electronics

MDPI

Article

Dual-Level Viewpoint-Learning for Cross-Domain Vehicle Re-Identification

Ruihua Zhou [1], Qi Wang [2,*], Lei Cao [3,*], Jianqiang Xu [2,*], Xiaogang Zhu [4], Xin Xiong [3], Huiqi Zhang [5] and Yuling Zhong [1]

[1] School of Software, Nanchang University, Nanchang 330047, China; zrh@email.ncu.edu.cn (R.Z.); yulingzhong@email.ncu.edu.cn (Y.Z.)
[2] School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China
[3] Department of Information, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, China; xiongxinxx@ncu.edu.cn
[4] School of Public Policy and Administration, Nanchang University, Nanchang 330031, China; ncuzxg@ncu.edu.cn
[5] College of Food Science &Technology, Nanchang University, Nanchang 330047, China; 407900210107@email.ncu.edu.cn
[*] Correspondence: wangqi@ncu.edu.cn (Q.W.); ndyfy01955@ncu.edu.cn (L.C.); xjq@ncu.edu.cn (J.X.)

Abstract: The definition of vehicle viewpoint annotations is ambiguous due to human subjective judgment, which makes the cross-domain vehicle re-identification methods unable to learn the viewpoint invariance features during source domain pre-training. This will further lead to cross-view misalignment in downstream target domain tasks. To solve the above challenges, this paper presents a dual-level viewpoint-learning framework that contains an angle invariance pre-training method and a meta-orientation adaptation learning strategy. The dual-level viewpoint-annotation proposal is first designed to concretely redefine the vehicle viewpoint from two aspects (i.e., angle-level and orientation-level). An angle invariance pre-training method is then proposed to preserve identity similarity and difference across the cross-view; this consists of a part-level pyramidal network and an angle bias metric loss. Under the supervision of angle bias metric loss, the part-level pyramidal network, as the backbone, learns the subtle differences of vehicles from different angle-level viewpoints. Finally, a meta-orientation adaptation learning strategy is designed to extend the generalization ability of the re-identification model to the unseen orientation-level viewpoints. Simultaneously, the proposed meta-learning strategy enforces meta-orientation training and meta-orientation testing according to the orientation-level viewpoints in the target domain. Extensive experiments on public vehicle re-identification datasets demonstrate that the proposed method combines the redefined dual-level viewpoint-information and significantly outperforms other state-of-the-art methods in alleviating viewpoint variations.

Keywords: vehicle re-identification; dual-level viewpoint-annotation proposal; angle invariance pre-training; meta-orientation adaptation learning

check for updates

Citation: Zhou, R.; Wang, Q.; Cao, L.; Xu, J.; Zhu, X.; Xiong, X.; Zhang, H.; Zhong, Y. Dual-Level Viewpoint-Learning for Cross-Domain Vehicle Re-Identification. *Electronics* **2024**, *13*, 1823. https://doi.org/10.3390/electronics13101823

Academic Editor: Jenhui Chen

Received: 10 April 2024
Revised: 5 May 2024
Accepted: 6 May 2024
Published: 8 May 2024

1. Introduction

Vehicle re-identification (Re-ID) aims to retrieve specific target vehicles across a multi-camera surveillance system [1–3]. Although some supervised learning works have achieved remarkable performance, the Re-ID system still suffers from the challenge of massive manual annotations. Therefore, cross-domain Re-ID has been developed to alleviate the bottleneck of labor costs, aiming to obtain initial parameters through source domain pre-training and then generalize to the unseen target domain in an unsupervised manner. Compared with person-based Re-ID [4–6], vehicle-based Re-ID will face unique challenges in term of viewpoint annotations.

The primary challenge is that the coarse viewpoint annotation cannot accurately parse the vehicle viewpoint, as shown in Figure 1a. Vehicle Re-ID can be regarded as learning discriminative features in a cross-view matching process [7,8]. To achieve view alignment, several approaches divide the vehicle viewpoint into three categories (i.e., front, rear, and side) and then conduct metric learning to learn robust representation. This subjective division strategy will create ambiguity relative to the alignment of similar vehicle viewpoints and make it impossible to parse the subtle differences between different vehicles. The above issue motivates us to discover how to accurately calculate the vehicle viewpoint and obtain viewpoint-invariant representation. Thus, this paper redefines a novel dual-level viewpoint-annotation proposal for calculation of angle-level label θ and orientation-level label O, as shown in Figure 1b,c. The dotted line in Figure 1c represents dividing the angle-level annotations into multiple intervals. In the following, Section 3.2 will introduce the calculation method in detail.

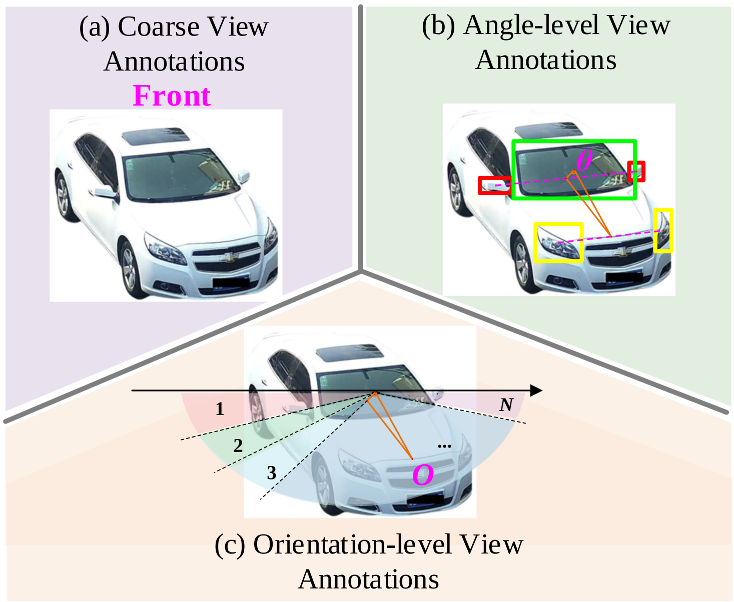
(a) Coarse View Annotations
Front
(b) Angle-level View Annotations
θ
(c) Orientation-level View Annotations
1
2
3
...
N
O

Figure 1. Illustration of different viewpoint annotations produced by (**a**) a coarse viewpoint, (**b**) an angle-level viewpoint, and (**c**) an orientation-level viewpoint.

Another challenge is that viewpoint variations make the pre-training model not well generalized to the unknown viewpoints in the target domain. Existing works focus on the use of metric learning [9,10], local representation learning [11–14], camera relevance [15,16], and the attention mechanism [17,18] to remedy the differences in cross-view variations. However, these works do not consider the differences between the viewpoint variations in the source domain and the target domain. As far as the labeled source domain is concerned, our motivation is to obtain a pre-training model sensitive to angle-level deviation. For the unlabeled target domain, the generalization ability as to unknown viewpoints is urgently needed for this amelioration. Therefore, it is also crucial to develop a meta-learning strategy based on cross-view data to adapt to unknown viewpoint variations.

Based on the aforementioned discussions, this paper proposes an angle invariance pre-training and meta-orientation adaptation learning method to fit viewpoint variations for cross-domain vehicle Re-ID. The major contributions of this paper are summarized in three aspects as follows:

1. To overcome the ambiguity of viewpoint information in human subjectivity, a dual-level viewpoint-annotation proposal is defined, one with a novel method for viewpoint measurement, and one which can alleviate the subjective error of manual annotations.
2. A method of angle invariance pre-training is designed to explore identity similarity and difference between vehicles across angle-level viewpoints. During the whole pre-training procedure, the part-level pyramidal network (PLPNet) with angle bias

metric loss is adopted to obtain the angle invariance feature, which provides more subtle angle-level discrimination for the downstream target domain.

3. A meta-orientation adaptation learning strategy is proposed for extending Re-ID model generalization in a meta-learning manner by utilizing orientation-level viewpoint annotations.

2. Related Work

Existing Re-ID methods face some challenges in the domain of the unseen. This section briefly reviews three main research directions related to our work: multi-view learning, cross-domain learning, and meta-learning.

2.1. Multi-View Learning for Vehicle Re-ID

Previous vehicle Re-ID methods [19–21] relied on manually labeled vehicle viewpoints to produce robust cross-view representations. Wang et al. [22] introduced a novel vehicle Re-ID framework, including an orientation-invariant feature-embedding module and a spatial–temporal regularization module The former module can better extract and align local region features, and the latter can make the retrieval results more refined. Zhou et al. [23] utilized two end-to-end deep architectures, named spatially concatenated convNet (SCCN) and CNN-LSTM bi-directional loop (CLBL), to solve the vehicle viewpoint uncertainty problem. Chu et al. [10] proposed a viewpoint-aware network (VANet), which learns two metrics for comparable viewpoints and disparate viewpoints in two feature spaces, respectively. Liu et al. [24] proposed a parsing guided cross-part reasoning network (PCRNet) which explores vehicle parsing to learn discriminative part-level features and models the correlation among vehicle parts to realize precise part alignments for vehicle Re-ID. Teng et al. [25] provided a novel dataset called unmanned aerial vehicles vehicle re-identification (UAV-VeID) and presented a viewpoint adversarial training technique and a multi-scale consensus loss to improve the robustness and discriminative capacity of learned deep features. However, the viewpoint annotation of the above methods is still a vague definition, and the viewpoint error of manual annotation will reduce the generalization ability of the Re-ID model relative to the unknown viewpoints of vehicles.

