electronics

Article

Unsupervised Vehicle Re-Identification Based on Cross-Style Semi-Supervised Pre-Training and Feature Cross-Division

Guowei Zhan [1], Qi Wang [1,2,3], Weidong Min [1,2,3,*], Qing Han [1,2,3], Haoyu Zhao [1] and Zitai Wei [1]

[1] School of Mathematics and Computer Science, Nanchang University, Nanchang 330031, China
[2] Institute of Metaverse, Nanchang University, Nanchang 330031, China
[3] Jiangxi Key Laboratory of Smart City, Nanchang 330031, China
* Correspondence: minweidong@ncu.edu.cn

Abstract: Vehicle Re-Identification (Re-ID) based on Unsupervised Domain Adaptation (UDA) has shown promising performance. However, two main issues still exist: (1) existing methods that use Generative Adversarial Networks (GANs) for domain gap alleviation combine supervised learning with hard labels of the source domain, resulting in a mismatch between style transfer data and hard labels; (2) pseudo label assignment in the fine-tuning stage is solely determined by similarity measures of global features using clustering algorithms, leading to inevitable label noise in generated pseudo labels. To tackle these issues, this paper proposes an unsupervised vehicle re-identification framework based on cross-style semi-supervised pre-training and feature cross-division. The framework consists of two parts: cross-style semi-supervised pre-training (CSP) and feature cross-division (FCD) for model fine-tuning. The CSP module generates style transfer data containing source domain content and target domain style using a style transfer network, and then pre-trains the model in a semi-supervised manner using both source domain and style transfer data. A pseudo-label reassignment strategy is designed to generate soft labels assigned to the style transfer data. The FCD module obtains feature partitions through a novel interactive division to reduce the dependence of pseudo-labels on global features, and the final similarity measurement combines the results of partition features and global features. Experimental results on the VehicleID and VeRi-776 datasets show that the proposed method outperforms existing unsupervised vehicle re-identification methods. Compared with the last best method on each dataset, the method proposed in this paper improves the mAP by 0.63% and the Rank-1 by 0.73% on the three sub-datasets of VehicleID on average, and it improves mAP by 0.9% and Rank-1 by 1% on VeRi-776 dataset.

Keywords: vehicle re-identification; cross-style semi-supervised pre-training; pseudo label reassignment strategy; feature cross-division

Citation: Zhan, G.; Wang, Q.; Min, W.; Han, Q.; Zhao, H.; Wei, Z. Unsupervised Vehicle Re-Identification Based on Cross-Style Semi-Supervised Pre-Training and Feature Cross-Division. *Electronics* **2023**, *12*, 2931. https://doi.org/10.3390/electronics12132931

Academic Editors: George A. Papakostas and Byung Cheol Song

Received: 22 March 2023
Revised: 16 April 2023
Accepted: 7 May 2023
Published: 3 July 2023

1. Introduction

Currently, the development of unsupervised vehicle Re-Identification (Re-ID) algorithms [1–4] for large-scale system monitoring systems [5,6] is predominantly reliant on clustering of unlabeled target domain data and knowledge transfer from labeled source domain data. However, the absence of labeled information to guide the clustering process poses a significant challenge in enabling the model to learn discriminative features. To address this limitation, unsupervised domain adaptation methods have been proposed. These methods typically involve pre-training the model using source domain data and subsequently fine-tuning the pre-trained model on the target domain. Despite the advancements brought about by unsupervised domain adaptation, the performance of vehicle Re-ID still falls short of that achieved by supervised learning methods [7–9]. This performance gap can be attributed to the existing domain gap [10] between the source and target domains, as well as the reliance on global features for pseudo-label assignment during fine-tuning. Therefore, there is a need for further refinement of unsupervised vehicle Re-ID algorithms

to bridge the performance gap and enhance the accuracy of vehicle Re-ID in large-scale system monitoring scenarios.

Many existing re-identification methods aim to reduce the domain gap between different datasets by utilizing generative adversarial networks. During the fine-tuning stage, some methods [11–13] incorporate intermediate-domain data or style transfer data to minimize the difference in data distribution between the pre-training dataset and the fine-tuning dataset. However, the introduction of source domain data during the fine-tuning stage can interfere with the model's ability to learn target domain information. To overcome this challenge, some methods [14] attempt to introduce style transfer data during the pre-training stage to obtain a well-initialized pre-trained model. However, these methods often assign hard labels from the source domain data directly to the style transfer data, resulting in a mismatch between the style transfer data and the hard labels. In this paper, we propose the cross-style semi-supervised pre-training (CSP) module, which adopts a semi-supervised approach that leverages both labeled source domain data and unlabeled style fusion data to alleviate the domain gap and enhance the model's generalization ability. During the semi-supervised learning process, the CSP module generates soft labels for the style transfer data, allowing for better learning of the distribution of the style transfer data and effective mining of the target domain information embedded in the style transfer images.

In current vehicle Re-ID methods, pseudo labels are assigned based solely on the clustering results of global features [15,16]. However, this approach often introduces label noise due to the limitations of clustering algorithms and the tendency of vehicle images with similar IDs to be assigned to the same category, as shown in Figure 1. If not addressed, label noise can amplify during the training process and negatively impact the model's performance. To mitigate this issue, some methods have proposed regional partitioning approaches [17,18]. For instance, Wang et al. [19] proposed a method that extracts local features from different parts of the object and assigns class labels to these local features before performing classification. Similarly, Cho [20] proposed a model that leverages the complementary relationship between global features and local features obtained after region segmentation to reduce label noise. However, these methods are more applicable to person Re-ID than to vehicle Re-ID, as person have a natural structural advantage allowing for segmentation based on head, upper body, and lower body regions, each containing sufficient discriminative features. In contrast, vehicles lack such structural advantages and have fewer discriminative features compared to person, resulting in some local regions containing limited discriminative information. To address this challenge, we propose a feature cross-division (FCD) method for model fine-tuning to obtain feature partitions. The FCD method ensures that each feature partition contains sufficient discriminative information while preserving the correlation between feature partitions. Specifically, in this paper, we perform cross-partitioning on the extracted whole feature during the fine-tuning stage, obtaining multiple edge-overlapping feature partitions. We then measure the similarity of these partitioned features separately, and ultimately, the similarity measurement results will be referenced by all partitioned features as well as the global features.

