and compute the similarity between groups for group re-ID. Intuitively, when two groups share the same ID, there is probable that certain correspondence exists among the individuals in both groups. Therefore, the higher the similarity of individual pairs, the higher the similarity at the group level. The important weights of inter-graph between $\mathbf{h}_{si}^{(t-1)}$ and $\mathbf{h}_{rb}^{(t-1)}$ in G_s, G_r are computed as follows:

$$z_{ib} = \varphi(\mathbf{W}_z^{(t-1)} \mathbf{h}_{si}^{(t-1)}, \mathbf{W}_z^{(t-1)} \mathbf{h}_{rb}^{(t-1)}), \quad (14)$$

where φ denotes an inner product layer and $\mathbf{W}_z^{(t-1)}$ denotes a projection matrix. The calculation of attention weights for inter-graph messages transmitted from background node b in graph G_r to person node i in graph G_s is in the same manner:

$$w_{ib} = \text{soft max}(z_{ib}), \quad (15)$$

$$\mathbf{f}_{sip}^{(t)} = \sum_q w_{ib} \mathbf{W}_z^{(t-1)} \mathbf{h}_{rbq}^{(t-1)}. \quad (16)$$

The above node features can be updated by Multi-Layer Perceptron (MLP) to aggregate features into a better feature space and fine-tune the model to better training:

$$\mathbf{h}_{sip}^{(t)} = \text{MLP}(\mathbf{h}_{sip}^{(t-1)}, \mathbf{m}_{sip}^{(t)}, \mathbf{n}_{sip}^{(t)}, \mathbf{o}_{sibp}^{(t)}, \mathbf{r}_{sib}^{(t)}, \mathbf{u}_{sip}^{(t)}, \mathbf{f}_{sip}^{(t)}). \quad (17)$$

3.4.3. Correspondence Learning

To improve the learning of group correspondence, the circle loss is employed to enforce feature similarity within the same group and promote the separation of different groups:

$$\begin{aligned} L_g^{circle} &= \log\left[1 + \sum_{i=1}^K \sum_{j=1}^L \exp(\gamma(a_s^i h_s^i - a_r^j h_r^j))\right] \\ &= \log\left[1 + \sum_{j=1}^L \exp(\gamma a_s^i h_s^i) \sum_{i=1}^K \exp(\gamma a_r^j h_r^j)\right], \end{aligned} \quad (18)$$

where a_s^i and a_r^j denote non-negative weighting factors of i -th and j -th persons in images I_s and I_r respectively, γ denotes as a scale factor.

Moreover, when building global correspondence, it is crucial to take into account local person- and background-level information. The pair-wise loss can be adopted for person- and background-level corresponding learning:

$$L_p^{pair} = \sum_{i \in v_s, j \in v_r} \sum_p \max\{0, m - y_p^{ij}(1 - \|\mathbf{h}_{sip} - \mathbf{h}_{rjp}\|^2)\}, \quad (19)$$

$$L_b^{pair} = \sum_{i \in v_s, b \in v_r} \sum_b \max\{0, m - y_b^{ib}(1 - \|\mathbf{h}_{sip} - \mathbf{h}_{rbp}\|^2)\}, \quad (20)$$

where m denotes the margin and y_p^{ij} , y_b^{ib} represent the pair labels. Finally, the value of forecast matrix S and ground reality matrix S^{gt} can be calculated through cross-entropy loss [50]:

$$L_{pce} = - \sum_{i \in v_s, j \in v_r} (S_{i,j}^{gt} \log S_{i,j} + (1 - S_{i,j}^{gt}) \log(1 - S_{i,j})), \quad (21)$$

where $S^{gt} \in R^{N_s \times N_r}$ denotes a binary matrix, $S_{i,j}^{gt} = 1$ when i -th and j -th persons show the same identity. The total loss is calculated by combining all the aforementioned loss functions in a linear manner:

$$L = L^{match} + L_g^{circle} + L_p^{pair} + L_b^{pair} + L_{pce}. \quad (22)$$

4. Experimental Results

4.1. Datasets and Experimental Settings

Datasets. (1) the PRAI-1581 dataset [51] contains 39,461 person images from two UAVs flying at heights between 20 and 60 m above ground level and the resolution of person images is 30–50 pixels; (2) the Road Group dataset (RG) [10] contains 162 pairs of group images taken by two cameras and the resolution of group images is 89–302 pixels; (3) the DukeMTMC Group dataset (DG) [11] based on DukeMTMC dataset contains 177 pairs of group images captured by eight cameras and the resolution of group images is about 180–1034 pixels; (4) the CUHK-SYSU-Group dataset (CSG) [14] based on CUHK-SYSU dataset contains 3839 pairs of group images and the resolution of group images is 600×800 pixels; (5) CM-group dataset [52] contains visible and infrared 30,946 images of 427 groups and the resolution of group images is 2560×1440 pixels; (6) the SYSU-MM01 dataset [53] contains 45,863 visible and infrared person images and the resolution of person images is about 83–548 pixels; (7) the Aerial Image dataset (AID) [54] contains 10,000 aerial scene images, featuring 30 distinct classes with each class containing between 220 to 420 images at an altitude from about half a meter to about 8 m and the resolution of aerial images is 600×600 pixels; and (8) RGB-NIR Scene dataset [55] contains 477 images in nine classes in RGB and Near-infrared(NIR) and the resolution of images is 1024×768 pixels. The PRAI-1581, CSG, RG, DG, AID datasets are visible images with a spectral resolution of 390 nm–780 nm, the CM-group, SYSU-MM01 datasets are visible-infrared images with a spectral resolution of 1 mm–750 nm, and RGB-NIR Scene dataset are

RGB-NIR images with a spectral resolution of 1000 nm–1700 nm. In addition, the PRAI-1581 and AID datasets contain the messages of geometric and spatial resolution, while the other datasets only contain geometric resolution messages. Figures 5 and 6 show the examples of images.



Figure 5. Examples of images in the (a) PRAI-1581, (b) CUHK-SYSU-Group, (c) Road Group, (d) DukeMTMC Group datasets, (e) CM-group and (f) SYSU-MM01 datasets.