2.2. Cross-Domain Learning for Vehicle Re-ID

Cross-domain Re-ID [26,27] transfers knowledge to the unlabeled target domain for unsupervised training through pre-training in the labeled source domain to reduce the labor cost of the new domain. Yu et al. [28] proposed an unsupervised vehicle Re-ID approach that uses label-free datasets through self-supervised metric learning (SSML) based on a feature dictionary. Wang et al. [29] presented a multiple semantic knowledge learning method to multi-categorize from different views automatically using various cues. Also, the hard triple center loss was proposed to solve the unreliability of the pseudo-label of cluster. Bashir et al. [30] introduced an approach that mainly includes a progressive two-step cascading framework to transform the vehicle re-ID issue into an unsupervised learning paradigm. Zheng et al. [31] addressed an original viewpoint-aware clustering algorithm for vehicle Re-ID to solve the problem of the appearance difference from various viewpoints. Peng et al. [32] developed a domain adaptation structure composed of a vehicle transfer generative adversarial network (VTGAN) and an attention-based feature learning network (ATTNet). Different from the above works, this paper uses two aspects of viewpoint annotations to conduct source domain pre-training and target domain unsupervised learning, respectively.

2.3. Meta-Learning for Vehicle Re-ID

Meta-learning aims to learn unknown tasks through the knowledge of existing tasks. Some cross-domain methods based on meta-learning have gradually been proposed to strengthen the robustness of the Re-ID model. Yang et al. [33] introduced a dynamic and symmetric cross-entropy loss (DSCE) to mitigate the impact of noisy samples and a camera-aware meta-learning algorithm (MetaCam) to reduce the effect of camera movement. Zhao et al. [34] utilized the memory-based multi-source meta-learning (M3L) framework to improve the generalization ability of the unseen domains network based on training. Yang et al. [35] attempted to use the meta-attack algorithm to deceive Re-ID models on transparent domains through adversarial interference. Bai et al. [36] designed a domain generalization Re-ID network, named the dual-meta generalization network (DMG-Net), which makes full use of the multiple advantages of meta-learning. Most of the above methods apply existing annotations for meta-learning. However, our motivation is to combine the redefined viewpoint annotations with meta-learning to adapt to the viewpoint variations of the unknown domain.

3. The Proposed Methods

In this section, we first describe the pipeline of the proposed method in Section 3.1. Then, Section 3.2 designs a dual-level viewpoint-annotation proposal to define vehicle viewpoint at the angle level and the orientation level. Based on the above novel viewpoint definition, an angle invariance pre-training and a meta-orientation adaptation learning strategy are presented in Sections 3.3 and 3.4, respectively.

3.1. The Overall Framework

Based on a previous analysis, coarse manual annotations are difficult to accurately parse for a vehicle viewpoint. Additionally, a pre-training model with poor generalization cannot adapt to unknown viewpoint variations. This section proposes a novel dual-level viewpoint-learning framework to overcome the above challenges. Figure 2 shows the overall process of the proposed method, which consists of three steps. The three vehicle parts are first extracted by a pre-trained object-detection model, which then calculates the dual-level viewpoint-annotations. Then, PLPNet is designed, which adopts pyramidal convolution to mine subtle differences from different angle-level viewpoints. The source domain pre-training adopts PLPNet with an angle bias metric loss to learn angle-invariant features. Finally, a meta-orientation adaptation learning strategy is proposed that enforces a meta-orientation training step and a meta-orientation test step according to the orientation-level annotations to generalize the unknown viewpoints.

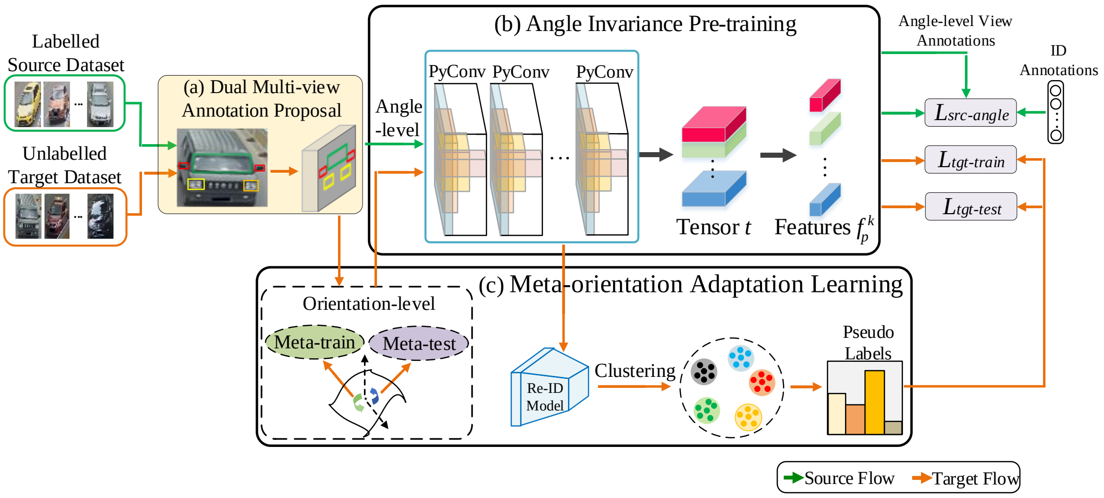

Figure 2. The overall of the dual-level viewpoint-learning framework.

3.2. Dual-Level Viewpoint-Annotation Proposal

In terms of the vehicle image captured by the two-dimensional space, annotating the accurate vehicle viewpoint is controversial and challenging. Existing viewpoint measurement methods are mainly used to manually annotate the vehicle viewpoint in the form of a subjective consciousness, which causes viewpoint deviation and suffers from excessive labor costs. To overcome the coarse definition of viewpoint, this paper defines a dual-level viewpoint-annotation proposal to accurately calculate vehicle viewpoint information at the angle level and orientation level, respectively.

To compute the vehicle angle-level viewpoint annotations from two-dimensional space, the crucial part-positioning of the training data is essential. Therefore, three pre-defined vehicle parts are selected for the key detection modules (i.e., rear-view mirror, window, and light). These parts meet the rigid requirements of the spatial geometry algorithm and to facilitate the calculation of the deflection angle of the vehicle, which explicitly reflects the various vehicle viewpoints. Subsequently, YOLOv4 [37] is adopted as our part region-localization framework to accurately localize the pre-defined vehicle parts in the source domain and target domain, respectively. Specifically, we randomly extracted 3000 images from the CompCars dataset for manual annotation of vehicle-part bounding boxes. Subsequently, the annotated data is used as an input for pre-training using YOLOv4 to obtain the vehicle-part detection model.

YOLO (You Only Look Once) v4 is a cutting-edge object detection system renowned for its efficiency and accuracy. YOLO v4 offers superior accuracy and speed in detecting objects within images or video streams. YOLO v4 excels in several crucial aspects. Firstly, it incorporates a feature pyramid network (FPN) to efficiently capture multi-scale features, ensuring robust detection of objects of various sizes within images or video frames. This guarantees that no object is missed, regardless of its scale or context. Moreover, YOLO v4 effectively extracts high-level features essential for precise object localization and classification. Additionally, the architecture utilizes optimization techniques such as Mish activation and spatial pyramid pooling (SPP) to boost model performance and efficiency. These optimizations not only enhance accuracy but also accelerate inference speed, enabling real-time object detection, even on devices with limited resources.

Given a vehicle image I, it is fed into YOLOv4 to obtain vehicle part detection results. Figure 3 shows the five detection bounding boxes, which contain one window box D_w, two rear-view mirror boxes D_m^n, and two light boxes D_l^n, where $n \in \{right, left\}$ indicates the direction of vehicle light or rear-view mirror.

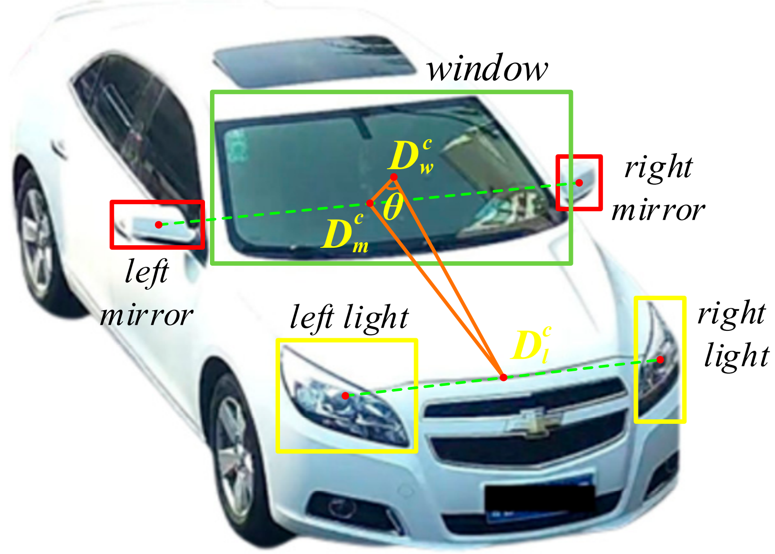

Figure 3. The calculation process of the angle-level viewpoint annotation. The dotted line indicates connecting the center points of two part bounding boxes.