In summary, this paper makes the following contributions:

(1) Addressing the problem of mismatch between the style transfer data and the hard labels of the source domain in the pre-training stage of existing methods for solving domain gap. CSP proposes a semi-supervised training approach where the source domain and the style transfer data with the target domain style are jointly used, improving the generalization ability of the pre-trained model. During training, soft labels are generated for the style transfer data, with a portion of the weight assigned to the clustering categories of the target domain. This allows the pre-trained model to fully learn the information of the target domain and obtain a better initialized pre-trained model.

(2) Addressing the problem of severe noise in pseudo-labels caused by excessive reliance on global features in existing vehicle Re-ID methods. FCD obtains feature partitions by cross-division of the overall features, retaining some edge-overlapping features. The significance of setting feature partitions in this paper is that different feature partitions will yield different similarity measurement results, and measuring different results can enhance the confidence of pseudo-labels. This approach helps to mitigate label noise and improve the accuracy of pseudo-labels in vehicle re-identification.

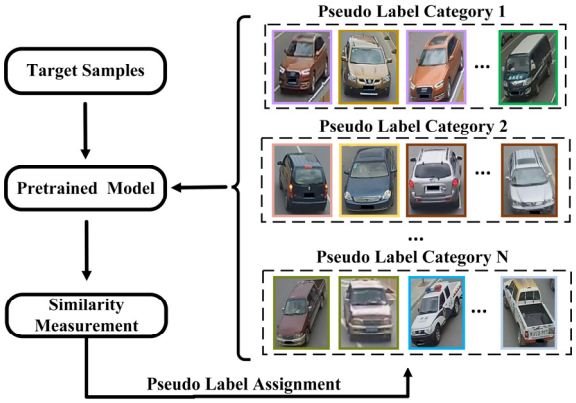

Figure 1. Pseudo label noise in fine-tuning process. Pseudo label noise refers to the problem of clustering errors in the process of pseudo-label assignment.

2. Related Works

2.1. Semi-Supervised Vehicle Re-ID

Semi-supervision aims to train models using both labeled and unlabeled data to alleviate the shortage of labeled data. The existing methods using semi-supervised learning have achieved good results. Wu et al. [21] developed a semi-supervised learning framework which relies on the Convolutional Neural Network (CNN) and re-ranking algorithm in the field of vehicle Re-ID. This network trains the Re-ID model in a semi-supervised way. Liu et al. [22] raised a novel semi-supervised Bayesian attribute learning (SBAL) algorithm for person re-ID, which enhances the feature-grabbing ability and improves the accuracy of label prediction. Qi et al. [23] provided an original progressive cross-camera soft-label learning framework, which aims to deal with the shortage of cross-camera label annotation information. Considering that semi-supervised learning can utilize both labeled and unlabeled data, this paper uses semi-supervised learning in the pre-training stage to mitigate domain gap and provide a better initialization model for the fine-tuning stage.

2.2. Unsupervised Vehicle Re-ID

The unsupervised vehicle Re-ID aims to minimize the shortage of labeled data and fully use the unlabeled data. For example, Bashir et al. [24] proposed a two-step cascaded framework based on unsupervised learning, which applies color information to obtain the reliable selection of clusters. Yu et al. [25] established a self-supervised metric learning (SSML) method condition on the feature dictionary. SSML designed a dictionary-based positive mining (DPML) to search the positive label of input images by calculating the feature's pairwise similarity, relative-rank consistency, and adjacent feature distribution similarity. Bashir et al. [26] construct a self-paced progressive unsupervised learning architecture which utilizes the clustering and filtering to group and filter the deep features extracted using the CNN to achieve a reliable selection of the input. Marin-Reyes et al. [27] proposed a metric learning model supervised on local constraints. They leveraged pairwise and triplet constraints for training a triplet network in a weakly-supervised fashion, where samples that share the same identity are close together, whilst different identities are maintained as distant. Nevertheless, obtaining more distinguishing features from unknown data and enhancing the discrimination ability of the model is still an urgent problem to be

solved. Therefore, UDA is used in this paper to achieve unsupervised vehicle Re-ID and make full use of labeled data.

2.3. Unsupervised Domain Adaptation

Unsupervised domain adaptation is a method to solve the domain gap of source domaint and target domain, which has achieved good performance in vehicle Re-ID. Wu et al. [28] introduced an original method for joint learning of 3D shapes and 2D images with a domain adaptation algorithm which establishes a connection among the feature spaces of 2D images and 3D shapes. Guo et al. [29] proposed a stable median center clustering (SMCC) for mining positive samples and reducing the impact of label noise. Xiao et al. [30] proposed a novel dynamic weighted learning method (DWL) for unsupervised domain adaptation. The weight of alignment learning and discriminability learning is dynamically adjusted to resolve excessive alignment or excessive pursuit of discriminability. Wang et al. [31] enhanced the accuracy of pseudo-labels during the unsupervised domain adaptation process through structured prediction and progressive selection. Wang et al. [32] designed a novel generative model norm-AE to generate synthetic features. The generated samples were applied to obtain a better classifier. Currently, UDA is widely used in both person and vehicle Re-ID. However, it is still necessary to optimize the training data and find ways to obtain higher confidence pseudo-labels for further improvement of the performance of UDA.

3. Materials and Methods

The definition of the UDA tasks for Re-ID: Generally, UDA tasks for Re-ID need two datasets: a source domain dataset $S = \{(x^1, y^1), \ldots, (x^{N_S}, y^{N_S})\}$, where N_S is the number of samples in the source domain, x^{N_S} is the N_S-*th* sample data, and y^{N_S} is its corresponding label, and a target domain dataset $T = \{t^1, \ldots, t^{N_T}\}$, where N_T is the number of samples in the target domain, and there is no label information for the target domain data. The traditional UDA task pre-trains the model through the dataset S, and then uses the obtained pre-trained model to extract the features of the target domain T. Finally, pseudo-labels are generated for the unlabeled data T through clustering, and the target domain data T carrying the pseudo labels are used to continue training the Re-ID model until convergence.

Description of the overall framework of this paper: The overall framework of this paper is shown in Figure 2. The implementation of the model includes the following three parts: (1) Firstly, the style transfer network is used to generate cross-domain style data, which are then used as unlabeled data for subsequent semi-supervised pre-training. (2) In the semi-supervised pre-training process, an initial network model is trained using the source domain data, and label prediction is performed on the generated cross-style data. At the same time, a pseudo-label reassignment strategy is designed, which replaces traditional hard labels with soft labels weighted by the target domain. (3) In the formal training process, the image similarity measurement is carried out by combining local features and partition features to more accurately predict the pseudo-labels of target samples.