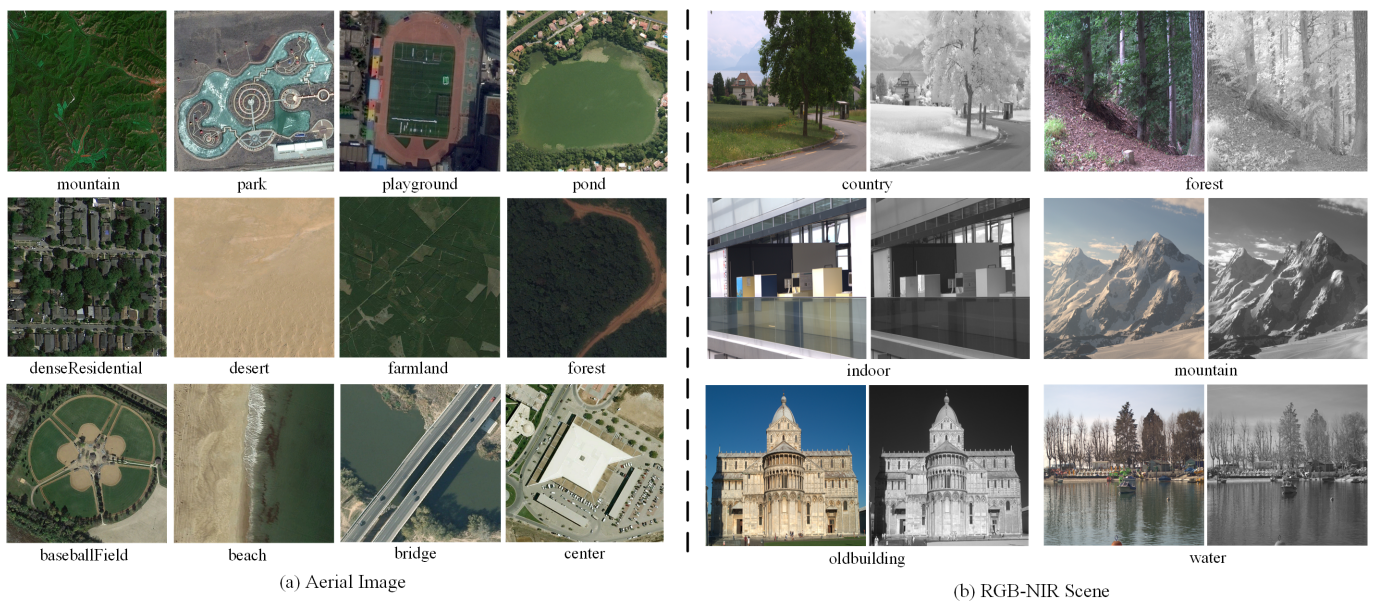


Figure 6. Examples of images in the (a) Aerial Image and (b) RGB-IR Scene datasets.

Experimental Settings. The datasets are divided into 60% training set and 40% testing set in a random manner. The model utilize ResNet50 as our backbone and train 300 epochs with 0.0001 initial learning rate. Each 100 epochs of training, the learning rate decreases by 10%. To facilitate implementation, virtual nodes are added to graphs with different numbers of nodes. The model employs two-layer GNNs and trains on a 2080ti GPU, which takes approximately 60 h to convergence on CSG dataset.

4.2. Comparison with Other Group Re-ID Methods

Because of absence of UAV group re-ID datasets, it is hard to perform our method on such datasets is challenging. However, the images of group re-ID datasets are similar to UAV person re-ID datasets, both from a top-down perspective, as shown in Figure 5a,b. Therefore, the several existing group re-ID methods are compared with our model on RG, DG and CSG datasets, including crowd matching-based and features aggregation-based methods, experimental results are shown in Table 2. It is readily apparent our approach reaches competitive results in all three datasets. More precisely, our approach reaches Rank-1 accuracies of 71.8%, 86.4% and 57.8% on the CSG, RG and DG datasets, respectively.

Table 2. In comparison to several group re-ID methods. The best results are shown are in black boldface font.

Methods	CUHK-SYSU-Group				Road Group				DukeMTMC Group			
	R1	R5	R10	R20	R1	R5	R10	R20	R1	R5	R10	R20
Crowd matching-based methods												
CRRRO-BRO (BMVC2009) [10]	10.4	25.8	37.5	-	17.8	34.6	48.1	-	9.9	26.1	40.2	-
Covariance (ICPR2010) [8]	16.5	34.1	47.9	67.0	38.0	61.0	73.1	82.5	21.3	43.6	60.4	78.2
BSC+CM (ICIP2016) [9]	24.6	38.5	55.1	73.8	58.6	80.6	87.4	92.1	23.1	44.3	56.4	70.4
PREF (CVPR2017) [16]	19.2	36.4	51.8	70.7	43.0	68.7	77.9	85.2	22.3	44.3	58.5	74.4
LIMI (MM2018) [41]	-	-	-	-	72.3	90.6	94.1	-	47.4	68.1	77.3	-
MGR (TCYB2021) [12]	57.8	71.6	76.5	82.3	80.2	93.8	96.3	97.5	48.4	75.2	89.9	94.4
Features aggregation-based methods												
DotGNN (MM2019) [42]	-	-	-	-	74.1	90.1	92.6	-	53.4	72.7	80.7	-
GCGNN (TMM2020) [11]	-	-	-	-	81.7	94.3	96.5	97.8	53.6	77.0	91.4	94.8
MACG(TPAMI2020) [14]	63.2	75.4	79.7	84.4	84.5	95.0	96.9	98.1	57.4	79.0	90.3	94.3
DotSCN(TCSVT2021) [15]	-	-	-	-	84.0	95.1	96.3	98.8	86.4	98.8	98.8	98.8
PRM (IEEEAccess2021) [43]	-	-	-	-	62.4	75.9	82.1	88.3	-	-	-	-
Ours	71.8	81.9	86.0	88.9	86.4	95.5	97.0	98.5	57.8	80.2	88.5	94.4

Comparison with crowd matching-based methods: Crowd matching-based methods adopt a single or multi-person matching strategy and assign important weights to people to capture the dynamic changes within the group. However, these methods (e.g., PREF [16], MGR [12]) cannot solve the challenge of varying group members well. Our novel approach demonstrates superior performance in comparison to all the methods evaluated. For example, compared to the MGR method, our proposed approach reaches a Rank-1 improvement of 14% on CSG dataset, a Rank-1 improvement of 6.2% on the RG dataset and a Rank-1 improvement of 9.4% on the DG dataset.

