remote sensing

Article

An Adaptively Attention-Driven Cascade Part-Based Graph Embedding Framework for UAV Object Re-Identification

Bo Shen [1,2], Rui Zhang [1,2,*] and Hao Chen [3]

[1] State Key Laboratory of Information Security, Institute of Information Engineering, CAS, Beijing 100093, China; shenbo@iie.ac.cn
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100093, China
[3] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China; hit_hao@hit.edu.cn
* Correspondence: r-zhang@iie.ac.cn; Tel.: +86-186-1173-3471

Abstract: With the rapid development of unmanned aerial vehicles (UAVs), object re-identification (Re-ID) based on the UAV platforms has attracted increasing attention, and several excellent achievements have been shown in the traditional scenarios. However, object Re-ID in aerial imagery acquired from the UAVs is still a challenging task, which is mainly due to the reason that variable locations and diverse viewpoints in UAVs platform are always resulting in more appearance ambiguities among the intra-objects and inter-objects. To address the above issues, in this paper, we proposed an adaptively attention-driven cascade part-based graph embedding framework (AAD-CPGE) for UAV object Re-ID. The AAD-CPGE aims to optimally fuse node features and their topological characteristics on the multi-scale structured graphs of parts-based objects, and then adaptively learn the most correlated information for improving the object Re-ID performance. Specifically, we first executed GCNs on the parts-based cascade node feature graphs and topological feature graphs for acquiring multi-scale structured-graph feature representations. After that, we designed a self-attention-based module for adaptive node and topological features fusion on the constructed hierarchical parts-based graphs. Finally, these learning hybrid graph-structured features with the most correlation discriminative capability were applied for object Re-ID. Several experimental verifications on three widely used UAVs-based benchmark datasets were carried out, and comparison with some state-of-the-art object Re-ID approaches validated the effectiveness and benefits of our proposed AAD-CPGE Re-ID framework.

Keywords: object re-identification; graph convolutional networks; unmanned aerial vehicle; attention mechanism; embedding learning

check for updates

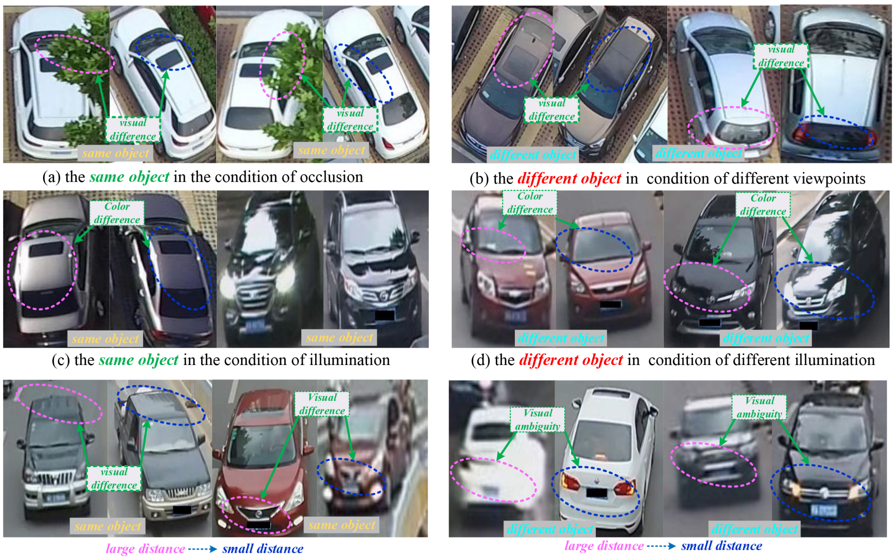
Citation: Shen, B.; Zhang, R.; Chen, H. An Adaptively Attention-Driven Cascade Part-Based Graph Embedding Framework for UAV Object Re-Identification. *Remote Sens.* **2022**, *14*, 1436. https://doi.org/10.3390/rs14061436

Academic Editors: Chunhui Zhao, Xiuping Jia, Wei Li, Shou Feng, Nan Su and Yiming Yan

Received: 9 February 2022
Accepted: 14 March 2022
Published: 16 March 2022

1. Introduction

Object re-identification (Re-ID), which is intended to re-identify the same individual from other non-overlapping cameras, has been drawing significant attention in the area of intelligent video surveillance [1], as well as playing a crucial role in a large variety of remote sensing and monitoring applications, such as urban planning [2], intelligent transportation [3], security monitoring, cross-camera tracking [4] and so on. Meanwhile, with the rapid development of modern intelligent technologies, unmanned aerial vehicles (UAVs) have gradually expanded their application fields, which is mainly due to their advantages, such as high mobility, flexible deployment, elastic service, etc. [5]. In addition, the remarkable achievements of the current UAV systems in the perspective of accuracy, efficiency and tracking have also been boosting the process and prosperity in adding autonomous vehicles for navigation, surveillance and more applications [6]. With the widespread availability of aerial images from the UAV platform, object Re-ID has become an essential fundamental task for several visual surveillance applications. As an excellent supplementary tactic for the conventional surveillance scenarios, UAV-based object Re-ID has been gaining interest and exhibits a prosperous development trend, which also facilitates more and more efforts devoted to the study of UAV-based object Re-ID from both industrial and academia [7,8]. Compared to object Re-ID in traditional scenarios where the imagery acquisition relied on cameras in fixed location and viewpoint, object Re-ID in aerial imagery taken from the UAV platform generally faces more challenges; some typical challenges are shown in Figure 1. It is worth noting that the UAV platform possesses the full coverage capability of complete viewpoints and a more comprehensive range of flight altitudes, which brings with it large diversity, as well as object ambiguity in the aerial images caused by circumstances, variations of viewpoints, poses, illumination [9–11], etc. For example, in some special imaging conditions, such as the narrow oblique perspective of UAVs, different objects of the collected aerial imagery always share identical contour characteristics and similar visual appearance. As we know, the core of Re-ID is to find an embedded feature space in which the objects within the same category gathers, but objects belonging to different categories diverge [12,13]. In this situation, the implementation of UAV-based object Re-ID requires feature representation, which enables it to develop a more compact and robust feature to distinguish each object. In addition, different from object detection, classification and tracking tasks, object Re-ID pays more attention to local regions that contain fine-grained discriminative information, which is required for a more powerful ability for feature discrimination. All of these, taken into account together, make object Re-ID in the aerial imagery of UAVs a more challenging task.



(a) the *same object* in the condition of occlusion (b) the *different object* in condition of different viewpoints

(c) the *same object* in the condition of illumination (d) the *different object* in condition of different illumination

(e) the *same object* acquired from different distance ranges (f) the *different object* acquired from different distance ranges

Figure 1. Illustration of parts of the challenges faced by UAV-based object Re-ID platform: (**a**) object appearance ambiguity by occlusions; (**b**) viewpoint variance resulted in dramatically different appearances of the same object; (**c**) illumination variance led to appearance ambiguity of the same objects; (**d**) illumination variances led to appearance ambiguity of the different objects; (**e**) resolution variances led to a visual difference of the same objects in different distances; (**f**) resolution variances led to the visual difference of the different objects in different distances.

Essentially, one main issue for UAV-based Re-ID is to learn discriminative and invariant features, which aims to ensure that the intra-object variations usually are more significant than inter-object similarities. Following this, the executions of Re-ID methods try to carry out a mapping operation for embedding object images into discriminative and compact feature space, then compute the similarity between the query and gallery images. In recent years, deep-learning-based approaches, such as feature representation learning based [14], metric-learning-based [15] and adversarial-learning-based [16] methods, have been achieved in some UAV-based Re-ID applications. Most of these kinds of Re-ID approaches depend on the supervised learning strategy, which is generally implemented by large-scale labeled training data to distinguish between objects with different identities. For example, Zhang et al. [17] constructed a large-scale UAV-based Re-ID dataset (Person Re-ID in Aerial Imagery, PRAI-1581) and then proposed to make use of subspace pooling operation on the feature maps that work by exacting the convolutional neural network (CNN) to represent the object instances in the aerial imagery for Re-ID. Yu et al. [18] proposed an asymmetric metric-based unsupervised learning framework for object Re-ID, which derived an asymmetric distance metric based on cross-viewing clustering and also combined it with a novel loss function for feature embedding under the deep neural network. In such a way, this method can effectively learn the compact feature representations for alleviating the view-specific distortions of object appearances and then further the Re-ID performance using these highly discriminative features. Guo et al. [19] present a pedestrian multiview generative adversarial networks (GAN)-based Re-ID method, which introduced the Monte Carlo search (MCS) and the attention mechanism into the adversarial learning process of the generator and the discriminator for learning enough detailed semantic features with high discriminative ability.

Although these approaches have effectively enhanced the object representation capability, their performance often lies in a substantial amount of data annotations acquired by tedious data collection and time-consuming processes. The robustness of these methods always drops dramatically in some special UAV-based Re-ID applications. In addition, most of the existing object Re-ID works mainly focus on identifying the intra-object variations and inter-object similarities from the perspective of visual appearance and neglect the exploration of spatial information for object Re-ID; however, on the UAV platform, the same object shows different visual appearances on the different spatial locations in some special imaging conditions, which is mainly due to the variations of spatial resolutions and different spatial implications. Furthermore, subtle cues deriving from implicitly and explicitly spatial relations in global and local regions can help identify different categories of objects with highly similar visual characteristics.

To further improve the performance of UAV-based object Re-ID platforms, in this paper, we propose an adaptively attention-driven cascade part-based graph embedding framework (AAD-CPGE). The AAD-CPEG Re-ID consisted of multiple part-based cascade graph convolutional network (GCN) branches incorporated by the self attention-based driven multi-graphs fusion module. The contributions of this paper can be summarized as follows:

(1) We designed a hierarchical part-based graph construction module to effectively derive multi-scale (or channel) spatial relations among different feature maps extracted from the pre-trained CNN-based model. Then, two sets of parts-based cascade node feature graphs and topological feature or (structure feature) graphs are constructed for exploring the global and local spatial relation representations from the different perspectives of the spatial domain.

(2) An self attention-driven based module was designed for adaptively fusing the node and topological features of the constructed complex hierarchical subgraphs by imposing the consistency and disparity constraints in the embedding feature space, as well as yielding highly discriminative features for disguising the spatial variations and appearance ambiguity among inter-objects and intra-objects.

(3) We also designed a novel loss function combined with the discrimination of graph node feature, topological structures and their combination in the corresponding embedding spaces, which can learn the explicitly and implicitly graph-based relations for further boosting object performance Re-ID.

Extensive experimental verifications on several benchmark datasets indicated the effectiveness and superiority of the proposed method. The remainder of this paper can be organized as follows. Section 2 introduced the related work. Section 3 described the proposed AAD-CPGE framework. Section 4 presented a series of experimental verifications and analyses to testify to the superiority. Section 5 gives the conclusion of this paper.

2. Related Work

Over the past years, UAVs have been regarded as a potential solution to surveil public spaces to refine public security and save labor costs. This kind of solution can be particularly effective in a lot of real-world visual applications [20,21], especially for the object Re-ID task. Meanwhile, a lot of the UAV-based datasets are available to the research community, which allows further blooming of the popularization and development of object Re-ID in a broader range of UAV-based applications [22–25]. Most of the existing approaches try to address the problems of object Re-ID approaches from the perspectives of feature representation and metric learning problem [26,27]. Before the emergence of deep learning, the Re-ID approaches mainly resorted to various types of conventional hand-crafted features, such as color, texture and gradient; however, these low-level feature representations are generally limited for large-scale searches and lack the ability to capture more semantic information. With the help of deep learning technologies, the traditional hand-crafted features are gradually replaced by some more advanced learning-based features, which directly learn highly discriminative representations from a large number of training data automatically.

Among the exiting deep-learning-based object Re-ID approaches, the unsupervised-learning-based approach was the mainstream for dealing with the Re-ID problem [28,29]. The main reason is that unsupervised learning methods enable potentially addressing the issues of Re-ID by drawing efficient deductive cues directly from the unlabeled image data. For example, Deng et al. [30] proposed an unsupervised learning-based Re-ID framework, which adopted the self-similarity and domain-dissimilarity measures for exploring the underlying discriminative information. Wang et al. [31] presented an attribute-identity-aware unsupervised learning approach, which learned high-level semantic information and attributed characteristics from the source domain and transferred them to the target domain for object Re-ID. Yu et al. [32] introduced a deep model into an unsupervised multi-label learning process, whose execution is realized by a complex negative mining strategy with the guidance of the soft multi-labeling procedure. Wu et al. [33] adopted a GAN to generate unlabeled images for the dataset augmentation and utilized a CNN-based semi-supervised learning strategy for object Re-ID; however, these approaches have achieved promising results on the recent object Re-ID task. This kind of Re-ID approach can fully utilize the annotation on color information; on the other hand, this approach requires many object identities, which are hard to acquire in real-world scenarios.