3.1. Review of Generative Methods

Currently, the image transfer [33] is a popular method used for achieving unsupervised domain adaptation that can automatically perform image-to-image translation without paired samples.

In the process of image style transfer, it is expected that the image transfer network can achieve the following operations on datasets X and Y. Firstly, training a generator G that can convert the image style from the X domain to the Y domain, i.e., $G(x) = y', x \in X$, is achieved. Meanwhile, the image transfer network trains another generator F that can learn the opposite mapping process, so that images from dataset X can learn the style of dataset Y, i.e., $G(y) = x\prime, y \in Y$. Secondly, two discriminators are used to identify the quality of generated images. If the image y′ generated by generator G from x is different from the

image y, then the discriminator D_Y will give a low score; otherwise, the opposite occurs. Finally, to ensure that the image x still retains its own content and only learns the style of Y domain, the image transfer network designs a cycle-consistency loss. In other words, the generated image $y\prime$ will be input into generator F and compared with image x to ensure that the two images are as similar as possible.

Using the aforementioned features of the image transfer network, this paper performs a style transfer between the labeled source domain and unlabeled target domain to obtain datasets $G(x)$ carrying the target domain style and $F(y)$ carrying the source domain style.

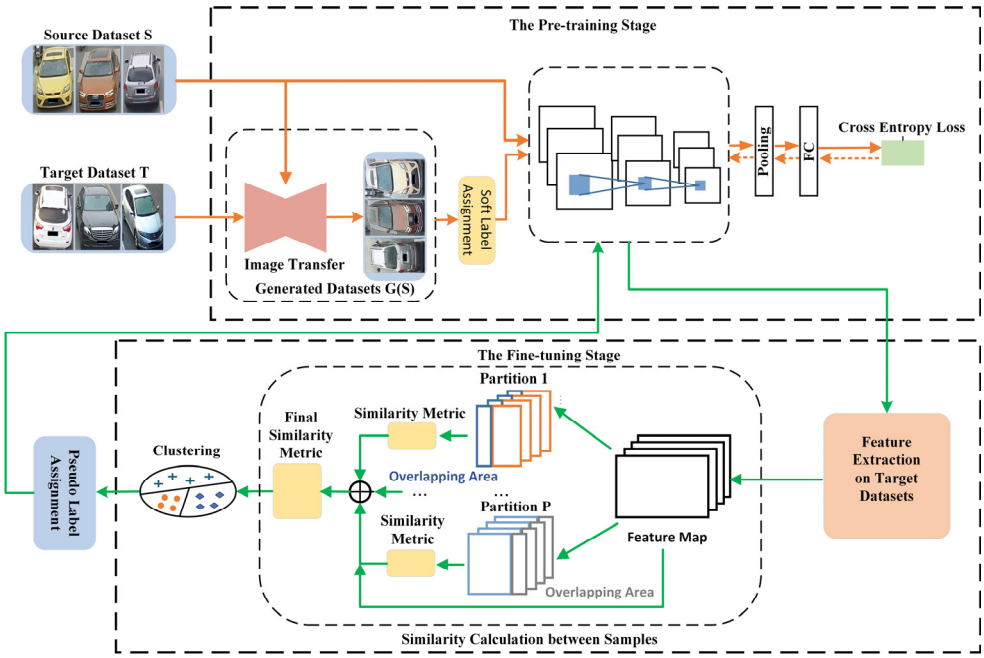

Figure 2. The overall framework of the method proposed in this paper. During the pre-training stage, style transfer network is used to generate data that carries the target domain style while preserving the source data content. The generated data and source domain data are then used together to train the pre-trained model, and soft labels are assigned to the generated data during this process. In the fine-tuning stage, the pre-trained model is used to extract features from the target domain, and the features are cross-partitioned. The results of the similarity measurement will ultimately combine the similarity measurement of global features and partitioned features.

3.2. Cross-Style Semi-Supervised Pre-Training

To address the domain gap problem in unsupervised domain adaptation, this section proposes a semi-supervised pre-training method based on cross-style learning.

There are various reasons for a domain gap between different datasets, such as differences in camera equipment and sample selection gap, which seriously affect the performance of model generalization. The key step in UDA is to transfer the pre-trained model on the source domain data to the unlabeled target domain, but, due to the existence of domain gap, the model's performance will be greatly reduced. To alleviate this problem, this paper attempts to introduce data with target domain style generated by the image transfer network, denoted as $G(x)$, during the pre-training stage. It is hoped that multi-style data pre-training can reduce the model's sensitivity to the target domain data and alleviate the impact of domain gap.

It should be noted that, in this paper, dataset $G(x)$ is generated from dataset X as unlabeled data. First, a model is trained on the existing labeled data X to obtain an initial model, which is then used to predict labels $L_{G(x)}$ for $G(x)$. Second, the labeled X data and the unlabeled data $G(x)$ with assigned pseudo-labels $L_{G(x)}$ are combined as a new training set to train the Re-ID model, and then the labels of dataset $G(x)$ are predicted

again. Meanwhile, more accurate pseudo labels can be obtained during this training process. Finally, the second process is reiterated until the model converges. Significantly, instead of using traditional hard labels as pseudo-labels for $G(x)$, this paper designs a soft label assignment method for the features of $G(x)$ data, namely, the dual-domain style fusion pseudo-label reassignment strategy, which will be introduced in Section 3.3.

The cross-entropy loss function is used for training the pre-training model, as shown in Equation (1).

$$L_{EC} = -\sum_{I=1}^{N_L} y_i log y_i' \tag{1}$$

where N_L represents the number of training samples, y_i represents the observed label, and y_i' represents the predicted label, in the process of predicting soft labels for unlabeled data.

3.3. Pseudo-Label Reassignment Strategy Based on Dual-Domain Style Fusion

To fully utilize the label information of labeled samples and preliminarily understand the geometric feature distribution of the target data, this section proposes a pseudo-label reassignment strategy that combines target style learning. The weight distribution strategy is shown in Figure 3.