Comparison with features aggregation-based methods: Features aggregation-based methods aggregate the contextual information from neighbors to re-identify each individual within groups as well as the entire group. However, these methods (e.g., GCGNN [11], MACG [14]) ignore the importance of background semantic information for messages transmission. Our proposed method significantly outperforms most compared methods. For example, compared to MACG method, our method reaches a Rank-1 improvement of 8.6% on the CSG dataset and a Rank-1 improvement of 1.9% on the RG dataset.

4.3. Person Re-ID in Groups

Utilizing group context information has been demonstrated for single person re-ID [10,12,13]. In our method, the background information and screening strategy can be utilized to enhance group and individual correspondence learning, which can directly apply to single person re-ID. Our method is compared with the person re-ID model without group context information, such as our baseline CNN, Strong-Baseline [56], PCB [57] and OSNet [58], and with group context information, such as MGR [12], Context Graph

(CG) [13] and MACG [14]. Table 3 shows the experimental results. In comparison to our baseline CNN, our model reaches a 4.7% increase in Rank-1 accuracy. Compared with MACG, our model reaches a 2.6% increase in Rank-1 accuracy. The above results reveal our model can effectively enhance the accuracy of single-person re-ID. The experimental results above illustrate that using group information not only helps in groups but also makes it easier to re-identify a single person. Therefore, the proposed DSFG framework can also re-identify a single person, even a UAV person.

Table 3. Comparison of results for single person re-ID with the inclusion of group context information. ✓ denotes the model uses group context information.

Model	Group	Rank-1	Rank-5	Rank-10
Our CNN	-	63.5	74.2	79.9
Strong-Baseline [56]	-	63.1	80.3	84.3
PCB [57]	-	63.7	80.2	83.9
OSNet [58]	-	62.4	77.0	81.0
MGR [12]	✓	63.8	79.9	83.8
CG [13]	✓	62.7	78.4	82.6
MACG [14]	✓	65.6	80.5	84.6
DSFG	✓	68.2	81.2	84.9

To assess the generalization capability of our framework for UAV person re-ID, the model trained on CSG is directly applied to the UAV PRAI-1581 dataset, as shown in Table 4. It is easy to observe that our method can effectively improve the accuracy of UAV person re-ID. For example, compared with PCB, our model reaches a 3.25% increase in Rank-1 accuracy. Since PRAI-1581 dataset lacks group context, a single person need to be replicated to construct a group for building the context graph.

Table 4. CSG dataset transfer results for PRAI-1581 dataset.

Model	CSG → PRAI-1581		
	Rank-1	Rank-5	Rank-10
Strong-Baseline [56]	46.10	49.23	52.34
PCB [57]	48.07	51.20	55.97
OSNet [58]	54.40	58.85	62.37
Ours	51.32	55.10	57.38

In addition, to explore the effectiveness of our model in infrared images, our models trained on the CSG and CM-Group datasets are also directly applied to the SYSU-MM01 dataset, respectively, as shown in Table 5. Note that only the infrared images are trained on CM-Group datasets. It can be observed to find that the transfer results of the model trained using CM-Group to SYSU-MM01 dataset is much higher than the model trained using CSG. It is probably that there is less useful information in infrared images. Therefore, our model works separately for visible and infrared images, but it is best to transfer in the same modality.

Table 5. CSG and CM-Group datasets transfer results for SYSU-MM01 dataset.

Model	CSG → SYSU-MM01		
	Rank-1	Rank-5	Rank-10
Strong-Baseline [56]	34.39	36.34	37.83
PCB [57]	36.54	38.47	40.61
OSNet [58]	40.18	43.64	45.59
Ours	38.12	41.22	43.67
Ours	52.24	55.83	57.92

4.4. Model Analysis

Backbone network. The impact of different backbone networks are analyzed on model performance. In this study, the ResNet50 backbone is replaced by ResNet18, ResNet34, ResNet101, DenseNet121 and DenseNet169. Figure 7a shows the experiment results. Significant performance decline is observed when utilizing shallow networks, especially in Rank-1. Furthermore, the utilization of ResNet101 only resulted in minor improvements when compared to ResNet50. These results highlight the substantial influence of distinct backbone networks on the performance with our model. Ultimately, it is clearly easy to observe ResNet50 strikes an optimal balance between model complexity and performance, making it a suitable choice for the backbone in our setup.

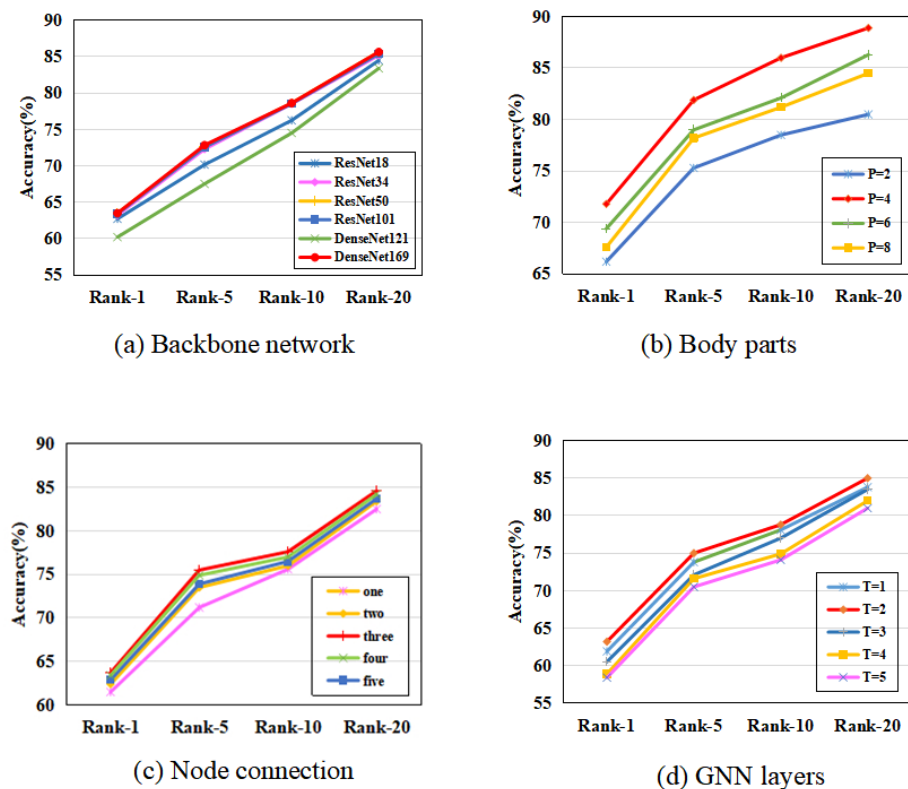


Figure 7. The analysis of model under various settings on the CSG dataset.