We calculate the center positions D_w^c, D_m^c, and D_l^c of the three vehicle parts, where D_w^c is calculated from the center point of the coordinates of D_w. D_m^c and D_l^c are obtained from the center point of the line (seen in the green dotted line of Figure 3) connecting the center positions of the left and right boxes. Subsequently, we adopt D_w^c, D_m^c, and D_l^c to construct three edges and employ D_w^c as the vertex of the angle to calculate the angle-level viewpoint θ. The angle $\theta_I\left(D_w^c, D_m^c, D_l^c\right)$ of image I can be written as

$$\theta_I(D_w^c, D_m^c, D_l^c) = \begin{cases} \left[\frac{\left|(x_m^c - x_w^c)\times\left(x_l^c - x_w^c\right) + (y_m^c - y_w^c) + \left(y_l^c - y_w^c\right)\right|}{\sqrt{(x_m^c - x_w^c)^2 + (y_m^c - y_w^c)^2} \times \sqrt{\left(x_l^c - x_w^c\right)^2 + \left(y_l^c - y_w^c\right)^2}}\right]; if \ \ p = 5 \\ 181 \qquad\qquad\qquad ; else \ \ p \ \neq 5 \end{cases} \tag{1}$$

where p denotes the number of bounding boxes. Values x and y are abscissa and ordinate values, respectively. Most notably, it has been discovered that when the camera captures the vehicle images from the side, the vehicle-part detection fails to locate all of the ideal five-part bounding boxes. To cope with the problems of few bounding boxes or no bounding boxes, the angle will be uniformly set to 181 when p is not equal to 5. Thus, the whole $\theta_I\left(D_w^c, D_m^c, D_l^c\right)$ ranges from 0 to 181.

Then, to obtain the orientation-level viewpoint, the entire angle range is evenly partitioned into N parts. Each vehicle image obtains the corresponding orientation-level viewpoint annotation according to the range of its own angle $\mathbb{R}_N$. For the same reason, cases with fewer than five bounding boxes will be annotated with the maximum value of N plus 1 as the orientation-level annotations. The orientation-level viewpoint annotation O_I of image I is defined as

$$O_I = \begin{cases} N; \ if \ \ p = 5 \ and \ \theta_I\left(D_w^c, D_m^c, D_l^c\right) \in \mathbb{R}_N \\ \qquad \max(N) + 1 \, ; \ else \ \ p \ \neq 5 \end{cases} \tag{2}$$

$\mathbb{R}_N$ represents the range of angle range in the N-th orientation-level viewpoint. In subsequent experiments, N is set to be 18.

This section creates a concrete dual-level definition of the previously ambiguous viewpoint annotation, and the following sections will enforce the two-level viewpoint information in source domain pre-training and target domain meta-learning.

3.3. Angle Invariance Pre-Training for Source Domain

In the cross-domain vehicle Re-ID task, the sensitivity of the source domain pre-training model to the angle bias will directly affect the performance of the downstream task. Therefore, a pre-training network, PLPNet, and an angle bias metric loss are developed to discover identity, similarity, and difference under various angles in this section.

The architecture of PLPNet is shown in Figure 2. A pyramidal convolution (PyConv) is introduced into each convolution layer of PLPNet, inspired by [38]. PyConv constructs n-levels of different size kernels, which have different spatial resolutions and depths. It can parse the vehicle image at multiple scales and capture more subtle features under different receptive fields. The kernels with smaller receptive fields (3 × 3 and 5 × 5 size) can obtain partial feature information of the vehicle, while the kernels with larger size (7 × 7 and 9 × 9 sized) can learn more reliable details about the context information of vehicle. The output tensor t is then divided into K horizontal stripes and the vectors in each stripe are averaged into a single partial-level vector $f_p^k(k = 1, 2, \ldots, K)$. Each partial-level vector can provide fine-grained feature information for vehicle image description, which improves the Re-ID accuracy of the same vehicle from different angle-level viewpoints.

Different appearances under varying angle-level viewpoints lead to large differences in extracted features. The angle variations drive the identical vehicles with various angles to be assigned to mismatch. The angle bias metric loss is then proposed to distinguish the subtle differences of different vehicles from nearby angles and identify the same vehicles from different angles. The loss function $L_{src-angle}$ is given as

$$L_{src-angle}(I_1, I_2, z, \beta, \Delta\theta) = (1 - z)\cdot\frac{1}{2}\left(D_E\cdot e^{-\Delta\theta}\right)^2 + z\cdot\frac{1}{2}max(0, margin - D_E)^2 + z\cdot D_E\cdot min(1, \beta - \Delta\theta)^2 \tag{3}$$

where I_1 and I_2 denote a pair of input images, and z is a symbol to determine whether the image pair shares the same ID. When the image pair belongs to the same ID, z = 1, and otherwise z = 0. $\Delta\theta$ represents the difference in image pair angles. D_E denotes the Euclidean distance of feature vectors extracted from samples I_1 and I_2. Values *margin* and β are the thresholds of Euclidean distance and angle bias, respectively. It can be observed that the angle bias metric loss draws together the same ID image pairs with large angle bias and pushes away different ID image pairs with small angle bias.

During the source domain pre-training, the vehicle images to PLPNet go through PyConv, which captures partial-level fine-grained information with higher connectivity. Simultaneously, the pre-training model learns angle invariance features under the supervision of the angle bias metric loss and provides robust initial parameters for downstream tasks.

3.4. Meta-Orientation Adaptation Learning for Target Domain

Although the proposed pre-training model can distinguish subtle differences at different angle-level viewpoints, it is insensitive to unlearned viewpoints in the target domain. Due to this issue, the pre-training model will produce excessive orientation-level viewpoint noises in the clustering process which reduce the downstream clustering quality. To enforce the downstream target tasks to obtain the "learning to generalize unlearned orientation-level viewpoints" capability, a meta-orientation adaptation learning strategy is proposed to optimize iterations of the downstream model in the target domain. The whole meta-learning strategy is divided into three steps: meta-orientation training, meta-orientation testing, meta-update.

Meta-orientation training. Assume that the target domain contains N categories of orientation-level viewpoints. We partition the target domain data into a meta-orientation training set O_{train} and a meta-orientation testing set O_{test} according to the orientation-level viewpoint annotations. The split ratio of the orientation-level viewpoint of O_{train} and O_{test} is controlled by λ (Section 4.3 will discuss the impact of the split ratio λ on meta-learning). To simultaneously optimize the downstream Re-ID model for both better generalization and discrimination capabilities, we also introduce angle-level viewpoint annotations of the target domain to supervise the meta-learning procedure. The meta-orientation train loss $L_{tgt-train}$ on the mini-batch samples N_{batch} is formulated as:

$$L_{tgt-train}(O_{train}, \varpi_{train}) = \sum_{i=1}^{N_{batch}} \left(L_{tgt-angle} + L_{tgt-cross}\right) \tag{4}$$

where $L_{tgt-angle}$ means using the angle bias metric loss in the target domain. $L_{tgt-cross}$ indicates cross-entropy loss. ϖ_{train} is the temporary model parameter of the current epoch.

Meta-orientation test. The SGD optimizer updates the model parameter ϖ_{train} to a temporary parameter ϖ_{test}, and then the meta-orientation test step is combined to calculate the meta-orientation test loss $L_{tgt-test}$, which is formulated as

$$L_{tgt-test}(O_{test}, \varpi_{test}) = \sum_{i=1}^{N_{batch}} \left(L_{tgt-angle} + L_{tgt-cross}\right) \tag{5}$$

Meta-update. The update of the whole meta-learning procedure combines the supervision information of meta-orientation train loss and meta-orientation test loss. The final loss function L_{meta} of downstream target tasks is formulated as

$$L_{meta} = L_{tgt-train} + L_{tgt-test} \tag{6}$$

The proposed meta-learning strategy in this section adopts orientation-level annotations to improve the generalization ability to unknown viewpoints and, additionally, introduces angle-level annotations to learn subtle discrimination in each step.

3.5. Discussion

Limitations of the Proposed Method. Although the proposed dual-level viewpoint-learning framework can effectively alleviate subjective judgment errors, the accuracy of viewpoint calculation will still depend on the performance of prior object detection models. The limitations of the proposed method are summarized as follows:

1. The bounding box level of vehicle-part coordinates still belongs to the coarse-grained perspective calculation method as the basis for viewpoint calculation. There is a lack of utilization of pixel-level part detection for fine-grained viewpoint calculation.
2. The proposed dual-level viewpoint-framework focuses on calculating viewpoint in scenarios where vehicle parts are visible, but does not fully consider the method of calculating vehicle viewpoint in occluded scenes. Furthermore, there are also cases where several bounding boxes are not detected.

Overfitting Analysis of the Proposed Method. In order to overcome the overfitting problem in cross-domain Re-ID tasks, our method has designed the following two points during the training process:

1. In terms of data augmentation used to alleviate overfitting issues, we use random cropping, horizontal flipping, and erasing to expand the training set during the training process. This operation can encourage the Re-ID model to continuously learn more challenging generated data during each epoch process, thereby overcoming the overfitting issues.
2. In terms of the proposed meta-learning strategy for alleviating the overfitting issues, we randomly divide different orientation-level viewpoint annotations according to the split ratio in each epoch of meta-learning. That is to say, the partitioning of the meta-orientation training set and meta-orientation testing set during each epoch process is dynamically changed based on directional viewpoint annotations. This design can continuously improve the generalization ability of the Re-ID model to changes in viewpoint, thereby alleviating overfitting issues during the training process.

4. Experiments

We elaborate on the datasets and experimental settings, respectively. Subsequently, the proposed method is validated from Sections 4.3–4.6 through detailed analyses of the sufficient experimental results.

4.1. Dataset and Evaluation Protocols

VeRi-776 [39] is collected from 20 real-world surveillance cameras in an urban district; there are more than 50,000 images of 776 vehicles in total. The images have diverse labels containing identity annotations, vehicle attributes and spatio-temporal information. The dataset is divided into two subsets for training and testing. The training set includes 37,781 images of 576 vehicles and the test set includes 11,579 images of 200 vehicles.