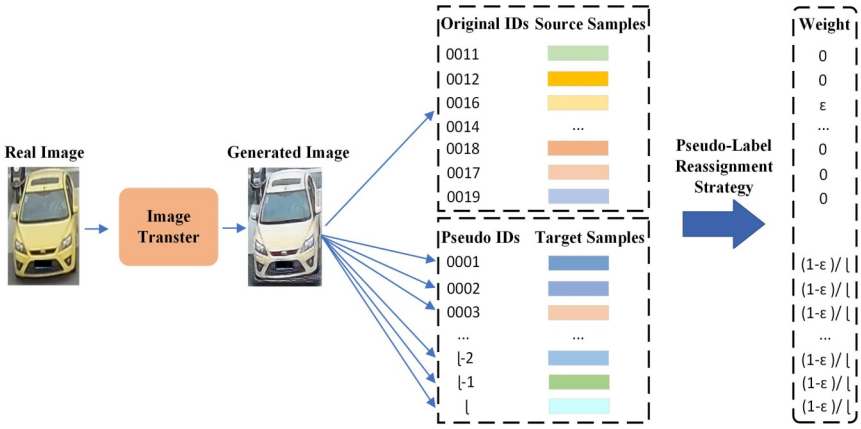

Figure 3. The weight distribution of the pseudo label reassignment strategy.

This method abandons traditional hard labels and instead assigns a certain weight to each target domain sample class, allowing the labels to exist in the form of soft labels. The new samples generated by the style transfer network, $G(x)$, participate in pre-training as an unlabeled dataset and obtain the soft labels assigned by the system. Inspired by the previous approach [4], this section encourages the network to assign a small portion of weight to each target domain class and treat each target image as a separate class for weight assignment. In other words, this section guides the model to initially learn the unlabeled samples in the target domain by assigning weights to each target domain class. In this process, the computational complexity will increase but the increase is valuable because the soft label reassignment can solve the problem that images generated by style transfer network do not match with the hard label. For each generated image, the soft label generation strategy is shown as Formula (2).

$$q(\delta.i,t) = \begin{cases} 0, \ i \in S \ and \ t \neq \ y_s^c \\ \delta, t = y_s^c \\ \frac{1-\delta}{l}, t \in T \end{cases} \tag{2}$$

In this formula, i represents the i-th unknown sample, t represents the class for which the weight is being calculated, S represents the source domain, and y_s^c represents the

original label. For any generated image $G(x)$, the corresponding loss function L_{PLR} for the pseudo-label reassignment process is shown as Formula (3):

$$L_{PLR} = -\frac{1}{l}\sum_{1}^{l}\log(p(t)) \tag{3}$$

In the initialization phase, l represents the number of images in the target domain and, as the iteration proceeds, l represents the number of clusters in the target domain. $p(t)$ is the predicted possibility of the training sample belonging to label t.

Based on the above analysis, the overall loss function during the pre-training stage is shown as Formula (4):

$$L_{Pre} = L_{EC} + L_{PLR} \tag{4}$$

where L_{EC} represents cross-entropy loss function, L_{PLR} is the loss function for the pseudo-label reassignment process mentioned in Equation (3), and L_{Pre} represents the loss function of the pre-training stage.

3.4. Fine-Tuning Based on Feature Cross-Devision

Because of the unknown total number of target classes and the lack of feature information mining, existing target domain fine-tuning methods based on complete features cannot effectively reduce pseudo-label noise. The main reason for the generation of pseudo-label noise during unsupervised fine-tuning is that the measurement of similarity between images is not accurate enough, leading to more errors in assigning pseudo-labels. To address this issue, this section proposes a fine-tuning method based on feature cross-division, which uses a more comprehensive similarity measure of the overall feature and partition features. When the similarity measurement of the overall features is incorrect and attempts to bring samples that do not belong to the same class closer together, the partitioned features may correct error. The partitioned features are more likely to discover more discriminative detailed features in the deep convolutional neural network due to the extracted partial features, thereby reducing the similarity scores between images of different categories and making images of different classes distinct from each other. The proposed method is shown in Figure 4.

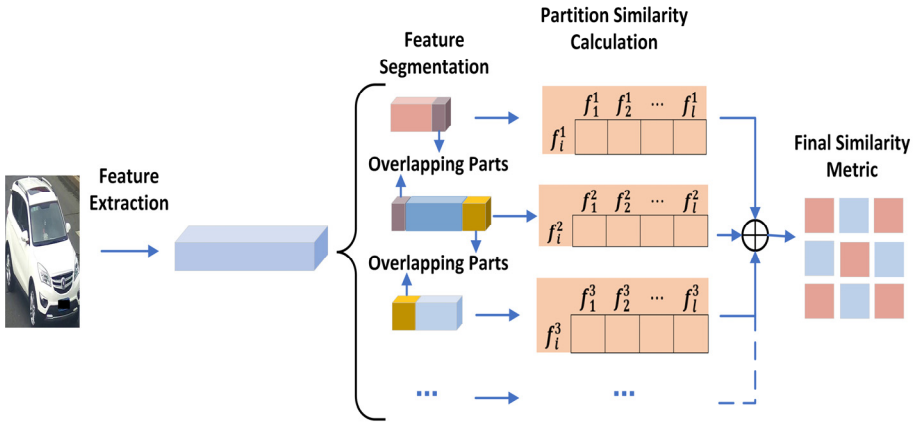

Figure 4. Similarity calculation of partition features.

First, the features $F_T = \left\{F_T^1, F_T^2, \cdots F_T^l\right\}$ extracted by the convolutional neural network model from the target dataset are divided into N regions (i.e., f_i^1 , $f_i^2 \cdots f_i^N$) for each image i. Compared with traditional feature partitioning methods, this method does not divide the feature map into independent partitions, but rather cross-partitions the feature map into N parts, meaning that there are overlapping regions between adjacent partitions. Through this approach, this section hopes to preserve the relationships between features to a greater extent and explore more similarities between features.

The similarity vectors corresponding to each partition are shown in Equation (5):

$$\begin{cases} S_i^1 = \left(s\left(f_i^1, f_1^1\right) \cdots s\left(f_i^1, f_{i-1}^1\right), s\left(f_i^1, f_{i+1}^1\right) \cdots s\left(f_i^1, f_l^1\right)\right); \\ S_i^2 = \left(s\left(f_i^2, f_1^2\right) \cdots s\left(f_i^2, f_{i-1}^2\right), s\left(f_i^2, f_{i+1}^2\right) \cdots s\left(f_i^2, f_l^2\right)\right); \\ \cdots \\ S_i^N = \left(s\left(f_i^N, f_1^N\right) \cdots s\left(f_i^N, f_{i-1}^N\right), s\left(f_i^N, f_{i+1}^N\right) \cdots s\left(f_i^N, f_l^N\right)\right) \end{cases} \tag{5}$$

where $s\left(f_i^d, f_j^d\right)(d = 1, 2 \cdots N)$ $(j = 1, 2, \cdots l)$ represents the similarity vector between the feature partition of the i-th image and that of other samples. S_i^N represents the similarity measurement result between the N-th feature partition of i-th image and the same partition of all other images.