Body parts. The effects of various human body and background information partitions are evaluated. The results of various partitions are shown in Figure 7b. It is evident from the results that a $P = 2$ partition yields notably lower performance compared to other partitions, primarily due to its coarse segmentation. Conversely, the most favorable performance is attained when employing $P = 4$ partitions.

Node connection. In our framework, the context graphs are constructed with connecting all members in a group. Figure 7c shows the experiment results, where 'Neighbor = K ' means that an individual is exclusively linked to its K closest neighbors. The observed trend indicates a consistent improvement in performance as K increases in size, i.e., with denser connections. In the case of 'Neighbor = 3' and 'Neighbor = 4', the accuracy is already at its highest. In the case of 'Neighbor = 5', the accuracy is lower than 'Neighbor = 3' and 'Neighbor = 4'. It is probably that most of groups the existing datasets are four-persons or five-persons. In addition, when constructing the context graph with empty nodes, the model may create redundancy to result in performance decrease. It can be observed that 'Neighbor = 3' and 'Neighbor = 4' graph configurations facilitate effective passing of messages in the group, thereby aiding model in acquiring contextual information and screening the other person not belonging to the current group.

GNN layers. Furthermore, the impact of utilizing different numbers of GNN layers ($T = 1, \dots, 5$) is analyzed. The experimental results are shown in Figure 7d. It is easy to find that the use of two-layer GNN results in improved performance compared to GNNs with other layer numbers. This improvement can be attributed to the fact that single-layer GNN may not provide adequate representational capacity, whereas GNNs with more layers may introduce excessive parameters, increasing the risk of overfitting. This observation could be linked to limited size of the dataset, and the need for further dataset expansion might require the utilization of deeper GNNs to improve performance.

4.5. The Visualization of the Results

Graph-based Feature Matching Module. To confirm the GFM module's effectiveness, the image matching results are visualized on CSG dataset by using the original image and removing the person area. The visualization results are shown in Figure 8. It can be observed that keypoints matched in Figure 8b are much less than those in Figure 8a, effectively reducing the keypoints of pedestrians. It is more helpful for us to find additional useful features that do not contain the object (person).

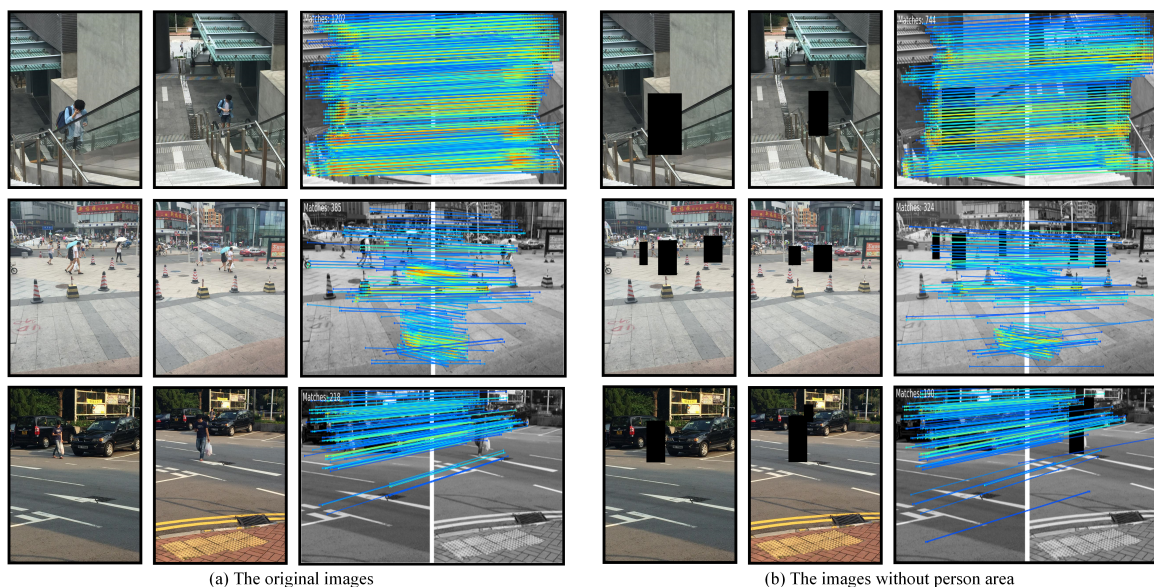


Figure 8. The visualization of keypoints matching results on the CSG dataset.

In addition, the GFM module is performed to feature-match and re-identify the images of same class and different classes on AID and RGB-NIR Scene datasets, respectively. The visualization results are depicted in Figures 9 and 10. In Figure 9, the first two lines are visible images matching on AID dataset. It can be observed that there are far more matching points in the same class of images than in different classes of images. The last two lines are NIR images matching on RGB-NIR Scene dataset. It can be found that our proposed module not only adapts to visible images, but also NIR images. In Figure 10, the first two lines are visible images re-ID of the same class on AID dataset, the last two lines are NIR images re-ID of the same class on the RGB-NIR Scene dataset. It is evident to observe that the GFM module can effectively retrieve the same class.

Pre-Processing Module. As the group moves, the members of the group will change. To solve the above issue, the PREP module presented aims to remove additional pedestrians. The experimental results of the pruning process in six-persons groups are shown in Figure 11. The *SSIM* algorithm is utilized to determine the structural similarity between the pruned group and contrast group. It can be observed that the group similarity is higher after each pruning operation. In addition, the different structures are designed with five-persons groups and four-persons groups and the situations in which they are judged as the same group and different groups, as shown in Figure 12a,b. The experimental results

indicate that compared with the contrast groups, groups judged as different groups exhibit lower similarity than groups judged as the same.