VehicleID [40] is a dataset of vehicle images captured by real-world cameras during the daytime. Each subject in VehicleID has a massive number of images collected from the front and back, and some of the images are annotated with various aspects of model information to facilitate the Ve-ID. Its training set includes 110,178 images of 13,134 vehicles. Its test set is divided into three sections; they are as follows: Test800 is made up of 6532 probe images and 800 gallery images of 800 vehicles, Test1600 is comprised of 11,385 probe images and 1600 gallery images of 1600 vehicles, and Test2400 is composed of 17,638 probe images and 2400 gallery images of 2400 vehicles.

VERI-Wild [41] contains 416,314 images of 40,671 subjects. The vehicle images are captured by 174 high-definition cameras scattered randomly in the wild. Different from the training set with 277,797 images of 30,671 identities, the testing set is divided as follows: Test3000 with 41,816 images, Test5000 with 69,389 images, and Test10000 with 138,517 images, respectively.

Probe refers to the collection of vehicle images which need to be queried. Gallery is a candidate image set that contains all identity vehicles. The task of vehicle Re-ID is the process of matching vehicle images with the same identity as the probe in the gallery through a probe. That is to say, the difference between the two is that the probe is the target vehicle to be retrieved, and the gallery is used to provide the probe for matching. Due to limitations in computing power, we extracted 30,000 images from the training sets of VeRi-776 and VERI-Wild as the target domains for training, respectively.

The detailed statistics of the above-mentioned three vehicle Re-ID datasets are shown in Table 1.

Table 1. The Statistics of different vehicle Re-ID datasets.

Datasets	Image Size	Number of Cameras	Number of Images (Number of IDs)				
			Total Set	**Training Set**	**Test Set**		
VeRi-776	224 × 224	20	50,117 (776)	37,778 (576)	12,339 (200)		
VehicleID	224 × 224	-	221,763 (26,267)	110,178 (13,134)	Test800 6532 (800)	Test1600 11,385 (1600)	Test2400 17,638 (2400)
VeRi-Wild	224 × 224	176	416,314 (40,671)	277,794 (30,671)	Test3000 41,816 (3000)	Test5000 69,389 (5000)	Test10000 138,517 (10,000)

For the cross-domain vehicle Re-ID task, the Rank-n accuracy (i.e., n = 1 or 5), and the mean average precision (mAP) are utilized to evaluate overall performance for test images.

Calculation of Rank-*n*. The Rank-n is used to represent the hit probability of the vehicle Re-ID ranking result, which represents the probability that the probe image i finds the positive candidate sample within the top-n retrieval results,

$$\text{Rank-}n = \frac{\sum_{i=1}^{M} gt(i,n)}{M} \tag{7}$$

where M represents the total number from the probe set to be queried, and $gt(i,n)$ is a two-value logic function. When there are positive samples i in the top-n ranking results, the value of $gt(i,n)$ is equal to 1, otherwise it is 0.

Calculation of mAP. For each probe image, the average precision (AP) is computed as

$$\text{AP} = \frac{\sum_{j=1}^{N} p(j) \times gt(j)}{N} \tag{8}$$

where N is the total number of images in the gallery set. Values $p(j)$ and $gt(j)$ represent the precision at the j-th position in the ranking list and a two-value logic function, respectively. If a probe matches the j-th element, then $gt(j)$ = 1; otherwise, $gt(j)$ = 0.

Then, the average accuracy mAP of each probe image by the value of AP can be calculated as

$$\text{mAP} = \frac{\sum_{i=1}^{M} \text{AP}(i)}{M} \tag{9}$$

where M represents the total number from the probe set to be queried, and AP(i) represents the accuracy AP calculated for each probe image i.

Calculation of Macro-averaged F1 score. The F1 score is the harmonic mean of precision and recall, and it is commonly used to determine the accuracy of classification tasks. Precision refers to the proportion of correctly identified positive items to all identified positive items, while recall refers to the proportion of correctly identified positive items to all actual positive items. The F1 score can be calculated as

$$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \tag{10}$$

where TP (i.e., true positive) represents items correctly identified as positive, FP (i.e., false positive) represents items incorrectly identified as positive, and FN (i.e., false negative) represents items incorrectly identified as negative.

The Macro-F1 score used in this paper evaluates the overall classification performance of vehicle Re-ID by calculating the arithmetic mean of all label F1 scores in the gallery. The formula for Macro-F1 score is written as

$$\text{Macro-F1} = \frac{1}{L}\sum_{l=1}^{L}\frac{2TP_l}{2TP_l + FP_l + FN_l} \tag{11}$$

where TP_l, FP_l, and FN_l represent the number of true positives, false positives, and false negatives for class l in all samples in the gallery set, respectively. L represents the number of categories for all vehicles in the gallery set. The larger the value of Macro-F1, the better the classification performance of the vehicle.

4.2. Experiment Settings

The experimental running environment is the Ubuntu 18.04 LTS operating system. The proposed PLPNet is adopted as the backbone network during the entire UDA task. The learning rate was set to 0.00035, the batch size to 64, and the updating rate to 0.2. All training images were resized to 224 × 224. The Re-ID model is updated by the stochastic gradient descent (SGD) optimizer and the total epoch is equal to 50. Significantly, for the VehicleID, only coarse annotations of the front and rear views are available. To verify that the proposed angle-level viewpoint annotations can alleviate the ambiguity definition, VehicleID is unified as the source domain, and VeRi-776 and VERI-Wild are adopted as the target domains in the subsequent cross-domain Re-ID experiments.

4.3. Ablation Studies

To evaluate the effectiveness of the proposed method in cross-domain tasks, we conducted a sequence of detailed ablation analyses, as described in this section.

Effectiveness of each module. To verify the contribution of each individual module, Table 2 reports the performance of different modules in cross-domain vehicle Re-ID. Each module is explained as follows:

- "Direct Transfer" means adopting the traditional cross-domain Re-ID method, using ResNet-50 as the backbone.
- The term "w/o (O + A)" means adopting the traditional cross-domain Re-ID method, using PLPNet as the backbone.
- The term "w/o O" means not using the meta-orientation adaptation learning strategy and only using the angle invariance pre-training method for cross-domain Re-ID tasks.
- The term "w/o A" means not using the angle invariance pre-training method and only using the meta-orientation adaptation learning strategy for cross-domain Re-ID tasks.
- "Ours" means using all proposed modules.

As can be seen from the experimental results, the performance of the "Direct Transfer" method is the worst. The key reason is that the negative impacts of viewpoint variations on pre-training and downstream tasks are not considered. Compared with "Direct Transfer," the "w/o (O + A)" method greatly improves the accuracy of mAP and Rank. It is further confirmed that the fine-grained features extracted by PLPNet in the source domain pre-training can improve the accuracy of downstream tasks. It is noteworthy that "Our" method achieves better performance than "w/o O" and "w/o A". This result proves that our method combines the advantages of angle-level viewpoint invariance and orientation-level viewpoint generalization.

Table 2. Comparison of each individual module when tested on two datasets. "R1" and "R5" represent the accuracy rates of Rank-1 and Rank-5, respectively, with numerical units in percentages. The highest accuracy is marked in bold.

Different Modules	**VeRi-776**			**VERI-Wild**								
				Test3000			**Test5000**			**Test10000**		
	R1	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**
Direct Transfer	65.40	73.50	25.50	50.97	70.57	20.74	29.34	49.72	12.06	27.46	46.25	9.36
Ours w/o (O + A)	77.80	84.90	34.10	52.10	74.80	27.00	45.10	68.20	23.10	35.30	58.30	18.10
Ours w/o O	78.10	85.60	34.20	53.60	75.50	27.50	46.80	69.80	23.90	35.50	58.00	17.60
Ours w/o A	80.80	87.10	36.80	55.80	77.50	28.70	48.70	71.10	24.90	39.90	62.90	20.90
Ours	**83.10**	**89.00**	**37.80**	**59.90**	**80.70**	**31.40**	**51.90**	**74.90**	**27.30**	**41.80**	**65.80**	**21.70**

Effectiveness of different pre-training models. To demonstrate the superiority of PLPNet as a pre-training model, we further explore the three backbones of ResNet-50, PCB, and DenseNet for pre-training. Specifically, the ResNet-50 is a widely used traditional baseline for feature extraction tasks, one which achieved high accuracy on each large target classification dataset. PCB is a strong baseline for learning part-informed features, which can better capture fine-grained features and context information between vehicle-part regions. Compared with ResNet, DenseNet has a smaller number of parameters, and its bypass enhances feature reuse with better resistance to fit. However, the complementary parsing of coarse-grained features and fine-grained feature information plays an important role in vehicle re-identification, one which is lacking in these three backbones above. PLPNet adopts the advantages of pyramidal convolution, which can both accurately capture fine-grained features and explore the relationship between different levels of feature information. Thus, PLPNet improves the resolution accuracy of different angle-level viewpoints of the same vehicle. As shown in Table 3, PLP-Net performs the best compared with all the other pre-training models in both mAP and Rank-1 accuracy in the two datasets, which verifies that PLPNet is a competitive pre-training model in the cross-domain Re-ID task.

Table 3. Comparison of different pre-training models when tested on two datasets. "R1" and "R5" represent the accuracy rates of Rank-1 and Rank-5, respectively, with numerical units in percentages. The highest accuracy is marked in bold.