The results of each partition are added together to obtain the final direct distance measurement. To ensure the accuracy of the pseudo labels, this paper also includes the similarity calculated from the global features in the final direct distance measurement. The paper uses the total similarity vector as shown in Equation (6) to measure similarity:

$$S_{total} = S_{ori} + S_1 + S_2 + \cdots + S_N \tag{6}$$

S_{ori} represents the similarity measure results based on global features and S_i represents the similarity measure results based on the i-th feature partition.

During the optimization stage, this paper uses the commonly used cross-entropy loss function to optimize the network. Therefore, the total loss function of the framework is shown in Equation (7). Driven by the total loss function, the model in this paper performs outstandingly in addressing the domain gap and reducing the noise carried by images in different datasets.

$$L = L_{Train} + L_{Pre} \tag{7}$$

where L_{Train} represents the loss function of the fine-tuning.

4. Results

4.1. Experimental Dataset and Evaluation Metrics

The effectiveness of the proposed method was validated using datasets from real-world large-scale surveillance scenes, namely VeRi-776 [34] and VehicleID [35]. A summary of VeRi-776 and VehicleID is shown in Table 1.

Table 1. Statistics of publicly available datasets. (Color refers to the number of appearances of colors of the vehicle and camera refers to the number of shooting cameras included in this dataset).

Datasets	Year	ID	Image	Color	Images per Vehicle	Camera
VeRi-776	2016	776	49,357	10	64.43	20
VehicleID	2017	26,328	221,567	6	8.44	2

VeRi-776 is a large-scale dataset for vehicle re-identification, consisting of 49,357 images from 776 different vehicle IDs. Among them, 37,778 images (from 576 vehicles) are used for the training stage, and 13,257 images (from 200 vehicles) are used for the testing stage, including 11,579 images belonging to the gallery set and 1679 images belonging to the query set. VehicleID is a vehicle Re-ID dataset consisting of images captured by two cameras, with approximately 110,178 images from around 13,134 different vehicle IDs in the training set and 111,585 images of 13,113 vehicles in the test set. The test set consists of three subsets with different scales which contain 800, 1600, and 2400 vehicles in this paper, respectively.

Given a target vehicle or non-vehicle image, the Re-ID model extracts the feature of this image and then matches some of the nearest features to perform metric ranking; the top features are identified as the same ID as the target image.

To accurately evaluate the performance of the model, this paper uses Rank-k and mean Average Precision (mAP) as evaluation metric for performance. Rank-k refers to the accuracy of the top k images sorted by similarity results that belong to the same ID as the query image. mAP refers to the average value of AP for each image category. AP refers to the ratio of the sum of accuracies in the target class in the test set to the number of images belonging to the target class, as shown in Formula (8).

$$AP = \sum_{k=1}^{n} \frac{p(k) \times f(k)}{n_{total}} \tag{8}$$

where n represents the number of vehicles involved in the calculation, k represents the order of vehicles retrieved, $p(k)$ represents the accuracy of the result at the k-th position, and $f(k)$ represents its value of 1 if the result at the k-th position is correct, otherwise 0.

The AP results are then averaged to obtain the mAP, as shown in Formula (9).

$$mAP = \sum_{t=1}^{T} \frac{AP(t)}{T} \tag{9}$$

where T denotes the number of query samples.

4.2. Experimental Settings

This paper has experimented with using the proposed model on a LINUX operating system and Pytorch 1.4.0 deep learning experimental environment. The hardware resources used in this experiment were Xeon(R) E5-2650 v4 processor at 2.20 GHz and NVIDIA-Tesla-P40 GPU.

ResNet-50 [36] was used as the baseline model in this paper, which was pre-trained on ImageNet [37]. All input images were uniformly processed into a size of 256 × 256. This paper used the stochastic gradient descent and set the initial learning rate and decay rate as 0.05 and 5×10^{-4}; the batch size was set to 16 and the epoch was set to 10. Since CycleGAN has already shown good performance and applicability in existing works, CycleGAN was used as the style transfer network to provide unlabeled data with the target domain style. For the training of CycleGAN, this paper is set according to its original experimental details [33].

4.3. Comparison with Existing Theoretical Methods

To confirm the performance of proposed model for vehicle Re-ID, this paper compares it with some recent research theories and shows the result in Tables 2 and 3. G(X) represents source domain data with a target domain style; X and Y represent the source and target domains, respectively. Source_G(X) represents pre-training the model using only G(X); Target_G(X) represents pre-training the model using source domain images and then using G(X) as the training set during fine-tuning. ST_G(X) represents using G(X) not only to train the pre-trained model but also for fine-tuning. The specific information is shown in Table 4. In all three processes, the FCD module proposed in this article is introduced during the fine-tuning stage. Part-based pseudo label refinement (PPLR) [20] proposes a model that leverages the complementary relationship between global features and locally extracted features derived from region segmentation, with the objective of mitigating label noise. Cluster Contrast for Unsupervised Person Re-Identification (CCUP) [15] proposes a new method called Cluster Contrast, which involves the storage of feature vectors and computation of contrastive loss at the cluster level. Additionally, this method introduces momentum update to strengthen the consistency of cluster-level features in the sequential space. Self-Paced Contrastive Learning framework (SPCL) [38] proposes a simple and effective Self-Paced Contrastive Learning framework, whose core idea is to use multiple forms of category prototypes to provide mixed supervision, in order to achieve sufficient mining of all training data.

Table 2. Comparison with the state-of-the-art methods (VeRi-776 is the source dataset and VehicleID is the target dataset; * stands for purely unsupervised method).

Methods	VeRi-776 to VehicleID (%)								
	Test Size = 800			Test Size = 1600			Test Size = 2400		
	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5
Direct transfer [36]	26.6	35.6	49.5	20.1	28.6	44.1	17.1	22.1	35.4
ST_G(X) [33]	36.1	43.7	58.4	28.8	36.0	52.5	25.9	30.3	42.2
Target_G(X) [33]	34.5	39.9	56.0	26.7	33.3	50.3	23.9	27.8	40.4
Source_G(X) [33]	37.3	47.8	61.0	31.5	39.2	54.3	27.2	31.6	43.8
PPLR * [20]	49.6	54.1	66.7	45.1	51.5	65.4	42.0	44.9	58.8
CCUP * [15]	49.1	53.7	65.8	44.6	51.1	64.3	41.4	44.3	58.0
SPCL [38]	51.2	53.7	67.0	45.8	51.9	65.1	42.2	45.2	60.0
Ours	**51.9**	**54.4**	**67.4**	**46.5**	**52.7**	**65.6**	**42.7**	**45.9**	**60.3**

Table 3. Comparison with the state-of-the-art methods (VehicleID is the source dataset and VeRi-776 is the target dataset; * stands for purely unsupervised method).