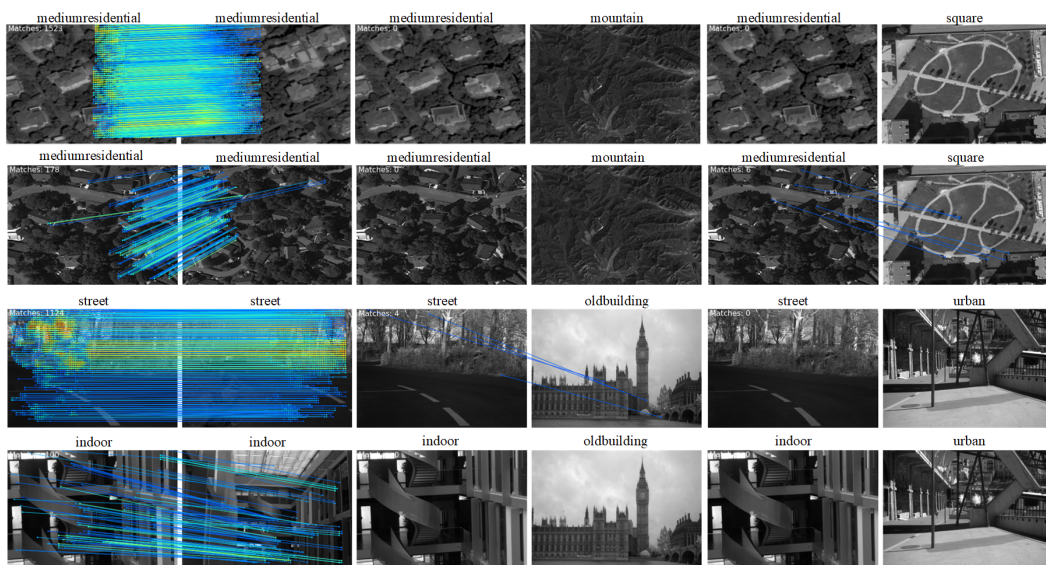


Figure 9. The visualization of keypoints matching results in the same and different classes on the AID and RGB-NIR Scene datasets.

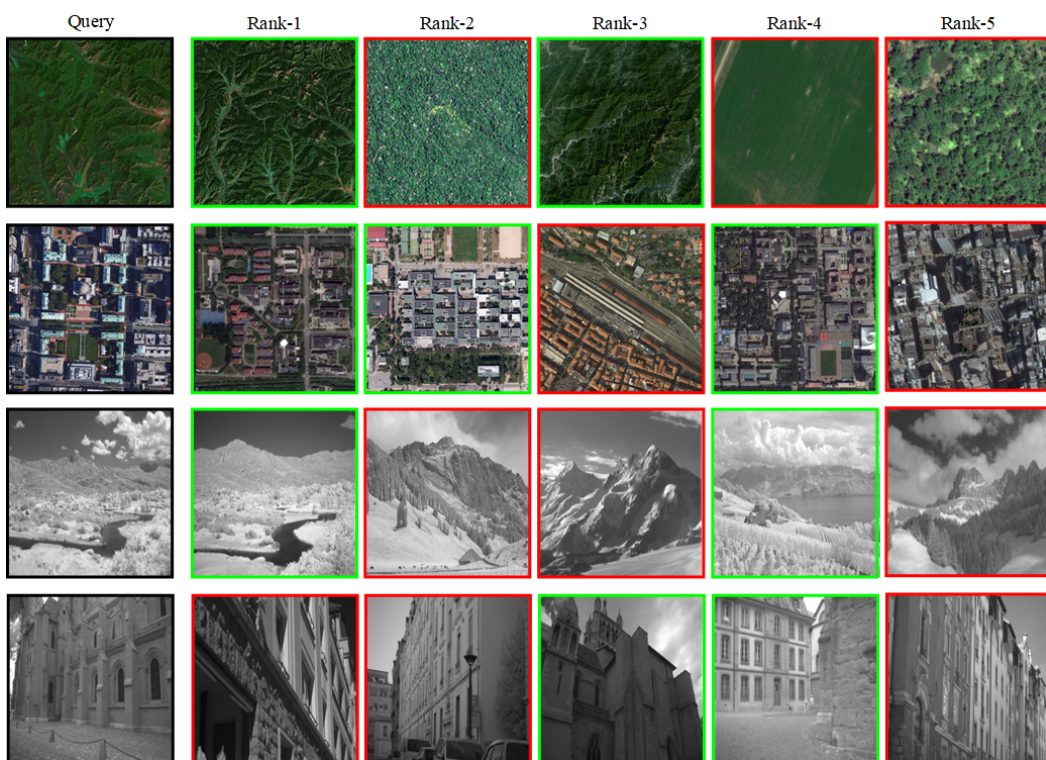


Figure 10. The visualization retrieved of visible and NIR images re-ID results. The black bounding boxes images are the query, the red and green bounding boxes denote wrong and right matches, respectively.