To overcome the above drawbacks, metric learning-based have been widely explored and achieved significant success in the Re-ID task. The core of metric learning is to learn the similarity of input images under two-stream or multi-stream parallel networks, which ensures that the distance of objects with the same category becomes closer, and the distance of objects with different categories becomes farther as much as possible [34,35]. For object Re-ID, the implementation of metric learning aims to make the distance between two vehicles with the same identity smaller than the distance between the vehicles with different identities. For this purpose, several metric learning-based have been explored for improving the object Re-ID performance. For example, Cui et al. [36] employed a dual-stream multi-DNN fusion Siamese neural network (MFSNN), which integrated with the color and structure information in the embedded feature space for feature discrimination. Bai et al. [37] proposed a group-sensitive-triplet embedding (GS-TRE) network, which embedded the group-sensitive triplets into the feature space for object group-based similarity discrimination. Zhang et al. [38] studied a triplet-wise-based training strategy that captures the relative similarity among each triplet unit and also designed an advanced classification-oriented loss function for improving the object Re-ID performance. Although these state-of-the-art metric learning approaches have achieved excellent performance for object Re-ID, there is still plenty of room for further enhancement. In addition, most of these measured learning-based Re-ID approaches always perform poorly when lacking training samples or imaging conditions changed drastically, etc.

In recent years, the attention mechanism has received huge concern and also achieved great successes in the task of object Re-ID [39]. The attention mechanism provides guidance information for the network and helps it focus on the discriminative local information. Several attention-based approaches for Re-ID generally employ hard attention, soft attention and some variants from these two basic attention mechanisms to address the misalignment issue for the task of object Re-ID, which is caused by the variations of object appearance, location, viewpoints and other imaging conditions. For example, Yao et al. [40] proposed an attention mask-based network for UAV vehicle Re-ID, which is combined with the principal component analysis (PCA) method for obtaining the color annotation-based attention masks (AMs), and also provided guidance for object Re-ID. Guo et al. [41] proposed a two-level attention network supervised by a multi-grain ranking loss (TAMR) for object Re-ID, this method fused hard part-level attention and soft pixel-level attention mechanism and introduced them into the framework of the backbone for feature discrimination. The kind of attention-based object Re-ID approach enables automatically learning the feature representations on high discriminative regions, resulting in the improvement of accuracy for object Re-ID; however, we find that most of the existing attention-based Re-ID methods mainly pay attention to the appearance information of image regions and neglect the underlying spatial relations of the set of more discriminative image regions. In this situation, these attention-based approaches always perform poorly when the datasets are less labeled and the background is more complex.

More recently, graph convolutional networks (GCNs) [42], as an extension of the traditional CNN, have been demonstrated efficiently for graph-structured data representation and intelligent learning areas. The GCNs initially applied for learning features on non-Euclidean data because it enables the flexibly aggregate of the information that passes on the graph nodes. The operation of graph convolution is directly executed on the graph nodes and their spatial neighbors. Recent works also show that the GCNs owns high potential ability for several visual applications of computer vision, such as multi-label recognition [43], action recognition [44], object Re-ID [45–48], etc.; however, these existing GCN-based object Re-ID approaches mainly focus on learning the individual relations of object categories and ignore the exploration for implicit semantic relations. The main differences between the methods above of our proposed method can be summarized as the following points: First, we present a method to construct the part-based hierarchical graph model, which completely considers the spatial relations across multiple scales feature maps among different parts of the inter-objects and inter-objects. Second, our method captures the spatial relations among the part-based hierarchical graph and derives the underlying topological relationships of the graph nodes in the embedding feature space. Meanwhile, the similarity between graph node features and that are inferred by topological relations are complementary and can be fused to derive deeper correlation information for learning more discriminative features. Third, compared with other fusion or concatenation models, our proposed method introduced the self-attention mechanism into the spatial and topological feature fusion process, which can select the important features on the constructed hierarchical graph and make the context-aware hybrid feature representations more robust for improving the object Re-ID performance.

3. Proposed Method

Due to the attributes of variable locations, flexible altitude-flights and adjustable viewpoints in UAVs platforms, the object appearances in aerial imagery naturally exhibit high ambiguity among different object categories. In this situation, learning discriminative and robust feature representation is crucial for the UAV-based object Re-ID task, where intra-object visual variations are always more considerable than inter-object similarities. To address these drawbacks, we proposed the called AAD-CPGE framework for UAV-based object Re-ID. The motivation of the proposed method is to derive more discriminative and compact features from the varying spatial cues of obtained multi-scale feature maps, which can effectively improve Re-ID performance by exploring the spatial significance of hierarchical feature maps. As shown in Figure 2, the implementation of our proposed AAD-CPGE framework includes three stages, which executes the task for UAV-based object re-identification (Re-ID). In the first stage, we adopt the pre-training CNN model to obtain each imagery's feature map and then extract the series of part-based features. In the second stage, the cascade topological graphs and node feature graphs are constructed, respectively, designed to explicitly represent hierarchical spatial relations and topological structures of node features. Meanwhile, based on the cascade node feature graphs and their topological graphs, we exploited a dual-stream multiscale graph convolutional operation on the graph-structured feature space and topological space of two kinds of constructed graphs to learn spatial topological embedding of the node features. We further utilized the advantage of the attention mechanism for adaptively fusing the learned node features and topological structures, which is achieved by automatically learning the importance weights for the above two embeddings. In the last stage, the fused features with the most correlations and highly discrimination are fed into the perceptron layer to enhance the object Re-ID performance.

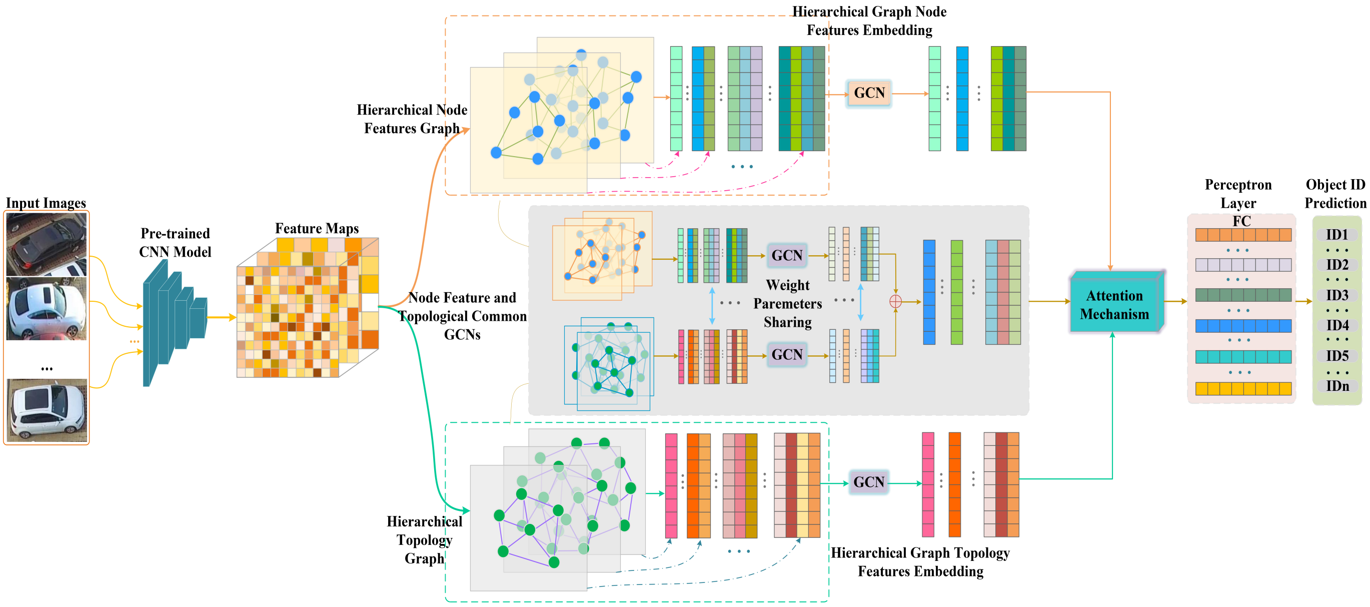

Hierarchical Graph Node Features Embedding
Hierarchical Node Features Graph
GCN
Input Images
Pre-trained CNN Model
Feature Maps
Node Feature and Topological Common GCNs
GCN
Weight Paremeters Sharing
GCN
Attention Mechanism
Perceptron Layer
FC
Object ID Prediction
ID1
ID2
ID3
ID4
ID5
IDn
Hierarchical Topology Graph
GCN
Hierarchical Graph Topology Features Embedding

Figure 2. The whole architecture of our proposed AAD-CPGE framework of object Re-ID.

3.1. Part-Based Feature Extraction Using the Pre-Trained CNN Model

For the given series of object images, $I = \{I_1, I_2, \ldots, I_N\}$, we first utilized the modified ResNet50 [49] network pre-trained on ImageNet [50] for extracting the corresponding feature representation. Following the work in Reference [51], we first disregarded the global average pooling layer and the fully connected layer, and the stride of conv4_1 is empirically set to 1. In this way, for the extracted feature map, we adopted the pooling strategy that carried out spatial down-sampling operations on the uniform feature layers (partitions) by setting three different down-sampling factors, which were intended to acquire part-based multi-scale feature representations of each imagery. In here, the union of feature representations with different parts of the k-th partition of the imagery I can be denoted as $X^{(k)} = \left(x_1^k, x_2^k, \cdots, x_i^k, \cdots, x_{n_k}^k\right)$. Then, we leveraged one-layer orthodox convolutional for dimensionality reduction in the part-based visual feature in each of the partitions, whose formulation can be defined as:

$$f_i^k = \text{Conv}\left(x_i^k\right), \quad i = 1, 2, \ldots, n_k; \tag{1}$$

where $f_i^k \in \mathbb{R}^{d \times 1}$ denotes the feature representation of the i-th part of the k-th partition scale. d is the dimension of each part-based feature vector and n_k is the number of parts in the k-th partition scale. Based the series of part-based feature representations of feature maps with the corresponding spatial scale, we denoted the feature collection at each partition scale as $F^{(k)} = \left(f_1^k, f_2^k, \cdots, f_{n_k}^k\right) \in \mathbb{R}^{d \times n_k}$. In here, we note that the partition scale (or feature scale) is empirically set to $k = 1, 2, 3$ in our following experimental verifications.