Methods	VehicleID to VeRi-776 (%)		
	mAP	Rank-1	Rank-5
Direct transfer [36]	22.3	58.9	66.9
ST_G(X) [33]	32.8	58.7	68.1
Target_G(X) [33]	31.3	59.0	71.0
Source_G(X) [33]	34.1	58.7	64.4
PPLR * [20]	44.7	73.3	83.1
CCUP * [15]	43.8	72.1	81.3
SPCL [38]	44.1	73.3	82.2
Ours	**45.6**	**74.3**	**83.7**

Table 4. Specific information related to the comparison method.

Methods	Source Dataset	Target Dataset	CSP	FCD
ST_G(X)	G(X)	G(X)	×	√
Target_G(X)	X	G(X)	×	√
Source_G(X)	G(X)	Y	×	√

It can be observed from Tables 2 and 3 that on the VeRi-776 to VehicleID task, our method improves an average of 0.63% in mAP, 0.73% in Rank-1, and 0.4% in Rank-5 compared with the best overall performing SPCL method. On the VehicleID to VeRi-776(%) task, our method respectively improves mAP, Rank-1, and Rank-5 by 0.9%, 1%, and 0.6% compared with PPLR. Compared with the remaining comparison experiments, the superiority of our method is more obvious.

4.4. Ablation Studies

4.4.1. Discussion on the Parameters of Pseudo-Label Reassignment Strategy

The pseudo label reassignment strategy updates the parameters in the pseudo-label generation strategy $q(\delta, i, t)$ adaptively. The steps are as follows: (1) Initializing the parameters in the soft label generation strategy, i.e., $\delta = \frac{1}{D_n}$, where D_n represents the number of datasets involved in style transfer. (2) When training the model, the current soft label generation strategy is used to generate soft labels on the training data. Then the generated soft labels are used as the labels of the data generated by style transfer for model pre-training. (3) At the end of each epoch, the value l in the pseudo generation strategy is adjusted according to the number of clusters.

4.4.2. Discussion on the Number of Feature Partitions N

In the fine-tuning stage of the proposed framework, to improve the algorithm's performance, this paper divided the features extracted from each image into several cross partitions and used N to represent the number of partitions. The value of N plays a crucial role in the calculation of the final sample similarity. Therefore, this section shows the impact of N on the overall theoretical framework in Tables 5 and 6. From these two tables, it can be observed that when $N = 3$, the Re-ID model achieved the best performance on both datasets. In other words, it can be concluded that this paper improves the accuracy of unsupervised data classification by using the method of combining global features and partition features to replace the sole use of global features for image similarity measurement. This is because when two images are highly similar, measuring their similarities through global features may assign them high similarity scores, leading to pseudo-labeling noise; however, when partition features are introduced, they can focus on more detailed features and obtain different similarity scores from global features, allowing the model to assign pseudo-labels based on different measurement criteria, thereby increasing the confidence of pseudo-labels. In the subsequent experimental process, the model will uniformly set N = 3 to obtain better performance. It is worth noting that the reason for choosing cross-division and retaining the common areas between partitions in this paper is to preserve the similarity between partitions and obtain more convincing results.

Table 5. Effects of different N on model performance (VeRi-776 is the source dataset and VehicleID is the target dataset).

Parameter N	VeRi-776 to VehicleID (%)								
	Test Size = 800			Test Size = 1600			Test Size = 2400		
	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5
N = 1	49.8	53.6	66.1	44.8	51.4	63.7	41.1	44.8	58.4
N = 2	50.9	54.3	66.8	45.1	51.9	64.6	41.3	45.3	59.1
N = 3	51.9	54.4	67.4	46.5	52.7	65.6	42.7	45.9	60.3

Table 6. Effects of different N on proposed model performance (VehicleID is the source dataset and VeRi-776).

Parameter N	VehicleID to VeRi-776 (%)		
	mAP	Rank-1	Rank-5
N = 1	43.9	73.4	80.1
N = 2	44.5	73.8	81.7
N = 3	45.6	74.3	83.7

The specific description of feature partitioning in this paper is shown in Table 7. Before feature partitioning, the dimension of all features was 1×2048, and the features were only partitioned based on the length of the feature dimension.

Table 7. Specific division of feature dimensions.

Parameter N	Feature Partition		
	Partition1	Partition2	Partition3
N = 1	1:2048	×	×
N = 2	1:1366	683:2048	×
N = 3	1:1024	512:1536	1024:2048

4.4.3. Effects of CSP and FCD on the Re-ID Model

To verify the effectiveness of cross-style semi-supervised pre-training (CSP) and feature cross-division (FCD) for fine-tuning, this paper conducted related experiments, and the experimental results are shown in Tables 8 and 9.

Table 8. The impact of CSP and FCD in the proposed framework on the trained model (VeRi-776 is the source dataset and VehicleID is the target dataset).

Methods	VeRi-776 to VehicleID (%)								
	Test Size = 800			Test Size = 1600			Test Size = 2400		
	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5
Direct transfer	26.6	35.6	49.5	20.1	28.6	44.1	17.1	22.1	35.4
Ours w/o CSP	49.9	53.0	65.7	44.0	50.4	63.3	40.9	44.9	58.9
Ours w/o FCD	49.8	53.6	66.1	44.8	51.4	63.7	41.1	44.8	58.4
Ours	51.9	54.4	67.4	46.5	52.7	65.6	42.7	45.9	60.3

Table 9. The impact of CSP and FCD in the proposed framework on the trained model (VehicleID is the source dataset and VeRi-776 is the target dataset).

Methods	VehicleID to VeRi-776 (%)		
	mAP	Rank-1	Rank-5
Direct transfer	22.3	58.9	66.9
Ours w/o CSP	44.5	73.5	82.9
Ours w/o FCD	43,9	73.4	80.1
Ours	45.6	74.3	83.7

This section mainly analyzes the following situations. In the first case, the direct transfer [36] means the pre-trained model based on the source domain is directly used for classification of the target dataset. In the second case, the labeled source data are still used for pre-training the model, but FCD is used to calculate the similarity of images during fine-tuning. In the third case, pre-training is carried out according to the CSP proposed in this paper, and in the fine-tuning stage, the similarity of images is directly measured using global features. In the fourth case, the model proposed in this paper is used to realize the Re-ID task. The conclusion that can be drawn is that the application of each module proposed in this section improves the performance of the model compared to direct transfer. Moreover, the overall framework, including CSP and FCD, performs better than the single use of each module.