Different Pre-Training Models	**VeRi-776**			**VERI-Wild**								
				Test3000			**Test5000**			**Test10000**		
	R1	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**
ResNet-50	81.30	87.90	37.00	57.10	78.50	29.80	49.00	72.80	25.60	38.50	62.40	20.20
PCB	81.10	87.70	36.80	58.20	80.00	30.80	50.40	73.70	26.60	40.10	63.90	21.10
DenseNet	80.00	86.80	35.10	58.80	80.60	31.10	51.40	74.10	27.10	41.20	64.40	21.20
PLPNet (Ours)	**83.10**	**89.00**	**37.80**	**59.90**	**80.70**	**31.40**	**51.90**	**74.90**	**27.30**	**41.80**	**65.80**	**21.70**

The influence of different orientation-level view partitions *N*. The selection of different orientation-level view partitions will be analyzed first, as it will directly affect the precision of dual-level view definition at the orientation level. We evenly divide the orientation-level view labels into 9, 12, 18, and 36 parts in the entire angle-level range. Table 4 shows the influence of different selections on UDA Re-ID performance. In the experimental results, it can be seen that our method yields the best Rank-1, Rank-5, and mAP when the partition N is equal to 18. Therefore, the more orientation-level view labels are divided, the performance of Re-ID may be increasingly affected, and it should be maintained within an appropriate range.

Table 4. Influence analyses of different orientation-level view partitions when tested on two datasets. "R1" and "R5" represent the accuracy rates of Rank-1 and Rank-5, respectively, with numerical units in percentages. The highest accuracy is marked in bold.

Orientation-Level View Partition *N*	VeRi-776			VERI-Wild		
	R1	R5	mAP	R1	R5	mAP
9	82.20	87.90	36.90	57.10	78.50	29.70
12	81.00	85.90	35.60	56.00	78.60	29.70
18	**83.10**	**89.00**	**37.80**	**59.90**	**80.70**	**31.40**
36	82.20	88.20	36.80	59.00	79.80	31.00

Sensitivity analysis on meta-learning split ratio. To find the best split ratio of meta-orientation training step and meta-orientation testing step, we set a parameter λ between 0 and 1. Table 5 compares the Rank-1, Rank-5, and mAP of different λ in a meta-learning manner. Notably, the proposed method achieves 83.10% Rank-1 and 37.80% mAP accuracy when λ is set to 0.6 on VeRi-776. The interval accuracy of the value of parameter λ is close to 0.6. It can be observed that during the training process of using two datasets as target domains, the highest Re-ID performance is achieved when parameter λ is in a range close to 0.5 to 0.7, and when parameter λ is far from this range, it will make the Re-ID performance more sensitive. That is to say, the selection of parameter λ will change the distribution of vehicle viewpoint in the meta-learning process, thereby affecting the Re-ID performance. On the other hand, the closer the value of parameter λ is to the range of 0.5 to 0.7, the more evenly distributed the viewpoint information contained in the partitioned meta-orientation training and meta-orientation testing. This also fully verifies the robustness of our dual-level viewpoint-annotation proposal method, which can calculate the viewpoint information of each sample in datasets with different data distributions.

Table 5. Influence analyses of different split ratios λ when tested on two datasets. "R1" and "R5" represent the accuracy rates of Rank-1 and Rank-5, respectively, with numerical units in percentages. The highest accuracy is marked in bold.

Meta-Learning Split Ratio λ	VeRi-776			VERI-Wild		
	R1	R5	mAP	R1	R5	mAP
0.1	79.30	87.90	36.20	56.40	78.20	29.90
0.2	81.10	87.60	36.10	59.30	79.60	31.00
0.3	78.90	88.60	35.00	58.60	80.10	31.40
0.4	79.70	88.10	35.90	59.80	80.60	31.40
0.5	81.20	88.60	36.10	59.40	80.30	31.30
0.6	**83.10**	**89.00**	**37.80**	**59.90**	**80.70**	**31.40**
0.7	81.60	87.90	36.70	58.70	80.30	31.30
0.8	79.50	87.70	36.30	57.80	79.30	30.70
0.9	78.80	87.80	35.90	57.40	79.20	30.60

4.4. Qualitative Visualization Analysis

Comparison of vehicle-part detection under different lighting conditions. To verify the robustness of YOLOv4 under different lighting conditions, we visualize the detection results of vehicle parts under bright and dark lighting conditions, as shown in Figure 4. It can be observed that the YOLOv4 model can accurately detect predefined vehicle-part bounding boxes under different lighting conditions. For instance, Example B can accurately detect car windows, even if the surrounding background and reflective window are extremely similar in color, under bright lighting conditions.

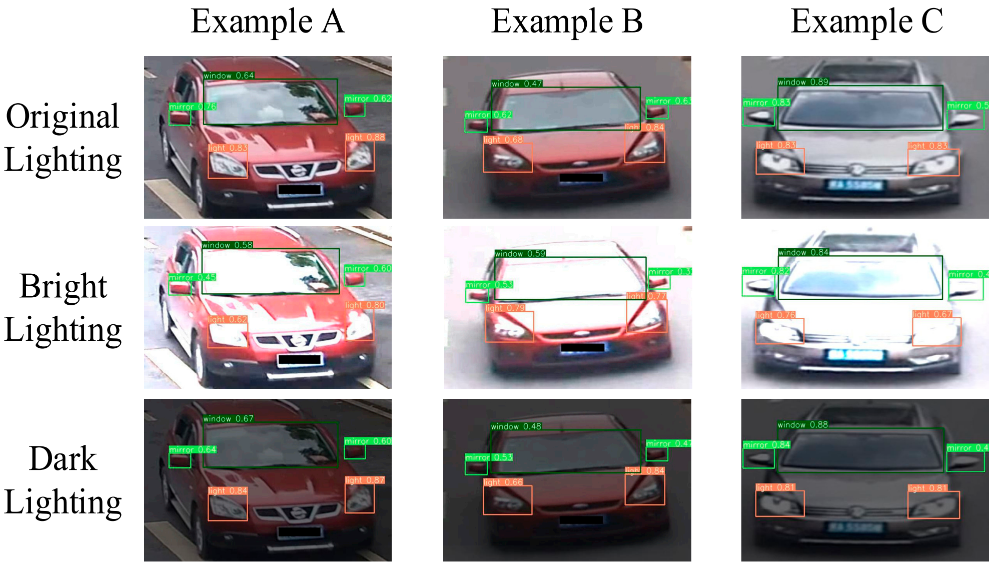

Figure 4. Detection results of YOLO v4 on vehicle parts under different lighting conditions.

The t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization. We visualize the t-SNE feature map of 10 classes of vehicles randomly sampled from the VeRi-776 dataset under different settings. It is worth noting that t-SNE is a non-linear dimensionality reduction technique used to map high-dimensional data to a two-dimensional spatial coordinate system. Therefore, the x-axis and y-axis represent the position information of the vehicle sample in the reduced two-dimensional space, respectively. As shown in Figure 5, different colors denote vehicle instance examples with different IDs, and different shapes represent different viewpoints. "Ours" means using all proposed modules. "Baseline" and "PLPNet" indicate that ResNet-50 and PLPNet are adopted as backbones for the traditional cross-domain Re-ID method. It can be seen that PLPNet and our model achieve more separable results among different classes than does the baseline model. Specifically, the features of the same category extracted by PLPNet are most compact in the feature space. What is more, the clusters of different categories are pulled away from each other. This indicates that our model is not only able to learn angle invariance features by capturing the fine-grained features of vehicles, but also has strong generalization ability to vehicle viewpoint variations.

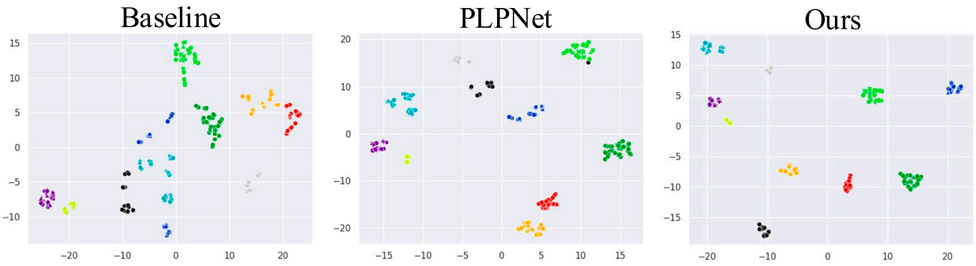

Figure 5. The t-SNE visualization for the outer features on VeRi-776, as extracted by the three methods.

Feature attention map visualizations. To further evaluate the validity of the features generated from PLPNet in the "Ours" method, we randomly selected three images from the VeRi-776 and compared them with the feature attention maps obtained from the baseline (i.e., ResNet-50). As shown in Figure 6, it can be observed that the feature maps obtained from the baseline only concentrate on some relatively fuzzy regions. The "Ours" method using PLPNet focuses on more discriminative and fine-grained details of the vehicle, like headlights and wheels, which are superior in capturing discriminative cues from the vehicle image.

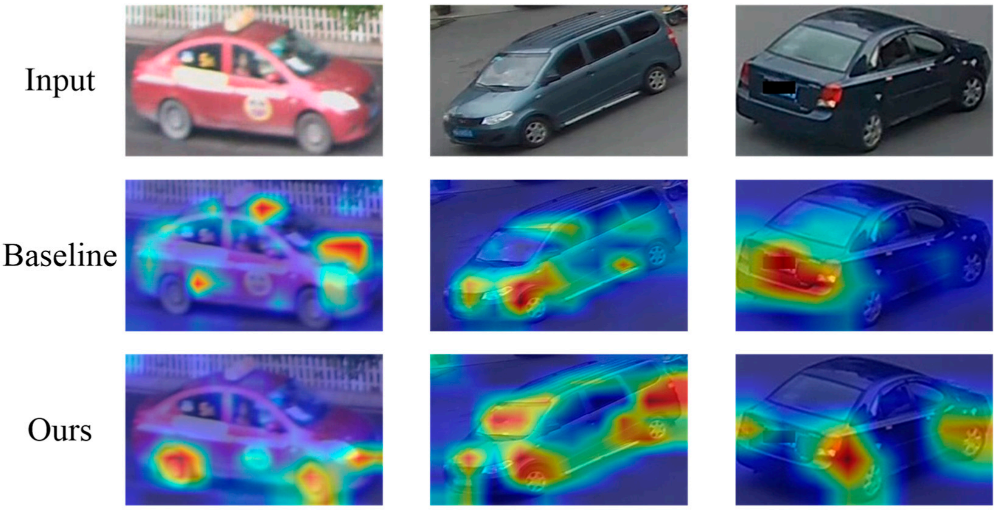

Figure 6. Visualized attention map produced by baseline and "Ours" using PLPNet on the same raw image. These feature maps highlight distinctive semantic features obtained from each pre-training model.