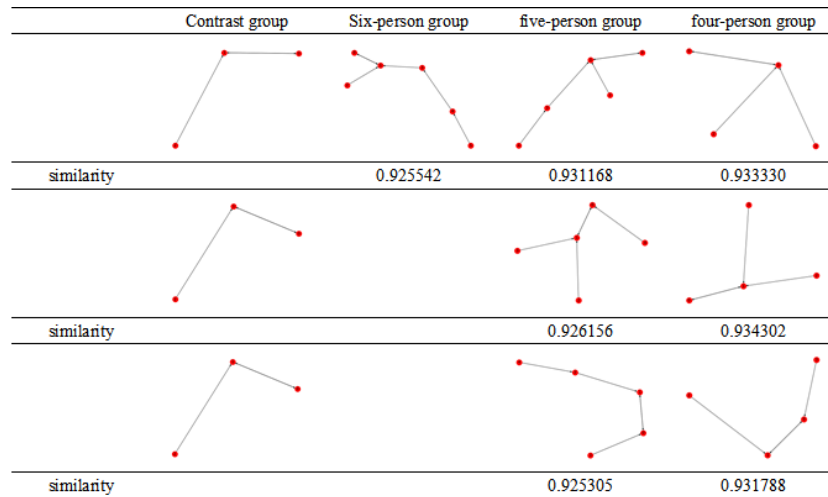
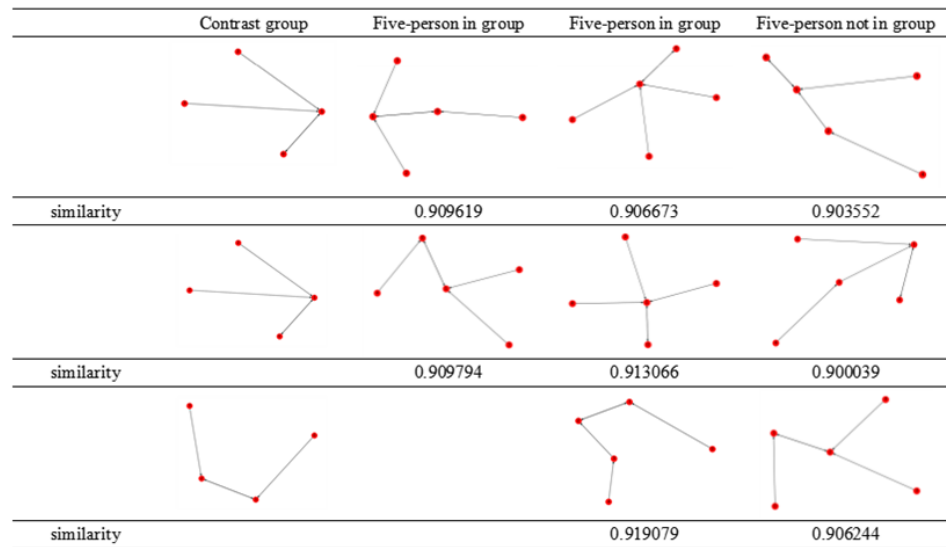
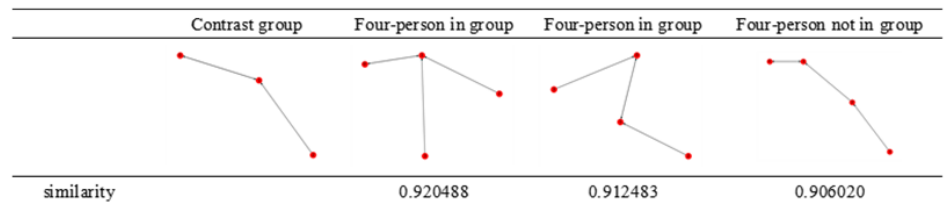


Figure 11. The visualization of the pruning process in six-person group and their similarity to the contrast group.



(a) The different structures of five-person groups



(b) The different structures of four-person groups

Figure 12. The visualization of different structures in five-person and four-person groups and their similarity to the contrast group.

The visualization of group and UAV person re-ID results. The visualization results are depicted in Figure 13 and Figure 14, respectively. The images from the test set can be retrieved in the gallery set and it is easy to observe that persons sharing a similar appearance have an impact on the retrieval results, which may make it hard to retrieve the correct match with the model.



Figure 13. The retrieved visualization of group re-ID results. The black bounding boxes images are the query, the red and green bounding boxes denote wrong and right retrieval results, respectively.



Figure 14. The retrieved visualization of UAV person re-ID results. The black bounding boxes images are the query, the red and green bounding boxes denote wrong and right retrieval results, respectively.

4.6. Ablation Study

Module Ablation. The ablation experiments are performed on three datasets to confirm the effectiveness of our proposed module, as shown in Table 6. The model regards the method of MACG as our baseline and proposes the MOCG module based on the GFM module to improve the MACG method. It can be understood that after using the GFM module to obtain background features, the MOCG method proposed needs to be used.

It is easy to observe the images of CSG datasets exist overlapping areas under different cameras, so adding the GFM module can effectively improve the performance. In the RG dataset, almost every image has the same ground background information. Therefore, adding the GBFM module has an insignificant impact and it cannot be used as an effective way to associate two images. In the DG dataset, the same group of people will move to another scene with completely different background information (across background scenes). In this case, the background semantic information does not match and adding it may introduce noise to group re-ID, thus reducing the performance extensively. Therefore, the method of the GFM module still has limitations on some specific datasets. In addition, experimental results prove that the PREP module and circle loss [59] (CL) function have significant improvements on three datasets.

Table 6. Examination of the components of our proposed approach on three datasets. Bold indicates the best performance.

	CSG	RG	DG
MACG	63.2	84.5	57.4
MOCG + GFM	64.9	84.4	44.7
MACG + PREP	65.2	84.9	57.8
MOCG + GFM + PREP	66.5	85.0	46.8
MOCG + GFM + PREP + CL	71.8	86.4	49.5

In addition, the distinct variants of the framework are examined on CSG dataset, such as Graph-based Feature Matching (GFM), Pre-processing (PREP) module and circle loss (CL) [59] function, as shown in Table 7. The ultimate model performance can be determined by rearranging elements in diverse arrangements.

Table 7. Ablation study of various combination with our proposed modules on the CSG dataset. ✓ denotes the model chooses the various modules.

Modules	Base	Variants	Variants	Variants	Variants	Variants	Variants
GFM		✓			✓		✓
PREP			✓			✓	✓
CL				✓	✓	✓	✓
Rank-1	63.2	64.9	65.2	65.4	68.8	69.2	71.8

Distance ablation. To confirm the efficacy of utilizing Euclidean distance in calculating the distance between keypoints and person, the Euclidean distance is compared with Cosine and Manhattan, respectively, as shown in the Table 8. It can be found that the Rank-1 of the Euclidean distance is higher than the others. It is probably that the relative distance between person and the same background matching points is unchanging, so the optimal background matching point can be calculated with person and background points by using Euclidean distance.

Table 8. Ablation study of various distance on CSG dataset. Bold indicates the best performance.

	Rank-1	Rank-5	Rank-10
Manhattan distance	55.2	62.3	66.9
Cosine distance	60.7	70.4	74.1
Euclidean distance	71.8	81.9	86.0

Loss Fuction Ablation. The impact of employing various loss functions on our model's effectiveness is illustrated by employing cross-entropy loss, triple loss, and circle loss functions. The circle loss has two different parameters, i.e., 32 and 64. The experimental results are reported on CSG dataset in Table 9. Our findings indicate that the utilization of circle

loss function effectively minimizes feature distance within a class while simultaneously maximizing the distance between classes.

Table 9. The experiment of circle loss replacing cross entropy loss or triple loss function on CSG dataset. Bold indicates the best performance.