To demonstrate the role of each module in more detail, this article shows the accuracy changes in the last few iterations of the specific experimental iteration process in Figure 5. It can be observed that during each iteration, the model using CSP+FCD has higher accuracy than the model using the two modules alone. At the same time, this paper also visualized the rank list during the last training process to support the role of the FCD module, as shown in Figure 6.

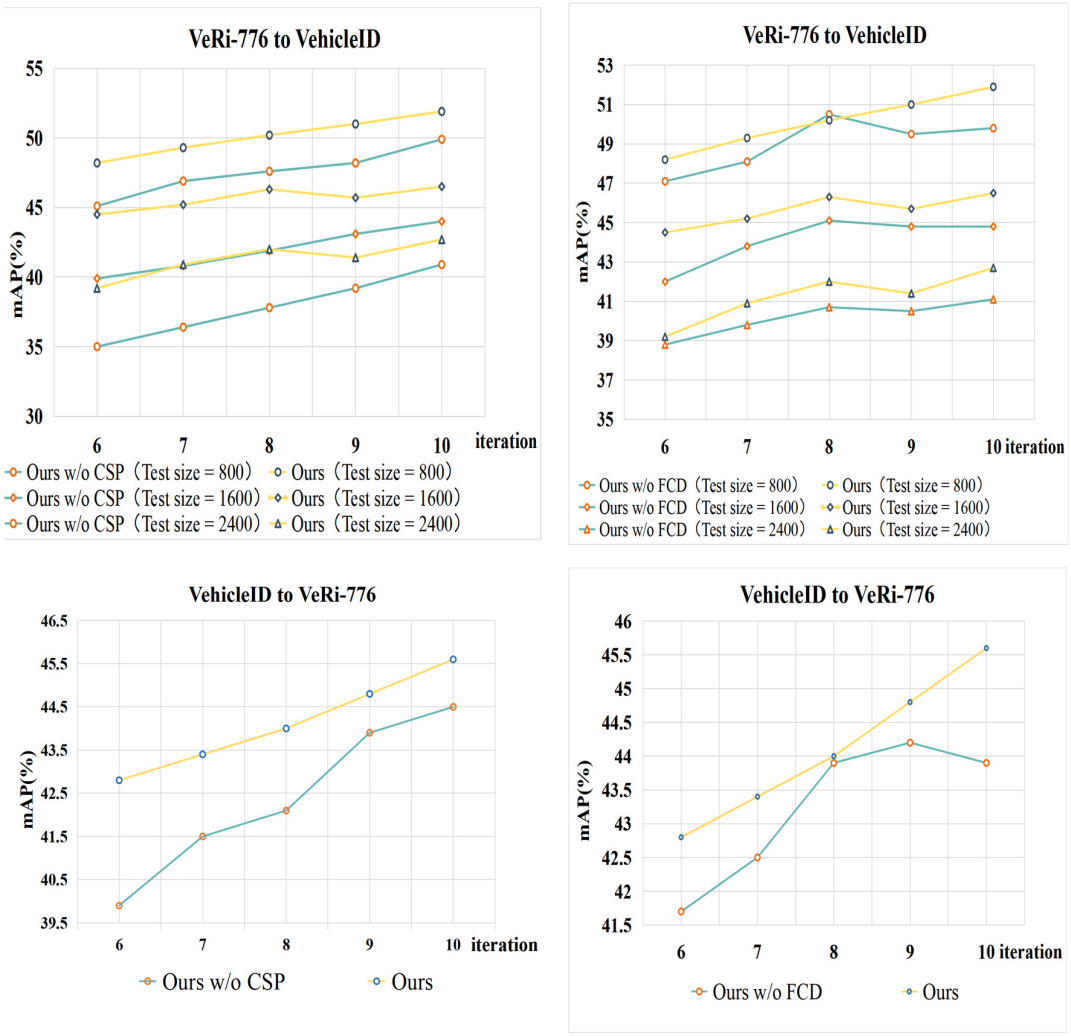

Figure 5. Changes in accuracy with iterations.

Figure 6. The rank list during fine-tuning (different colors of image boxes represent different sample classes).

5. Conclusions

To address the issues of the mismatch of data and label in the pre-training stage and the pseudo-labels being overly reliant on global features in the fine-tuning stage of unsupervised vehicle domain adaptation based on GANs, this paper proposes two modules: cross-style semi-supervised pre-training and feature cross-division. CSP conducts semi-supervised training on the pre-training model using both style transfer data and source domain data to improve the model's generalization ability. In this process, CSP generates soft labels that correspond to the data distribution for style transfer data during the pre-training stage and mines more target information. Additionally, the FCD module obtains partition features during fine-tuning to improve the confidence of pseudo-labels and reduce the model's reliance on global features. The superiority of the proposed method is fully verified through experiments on two large public datasets. However, there is still a drawback to our work: the proposed method of this paper requires a long time for training; thus, it is not possible to have a model immediately put into testing in a short time. In the future, we will further study the work of light weight to reduce the training time of the model. In addition to this, we will focus on vehicle Re-ID algorithms for other types of noise such as random noise of input data or parametric noise of the model in the future.

Author Contributions: Conceptualization, G.Z., Q.W. and W.M.; methodology, G.Z., Q.W. and W.M.; software, G.Z., Q.W., Z.W. and Q.H.; formal analysis, G.Z., H.Z. and Q.H.; writing—original draft preparation, G.Z. and H.Z.; writing—review and editing, G.Z., Q.H. and Z.W.; supervision, W.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 62076117 and No. 62166026) and Jiangxi Key Laboratory of Smart City (Grant No. 20192BCD40002).

Data Availability Statement: Not applicable.