3.2. Part-Based Multi-Scale Graph Construction

To explore and utilize the spatial and topological relations among hierarchical part-based visual feature maps, we adopted an part-based multi-scale graph construction strategy for hierarchical graph-structured representation, and then applied it for graph embedding in node feature space and the topology space. Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ denote the constructed multi-scale (hierarchical) spatial graphs with three feature channels, where the collection of node features in all the hierarchical graphs can be described as $V = \{V^{(k)}, k = 1, 2, 3\}$, $V^{(k)}$ represented the graph nodes at the k-th channel of the feature map, $E = \{e_{ij}^{(k)}, k = 1, 2, 3\}$. In each hierarchical feature map, $v_i^{(k)} \in V^{(k)}$ denotes each part-based graph node, which is assigned with a corresponding feature vector $v_i^{(k)}$, the concatenation of node features in each channel of feature map can be denoted as $X = [X^{(1)} \| X^{(2)} \| X^{(3)}]$, where $X^{(k)} = (x_1^k, x_2^k \cdots x_{n_k}^k)$ is the collection of all the part-based features vectors that extracted from the k-th feature map. Meanwhile, among the edges collection E, each edge depicted the pairwise relations between every two patches in the hierarchical graphs. Inspired by [52], the pairwise relations between every two parts of the feature maps can be formulated as:

$$e(x_i, x_j) = \phi(x_i)^{\mathrm{T}} \varphi(x_j); \tag{2}$$

where ϕ and φ indicated, two symmetrical transformations applied for mapping the original part-based appearance features into the latent feature spaces. More specifically, the above two kinds of mapping functions can be defined as $\phi(\mathbf{x}) = \mathbf{wx}$ and $\phi'(\mathbf{x}) = \mathbf{w}'\mathbf{x}$. The matrices of weight parameters $\mathbf{W}$ and $\mathbf{W}'$ owned some dimensions with $d \times d$, which can be achieved by learning the GCNs via backpropagation. Introducing these transformations into the procedure for relations modeling allows us to learn the correlations among different parts of the objects within the same channel of the feature map and derive the relations of different object parts based across different channels of feature maps. Following this way, we could acquire the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ associated to the hierarchical graph G, where each entry of $\mathbf{A}_{ij}$ indicated the relations of each pair of node v_i and v_j, then the adjacency can be regarded as the key component for relations learning through the GCNs. As we know, the affinity matrix $\mathbf{A}_{(i,j)}$ should satisfy two conditions for the implementation of matrix-based graph convolutional operations: (1) In each row of the constructed affinity matrix, the sum of all the edge values that indicated the connectivity of different parts should be 1; (2) each entry of the adjacency should be non-negative, the weight coefficient should be in the range of (0, 1). For achieving the aforementioned points, we exploited the normalization operation on each row of the adjacent matrix $\mathbf{A}$ by the following formulations:

$$\mathbf{A}_{(i,j)} = \frac{e^2(x_i, x_j)}{\sum_{j=1}^{N} e^2(x_i, x_j)}; \tag{3}$$

To construct the node embedding of hierarchical graph $\mathbf{L}$, spectral filtering on graph is defined as a signal x filtered by $g_{\boldsymbol{\theta}} = \mathrm{diag}(\boldsymbol{\theta})$ in the Fourier domain, namely:

$$g_{\boldsymbol{\theta}} \star \mathbf{x} = \mathbf{U} g_{\boldsymbol{\theta}} \mathbf{U}^{\top} \mathbf{x}; \tag{4}$$

where $\mathbf{U}$ is the matrix composed of the eigenvectors of the normalized graph Laplacian matrix $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} = \mathbf{U} \Lambda \mathbf{U}^{\top}$. Λ denotes a diagonal matrix containing the eigenvalues of $\mathbf{L}$, $\mathbf{D}$ is the degree matrix $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ and $\mathbf{I}_n$ is the identity matrix $I_n = \mathrm{diag}(1, 1, \ldots, 1)$. In order to reduce the computational consumption of eigenvector decomposition of Equation (4), Hanmond et al. [53] simplified it as the following approximation:

$$g_{\theta' \star \mathbf{x}} \approx \sum_{k=0}^{K} \theta'_k T_k(\tilde{\mathbf{L}})\mathbf{x}; \tag{5}$$

where $T_k(\mathbf{x}) = 2\mathbf{x}T_{k-1}(\mathbf{x}) - T_{k-2}(\mathbf{x})$ denotes the Chebyshev polynomials with $T_0(\mathbf{x}) = 1$ and $T_1(\mathbf{x}) = x$, $\widetilde{\mathbf{L}} = 2/(\mathbf{r}_{\max})\mathbf{L} - \mathbf{I}_n$ is the scaled Laplacian matrix, $\mathbf{r}_{\max}$ is the largest eigenvalue of $\mathbf{L}$; therefore, the above equation can be verified by using the formulation $\left(\mathbf{U}\Lambda\mathbf{U}^{\top}\right)^k = \mathbf{U}\Lambda^k\mathbf{U}^{\top}$. As can be seen, this expression is a kth-order polynomial regarding the Laplacian. In this paper, we only considered the first-order neighborhood, i.e., $k = 1$, and thus, Equation (5) can be defined as a linear function on the graph Laplacian spectrum. Then, inspired by the work in reference [54], Equation (5) can be further simplified to the following definition:

$$g_{\theta'} \star \mathbf{x} \approx \theta'_0\mathbf{x} + \theta'_1(\mathbf{L} - \mathbf{I})\mathbf{x} = \theta'_0\mathbf{x} - \theta'_1\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{x}; \tag{6}$$

where θ'_0 and θ'_1 denotes two free parameters. For reducing the number of parameters to address overfitting, the above equation can be converted to:

$$g_{\theta} \star \mathbf{x} \approx \theta\left(\mathbf{I} + \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\right)\mathbf{x}; \tag{7}$$

In here, let $\theta = \theta'_0 = -\theta'_1$. To improve the robustness of the graph learning, we adopted a re-normalization trick to approximate the graph-Laplacian as follows:

$$\widehat{\mathbf{A}} = \widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}; \tag{8}$$

where $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$ indicated the self-loop adjacency matrix, $\mathbf{I_n} \in \mathbb{R}^{N\times N}$ is the identity matrix and $I_n = \mathrm{diag}(1, 1, \ldots, 1)$, $\widetilde{\mathbf{D}}_{(i,i)} = \sum_j \widetilde{\mathbf{A}}_{(i,j)}$ is the diagonal degree matrix of the adjacency. Then, we can acquire the whole adjacent matrix $\widehat{\mathbf{A}}_f$ of the node features in the above hierarchical graph $G(A, X)$. Since in the topology space, the input for constructing hierarchical topological graphs is similar to the node feature graphs, that is, let the topological graph be $G_t = (\mathbf{A}_t, \mathbf{X}_t)$, which satisfies the conditions $A_t = A$ and $\mathbf{X}_t = \mathbf{X}$. Following this perspective, the strategy for graph embedding on the topological graph can be implemented in the same way as the node feature space; therefore, the part-based specific information encoded in topology space can be obtained. After that, we can further explore the company relation's learning and the most optimal integration between node features and topological structures of the part-based hierarchical graphs in the embedded feature space and the topology space.

4. Graph Embedding on Features and Topology Spaces of Hierarchical Graphs

The primary purpose of our proposed AAD-CPGE framework is to learn a contextual and compact representation for each object by exploring different parts-based hierarchical spatial and topological graphs. It can result in more discriminative feature representations, which own the most correlated information from both node features and topological structures, thus improving the performance of object Re-ID; therefore, we designed the node feature hierarchical graph-based GCNs module and hierarchical topological graph GCN module to capture the compact relationships between parts in the feature space and topology space.

4.1. Graph Embedding in Feature and Topology Spaces

For the given hierarchical feature graph $G = (\mathbf{A}, \mathbf{X})$, we executed the graph convolutional operation to mine the relations of the parts among all the feature maps. Concretely, the implementation of the graph convolutional operations were carried out on L-layers GCNs, and the graph operations of the whole GCNs can be defined as:

$$\mathbf{Z}^l_f = \mathrm{LeakyReLU}(\widehat{\mathbf{A}}\mathbf{Z}^{l-1}_f\mathbf{W}^l_f); \tag{9}$$

where $\mathbf{Z}_f^{(l)} \in \mathbb{R}^{N \times d_l}$ denotes the set of hidden features for all the parts at l-th layer where $(1 \leq l \leq L)$, and d_l indicates the dimension of node features, $\mathbf{Z}_f^0 \in \mathbb{R}^{N \times d}$ represents the initial parts-based features that are extracted by the pre-trained CNN model. $\mathbf{W}_f^{(l)} \in \mathbb{R}^{d_l \times d_l}$ denotes the parameter matrix to be learned in each graph convolution layer. In here, we utilized LeakyReLU as the activation function whose formulation can be defined as:

$$\text{LeakyReLU}(Z_f) = \begin{cases} \alpha Z_f & \text{if } Z_f < 0 \\ Z_f & \text{otherwise} \end{cases}; \tag{10}$$

where the slope parameter was empirically set to $\alpha = 0.1$, the LeakyReLU function is adopted for non-linear transformation on the feature maps of GCNs. We note that the following types of GCNs are all set to the same α values for the non-linear transformations. After the graph convolutional operations in every layer of GCNs, a normalization layer and the LeakyReLU activation function are connected following the final output layer of GCNs. Motivated by the works in Reference [49], we exploited shortcut connection to ensure the effectiveness and robustness of the whole operation in GCNs as:

$$\mathbf{Z}_f^l := \mathbf{Z}_f^l + \mathbf{Z}_f^{l-1}, 2 \leq l \leq L; \tag{11}$$

Afterward, we can achieve the final layer output embedding of hierarchical features as $\mathbf{Z}_f$.

In addition, given the above operations, we obtain the topology embedding output $\mathbf{Z}_t$ by changing the input of the GCNs as follows:

$$\mathbf{Z}_t^l = \text{LeakyReLU}(\widehat{\mathbf{A}}_t \mathbf{Z}_t^{l-1} \mathbf{W}_t^l); \tag{12}$$

$$\mathbf{Z}_t^l := \mathbf{Z}_t^l + \mathbf{Z}_t^{l-1}, 2 \leq l \leq L; \tag{13}$$

In this way, we can also learn the topology embedding that captured the topological relations $\mathbf{Z}_t$ of the hierarchical part-based graphs in the topology space, $\mathbf{W}_t^l$ denotes the l-th layer weight matrix.

It is widely recognized that the spaces of the structured-graph node features and their topology spaces are not completely irrelevant. Recent studies [55,56] on GCNs have shown that the correlation between graph node features and their topological features is a critical factor that affects the ability of GCN-based embedded feature learning. From this point, we believe that the similarity between hierarchical graph node features and that derived from their topological structure are complementary to each other, and the optimal fusion among the node features, topological structures and their combinations are also significant for improving the object Re-ID performance; therefore, we exploited a Siamese architecture-based common-GCN module [57] on our constructed hierarchical graphs, which aimed to extract the most correlated information between graph node features and their topological features. The implementation of the common-GCN module made use of a parameter sharing strategy to achieve the embedding shared in the above two spaces. Specifically, we first acquired the node embedding $\mathbf{Z}_t^l$ from the hierarchical topology graph $G = (\mathbf{A}_t, \mathbf{X})$ by utilizing Equation (7), and the l-th layer weight matrix $\mathbf{W}_t^l$ was also obtained. For ease of the common-GCN embedding description, we replaced the common node topology embedding $\mathbf{Z}_t^l$ and the corresponding weight matrix $\mathbf{W}_t^l$ by $\mathbf{Z}_{St}^l$ and $\mathbf{W}_{St}^l$, respectively. Then, to obtain the shared information between the topology and feature spaces, we introduced the common weight matrix $\mathbf{W}_S^l$ into the node embedding procedure of the common module from the hierarchical node feature graph whose implementation can be formulated as:

$$\mathbf{Z}_{Sf}^l = \text{LeakyReLU}(\widehat{\mathbf{A}}_{St} \mathbf{Z}_{Sf}^{l-1} \mathbf{W}_S^l); \tag{14}$$

$$\mathbf{Z}_{Sf}^l := \mathbf{Z}_{Sf}^l + \mathbf{Z}_{Sf}^{l-1}; \tag{15}$$

where $\mathbf{Z}^{l}_{Sf}$ indicates the l-th layer output embedding and $\mathbf{Z}^{(0)}_{Sf} = \mathbf{X}$. Following this, the shared weight matrix can efficiently filter out the shared characteristics from the feature and topology spaces. By changing different hierarchical graphs to apply the above weight parameter sharing strategy, we can acquire the node feature embedding $\mathbf{Z}^{l}_{Sf}$, topology embedding $\mathbf{Z}^{l}_{St}$. Then, their common embedding $\mathbf{Z}^{l}_{S}$ can also be obtained as follows:

$$\mathbf{Z}_{S} = \left(\mathbf{Z}_{Sf} + \mathbf{Z}_{St}\right)/2; \tag{16}$$

4.2. Attention-Driven Embedded Features Fusion in the Uniform Latent Spaces

Based on the aforementioned three categories of embedding, we take advantage of the self-attention mechanism to adaptively learn the importance of their weights for assigning the corresponding optimal combination weights for feature fusion. The mechanism of our adopted attention module is shown in Figure 3.

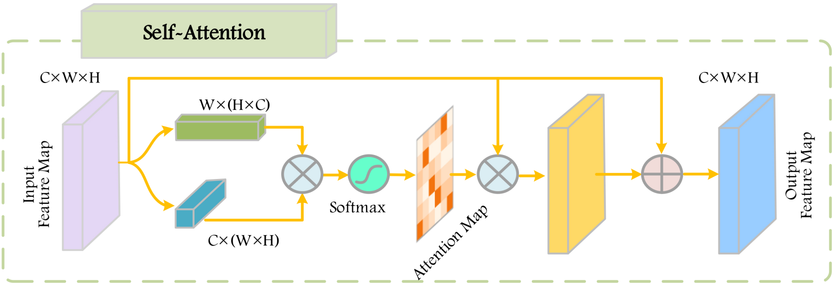


Figure 3. The diagram of the attention module.