The ranking results visualization. For each group, we list the top-10 ranking results produced by the "Baseline", "PLPNet", and "Our" models in Figure 7, respectively. The green bounding boxes indicate correct results and red ones correspond to false results. Clearly, the "Baseline" model mainly identifies the different vehicles with the same viewpoint, as the correct matchings are evident in both examples. However, "Our" model can hit the highest number of correctly matched vehicles with different viewpoints in the earlier ranking, and also achieves better performance than PLPNet. The main reason is that there are frequent cases of the same vehicle with different viewpoints, which easily confuses the Re-ID model. For Example B, the proposed dual-level viewpoint-proposal method is used to calculate the viewpoint of the rear vehicle image. Subsequently, during the training process, the appearance of vehicles from different viewpoints is learned to obtain the ability to identify and track vehicles in different directions. Finally, the similarities of feature vectors extracted between Example B and the candidate image sets are used to determine whether it is the tracked target. Previous methods based on vehicle-part detection mainly tracked specific targets via the region information of vehicle-part bounding boxes. Unlike these methods, our method utilizes the position information of vehicle-part bounding boxes to assist the Re-ID model in learning viewpoint-sensitive appearance features for tracking targets, rather than relying on the appearance information inside the part bounding boxes. Therefore, the "Ours" model combines PLPNet with a meta-learning strategy to further improve the performance. In cases where different vehicles have similar viewpoints, PLPNet can notice more subtle differences in vehicles and discover additional clues, and the meta-learning strategy makes the Re-ID model more sensitive to unseen-viewpoint perception.

Sample pair distances under different image attribute settings. To verify the robustness of this article, we randomly selected 1500 pairs of positive and negative sample pairs from the VeRi-776 dataset for Euclidean distance calculation. Subsequently, we randomly increased or decreased brightness, contrast, and saturation by 20% for these sample pairs to simulate different image attribute variations. Figure 8 shows the similarity distributions of these sample pairs in the context of image brightness, contrast, and saturation. From Figure 8a, it can be observed that the distance between the positive sample pairs of the original image and the distance between the positive sample pairs after changing brightness, contrast, and saturation are very similar, and the Euclidean distance value is small. The distance distributions of the negative sample pairs after changing various image attributes in Figure 8b are also similar and have a large Euclidean distance value. This indicates that the proposed method can robustly bring positive sample pairs closer and push negative sample pairs farther in an image context that changes brightness, contrast, and saturation.

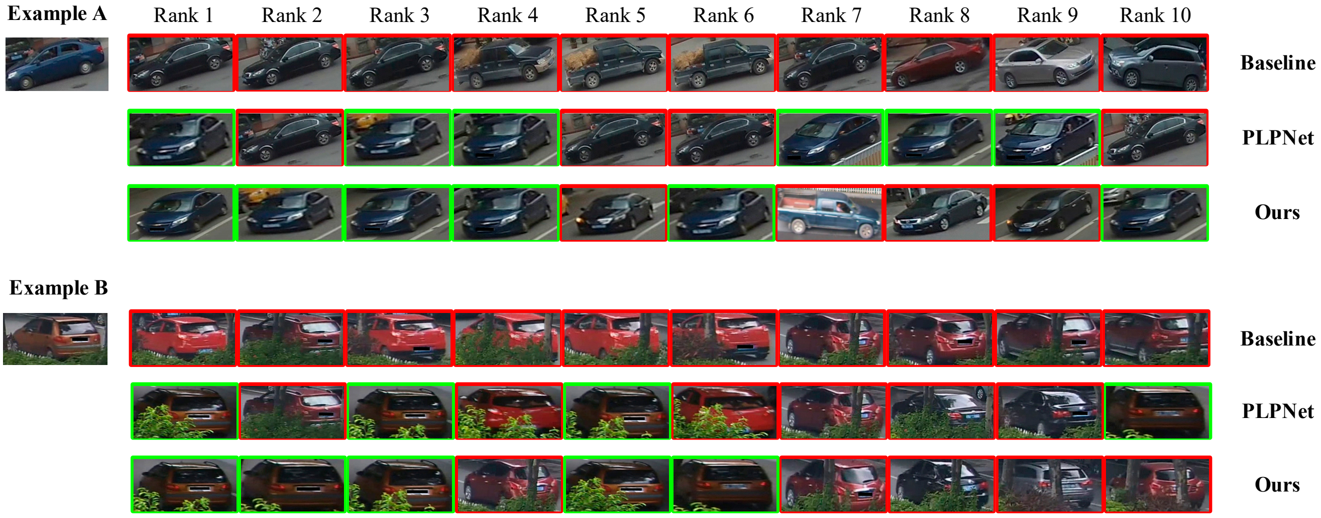

Figure 7. Two groups of ranked visualizations of challenging samples are selected from the VeRi-776 test data.

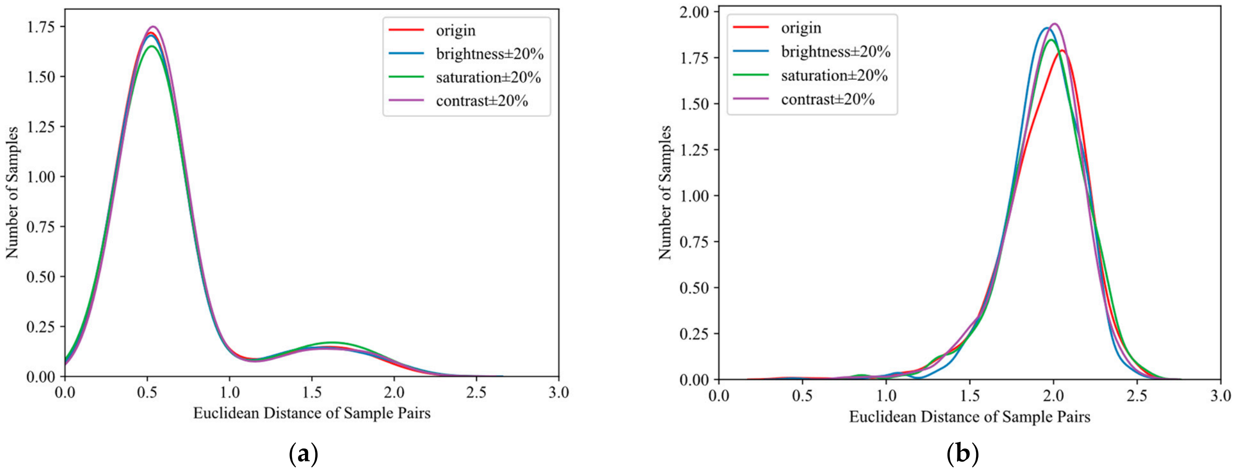

Figure 8. Visualization of sample pair similarity distribution. (**a**) Euclidean distance distributions between positive sample pairs under different image attribute settings; (**b**) Euclidean distance distributions between negative sample pairs under different image attribute settings; “origin” represents the original image; “brightness ± 20%”, “contrast ± 20% ”, and “saturation ± 20%” represent randomly increasing or decreasing the brightness, contrast, and saturation of sample pairs by 20%.

4.5. Comparison with State-of-the-Art Approaches

To demonstrate the superiority of the dual-level viewpoint-learning framework, we compare the proposed method with the state-of-the-art methods on two UDA Re-ID tasks, including VehicleID-to-VeRi-776 and VehicleID-to-VERI-Wild. The experimental results are summarized in Table 6. Our method achieves a performance of 37.80% on mAP and 83.10% on Rank-1 accuracy with VehicleID-to-VeRi-776. MMT combines hard and soft pseudo labels in a collaborative training manner to tackle the problem of noisy pseudo labels in the clustering phase. Although MMT has achieved remarkable clustering results, the noise introduced by changes in vehicle perspective still reduces the accuracy of clustering. Compared with the MMT, our method shows increases of 14.71% on mAP and 22.73% on Rank-1 accuracy with VehicleID-to-VeRi-776. Additionally, our method gains improvements of 8.91% mAP and 14.60% in Rank-1 accuracy over the second-best performance method AE when tested on VeRi-776. Although AE minimizes distances between similar identity instances to address the domain-shift problem, it does not consider the differences between different views of the same identity. For this purpose, our method overcomes the shortcomings of the AE by combining dual-level viewpoint-information during the training process.

Table 6. Comparison of the state-of-the-art cross-domain methods when tested on two datasets. "R1" and "R5" represent the accuracy rates of Rank-1 and Rank-5, respectively, with numerical units in percentages. The highest accuracy is marked in bold.