	Rank-1	Rank-5	Rank-10	Rank-20
Ours	66.5	72.5	77.3	80.2
Ours + cross entropy loss	52.6	67.8	74.2	78.5
Ours + triple loss	68.4	75.6	82.3	86.2
Ours + circle loss_32	70.1	81.1	84.4	88.0
Ours + circle loss_64	71.8	81.9	86.0	88.9

5. Discussion

In the group re-ID tasks, the model performs with high accuracy using our proposed algorithm on three datasets. And it can be directly applied to the UAV visible person re-ID PRAI-1581 dataset and infrared SUSY-MM01 dataset for the first time. However, our proposed algorithm needs too much time to train. This is probably because after new background features nodes are added to the graph, the model requires a lot of time to transfer context information between groups during the matching process. In addition, although graph structure has achieved great achievement in solving the issues of group re-ID, no solution based on spatio-temporal information has been found. It is probably that there is a lack of appropriate data. It can be argued that the spatio-temporal problem may effectively enhance not only the transmission of information between group members, but also the problem of member changes (i.e., the time when persons enter and exit the images). Finally, given the scarcity of datasets for UAV group re-ID, it is worthwhile to prioritize research efforts in the future.

6. Conclusions

In this paper, a novel framework is proposed to solve the challenge of group re-ID, containing three modules: Graph-based Feature Matching (GFM), Pre-Processing (PREP) and Multi-Object Centext Graph (MOCG). The GFM module enhances the transmission of group contextual information by adding background matched features for the first time. The PREP module solves the challenge of group member changes in the group re-ID, which can remove group members that do not belong to the current group through pruning operations. And the MOCG module aggregates the messages from background features to person features and updates the person and group features maximizely. Experiments show that outstanding results are obtained on three datasets for group re-ID, demonstrating the efficacy of the suggested methodology. The Rank-1 on CUHK-SYSU-Group, Road Group and DukeMTMC Group datasets reach 71.8%, 86.4% and 57.8%, respectively. In addition, the models trained on CSG and CM-Group datasets to the UAV visible person re-ID PRAI-1581, infrared SUSY-MM01 and RGB-NIR Scene datasets, which achieve certain results.

Author Contributions: Conceptualization, methodology, writing, funding acquisition, and supervision, G.Z. and T.L.; software, validation, and data curation, T.L., Z.Y. and G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant number: 62172231); Natural Science Foundation of Jiangsu Province of China (Grant number: BK20220107).

Data Availability Statement: This study is available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Z.; Liu, W.; Matsui, Y.; Satoh, S. Effective and efficient: Toward open-world instance re-identification. In Proceedings of the 28th ACM International Conference on Multimedia, Virtual, 12–16 October 2020; pp. 4789–4790.
2. Almazan, J.; Gajic, B.; Murray, N.; Larlus, D. Re-id done right: Towards good practices for person re-identification. *arXiv* **2018**, arXiv:1801.05339.
3. Wang, Y.; Wang, L.; You, Y.; Zou, X.; Chen, V.; Li, S.; Huang, G.; Hariharan, B.; Weinberger, K. Resource aware person re-identification across multiple resolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8042–8051.
4. Zhang, G.; Ge, Y.; Dong, Z.; Wang, H.; Zheng, Y.; Chen, S. Deep high-resolution representation learning for cross-resolution person re-identification. *IEEE Trans. Image Process.* **2021**, *30*, 8913–8925. [[CrossRef](#)] [[PubMed](#)]
5. Lin, W.; Shen, Y.; Yan, J.; Xu, M.; Wu, J.; Wang, J.; Lu, K. Learning correspondence structures for person re-identification. *IEEE Trans. Image Process.* **2017**, *26*, 2438–2453. [[CrossRef](#)] [[PubMed](#)]
6. Zhang, G.; Chen, Y.; Lin, W. Low resolution information also matters: Learning multi-resolution representations for person re-identification. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 19–26 August 2021; pp. 1295–1301.
7. Chen, S.; Guo, C.; Lai, J. Deep ranking for person re-identification via joint representation learning. *IEEE Trans. Image Process.* **2016**, *25*, 2353–2367. [[CrossRef](#)] [[PubMed](#)]
8. Cai, Y.; Takala, V.; Pietikainen, M. Matching groups of people by covariance descriptor. In Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2744–2747.
9. Zhu, F.; Chu, Q.; Yu, N. Consistent matching based on boosted salience channels for group re-identification. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 4279–4283.
10. Zheng, W.; Gong, S.; Xiang, T. Associating groups of people. *BMVC* **2009**, *6*, 1–11.
11. Zhu, J.; Yang, H.; Lin, W.; Liu, N.; Wang, J.; Zhang, W. Group re-identification with group context graph neural networks. *IEEE Trans. Multimed.* **2020**, *23*, 2614–2626. [[CrossRef](#)]
12. Lin, W.; Li, Y.; Xiao, H.; See, J.; Zou, J.; Xiong, H.; Wang, J.; Mei, T. Group reidentification with multigrained matching and integration. *IEEE Trans. Cybern.* **2019**, *51*, 1478–1492. [[CrossRef](#)] [[PubMed](#)]
13. Yan, Y.; Zhang, Q.; Ni, B.; Zhang, W.; Xu, M.; Yang, X. Learning context graph for person search. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
14. Yan, Y.; Qin, J.; Ni, B.; Chen, J.; Liu, L.; Zhu, F.; Zheng, W.; Yang, X.; Shao, L. Learning multi-attention context graph for group-based re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *45*, 7001–7018. [[CrossRef](#)]
15. Huang, Z.; Wang, Z.; Tsai, C.; Satoh, S.; Lin, C. Dotscn: Group re-identification via domain-transferred single and couple representation learning. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 2739–2750. [[CrossRef](#)]
16. Lisanti, G.; Martinel, N.; Del Bimbo, A.; Luca Foresti, G. Group re-identification via unsupervised transfer of sparse features encoding. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2449–2458.
17. Zhang, G.; Luo, Z.; Chen, Y.; Zheng, Y.; Lin, W. Illumination unification for person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 6766–6777. [[CrossRef](#)]
18. Zheng, L.; Yang, Y.; Hauptmann, A. Person re-identification: Past, present and future. *arXiv* **2016**, arXiv:1610.02984.
19. Karanam, S.; Li, Y.; Radke, R. Person re-identification with discriminatively trained viewpoint invariant dictionaries. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4516–4524.
20. Bak, S.; Zaidenberg, S.; Boulay, B. Improving person re-identification by viewpoint cues. In Proceedings of the 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, Republic of Korea, 26–29 August 2014; pp. 175–180.
21. Li, X.; Zheng, W.; Wang, X.; Xiang, T.; Gong, S. Multi-scale learning for low-resolution person re-identification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3765–3773.
22. Huang, Y.; Zha, Z.; Fu, X.; Zhang, W. Illumination-invariant person re-identification. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 365–373.
23. Cho, Y.; Yoon, K. Improving person re-identification via pose-aware multi-shot matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1354–1362.
24. Zhao, H.; Tian, M.; Sun, S.; Shao, J.; Yan, J.; Yi, S.; Wang, X.; Tang, X. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1077–1085.
25. Sarfraz, M.; Schumann, A.; Eberle, A.; Stiefelagen, R. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 420–429.
26. Zhang, G.; Fang, W.; Zheng, Y.; Wang, R. SDBAD-Net: A Spatial Dual-Branch Attention Dehazing Network based on Meta-Former Paradigm. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 60–70. [[CrossRef](#)]
27. Chen, Y.; Zhang, G.; Lu, Y.; Wang, Z.; Zheng, Y. Tipcb: A simple but effective part-based convolutional baseline for text-based person search. *Neurocomputing* **2022**, *494*, 171–181. [[CrossRef](#)]