The conventional way to implement feature fusion is regarding every relation graph embedding equally by conducting the element-wise, mean operation or concatenating them using a linear transformation. Nevertheless, every category of graph embedding that contains their specific underlying meaning should not be dealt with identically; therefore, we explored strategies to adaptively learn an important weight for every graph embedding and exploited the attention mechanism for implementation. Concretely, for each category of graph embedding, we performed a three-step process to learn their corresponding importance: (1) we first retrieved their corresponding embedding of all related graph nodes and transformed them by a non-linear transformation; (2) then, we aggregated the concatenations of their embeddings by utilizing the averaging function for their transformed embeddings; (3) finally, we computed their corresponding attention coefficients by measuring the similarity between their aggregated embedding with their attention vectors. This whole learning process can be formulated as:

$$u^{l}_{Sf} = \left[\mathbf{Q}^{l}_{Sf}\right]^{\top} \cdot \frac{1}{\left|V_{Sf}\right|} \sum_{v' \in V^{l}_{Sf}} \left[\tanh\left(\mathbf{W}^{l}_{Sf} \cdot \mathbf{h}^{l}_{v',Sf} + \mathbf{b}\right)\right]; \tag{17}$$

where u^{l}_{Sf} denotes the attention coefficients, V_{Sf} represents the series of graph nodes connected by the relation S_f, $\mathbf{b} \in \mathbb{R}^{d}$ is the bias vector, $\mathbf{W}^{l}_{Sf} \in \mathbb{R}^{d \times d}$ and $\mathbf{Q}^{l}_{Sf} \in \mathbb{R}^{d}$ are the corresponding trainable transformation matrix and attention vector, respectively. Using the same approach, we can achieve the attention coefficients u^{l}_{St} and u^{l}_{S} in the embedding adjacency matrices $\mathbf{Z}^{l}_{St}$ and $\mathbf{Z}^{l}_{S}$, respectively. Then, the normalized importance of node feature embedding S_f with regard to v can be obtained by:

$$\eta_{Sf} = \mathrm{softmax}\left(u^{l}_{sf}\right) = \frac{\exp\left(u^{l}_{sf}\right)}{\exp\left(u^{l}_{Sf}\right) + \exp\left(u^{l}_{St}\right) + \exp\left(u^{l}_{S}\right)}; \tag{18}$$

Similarly, we can achieve the normalized attention weights $\eta_{St} = \text{softmax}(u^l_{St})$ and $\eta_S = \text{softmax}(u^l_S)$. Then, we incorporated the above three embeddings to obtain the final embedding Z_E as follows:

$$\mathbf{Z}_E = \boldsymbol{\eta}_{Sf} \cdot \mathbf{Z}_{Sf} + \boldsymbol{\eta}_{St} \cdot \mathbf{Z}_{St} + \boldsymbol{\eta}_S \cdot \mathbf{Z}_S; \tag{19}$$

Then, for the concatenation of the above three types of hierarchical part-based graph embedding features, we utilized it to optimize our proposed AAD-CPGE framework and predicted the identity label of the corresponding object for the task of Re-ID.

4.3. Objective Functions

As depicted in Figure 2, the entire architecture of our proposed AAD-CPGE model consisted of three branches: the graph node feature embedding branch, a graph topology embedding branch and a branch of their combination that was designed to learn the most correlation information between the former two branches. Specifically, the graph feature embedding branch was designed for modeling the hierarchical spatial relations of parts among different scales of feature maps, which can effectively capture the global and local information among hierarchical part-based graphs. Then, the topology embedding branch is proposed for extracting the underlying complete structural information. Since the similarity between hierarchical graph node feature and their derived topological structures are complementary to each other, then we employed the combination of these two types of embeddings for adaptively deriving more deeper correlation information for the task of object Re-ID. Following this way, we utilized the softmax cross-entropy loss function and the batch hard triplet loss function to train the proposed AAD-CPGE network. For the definition of softmax cross-entropy loss, we first denoted the object label predictions for n hierarchical graph nodes as $\widehat{\mathbf{Y}} = [\hat{y}_{v_{im}}] \in \mathbb{R}^{n \times C}$, where $\hat{y}_{v_{im}}$ indicates the probability of the i-th graph node v_i belonging to the object class m. Then, the predictions for each class of graph node $\widehat{\mathbf{Y}}_m$ can be calculated as:

$$\widehat{\mathbf{Y}}_m = \text{softmax}(\mathbf{W_m} \cdot \mathbf{Z}_E + \mathbf{b_m}); \tag{20}$$

where $\text{softmax}(x) = \frac{\exp(x)}{\sum_{m=1}^{M} \exp(x_m)}$ denotes a normalizer across all the object identity classes; therefore, for the given selected training set C, where each $c \in C$, the real object label class is $\mathbf{Y}_c$ and the predicted object label class is $\widehat{\mathbf{Y}}_C$. By this way, the cross-entropy loss for hierarchical graph node class prediction over all the series of training nodes is defined as $\mathcal{L}_p$, which can be formulated as:

$$\mathcal{L}_p = -\sum_{c \in C} \sum_{m=1}^{M} \mathbf{Y}_C \ln \widehat{\mathbf{Y}}_C; \tag{21}$$

In addition, for the definition of the triplet loss function, we defined the triplet loss for three types of embedding features and the whole triplet loss can be represented as the sum of them. The uniform form of soft hard triplet loss can be defined as:

$$\begin{aligned} L_{triplet}(x) = \sum_{i=1}^{P} \sum_{a=1}^{K} \ln(1 + exp&(\overbrace{\max_{p=1,\cdots,K} D(\mathbf{x}_{i,a}, \mathbf{x}_{i,p}))}^{\text{hardest positive}} \\ &\underbrace{- \min_{\substack{n=1,\ldots,K \\ j=1,\ldots,P \\ j \neq i}} D(\mathbf{x}_{i,a}, \mathbf{x}_{j,n})))}_{\text{hardest negative}} \end{aligned} \tag{22}$$

where P and K denote the number of identity labels and sampled imagery of every object identity label. Supposed that there are $P \cdot K$ images in a mini-batch, $\mathbf{x}_{i,a}$, $\mathbf{x}_{i,p}$ and $\mathbf{x}_{j,n}$

indicate the features that obtained from anchor, positive and negative samples, respectively, $D(\cdot)$ represents the L2-norm distance between two feature vectors. By adjusting different inputs of the above uniform soft triplet loss function, we can achieve three types of triplet loss function, including graph node feature embedding loss L^{Sf}_{triplet}, topological structures embedding loss L^{St}_{triplet} and their combination embedding $L^{\text{E}}_{\text{triplet}}$. Then, the final triplet loss can be formulated as:

$$L^{S}_{\text{triplet}} = L^{\text{E}}_{\text{triplet}} + L^{Sf}_{\text{triplet}} + L^{St}_{\text{triplet}}; \tag{23}$$

Thus, the total loss function L_{total} can be defined as the combination of the above loss functions as:

$$L_{\text{total}} = L_{\text{p}} + \lambda L^{S}_{\text{triplet}}; \tag{24}$$

where λ is a balancing hyper parameter. After that, with the guidance of the set of labeled object samples, we were able to train the whole AAD-CPGE framework and optimize its model via the back-propagation algorithm, as well as learning all types of embeddings based on our constructed hierarchical part-based graphs for the task of object Re-ID.

5. Experiments and Results

In this section, we present comprehensive experimental validation and analysis, consisting of the datasets description, evaluation measure and the implementation details, which are first introduced in detail. After that, a series of ablation experiments were performed by adjusting different combinations of sub-networks, whose implementations aimed to demonstrate the contributions of each component of the whole architecture of our proposed Re-ID framework. We also carried out quantitative comparisons with several state-of-the-art object Re-ID approaches and showed their corresponding qualitative results.

5.1. Datasets and Evaluation Metrics

(1) **Dataset**: The experimental verifications of our proposed object Re-ID were mainly conducted on two publicly available datasets collected on the UAV platform, including the benchmark on unmanned aerial vehicle re-identification in video imagery (UAV-VeID) [8] and person Re-ID in aerial imagery (PRAI-1581) [17]. In addition, we also implemented a group of experiments on the traditional dataset VeRi-776 [58] for testifying the robustness of our proposed method. In the following, we first review the necessary information about the above two datasets that were adopted in our verified experiments in this paper.

The **UAV-VeID** [8] dataset consists of 41,917 images of 4601 vehicles, which are split into three subsets, the training set, testing set and validation set. The numbers of the images in the training set, testing set and validation set are 5862, 11,738, and 5683, respectively. The images of UAV-VeID are collected from video sequences that are acquired on the UAV platform, and these images are captured by UAV-mounted cameras from different locations with various backgrounds and lighting conditions, e.g., including the crossroads in urban areas, the highway intersections, parking lots and so on. The flying altitude of the UAV platform ranges from 15 to 60 m, and the vertical angle of the UAV camera is set in the range from 40 to 80°, which results in multi-scale objects, and also various viewpoints of the vehicle objects in the images.

PRAI-1581 [17] collects images by using two UAVs whose flight altitudes range from 20 to 60 m above the ground; these two UAV-based platforms are controlled by two different pilots to allow effectively monitoring different non-overlapping areas and covering most of the complex surveillance scenes. To capture enough videos in complete imaging conditions, such as more diverse viewpoints and backgrounds, the imagery collection process of PRAI-158 is adopted as the hovering, curing and rotating sports models to control the two UAVs. In this way, the PRAI-158 dataset contains a total number of 39,461 images for 1581 individual identities. During the process of our experimental verifications, we randomly divided the dataset into the training set and testing set. For a fair comparison, the ratio of image numbers between the training set and testing set is set to 1:1. Concretely, the training set contains 19,523 images of 782 identities, the remaining images of the dataset are regarded as the testing set with 19,938 images of 799 identities.

VeRi-776 [59,60] is constructed from unconstrained traffic scenarios where the object images are captured by 20 cameras. Similar to the former dataset, VeRi-776 has also been divided into a training set and a testing set. Specifically, the training set consists of 37,746 images of 576 objects, the testing set contains a query subset with 1678 images of 200 objects and a gallery subset with 11,579 images with the same 200 objects. It is worth recognizing that each object imagery in the VeRi-776 has been equipped with color annotation. During the testing process, we adopted the default data split ratio for experimental validation and analysis.

During the process of experimental verifications, we have given the specific acquired conditions for each experimental dataset and detailed information about the imagery information of the aforementioned dataset as listed in Table 1.

Table 1. Detailed setting comparisons between different experimental datasets.

Dataset	Cameras	Viewpoint	Total Image Number	Object Identities	Training Images	Testing Image
UAV-VeID (Experiment.1)	Mobile UAV	Flexible	41,917	4061	18,709	3742
PRAI-1581 (Experiment.2)	Mobile UAV	Flexible	39,461	1581	19,523	19,938
VeRi-776 (Experiment.3)	Fixed	3	51,035	776	37,746	11,579

(2) **Evaluation Metric**: To validate the superiority of our proposed method, we adopted the mean average precision (mAP) and the cumulative matching precision (CMC) for quantitative comparison. During the process of our experimental verification, each object imagery in the query sequence aimed to retrieve the same object in the gallery sequence in terms of the Euclidean distance, which is computed among the embedding features of queries and galleries. Note that there is only one ground truth that matches for an arbitrary input query in the testing set of UAV-VeID, we employed the CMC-k to estimate the Re-ID performance of the series of compared approaches, and indicate the probability of correct matching in the top-k ranked retrieved results. For the PRAI-158 and VeRi-776 datasets, the CMC-k and mAP are utilized as the evaluating metric for object Re-ID.