Methods	**VeRi-776**			**VERI-Wild**								
				Test3000			**Test5000**			**Test10000**		
	R1	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**
Direct Transfer	43.56	54.11	14.77	50.97	70.57	20.74	29.34	49.72	12.06	27.46	46.25	9.36
SPGAN [42]	50.72	62.63	15.83	28.77	48.27	10.63	24.60	42.96	8.97	23.40	40.78	7.13
UDA_TP [43]	51.85	64.54	18.12	46.30	59.20	12.30	16.20	29.00	5.21	17.65	29.99	4.53
ECN [44]	42.80	55.40	16.20	30.10	49.20	13.30	25.60	43.60	10.90	19.40	35.50	8.00
DomainMix [45]	53.30	64.60	15.40	33.20	53.80	14.10	28.40	48.10	12.20	21.10	38.90	9.00
SpCL [46]	58.90	68.00	24.40	48.80	72.80	25.10	42.00	66.10	21.50	32.70	55.70	16.60
MMT [47]	60.37	70.14	23.09	55.63	77.43	27.71	47.70	71.46	23.63	40.24	64.98	18.00
AE [48]	68.50	78.60	28.89	55.60	76.60	28.00	50.90	73.60	24.60	41.50	64.70	18.90
Ours	**83.10**	**89.00**	**37.80**	**59.90**	**80.70**	**31.40**	**51.90**	**74.90**	**27.30**	**41.80**	**65.80**	**21.70**

When using VehicleID as the source domain and VERI-Wild as the target domain, we also achieve the best performance, of 31.40%, on mAP and 59.90% on Rank-1 accuracy, which are 3.40% and 4.30% higher than AE. To further verify the effectiveness of our proposed method, we also compare it with state-of-the-art ones, i.e., DomainMix and SpCL, for viewpoint-aware problems in meta-learning for UDA vehicle Re-ID. It should be noted that our method has a significant performance gain compared with them on each of the three subsets of VERI-Wild. The proposed method outperforms the previous methods by a considerable margin, which proves the superiority of dual-level viewpoint-learning framework in alleviating the challenge of viewpoint variations.

To further validate the performance and efficiency of the proposed method in cross-domain tasks, Table 7 reports the comparison of the proposed method with existing methods in Macro-averaged F1 score and time complexity. Compared to the well-performing SpCL and MMT methods, the proposed method still maintains the best performance on the Macro-averaged F1 score. The main reason is that these methods did not take into account the challenge of vehicle viewpoint variations during the training process, while the proposed method utilizes redefined dual-level viewpoint-information to fully explore the visual appearance variations within and between domains.

Table 7. Comparison of computation time and classification performance with the state-of-the-art cross-domain methods when tested on VeRi-776. "Time" and "Macro-F1" represent the computational times and Macro-averaged F1 score, respectively. The units of values for "Time" and "Macro-F1" are "hours: minutes: seconds" and percentages, respectively. The highest accuracy is marked in bold.

Methods	**VeRi-776**	
	Time	**Macro-F1**
Direct Transfer	5 h: 30 m: 03 s	72.36
SPGAN [42]	7 h: 11 m: 22 s	85.43
UDA_TP [43]	8 h: 22 m: 15 s	81.71
ECN [44]	6 h: 47 m: 43 s	78.16
DomainMix [45]	9 h: 28 m: 48 s	85.22
SpCL [46]	14 h: 40 m: 39 s	84.96
MMT [47]	12 h: 01 m: 30 s	85.93
AE [48]	7 h: 33 m: 37 s	74.98
Ours	**10 h: 31 m: 00 s**	**91.02**

In terms of computational time, the proposed method may not have the fastest training speed due to the time cost of viewpoint calculation. Compared to other methods, the computational time of our method is still within an acceptable range, and it is reasonable to sacrifice a small portion of time to improve the Re-ID performance. In summary, the proposed method can achieve excellent cross-domain vehicle Re-ID performance within a reasonable calculation time interval.

4.6. Further Studies in Unsupervised Setting

The proposed dual-level viewpoint-learning framework is not only applicable to UDA Re-ID tasks but also to unsupervised Re-ID ones. PLPNet is also employed as the backbone of unsupervised Re-ID tasks. Simultaneously, angle bias metric loss and meta-orientation adaptation learning jointly conduct the entire unsupervised training process to fuse dual-level viewpoint-information. Table 8 shows the proposed method performance on the two Re-ID datasets against state-of-the-art unsupervised methods. "Ours (Uns)" means that the proposed method adopts an unsupervised manner to train the Re-ID model. The comparison with these unsupervised methods shows that the dual-level viewpoint-learning framework has also achieved competitive performance in unsupervised Re-ID tasks.

Table 8. Comparison of the state-of-the-art unsupervised methods when tested on two datasets. "R1" and "R5" represent the accuracy rates of Rank-1 and Rank-5, respectively, with numerical units in percentages. The highest accuracy is marked in bold.

Methods	**VeRi-776**			**VERI-Wild**								
				Test3000			**Test5000**			**Test10000**		
	R1	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**	**R1**	**R5**	**mAP**
MMT [47]	25.40	61.70	71.60	23.30	46.70	70.90	19.80	39.70	64.20	15.10	30.10	53.20
SPCL [46]	25.80	65.60	74.30	27.80	52.60	76.50	23.60	45.30	69.70	18.20	34.70	59.30
GSMLP-SMLC [49]	13.30	44.30	51.60	15.80	37.60	54.10	13.60	32.60	49.90	10.30	25.40	41.90
MetaCam [33]	25.60	67.10	76.00	28.20	53.90	76.90	24.10	46.00	70.20	18.80	35.90	59.70
SSML [28]	20.20	60.90	69.80	13.90	35.80	57.20	11.70	30.70	50.10	8.70	23.20	41.10
RLCC [50]	25.60	64.00	73.30	28.20	53.80	78.10	24.00	45.60	71.60	18.70	35.50	60.50
CACL [51]	23.70	55.70	69.00	28.00	53.30	77.40	24.00	45.70	70.80	18.50	35.30	60.40
Ours (Uns)	**28.80**	**72.20**	**79.10**	**30.60**	**56.30**	**79.00**	**25.30**	**47.80**	**72.10**	**19.80**	**37.20**	**61.60**

5. Conclusions

The rough definition of viewpoint annotations in existing methods will make it difficult for Re-ID models to learn viewpoint-invariant features, leading to cross-view misalignment. In this paper, our motivation is to redefine viewpoint annotations accurately to obtain a cross-domain Re-ID model that can adapt to various viewpoint variations. Thus, a dual-level viewpoint-learning framework is proposed to alleviate the viewpoint variations in cross-domain Re-ID tasks. For the source domain pre-training, the proposed PLPNet captures the subtle differences of vehicles from different angle-level viewpoints to gain angle invariance features. Based on the pre-training model, we develop a meta-orientation adaptation learning strategy to enhance generalization ability as to unknown viewpoints in the target domain.

Compared to existing cross-domain methods, the strength of the proposed method is that it can use two novel defined viewpoint annotations to learn vehicle viewpoint variations at the angle and direction levels, respectively. That is to say, the proposed method can not only achieve angle-level discrimination capacity in the source domain pre-training but also extend generalization for the unknown orientation-level viewpoint in the target domain. Extensive experimental results in VeRi-776 and VERI-Wild demonstrate the superiority of the proposed method. In future work, pixel-level information will be considered to determine identity relevance between diverse viewpoints.

Author Contributions: Conceptualization, R.Z.; methodology, R.Z. and Q.W.; software, R.Z. and L.C.; validation, R.Z., Y.Z., X.Z. and J.X.; formal analysis, R.Z., L.C. and X.X.; investigation, X.Z., J.X. and X.X.; resources, Q.W. and L.C.; data curation, J.X., X.X. and H.Z.; writing—original draft preparation, R.Z. and Q.W.; writing—review and editing, L.C., X.Z. and H.Z.; visualization, X.Z. and Y.Z.; supervision, Q.W. and J.X.; project administration, Q.W.; funding acquisition, R.Z., Q.W., J.X. and L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, under Grant No. 82360341; the Jiangxi Provincial Natural Science Foundation, under Grant Nos. 20224BAB212011, 20232BAB202051, and 20232BAB212008; the Jiangxi Science and Technology Program, under Grant No. 20232ABC03A32; the Science and Technology Program of National Health Commission, under Grant No. NIHA23JXH023; and the National College Students Innovation and Entrepreneurship Training Program, under Grant No. 202310403001S.

Data Availability Statement: Data will be made available on request.