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

References

1. Wang, Y.; Wei, Y.; Ma, R.; Wang, L.; Wang, C. Unsupervised vehicle re-identification based on mixed sample contrastive learning. *Signal Image Video Process.* **2022**, *16*, 2083–2091. [CrossRef]
2. Dubourvieux, F.; Audigier, R.; Loesch, A.; Ainouz, S.; Canu, S. A formal approach to good practices in pseudo-labeling for unsupervised domain adaptive re-identification. *Comput. Vis. Image Underst.* **2022**, *223*, 103527. [CrossRef]
3. Xu, Y.; Guo, X.; Rong, L. A review of research on vehicle re-identification methods with unsupervised learning. *J. Front. Comput. Sci. Technol.* **2023**, *17*, 1017–1037.
4. Wang, Q.; Min, W.; Han, Q.; Yang, Z.; Xiong, X.; Zhu, M.; Zhao, H. Viewpoint adaptation learning with cross-view distance metric for robust vehicle re-identification. *Inf. Sci.* **2021**, *564*, 71–84. [CrossRef]
5. Min, W.; Fan, M.; Guo, X.; Han, Q. A new approach to track multiple vehicles with the combination of robust detection and two classifiers. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 174–186. [CrossRef]
6. Min, W.; Liu, R.; He, D.; Han, Q.; Wei, Q.; Wang, Q. Traffic sign recognition based on semantic scene understanding and structural traffic sign location. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15794–15807. [CrossRef]
7. Han, Q.; Liu, H.; Min, W.; Huang, T.; Lin, D.; Wang, Q. 3D Skeleton and two streams approach to person re-identification using optimized region matching. *ACM Trans. Multimed. Comput. Commun. Appl.* **2022**, *18*, 1–17. [CrossRef]
8. Liu, X.; Zhang, S.; Huang, Q.; Gao, W. Ram: A region-aware deep model for vehicle re-identification. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.
9. Chen, T.S.; Liu, C.T.; Wu, C.W.; Chien, S.Y. Orientation-aware vehicle re-identification with semantics-guided part attention network. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 330–346.
10. Wang, Y.; Peng, J.; Wang, H.; Wang, M. Progressive learning with multi-scale attention network for cross-domain vehicle re-identification. *Sci. China Inf. Sci.* **2022**, *65*, 160103. [CrossRef]
11. Liu, C.T.; Lee, M.Y.; Wu, C.W.; Chen, B.Y.; Chen, T.S.; Hsu, Y.T.; Chien, S.Y. Supervised joint domain learning for vehicle re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR) Workshops, Seoul, Korea, 27–28 October 2019; pp. 45–52.
12. Peng, J.; Wang, H.; Zhao, T.; Fu, X. Cross domain knowledge transfer for unsupervised vehicle re-identification. In Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China, 8–12 July 2019; pp. 453–458.

13. Guo, X.; Yang, C.; Li, B.; Yuan, Y. Metacorrection: Domain-aware meta loss correction for unsupervised domain adaptation in semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 3927–3936.
14. Peng, J.; Wang, Y.; Wang, H.; Zhang, Z.; Fu, X.; Wang, M. Unsupervised vehicle re-identification with progressive adaptation. *arXiv* **2020**, arXiv:2006.11486.
15. Dai, Z.; Wang, G.; Yuan, W.; Zhu, S.; Tan, P. Cluster contrast for unsupervised person re-identification. In Proceedings of the Asian Conference on Computer Vision(ACCV), Macao, China, 4–8 December 2022; pp. 1142–1160.
16. Zhu, W.; Peng, B. Manifold-based aggregation clustering for unsupervised vehicle re-identification. *Knowl. Based Syst.* **2022**, *235*, 107624. [CrossRef]
17. Chen, X.; Sui, H.; Fang, J.; Feng, W.; Zhou, M. Vehicle Re-identification using distance-based global and partial multi-regional feature learning. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 1276–1286. [CrossRef]
18. 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 IEEE International Conference Compution Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 379–387.
19. Wang, Y.; Qi, G.; Li, S.; Chai, Y.; Li, H. Body part-level domain alignment for domain-adaptive person re-identification with transformer framework. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 3321–3334. [CrossRef]
20. Cho, Y.; Kim, W.J.; Hong, S.; Yoon, S. Part-based pseudo label refinement for unsupervised person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 7308–7318.
21. Wu, A.; Zheng, W.S.; Lai, J.H. Distilled camera-aware self training for semi-supervised person re-identification. *IEEE Access* **2019**, *7*, 156752–156763. [CrossRef]
22. Liu, W.; Chang, X.; Chen, L.; Yang, Y. Semi-supervised bayesian attribute learning for person re-identification. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), New Orleans, LA, USA, 2–7 February 2018; pp. 7162–7169.
23. Qi, L.; Wang, L.; Huo, J.; Shi, Y.; Gao, Y. Progressive cross-camera soft-label learning for semi-supervised person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2815–2829. [CrossRef]
24. 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]
25. Yu, J.; Oh, H. Unsupervised vehicle re-identification via self-supervised metric learning using feature dictionary. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 3806–3813.
26. Bashir, R.M.S.; Shahzad, M.; Fraz, M.M. DUPL-VR: Deep unsupervised progressive learning for vehicle re-identification. In Proceedings of the International Symposium on Visual Computing (ISVC), Las Vegas, NV, USA, 19–21 November 2018; pp. 286–295.
27. Antonio Marin-Reyes, P.; Palazzi, A.; Bergamini, L.; Calderara, S.; Lorenzo-Navarro, J.; Cucchiara, R. Unsupervised vehicle re-identification using triplet networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 166–171.
28. Wu, Z.; Zhang, Y.; Zeng, M.; Qin, F.; Wang, Y. Joint analysis of shapes and images via deep domain adaptation. *Comput. Graph.* **2018**, *70*, 140–147. [CrossRef]
29. Guo, J.; Sun, W.; Pang, Z.; Fei, Y.; Chen, Y. Stable median centre clustering for unsupervised domain adaptation person re-identification. *Comput. Intell. Neurosci.* **2021**, *2021*, 2883559. [CrossRef] [PubMed]
30. Xiao, N.; Zhang, L. Dynamic weighted learning for unsupervised domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 15242–15251.
31. Wang, Q.; Breckon, T. Unsupervised domain adaptation via structured prediction based selective pseudo-labeling. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020; pp. 6243–6250.
32. Wang, Q.; Meng, F.; Breckon, T.P. Data augmentation with norm-AE and selective pseudo-labelling for unsupervised domain adaptation. *Neural Netw.* **2023**, *161*, 614–625. [CrossRef] [PubMed]
33. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 21–26 July 2017; pp. 2223–2232.
34. 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), Cham, Switzerland, 11–14 October 2016; pp. 869–884.
35. Liu, H.; Tian, Y.; Wang, Y.; Pang, L.; Huang, T. Deep relative distance learning: Tell the difference between similar vehicles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2167–2175.
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

37. Technicolor, T.; Related, S.; Technicolor, T.; Related, S. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90.
38. 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, Virtual, 6–12 December 2020; Volume 33, pp. 11309–11321.