5.2. Implementation Details

We first utilized ResNet as our backbone network that was pre-trained on the ImageNet to acquire a convolutional feature map for two-stream networks. Afterward, we adopted the one-layer orthodox convolutional operation and multi-layer graph convolutional operations for dealing with the hierarchical learned part-based features. During the training process, we resized all the input samples into the size of 256 × 256 with random horizontal flips for data augmentation, the total number of epochs is set to 800 and the batch size is set as 80 in general for all the datasets, and then to initialize the leaning rate of the backbone network, we set it to 0.01. Meanwhile, the initial learning rate of the GCNs is set to 0.0003, which then decayed by 10 for every 200 epochs; the Adam is selected for optimizing our proposed network. For the parameters of our proposed GCNs-based modules, the number of GCN layers is set to 3. In the training stage of our proposed model, we concatenated all types of the hierarchical part-based graph features for each query imagery to generate its corresponding final feature descriptions. All the experimental verifications were employed with two NVIDIA Titan X GPUs on the same machine for a fair comparison. In order to completely compare and verify the effectiveness of our proposed method, we compare the proposed AAD-CPGE framework with several state-of-the-art Re-ID approaches on different datasets, the whole diagram of the set of experimental verifications is depicted in Figure 4:

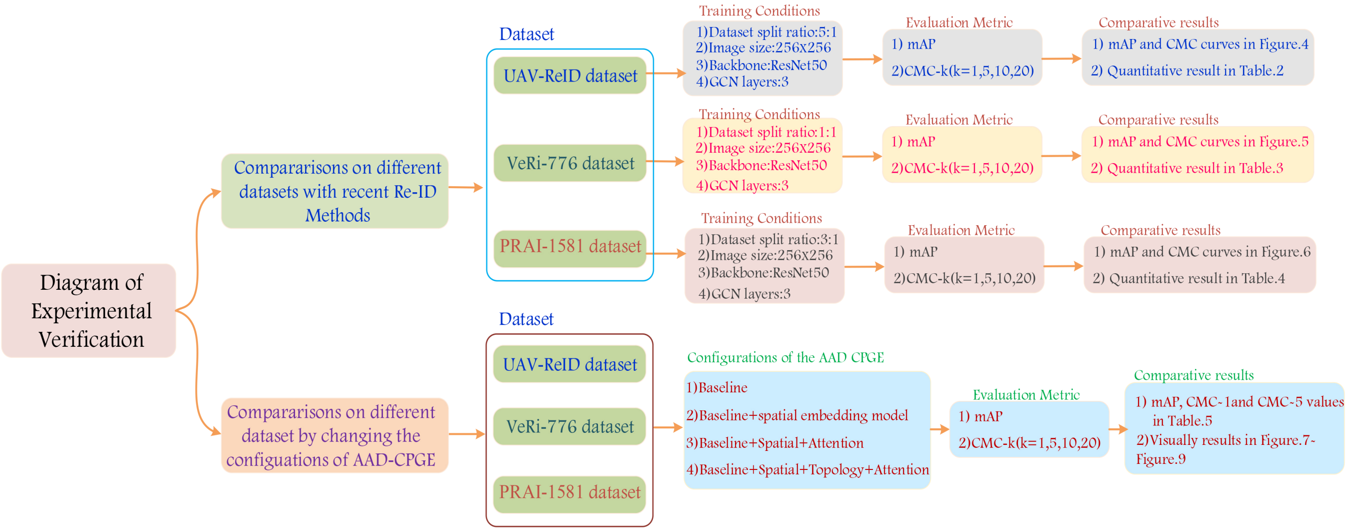


Diagram of Experimental Verification
Comparisons on different datasets with recent Re-ID Methods
Comparisons on different dataset by changing the configuations of AAD-CPGE
Dataset
UAV-ReID dataset
VeRi-776 dataset
PRAI-1581 dataset
Training Conditions
1)Dataset split ratio:5:1
2)Image size:256x256
3)Backbone:ResNet50
4)GCN layers:3
Evaluation Metric
1) mAP
2)CMC-k(k=1,5,10,20)
Comparative results
1) mAP and CMC curves in Figure.4
2) Quantitative result in Table.2
Training Conditions
1)Dataset split ratio:1:1
2)Image size:256x256
3)Backbone:ResNet50
4)GCN layers:3
Evaluation Metric
1) mAP
2)CMC-k(k=1,5,10,20)
Comparative results
1) mAP and CMC curves in Figure.5
2) Quantitative result in Table.3
Training Conditions
1)Dataset split ratio:3:1
2)Image size:256x256
3)Backbone:ResNet50
4)GCN layers:3
Evaluation Metric
1) mAP
2)CMC-k(k=1,5,10,20)
Comparative results
1) mAP and CMC curves in Figure.6
2) Quantitative result in Table.4
Dataset
UAV-ReID dataset
VeRi-776 dataset
PRAI-1581 dataset
Configurations of the AAD CPGE
1)Baseline
2)Baseline+spatial embedding model
3)Baseline+Spatial+Attention
4)Baseline+Spatial+Topology+Attention
Evaluation Metric
1) mAP
2)CMC-k(k=1,5,10,20)
Comparative results
1) mAP, CMC-1and CMC-5 values in Table.5
2)Visually results in Figure.7-Figure.9

Figure 4. The whole diagram of experimental verification on different datasets.

5.3. *Comparisons with State-of-the-Art Re-ID Approaches on Different Datasets*

To testify the effectiveness and the robustness of our proposed AAD-CPGE framework for object Re-ID, we compared our AAD-CPGE with several state-of-the-art object Re-ID approaches on the datasets of UAV-ReID, VeRi-776 and PRAI-1581, respectively. Specifically, we made comprehensive experimental comparisons with seven state-of-the-art Re-ID approaches, including the Bayes merging of multiple vocabularies (BoW-Shift) [61], the bag-of-words model with color names descriptors (Bow-CN) [62], part regularized model (PRM) [63], adaptive attention vehicle re-identification (AAVER) [14], self-supervised attention for vehicle re-identification (SAVER) [64], vehicle re-identification based on vehicle-orientation-camera (VOC-ReID) [65], parsing-based view-aware embedding network for vehicle re-identification (PVEN) [66], spatial–temporal graph convolutional networks (ST-GCN) [46], similarity-guided graph neural network (SG-GCN) [48] and hybrid pyramidal graph network (HPGN) [67].

5.3.1. Comparisons on the UAV-ReID Dataset

The comparisons of the object UAV-based Re-ID of different kinds of approaches on the UAV-ReID dataset are summarized in Table 2. As the set of quantitatively compared results in Table 2, we can find that our proposed AAD-CPGE framework using the hierarchical spatial and topological information generally achieves better performance than the traditional Re-ID methods with handcrafted features, and surpassed some of the recent unsupervised learning-based approaches and GCN-based Re-ID approaches.

Specifically, compared with the traditional hand-crafted features methods BoW-Shift and Bow-CN, the proposed AAD-CPGE, respectively, improved the CMC-1 matching accuracy and mAP by (56.90%, 66.59%), (54.25%, 61.51%). Similarly, compared with the baseline CNN-based methods (the PRM and VOC-ReID) by using mAP and CMC-1, the proposed method also performed obvious advantages for object Re-ID, which brought about an improvement of (32.1%, 53.52%) and (30.35%, 50.37%), respectively. Meanwhile, the AAD-CPGE also obtained notable improvement compared with the recent attention mechanism-based Re-ID methods including SAVER, PVEN and AAVER. We can see that in the comparisons in terms of mAP, our method achieved an improvement of 17.64%, 17.16% and 14.14% respectively. This indicated the effectiveness of our AAD-CPGE by exploiting the spatial and topological information via GCNs learning for the task of object Re-ID. In addition, compared to the kind of graph-based approaches, the proposed method achieved further improvement. The proposed AAD-CPGE outperformed SG-GCN by +12.14% in mAP and +27.01% in CMC-1, ST-GCN by +5.62% in mAP and +16.96% in CMC-1, and HPGN by +53.46% in mAP and +1.47% in CMC-1, respectively.

Table 2. Comparisons results(%) of object Re-ID on UAV-ReID dataset by using the metrics of mAP, CMC-1, CMC-5, CMC-10 and CMC-20, respectively.

Methods	mAP	CMC-1	CMC-5	CMC-10	CMC-20
BoW-Shift	27.69	30.56	36.87	40.12	50.14
BoW-CN	30.34	35.64	39.75	46.68	56.32
PRM	52.49	43.63	56.72	67.65	73.93
VOC-ReID	54.24	46.78	59.89	68.94	75.03
SAVER	66.95	62.37	69.85	71.74	77.86
PVEN	67.43	68.35	70.64	75.34	78.19
AAVER	70.45	71.21	74.16	78.95	84.57
SG-GCN	72.45	70.14	85.24	87.56	94.32
ST-GCN	78.97	80.19	84.67	88.46	96.34
HPGN	81.13	95.68	96.12	97.87	98.04
Ours	84.59	97.15	97.46	98.57	98.89

Although SG-GCN employed the spatial relation inferring between the set of gallery images for improving the feature discriminative ability, it only focused on modeling the inter-gallery-image relations and lacked considering the intra-object relations between each pair of gallery images. In contrast, the proposed framework jointly optimizes the correspondence learning among the series of intra-objects and inter-objects of gallery images, thus yielding better Re-ID performance. In addition, compared with the ST-GCN and HPGN, our proposed AAD-CPGE also jointly takes into account the node features and topological features in two embedding spaces, and also employs the attention mechanism for adaptively fusing the important features for improving the compact and completely discrimination for further enhancing the Re-ID performance. As a result, the proposed AAD-CPGE framework is capable of resulting in the best Re-ID performance compared with a lot of recent Re-ID approaches. In addition, several mAP curves and CMC curves comparisons that indicated the effectiveness of our proposed method are shown in Figure 5. These quantitative and qualitative compared results with some state-of-the-art Re-ID approaches demonstrated the effectiveness of the proposed AAD-CPGE framework by further exploiting the intrinsic spatial and structure information of object parts via an attention-driven embedding graph learning model. In addition, the AAD-CPEG performed better than some recent and traditional approaches, which demonstrates the superiority of our method for the object Re-ID task.

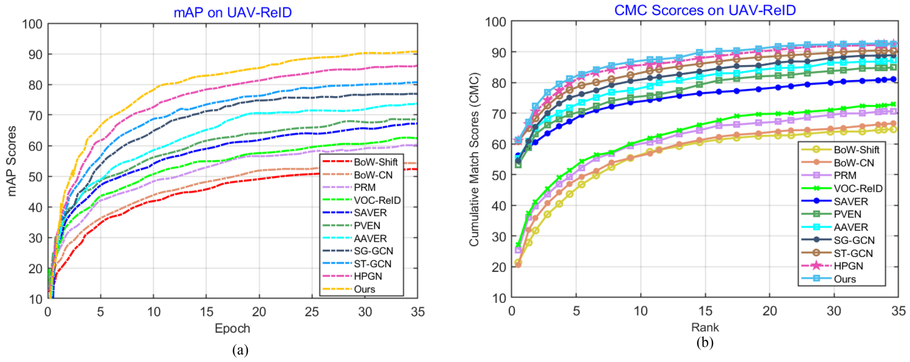


Figure 5. The mAP and CMC curve comparison of our method and several state-of-the-art approaches on UAV-ReID dataset. (**a**) mAP curve comparison on the UAV-ReID; (**b**) CMC curve comparison on the UAV-ReID.

5.3.2. Comparisons on the VeRi-776 Dataset

To demonstrate the effectiveness and robustness of our proposed AAD-CPGE framework for the task of object Re-ID, we also conducted another group of comparative experiments on the VeRi-776 dataset. Table 3 shows the quantitative results in terms of mAP and the CMC-k (k = 1, 5, 10, 20) values, which were aimed to evaluate the Re-ID performance of the approaches.

Table 3. Comparisons results(%) of object Re-ID on VeRi-776 dataset by using the metrics of mAP, CMC-1, CMC-5, CMC-10 and CMC-20, respectively.