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

References

1. Jiao, B.; Yang, L.; Gao, L.; Wang, P.; Zhang, S.; Zhang, Y. Vehicle Re-Identification in Aerial Images and Videos: Dataset and Approach. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 1586–1603. [CrossRef]
2. He, Q.; Lu, Z.; Wang, Z.; Hu, H. Graph-Based Progressive Fusion Network for Multi-Modality Vehicle Re-Identification. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 12431–12447. [CrossRef]
3. Wang, Q.; Min, W.; Han, Q.; Liu, Q.; Zha, C.; Zhao, H.; Wei, Z. Inter-Domain Adaptation Label for Data Augmentation in Vehicle Re-Identification. *IEEE Trans. Multimed.* **2022**, *24*, 1031–1041. [CrossRef]
4. Zhang, Z.; Lan, C.; Zeng, W.; Chen, Z.; Chang, S.-F. Beyond Triplet Loss: Meta Prototypical N-Tuple Loss for Person Re-Identification. *IEEE Trans. Multimed.* **2022**, *24*, 4158–4169. [CrossRef]
5. Wu, L.; Liu, D.; Zhang, W.; Chen, D.; Ge, Z.; Boussaid, F.; Bennamoun, M.; Shen, J. Pseudo-Pair Based Self-Similarity Learning for Unsupervised Person Re-Identification. *IEEE Trans. Image Process.* **2022**, *31*, 4803–4816. [CrossRef]
6. Liu, T.; Lin, Y.; Du, B. Unsupervised Person Re-Identification with Stochastic Training Strategy. *IEEE Trans. Image Process.* **2022**, *31*, 4240–4250. [CrossRef]
7. Liu, C.; Song, Y.; Chang, F.; Li, S.; Ke, R.; Wang, Y. Posture Calibration Based Cross-View & Hard-Sensitive Metric Learning for UAV-Based Vehicle Re-Identification. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19246–19257. [CrossRef]
8. Yang, L.; Liu, H.; Liu, L.; Zhou, J.; Zhang, L.; Wang, P.; Zhang, Y. Pluggable Weakly-Supervised Cross-View Learning for Accurate Vehicle Re-Identification. In Proceedings of the 2022 International Conference on Multimedia Retrieval, Newark, NJ, USA, 27 June 2022; pp. 81–89.
9. Jin, Y.; Li, C.; Li, Y.; Peng, P.; Giannopoulos, G.A. Model Latent Views with Multi-Center Metric Learning for Vehicle Re-Identification. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1919–1931. [CrossRef]
10. Chu, R.; Sun, Y.; Li, Y.; Liu, Z.; Zhang, C.; Wei, Y. Vehicle Re-Identification with Viewpoint-Aware Metric Learning. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8281–8290.
11. Tang, L.; Wang, Y.; Chau, L.-P. Weakly-Supervised Part-Attention and Mentored Networks for Vehicle Re-Identification. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 8887–8898. [CrossRef]
12. Lin, X.; Li, R.; Zheng, X.; Peng, P.; Wu, Y.; Huang, F.; Ji, R. Aggregating Global and Local Visual Representation for Vehicle Re-Identification. *IEEE Trans. Multimed.* **2021**, *23*, 3968–3977. [CrossRef]
13. Chen, X.; Yu, H.; Zhao, F.; Hu, Y.; Li, Z. Global–Local Discriminative Representation Learning Network for Viewpoint-Aware Vehicle Re-Identification in Intelligent Transportation. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–13. [CrossRef]
14. Qian, J.; Pan, M.; Tong, W.; Law, R.; Wu, E.Q. URRNet: A Unified Relational Reasoning Network for Vehicle Re-Identification. *IEEE Trans. Veh. Technol.* **2023**, *72*, 11156–11168. [CrossRef]
15. Zhu, X.; Luo, Z.; Fu, P.; Ji, X. VOC-ReID: Vehicle Re-Identification Based on Vehicle-Orientation-Camera. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 2566–2573.
16. Yu, J.; Kim, J.; Kim, M.; Oh, H. Camera-Tracklet-Aware Contrastive Learning for Unsupervised Vehicle Re-Identification. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23 May 2022; pp. 905–911.
17. Teng, S.; Zhang, S.; Huang, Q.; Sebe, N. Multi-View Spatial Attention Embedding for Vehicle Re-Identification. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 816–827. [CrossRef]
18. Li, M.; Huang, X.; Zhang, Z. Self-Supervised Geometric Features Discovery via Interpretable Attention for Vehicle Re-Identification and Beyond. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 194–204.

19. Li, M.; Liu, J.; Zheng, C.; Huang, X.; Zhang, Z. Exploiting Multi-View Part-Wise Correlation via an Efficient Transformer for Vehicle Re-Identification. *IEEE Trans. Multimed.* **2023**, *25*, 919–929. [CrossRef]
20. Zhang, C.; Wu, Y.; Shi, H.; Tu, Z. Multi-View Feature Complementary for Multi-Query Vehicle Re-Identification. In Proceedings of the 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 21 April 2023; IEEE Computer Society: Los Alamitos, CA, USA; pp. 1570–1573.
21. Meng, D.; Li, L.; Liu, X.; Gao, L.; Huang, Q. Viewpoint Alignment and Discriminative Parts Enhancement in 3D Space for Vehicle ReID. *IEEE Trans. Multimed.* **2023**, *25*, 2954–2965. [CrossRef]
22. Wang, Z.; Tang, L.; Liu, X.; Yao, Z.; Yi, S.; Shao, J.; Yan, J.; Wang, S.; Li, H.; Wang, X. Orientation Invariant Feature Embedding and Spatial Temporal Regularization for Vehicle Re-Identification. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 379–387.
23. Zhou, Y.; Liu, L.; Shao, L. Vehicle Re-Identification by Deep Hidden Multi-View Inference. *IEEE Trans. Image Process.* **2018**, *27*, 3275–3287. [CrossRef] [PubMed]
24. Liu, X.; Liu, W.; Zheng, J.; Yan, C.; Mei, T. Beyond the Parts: Learning Multi-View Cross-Part Correlation for Vehicle Re-Identification. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12 October 2020; pp. 907–915.
25. 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]
26. Zhou, Z.; Li, Y.; Li, J.; Yu, K.; Kou, G.; Wang, M.; Gupta, B.B. GAN-Siamese Network for Cross-Domain Vehicle Re-Identification in Intelligent Transport Systems. *IEEE Trans. Netw. Sci. Eng.* **2023**, *10*, 2779–2790. [CrossRef]
27. Wei, R.; Gu, J.; He, S.; Jiang, W. Transformer-Based Domain-Specific Representation for Unsupervised Domain Adaptive Vehicle Re-Identification. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 2935–2946. [CrossRef]
28. Yu, J.; Oh, H. Unsupervised Vehicle Re-Identification via Self-Supervised Metric Learning Using Feature Dictionary. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September 2021; pp. 3806–3813.
29. Wang, H.; Peng, J.; Jiang, G.; Fu, X. Learning Multiple Semantic Knowledge for Cross-Domain Unsupervised Vehicle Re-Identification. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Virtual, 5 July 2021; pp. 1–6.
30. Bashir, R.M.S.; Shahzad, M.; Fraz, M.M. Vr-Proud: Vehicle Re-Identification Using Progressive Unsupervised Deep Architecture. *Pattern Recognit.* **2019**, *90*, 52–65. [CrossRef]
31. Zheng, A.; Sun, X.; Li, C.; Tang, J. Aware Progressive Clustering for Unsupervised Vehicle Re-Identification. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11422–11435. [CrossRef]
32. Peng, J.; Wang, H.; Zhao, T.; Fu, X. Cross Domain Knowledge Transfer for Unsupervised Vehicle Re-Identification. In Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China, 8–12 July 2019; pp. 453–458.
33. Yang, F.; Zhong, Z.; Luo, Z.; Cai, Y.; Lin, Y.; Li, S.; Sebe, N. Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 4853–4862.
34. Zhao, Y.; Zhong, Z.; Yang, F.; Luo, Z.; Lin, Y.; Li, S.; Sebe, N. Learning to Generalize Unseen Domains via Memory-Based Multi-Source Meta-Learning for Person Re-Identification. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 6273–6282.
35. Yang, F.; Zhong, Z.; Liu, H.; Wang, Z.; Luo, Z.; Li, S.; Sebe, N.; Satoh, S. Learning to Attack Real-World Models for Person Re-Identification via Virtual-Guided Meta-Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 2–9 February 2021; Volume 35, pp. 3128–3135.
36. Bai, Y.; Jiao, J.; Ce, W.; Liu, J.; Lou, Y.; Feng, X.; Duan, L.-Y. Person30k: A Dual-Meta Generalization Network for Person Re-Identification. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 2123–2132.
37. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:200410934.
38. Duta, I.C.; Liu, L.; Zhu, F.; Shao, L. Pyramidal Convolution: Rethinking Convolutional Neural Networks for Visual Recognition. *arXiv* **2020**, arXiv:200611538.
39. 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; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; Volume 9906, pp. 869–884.
40. Liu, H.; Tian, Y.; Wang, Y.; Pang, L.; Huang, T. Deep Relative Distance Learning: Tell the Difference between Similar Vehicles. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2167–2175.
41. Lou, Y.; Bai, Y.; Liu, J.; Wang, S.; Duan, L. Veri-Wild: A Large Dataset and a New Method for Vehicle Re-Identification in the Wild. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3230–3238.

42. 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 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 994–1003.
43. Song, L.; Wang, C.; Zhang, L.; Du, B.; Zhang, Q.; Huang, C.; Wang, X. Unsupervised Domain Adaptive Reidentification: Theory and Practice. *Pattern Recognit.* **2020**, *102*, 107173. [CrossRef]
44. Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; Yang, Y. Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-Identification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 598–607.
45. Wang, W.; Liao, S.; Zhao, F.; Kang, C.; Shao, L. DomainMix: Learning Generalizable Person Re-Identification Without Human Annotations. In Proceedings of the 32nd British Machine Vision Conference 2021, BMVC 2021, Online, 22 November 2021; p. 355.
46. Ge, Y.; Zhu, F.; Chen, D.; Zhao, R.; Li, H. Self-Paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID. In Proceedings of the Advances in Neural Information Processing Systems; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H., Eds.; Curran Associates, Inc.: Scotland, UK, 2020; Volume 33, pp. 11309–11321.
47. Ge, Y.; Chen, D.; Li, H. Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-Identification. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26 April 2020.
48. Ding, Y.; Fan, H.; Xu, M.; Yang, Y. Adaptive Exploration for Unsupervised Person Re-Identification. *ACM Trans. Multimed. Comput. Commun. Appl.* **2020**, *16*, 1–19. [CrossRef]
49. Yu, J.; Oh, H. Unsupervised Person Re-Identification via Multi-Label Prediction and Classification Based on Graph-Structural Insight. *arXiv* **2021**, arXiv:210608798. [CrossRef]
50. Zhang, X.; Ge, Y.; Qiao, Y.; Li, H. Refining Pseudo Labels with Clustering Consensus over Generations for Unsupervised Object Re-Identification. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 3435–3444.
51. Li, M.; Li, C.-G.; Guo, J. Cluster-Guided Asymmetric Contrastive Learning for Unsupervised Person Re-Identification. *IEEE Trans. Image Process.* **2022**, *31*, 3606–3617. [CrossRef]

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