28. Liu, H.; Tang, X.; Shen, S. Depth-map completion for large indoor scene reconstruction. *Pattern Recognit.* **2020**, *99*, 107–112. [[CrossRef](#)]
29. Zhang, G.; Liu, J.; Chen, Y.; Zheng, Y.; Zhang, H. Multi-biometric unified network for cloth-changing person re-identification. *IEEE Trans. Image Process.* **2023**, *32*, 4555–4566. [[CrossRef](#)]
30. Gray, D.; Tao, H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 12–18 October 2008; pp. 262–275.
31. Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Person re-identification by symmetry-driven accumulation of local features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2360–2367.
32. Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Salient color names for person re-identification. In Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 536–551.
33. Zheng, W.; Gong, S.; Xiang, T. Global relation-aware contrast learning for unsupervised person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 8599–8610. [[CrossRef](#)]
34. Zhang, G.; Zhang, H.; Lin, W.; Chandran, A.K.; Jing, X. Camera contrast learning for unsupervised person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 4096–4107. [[CrossRef](#)]
35. Zhang, G.; Sun, H.; Zheng, Y.; Xia, G.; Feng, L.; Sun, Q. Optimal discriminative projection for sparse representation-based classification via bilevel optimization. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 1065–1077. [[CrossRef](#)]
36. Qiao, Y.; Cui, J.; Huang, F.; Liu, H.; Bao, C.; Li, X. Efficient style-corpus constrained learning for photorealistic style transfer. *IEEE Trans. Image Process.* **2021**, *30*, 3154–3166. [[CrossRef](#)]
37. Gao, X.; Zhu, L.; Xie, Z.; Liu, H.; Shen, S. Incremental rotation averaging. *Int. J. Comput. Vis.* **2021**, *129*, 1202–1216. [[CrossRef](#)]
38. Zheng, Z.; Zheng, L.; Yang, Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3754–3762.
39. Li, W.; Zhao, R.; Xiao, T.; Wang, X. Deepreid: Deep filter pairing neural network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 152–159.
40. Leng, Q.; Ye, M.; Tian, Q. A survey of open-world person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 1092–1108. [[CrossRef](#)]
41. Xiao, H.; Lin, W.; Sheng, B.; Lu, K.; Yan, J.; Wang, J.; Ding, E.; Zhang, Y.; Xiong, H. Group re-identification: Leveraging and integrating multi-grain information. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 192–200.
42. Huang, Z.; Wang, Z.; Hu, W.; Lin, C.; Satoh, S. DoT-GNN: Domain-transferred graph neural network for group re-identification. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1888–1896.
43. Hu, P.; Zheng, H.; Zheng, W. Part Relational Mean Model for Group Re-Identification. *IEEE Access* **2021**, *9*, 46265–46279. [[CrossRef](#)]
44. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
45. Battaglia, P.; Hamrick, J.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R. Relational inductive biases, deep learning, and graph networks. *arXiv* **2018**, arXiv:1806.01261.
46. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Philip, S. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [[CrossRef](#)]
47. Sarlin, P.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superglue: Learning feature matching with graph neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4938–4947.
48. Gilmer, J.; Schoenholz, S.; Riley, P.; Vinyals, O.; Dahl, G. Neural message passing for quantum chemistry. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1263–1272.
49. Ukita, N.; Moriguchi, Y.; Hagita, N. People re-identification across non-overlapping cameras using group features. *Comput. Vis. Image Underst.* **2016**, *144*, 228–236. [[CrossRef](#)]
50. Li, Y.; Gu, C.; Dullien, T.; Vinyals, O.; Kohli, P. Graph matching networks for learning the similarity of graph structured objects. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 3835–3845.
51. Zhang, S.; Zhang, Q.; Yang, Y.; Wei, X.; Wang, P.; Jiao, B.; Zhang, Y. Person re-identification in aerial imagery. *IEEE Trans. Multimed.* **2020**, *23*, 281–291. [[CrossRef](#)]
52. Xiong, J.; Lai, J. Similarity Metric Learning for RGB-Infrared Group Re-Identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 13662–13671.
53. Wu, A.; Zheng, W.; Yu, H.; Gong, S.; Lai, J. Rgb-infrared cross-modality person re-identification. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5380–5389.
54. Xia, G.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. [[CrossRef](#)]
55. Brown, M.; Süssstrunk, S. Multi-spectral SIFT for scene category recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 20–25 June 2011; pp. 177–184.

56. Luo, H.; Jiang, W.; Gu, Y.; Liu, F.; Liao, X.; Lai, S.; Gu, J. A strong baseline and batch normalization neck for deep person re-identification. *IEEE Trans. Multimed.* **2019**, *22*, 2597–2609. [[CrossRef](#)]
57. Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; Wang, S. Beyond part models: Person retrieval with refined part pooling (and A strong convolutional baseline). In Proceedings of the IEEE/CVF Conference on European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
58. Zhou, K.; Yang, Y.; Cavallaro, A.; Xiang, T. Omni-scale feature learning for person re-identification. In Proceedings of the IEEE/CVF Conference on International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3701–3711.
59. Sun, Y.; Cheng, C.; Zhang, Y.; Zhang, C.; Zheng, L.; Wang, Z.; Wei, Y. Circle loss: A unified perspective of pair similarity optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6398–6407.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.