Methods	mAP	CMC-1	CMC-5	CMC-10	CMC-20
BoW-Shift	31.37	37.09	44.91	48.57	60.86
BoW-CN	32.29	43.98	46.52	50.68	62.15
PRM	54.84	44.28	57.18	69.36	75.83
VOC-ReID	59.98	46.97	60.12	70.37	76.18
PVEN	68.79	52.65	62.83	75.47	80.56
AAVER	71.39	63.27	71.34	79.56	85.12
SAVER	75.96	73.57	86.79	88.24	95.43
SG-GCN	79.48	89.08	82.19	89.37	97.21
ST-GCN	79.60	92.43	97.08	97.35	98.01
HPGN	80.18	96.72	97.64	97.87	98.19
Ours	83.78	97.13	97.59	98.29	98.71

As can be seen in Table 3, the proposed AAD-CPGE can achieve 83.78% mAP and 93.13% on CMC-1 when jointly learning the embedded node features space and topology space derived from the hierarchical part-based graph, which performed a higher accuracy than the GCNs-based Re-ID approaches only exploring spatial relations in the feature space of graph nodes. It is obvious that our proposed AAD-CPGE brought about an improvement of 4.3%, 4.18% and 3.6% with SG-GCN, ST-GCN and HPGN, respectively, which were evaluated by the metric of mAP. Meanwhile, the AAD-CPGE also obtained slightly high CMC-1 matching accuracy against SG-GCN (97.13% vs. 89.08%), ST-GCN (97.13% vs. 92.43%) and HPGN (97.13% vs. 96.72%), respectively. Actually, the results are not difficult to understand, the main reason is that the AAD-CPEG introducing the attention module into the fusing the hierarchical graph nodes-based spatial and topological features, which was not only capturing the underlying spatial relations among different kinds intra-objects and inter-objects, but also provided incremental learning in the topology embedding spaces for further distinguish the ambiguity of objects with high correlations, and thus improving the performance for object Re-ID. Moreover, compared the traditional hand-crafted features-based approaches and the CNN baseline-based approaches for the task of Re-ID on the VeRi-776, our AAD-CPGE also superior to a large margin. Specifically, in terms of mAP, the proposed framework improved Re-ID performance by 52.1%, 51.49%, 28.94% and 23.8%, which were compared with BoW-Shift, BoW-CN, PRM and VOC-Re-ID, respectively. Similarly, based on the CMC-1 values, we also compared the Re-ID performances by using our proposed AAD-CPGE with the above hand-craft features-based methods and CNN-based methods, including BoW-Shift, BoW-CN, PRM and VOC-Re-ID, respectively. Concretely, collaborating with the BoW-Shift and BoW-CN, our AAD-CPGE brought about an improvement of 60.04% (97.13% vs. 37.09%) and 53.15% (97.13% vs. 43.98%), respectively. Comparing with the PRM and VOC-Re-ID, our proposed AAD-CPGE obtained an improvement of 52.85% (97.13% vs. 44.28%) and 50.16% (97.13% vs. 46.97%), respectively. These comparative results demonstrated the effectiveness of our method for the task of object Re-ID. After that, compared with the recent attention mechanism-based methods such as PVEN, AAVER and SAVER, our proposed AAD-CPGE framework also significantly outperformed them by 15.09%, 12.39% and 7.82% in terms mAP, as well as achieved an improvement of 44.48%, 33.86% and 23.56% in terms of CMC-1 matching accuracy, respectively. These comparison results reported in the above table demonstrated that our proposed AAD-CPGE Re-ID framework is superior to several state-of-the-art object Re-ID approaches. Moreover, the mAP curve and CMC curve as shown in Figure 6 also indicated the superiority of the proposed AAD-CPGE framework for the task of object Re-ID.

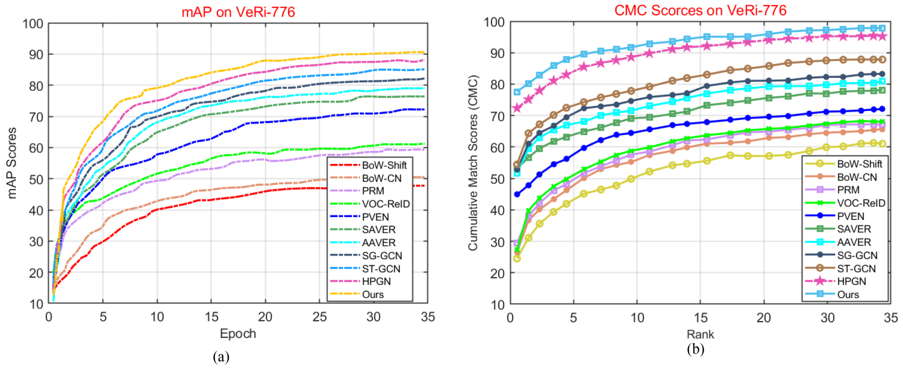


Figure 6. The mAP and CMC curve comparison of our method and several state-of-the-art approaches on VeRi-776 dataset. (**a**) mAP curve comparison on the VeRi-776; (**b**) CMC curve comparison on the VeRi-776.

5.3.3. Comparisons on the PRAI-1581 Dataset

To further testify the generality and the robustness of the proposed AAD-CPGE framework, we also provided quantitative comparisons in terms of mAP and CMC-k (*k* = 1, 5, 10, 20) values, which aimed at indicating the superiority of the proposed method on the larger and more challenging PRAI-1581 dataset UAV-based platform, where the images of object categories have ambiguous appearances, and also more variations of imaging conditions, such as the illumination, occlusion, background, viewpoint and so on. With the proposed method and also the compared object Re-ID approaches, we further carried out object Re-ID experiments on the PRAI-1581 dataset and demonstrated the quantitative comparisons with the existing state-of-the-art Re-ID approaches in Table 4.

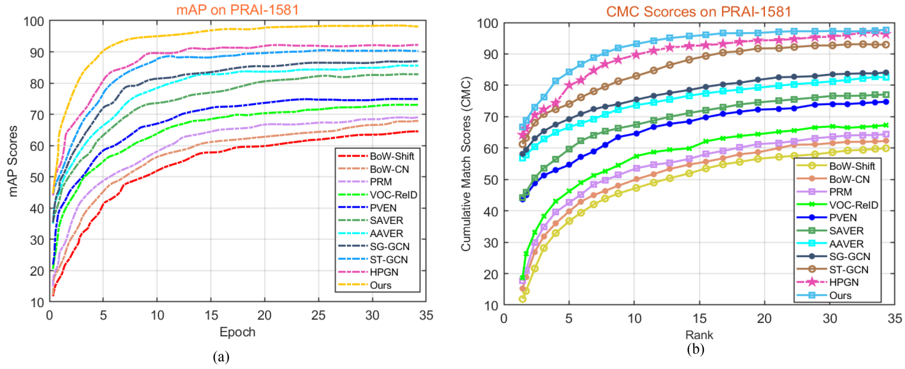
Table 4. Comparisons results(%) of object Re-ID on PRAI-1581 dataset by using the metrics of mAP, CMC-1, CMC-5, CMC-10 and CMC-20, respectively.

Methods	mAP	CMC-1	CMC-5	CMC-10	CMC-20
BoW-Shift	52.34	44.64	57.75	68.86	71.23
BoW-CN	54.94	48.98	56.65	70.32	74.15
PRM	55.84	50.98	62.43	70.46	75.94
VOC-ReID	62.98	52.97	63.42	74.56	77.16
PVEN	69.78	64.65	73.76	76.91	83.18
AAVER	78.45	64.76	78.39	84.36	86.19
SAVER	80.76	84.76	92.12	92.56	94.01
SG-GCN	83.70	89.95	96.41	97.06	97.48
ST-GCN	84.64	90.23	96.78	97.54	97.96
HPGN	85.12	92.45	97.32	97.65	98.21
Ours	90.34	95.23	97.68	98.53	98.89

As indicated in Table 4, among all kinds of the comparative Re-ID approaches, the proposed AAD-CPGE consistently boosted the state-of-the-art methods on average with 90.34% mAP and 95.23% CMC-1 values. Compared with the traditional methods that consist of the hand-crafted feature-based BoW-Shift, BoW-CN, the CNN-baseline PRM and VOC-ReID, the proposed AAD-CDGE, respectively, surpassed them in terms of mAP by 38.00% (90.34% vs. 52.34%), 35.40% (90.34% vs. 54.94%), 34.50% (90.34% vs. 55.84%) and

27.36% (90.34% vs. 62.98%), respectively. Meanwhile, the proposed method also brought about improvements of 50.59% (95.23% vs. 44.64%), 46.25% (95.23% vs. 48.98%), 44.25% (95.23% vs. 50.98%) and 42.26% (95.23% vs. 52.97%), which were notable margins based on the CMC-1 matching accuracy. The large improvements are because the proposed AAD-CPGE can effectively address the problems of the potential ambiguous appearance of part-based objects by exploring the spatial relations from the graph node feature space and their topology spaces. Meanwhile, compared with the recent attention model-based approaches, such as PVEN, AAVER and SAVER, our method also obtained large mAP and CMC gains, and the performance for both mAP and CMC-1 were significantly improved from 69.78% (PVEN) to 90.34% (AAD-CDGE) and 64.65% (PVEN) to 95.23% (AAD-CDGE).

Moreover, although some recent work utilized GCNs and attention models for training their network, our method still outperformed several GCNs-based approaches, including the SG-GCN, ST-GCN and HPGN. Specifically, the AAD-CDGE brought about an improvement of 6.64%, 5.7% and 5.22%, which were evaluated by the mAP, as well as achieving the set of gains of 5.28%, 5.00% and 2.78%, respectively. Meanwhile, the superiority of our proposed AAD-CPGE framework can be seen from the mAP curves and CMC curves as shown in Figure 7 , showing that the proposed approach can achieve better performance than several comparative state-of-the-art Re-ID methods. The aforementioned experimental comparisons provide strong evidence of the effectiveness and superiority of our proposed AAD-CDGE framework for the UAV-based object Re-ID. The significant improvements achieved by our AAD-CDGE can be attributed to the following two reasons: On the one hand, the AAD-CDGE can effectively learn the completely intrinsic spatial and topological relationships from the multi-scale feature maps of the constructed graphs. On the other hand, introducing an attention mechanism can highlight the more useful spatial and topological information for feature fusion, which also resulted in more powerful hybrid feature representations for effectively distinguishing the appearance ambiguities among different kinds of objects as enhance the performance for Re-ID task.



(a)

(b)

Figure 7. The mAP and CMC curve comparison of our method and several state-of-the-art approaches on PRAI-1581 dataset. (**a**) mAP curve comparison on the PRAI-1581; (**b**) CMC curve comparison on the PRAI-1581.

5.4. Object Re-ID Comparisons under Different Configurations of AAD-CPGE

To explicitly understand the contributions of different embedding modules designed for Re-ID under our proposed AAD-CPGE framework, we first demonstrated their Re-ID performances using different embedded learning network configurations. For the sake of completely assessing each contribution of the AAD-CPGE for object Re-ID, we adjusted the configurations of our method to verify their performance on the UAV-ReID, VeRi-776 and PARI-1581 datasets, respectively. Meanwhile, the mAP, CMC-1 and CMC-5 were adopted to evaluate their Re-ID performances. Concretely, during the implementation of the embedded learning of the AAD-CPGE framework, we employed its configurations by utilizing GCNs-based spatial embedding module (Baseline), hierarchical part-based graph spatial embedding module (HP-SE), attention-based driven hierarchical graph spatial embedding module (HP-SE-Attn), hierarchical part-based spatial and topological embedding module (HD-SET) and hierarchical part-based graph spatial and topological embedding driven by attention module (HP-SET-Attn). The quantitative comparisons under the above configurations on the three datasets are summarized in Table 5.

Table 5. Comparisons results(%) of object Re-ID on UAV-ReID, ReVi-776 and PRAI-1581 dataset by using the metrics of mAP, CMC-1 and CMC-5, respectively.

	Configuration				UAV-ReID			ReVi-776			PARI-1581		
	Baseline	**HP-SE**	**HP-SET**	**Atten**	**MAP**	**CMC-1**	**CMC-5**	**MAP**	**CMC-1**	**CMC-5**	**MAP**	**CMC-1**	**CMC-5**
AAD	✓	-	-	-	80.57	85.92	89.36	80.16	87.95	90.45	82.64	86.39	91.45
-CPGE	✓	✓	-	-	82.76	89.43	92.19	81.54	90.39	92.77	83.35	90.41	93.26
	✓	✓	-	✓	83.45	92.13	94.42	82.12	92.94	94.18	85.17	91.52	93.89
	✓	✓	✓	-	83.81	94.24	96.38	82.76	94.59	96.09	86.74	92.34	94.75
	✓	✓	✓	✓	84.59	97.15	97.46	83.78	97.13	97.59	90.34	95.23	97.68

5.4.1. Re-ID Performance Using Different Configurations on the UAV-ReID

We first constructed a group of experimental verifications using different configurations of the AAD-CPGE on the UAV-ReID. As per the quantitative results in Table 5, compared with the baseline module that obtained 80.57% mAP, 85.92% CMC-1 and 89.36% CMC-5, respectively, adapting the HP-SE module yielded 82.76% mAP, 89.43% CMC-1 and 92.19%, which brought about improvements of 2.19%, 3.51% and 2.83%, respectively.

Then, by adding the attention module in the HP-SE module, the HP-SE-Attn can precede the HP-SE by 0.69%, 2.7% and 2.23%, which were based on the mAP and CMC-K (k = 1, 5) values. These indicated that the hierarchical graph constructed strategy and the attention could effectively help the AAD-CPGE framework improve the Re-ID performance. After that, we introduced the topology embedding module into implementing the proposed method and achieved slight gains for the above three configurations for object Re-ID, which also demonstrated the advantage of exploring topological information for improving the Re-ID performance. In the final, we extracted the specific and joint embeddings from hierarchical part-based graph node features, topological features and their combinations attention-based simultaneously using the HP-SET-Attn module and made use of the fused features for the task of object Re-ID. As can be seen in Table 4, the HP-SET-Attn achieved the best mAP and CMC-k (k = 1, 5) values by 84.59%, 97.15% and 97.46%, respectively. By observing these results, we can find that the configuration of HP-SET-Attn achieved the best performance for object Re-ID, compared with other configurations in our proposed AAD-CPGE. Moreover, as shown in Figure 8, we have shown part of the visualization Re-ID results achieved by the baseline and the HP-SET-Attn configuration of the proposed AAD-CPGE framework, we can see that the HP-SET-Attn had the best capability to distinguish the match results from similar false-positive images in gallery set compared with the baseline model.

5.4.2. Re-ID Performance Using Different Configurations on the VeRi-776

To further indicate the effectiveness of each module of the AAD-CPGE framework for object Re-ID, we also explored another group of comparisons using different configurations of the VeRi-776. As can be observed in Table 5, compared with the baseline module that obtained 80.16% mAP, 87.95% CMC-1 and 90.46% CMC-5, respectively, employing the HP-SE module yielded 81.54% mAP, 90.39% CMC-1 and 92.77%, which achieved an improvement of 1.38%, 2.44% and 2.31%, respectively. The set of visual results of object VeRi-776 shown in Figure 9.

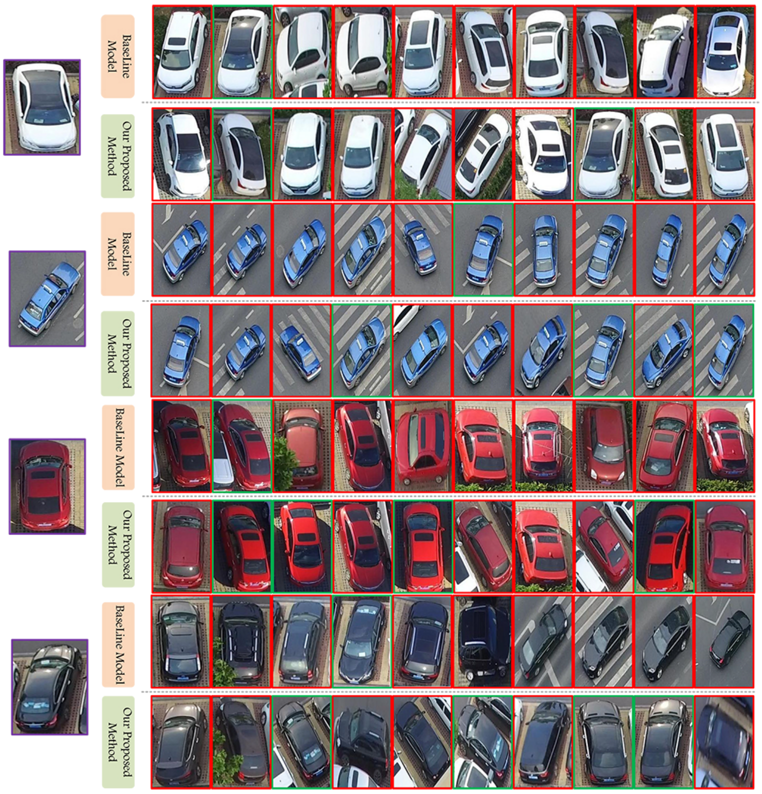


Figure 8. Parts of visual object Re-ID results on the UAV Re-ID using the baseline module and the HP-SET-Attn configuration of the proposed AAD-CPGE framework.

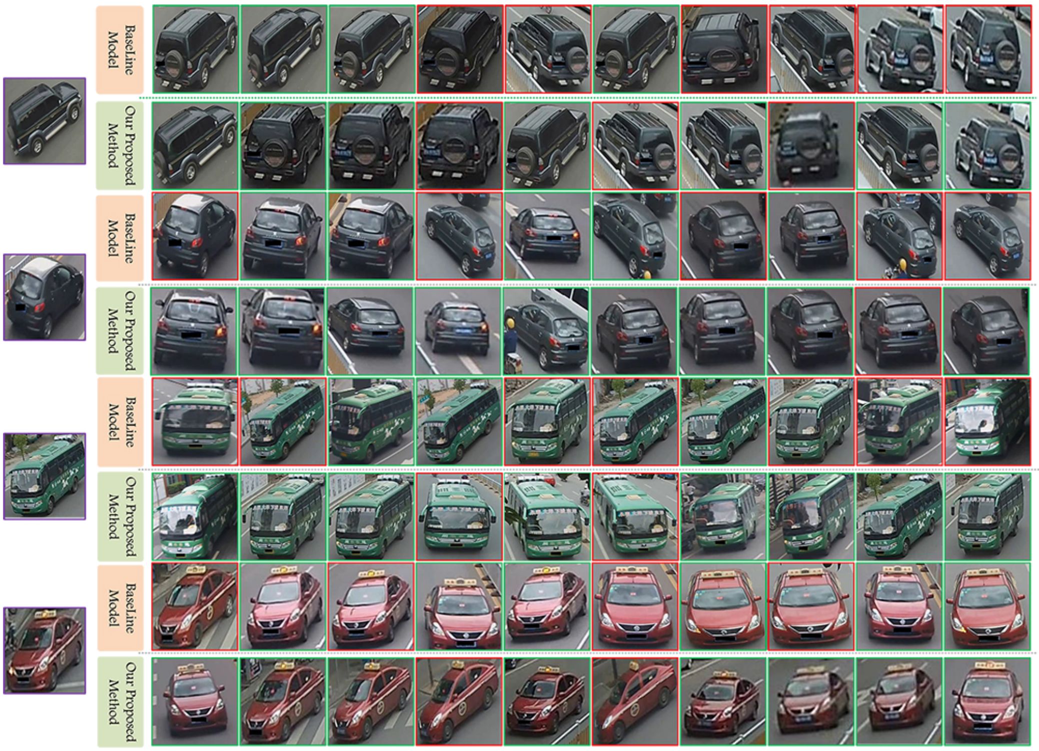


Figure 9. Parts of visual object Re-ID results on the VeRi-776 using the baseline module and the HP-SET-Attn configuration of the proposed AAD-CPGE framework.

5.4.3. Re-ID Performance Using Different Configurations on the PRAI-1581

At last, we carried out the last group of comparisons using different configurations of the PRAI-1581. As can be observed in Table 5, compared with the baseline module that obtained 82.64% mAP, 86.39% CMC-1 and 91.45% CMC-5, respectively, employing the HP-SE module yielded 83.35% mAP, 90.41% CMC-1 and 93.26%, which achieved an improvement of 0.71%, 4.02% and 1.81%, respectively.

Then, by adding the attention module in the HP-SE module, the HP-SE-Attn can precede the HP-SE by 1.82%, 1.11% and 0.63% terms of the mAP and CMC-K (k = 1, 5) values. These results demonstrated that the hierarchical graph constructed strategy and the attention could effectively help the AAD-CPGE framework to improve the Re-ID performance. After that, adding the topology embedding module into the whole learning process of the proposed framework also boosted the aforementioned three configurations for the Re-ID. Finally, we utilized the HP-SET-Attn module for the task of object Re-ID. As can be seen in Table 5, the HP-SET-Attn achieved the best mAP and CMC-k (k = 1, 5) values by 90.34%, 95.23% and 97.68%, respectively. By observing these results, we can find that the configuration of HP-SET-Attn achieved the best performance for object Re-ID, compared with other configurations in our proposed AAD-CPGE. Moreover, as shown in Figure 10, we also reported the visualization results achieved by the baseline and the HP-SET-Attn configuration of the proposed AAD-CPGE framework, which also indicate the superiority of HP-SET-Attn for object Re-ID.

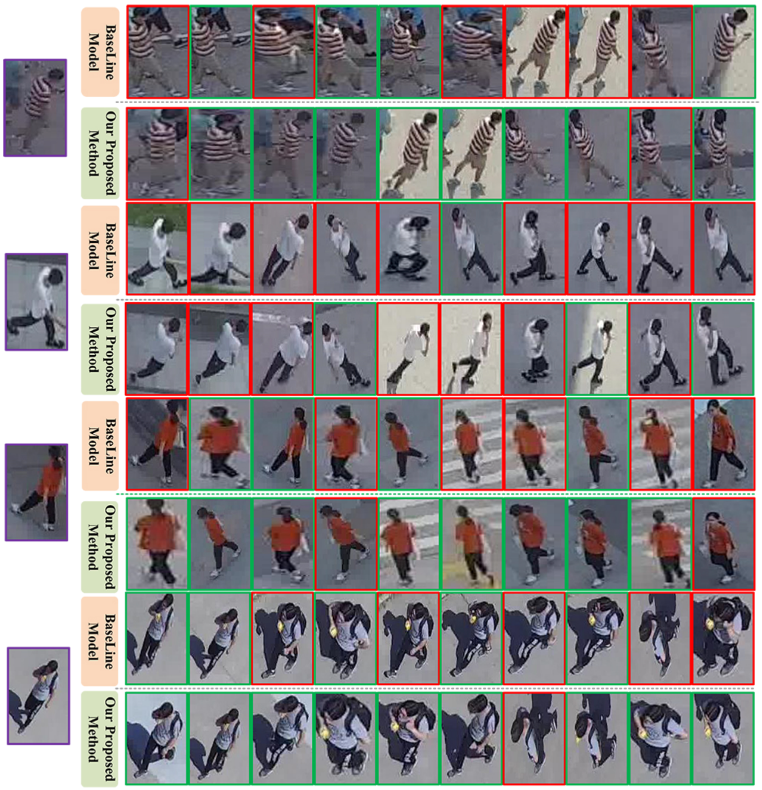


Figure 10. Parts of visual object Re-ID results on the PRAI-1581 using the baseline module and the HP-SET-Attn configuration of the proposed AAD-CPGE framework.

6. Conclusions

In this paper, we proposed an adaptively attention-driven cascade part-based graph embedding framework (AAD-CPGE) for the task of UAV-based object Re-ID platform. The AAD-CPEG Re-ID consisted of multiple part-based cascade graph convolutional networks (GCNs) branches, which are incorporated by the self attention-based driven multi-graphs fusion module. Different from the existing GCNs, which mainly consider the spatial cues for relation inferring, we explore the explicit spatial relations of the constructed hierarchical graphs and take into account their implicit topological characteristics to further improve the discriminative ability of the learning graph-structured feature representations. Meanwhile, we introduce the attention mechanism to adaptively fuse the graph features and their inferred topological features. In this way, the correlations discrimination between the graph features and the derived topological relations complement each other, which can effectively identify the variations and appearance ambiguities among inter-objects and intra-objects and then enhance the accuracy for object Re-ID. Our proposed AAD-CPGE starts from constructing two kinds of parts-based cascade node feature graphs and topological feature or (structure feature) graphs for exploring the global and local spatial-relations representations. After that, we designed the self attention-driven based module for adaptively fusing the node and topological features of the constructed graphs by imposing the consistency and disparity constraints in the embedding feature spaces. Then, we employed AAD-CPGE to learn the set of more discriminative features that derive from multi-channel feature maps. Finally, these learning hybrid graph-structured features with the most correlation discriminative capability are fed into a perceptron layer for object prediction and Re-ID. Experiments on three public datasets demonstrated that the proposed method outperformed state-of-the-art object Re-ID approaches.

Although our proposed AAD-CPGE framework has achieved desirable results in most situations for object Re-ID, there are also a few limitations, such as lacking the semantic relations discrimination between different kinds of multi-scale graphs, the high-dimension graph-structured feature always contains several invalid datapoints and reduces the accuracy and robustness for object Re-ID; therefore, in the future, we will impose some sparsity-based constraints into our proposed AAD-CPGE framework, which aims to improve the computational efficiency and ensure its performance for object Re-ID. In addition, we will explore more optimal GCNs, such as graph attention networks (GATs), to ensure more reliable parts for graph construction and learn more contextual and compact feature representations to further boost the performance of object Re-ID.

Author Contributions: Methdology, B.S.; writing—review and editing, R.Z. and H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Natural Science Foundation of China under Grant No.62172411 and 62172404.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Wang, Z.; Jiang, J.; Yu, Y.; Satoh, S. Incremental re-identification by cross-direction and cross-ranking adaption. *IEEE Trans. Multimed.* **2019**, *21*, 2376–2386. [CrossRef]
2. Qin, J.; Wang, B.; Wu, Y.; Lu, Q.; Zhu, H. Identifying Pine Wood Nematode Disease Using UAV Images and Deep Learning Algorithms. *Remote Sens.* **2021**, *13*, 162. [CrossRef]
3. Byun, S.; Shin I.K.; Moon, J.; Kang, J.; Choi, S.I. Road Traffic Monitoring from UAV Images Using Deep Learning Networks. *Remote Sens.* **2021**, *13*, 4027. [CrossRef]
4. Deng, C.; He, S.; Han, Y.; Zhao, B. Learning Dynamic Spatial-Temporal Regularization for UAV Object Tracking. *IEEE Signal Process. Lett.* **2021**, *28*, 1230–1234. [CrossRef]
5. Liu, Y.; Dai, H.; Wang, Q.; Shukla, M.K.; Imran, M. Unmanned aerial vehicle for internet of everything: Opportunities and challenges. *Comput. Commun.* **2020**, *155*, 66–83. [CrossRef]
6. Walambe, R.; Marathe, A.; Kotecha, K. Multiscale object detection from drone imagery using ensemble transfer learning. *Drones* **2021**, *5*, 66. [CrossRef]
7. Zhao, Y.; Shen, C.; Wang, H.; Chen, S. Structural analysis of attributes for vehicle re-identification and retrieval. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 723–734. [CrossRef]

8. Teng, S.; Zhang, S.; Huang, Q.; Sebe, N. Viewpoint and scale consistency reinforcement for UAV vehicle re-identification. *Int. J. Comput. Vis.* **2021**, *129*, 719–735. [CrossRef]
9. Ma, Y.; Li, Q.; Chu, L.; Zhou, Y.; Xu, C. Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision. *Remote Sens.* **2021**, *13*, 230. [CrossRef]
10. Jiang, S.; Jiang, W.; Huang, W.; Yang, L. UAV-based oblique photogrammetry for outdoor data acquisition and offsite visual inspection of transmission line. *Remote Sens.* **2017**, *9*, 278. [CrossRef]
11. Min, B.; Chala Urgessa, G.; Xing, M.; Han, L.; Chen, R. Toward More Robust and Real-Time Unmanned Aerial Vehicle Detection and Tracking via Cross-Scale Feature Aggregation Based on the Center Keypoint. *Remote Sens.* **2021**, *13*, 1416.
12. Fan, S.; Lin, M.; Jiang, J.; Kuo, Y. A Few-Shot Learning Method Using Feature Reparameterization and Dual-Distance Metric Learning for Object Re-Identification. *IEEE Access* **2021**, *9*, 133650–133662. [CrossRef]
13. Zhu, J.; Zeng, H.; Huang, J.; Liao, S; Lei, Z.; Cai, C.; Zheng, L. Vehicle re-identification using quadruple directional deep learning features. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 410–420. [CrossRef]
14. Zhu, J.; Zeng, H.; Huang, J.; Liao, S.; Lei, Z.; Cai, C.; Zheng, L. A dual-path model with adaptive attention for vehicle re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 20–26 October 2019; pp. 6132–6141.
15. Liu, X.; Zhang, X.; Wang, R.; Tian, Q. Group-group loss-based global-regional feature learning for vehicle re-identification. *IEEE Trans. Image Process.* **2019**, *29*, 2638–2652. [CrossRef]
16. Zhu, J.; Zeng, H.; Huang, J.; Liao, S.; Lei, Z.; Cai, C.; Zheng, L. Multiview image generation for vehicle reidentification. *Appl. Intell.* **2021**, *51*, 5665–5682.
17. Zhang, S.; Zhang, Q.; Yang, Y.; Xing, W.; Wang, P.; Jiao, B.; Zhang, Y. Person re-identification in aerial imagery. *IEEE Trans. Multimed.* **2020**, *23*, 281–291. [CrossRef]
18. Yu, H.; Wu, A.; Zheng, W. Unsupervised person re-identification by deep asymmetric metric embedding. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 956–973. [CrossRef]
19. Guo, J.; Pang, Z.; Yu, M.; Xie, P.; Liu, D. A Novel Pedestrian Reidentification Method Based on a Multiview Generative Adversarial Network. *IEEE Access* **2020**, *8*, 181943–181954. [CrossRef]
20. Mueller, M.; Smith, N.; Ghanem, B. A benchmark and simulator for uav tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Volume 8, pp. 445–461.
21. Robicquet, A.; Sadeghian, A.; Alahi, A.; Savarese, S. Learning social etiquette: Human trajectory prediction in crowded scenes. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 549–565.
22. Zhu, P.; Wen, L.; Bian, X.; Ling, H.; Hu, Q. Vision meets drones: A challenge. *arXiv* **2018**, arXiv:1804.07437.
23. Madeline, I.; Mygdalis, V.; Nikolaidis, N.; Montagnuolo, M.; Maurizio, N.; Negro, F.; Messina, A.; Pitas, I. High-level multiple-UAV cinematography tools for covering outdoor events. *IEEE Trans. Broadcast.* **2019**, *65*, 627–635.
24. Du, D.; Qi, Y.; Yu, H.; Yang, Y.; Duan, K.; Li, G.; Zhang, W.; Huang, Q.; Tian, Q. The unmanned aerial vehicle benchmark: Object detection and tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 370–386.
25. Kalra, I.; Singh, M.; Nagpal, S.; Singh, R.; Vatsa, M.; Li, G.; Sujit, P. Dronesurf: Benchmark dataset for drone-based face recognition. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–7.
26. Chao, D.; Liu, J.; Xu, F.; Liu, C. Ship detection from optical remote sensing images using multi-scale analysis and Fourier HOG descriptor. *Remote Sens.* **2019**, *13*, 1529.
27. Chao, D.; Liu, J.; Xu, F. Ship detection in optical remote sensing images based on saliency and a rotation-invariant descriptor. *Remote Sens.* **2018**, *10*, 400.
28. Wang, X.; Liu, M.; Raychaudhuri, D.; Paul, S.; Wang, Y.; Roy, C.; Amit, K. Learning Person Re-Identification Models From Videos With Weak Supervision. *IEEE Trans. Image Process.* **2021**, *30*, 3017–3028. [CrossRef] [PubMed]
29. Ye, M.; Li, J.; Ma, A.; Zheng, L.; Yuen, P. Dynamic graph co-matching for unsupervised video-based person re-identification. *IEEE Trans. Image Process.* **2019**, *29*, 2976–2990. [CrossRef]
30. Deng, W.; Zheng, L.; Ye, Q.; Kang, G.; Yang, Y.; Jiao, J. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 994–1003.
31. Wang, J.; Zhu, X.; Gong, S.; Li, W. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 2275–2284.
32. Yu, H.; Zheng, W.; Wu, A.; Guo, S.; Lai, J. Unsupervised person re-identification by soft multilabel learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 2148–2157.
33. Wu, F.; Yan, S.; Smith, J.; Zhang, B. vehicle re-identification in still images: Application of semi-supervised learning and re-ranking. *Signal Process. Image Commun.* **2019**, *76*, 261–271. [CrossRef]
34. Huang, Y.; Huang, Y.; Hu, H.; Chen, D.; Su, T. Deeply associative two-stage representations learning based on labels interval extension loss and group loss for person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 4526–4539. [CrossRef]

35. Wang, H.; Hou, J.; Chen, N. A survey of vehicle re-identification based on deep learning. *IEEE Access* **2019**, *7*, 172443–172469. [CrossRef]
36. Cui, C.; Sang, N.; Gao, C.; Zou, L. Vehicle re-identification by fusing multiple deep neural networks. In Proceedings of the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 28 November–1 December 2017; pp. 1–6.
37. Bai, Y.; Lou, Y.; Gao, F.; Wang, S.; Wu, Y.; Duan, L. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Trans. Multimed.* **2018**, *20*, 2385–2399. [CrossRef]
38. Zhang, Y.; Lou, D.; Zha, Z. Improving triplet-wise training of convolutional neural network for vehicle re-identification. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017; pp. 1386–1391.
39. Zhang, W.; He, X.; Yu, X.; Lu, W.; Zha, Z.; Tian, Q. A multi-scale spatial-temporal attention model for person re-identification in videos. *IEEE Trans. Image Process.* **2019**, *29*, 3365–3373. [CrossRef]
40. Yao, A.; Huang, M.; Qi, J.; Zhong, P. Attention Mask-Based Network with Simple Color Annotation for UAV Vehicle Re-Identification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
41. Guo, H.; Zhu, K.; Tang, M.; Wang, J. Two-level attention network with multi-grain ranking loss for vehicle re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 4328–4338. [CrossRef] [PubMed]
42. Zhang, M.; Cui, Z.; Neumann, M.; Chen, Y. An end-to-end deep learning architecture for graph classification. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018, New Orleans, LA, USA, 2–7 February 2018; Volume 32, pp. 1–8.
43. Chen, Z.; Wei, X.; Wang, P.; Guo, Y.; Wu, J. Multi-label image recognition with graph convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 5177–5186.
44. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 12026–12035.
45. Shen, Y.; Li, H.; Yi, S.; Chen, D.; Wang, X. Masked graph attention network for person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–20 June 2019; pp. 1496–1505.
46. Yang, J.; Zheng, W.; Yang, Q.; Chen, Y.; Tian, Q. Spatial-temporal graph convolutional network for video-based person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 3289–3299.
47. Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; Yang, Y. Learning to adapt invariance in memory for person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2723–2738. [CrossRef]
48. Shen, Y.; Li, H.; Yi, S.; Chen, D.; Wang, X. Person re-identification with deep similarity-guided graph neural network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 486–504.
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
50. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
51. Jiang, B.; Wang, X.; Zheng, A.; Tang, J.; Luo, B. Ph-GCN: Person retrieval with part-based hierarchical graph convolutional network. *IEEE Trans. Multimed.* 2021, *early access*. [CrossRef]
52. Wang, X.; Gupta, A. Videos as space-time region graphs. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 399–417.
53. Hammond, D.K.; Vandergheynst, P.; Gribonval, R. Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 129–150. [CrossRef]
54. Kopf, T.N.; Welling, X. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
55. Hoang, N.; Takanori, M. Revisiting graph neural networks: All we have is low-pass filters. *arXiv* **2019**, arXiv:1609.02907.
56. Wu, F.; Souza, A.; Zhang, T.; Fifty, C.; Yu, T.; Weinberger, K. Simplifying graph convolutional networks. In Proceedings of the International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6861–6871.
57. Wang, X.; Zhu, M.; Bo, D.; Cui, P.; Shi, C.; Pei, J. Am-GCN: Adaptive multi-channel graph convolutional networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual, 6–10 July 2020; pp. 1243–1253.
58. Liu, X.; Liu, W.; Mei, T.; Ma, H. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 869–884.
59. Liu, X.; Liu, W.; Ma, H.; Fu, H. Large-scale vehicle re-identification in urban surveillance videos. In Proceedings of the 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, USA, 11–15 July 2016; pp. 1–6.
60. Liu, X.; Liu, W.; Mei, T.; Ma, H. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Trans. Multimed.* **2017**, *20*, 645–658. [CrossRef]

61. Zheng, L.; Wang, S.; Zhou, W.; Tian, Q. Bayes merging of multiple vocabularies for scalable image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 20–23 June 2014; pp. 1955–1962.
62. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015; pp. 1116–1124.
63. He, B.; Li, J.; Zhao, Y.; Tian, Y. Part-regularized near-duplicate vehicle re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3997–4005.
64. Khorramshahi, P.; Peri, N.; Chen, J.; Chellappa, R. The devil is in the details: Self-supervised attention for vehicle re-identification. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 369–386.
65. Zhu, X.; Luo, Z.; Fu, P.; Ji, X. VOC-ReID: Vehicle re-identification based on vehicle-orientation-camera. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 602–603.
66. Meng, D.; Li, L.; Liu, X.; Li, Y.; Yang, S.; Zha, Z.; Gao, X.; Wang, S.; Huang, Q. Parsing-based view-aware embedding network for vehicle re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 7103–7112.
67. Shen, F.; Zhu, J.; Zhu, X.; Xie, Y.; Huang, J. Exploring spatial significance via hybrid pyramidal graph network for vehicle re-identification. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–12. [CrossRef]