

Article

Unsupervised Vehicle Re-Identification Method Based on Source-Free Knowledge Transfer

Zhigang Song, Daisong Li , Zhongyou Chen and Wenqin Yang * 

The Academy of Digital China (Fujian), Fuzhou University, Fuzhou 350108, China; zgsong@fzu.edu.cn (Z.S.); 211127138@fzu.edu.cn (D.L.); nnbwbyhx299@163.com (Z.C.)

* Correspondence: yangwenqin@fzu.edu.cn

Abstract: The unsupervised domain-adaptive vehicle re-identification approach aims to transfer knowledge from a labeled source domain to an unlabeled target domain; however, there are knowledge differences between the target domain and the source domain. To mitigate domain discrepancies, existing unsupervised domain-adaptive re-identification methods typically require access to source domain data to assist in retraining the target domain model. However, for security reasons, such as data privacy, data exchange between different domains is often infeasible in many scenarios. To this end, this paper proposes an unsupervised domain-adaptive vehicle re-identification method based on source-free knowledge transfer. First, by constructing a source-free domain knowledge migration module, the target domain is consistent with the source domain model output to train a generator to generate the “source-like samples”. Then, it can effectively reduce the model knowledge difference and improve the model’s generalization performance. In the experiment, two mainstream public datasets in this field, VeRi776 and VehicleID, are tested experimentally, and the obtained rank-k (the cumulative matching features) and mAP (the mean Average Precision) indicators are both improved, which are suitable for object re-identification tasks when data between domains cannot be interoperated.

Keywords: vehicle re-identification; unsupervised domain adaptation; source-free knowledge transfer; pseudo label; joint training



Citation: Song, Z.; Li, D.; Chen, Z.; Yang, W. Unsupervised Vehicle Re-Identification Method Based on Source-Free Knowledge Transfer. *Appl. Sci.* **2023**, *13*, 11013. <https://doi.org/10.3390/app131911013>

Academic Editors: Katia Lida Kermanidis, Phivos Mylonas and Manolis Maragoudakis

Received: 4 September 2023
Revised: 1 October 2023
Accepted: 2 October 2023
Published: 6 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vehicle re-identification (Re-ID) refers to judging whether vehicle images captured in non-overlapping areas belong to the same vehicle in a traffic monitoring scene within a specific range. Recently, vehicle re-identification methods based on supervised learning have made great progress [1–5]. However, the supervised learning method mainly has the following problems: (1) It is extremely dependent on complete labels, that is, the labels of training data from multiple non-overlapping cameras, annotating all large-scale unlabeled data, which is time-consuming and labor-intensive. (2) These methods perform well in the original task (source domain), but when deployed in a new environment (target domain), the performance will drop significantly due to the presence of domain bias.

To overcome these problems, researchers began to focus on the research of unsupervised domain-adaptive vehicle re-identification methods [6–11], that is, trying to transfer images from a well-marked source domain dataset to an unlabeled target domain dataset through knowledge transfer. Isola et al. [12] used generated images to train the ReID model by preserving the identity information from the well-labeled domain, while learning the style of the unlabeled domain to improve the performance of the model in the unlabeled target domain. Peng et al. [13] proposed a progressive adaptation learning algorithm for vehicle re-identification. This method utilizes the source domain to generate “pseudo-target samples” through the Generative Adversarial Network (GAN). It employs a dynamic sampling strategy during training to mitigate domain discrepancies.

Zheng et al. [14] proposed a viewpoint-aware clustering algorithm. It leverages a pre-trained vehicle orientation predictor to predict the orientations of vehicles and assign directional pseudo labels. First, it clusters vehicles with the same perspective. Subsequently, it clusters vehicles with different perspectives, thereby enhancing the performance of the vehicle re-identification model. Wang et al. [15] proposed a progressive learning method named PLM for vehicle re-identification in unknown domains. The method utilizes domain adaptation and a multi-scale attention network to smooth domain bias, trains a reID model, and introduces a weighted label smoothing loss to improve performance. The above methods preserve the identity information from a well-labeled source domain while learning the style of the unlabeled target domain, but usually face the problem that when learning an adaptive model on the target domain, it is unavoidable to visit the source domain to generate new samples for subsequent fine-tuning. However, due to issues related to data ownership and privacy concerns, such as the fact that vehicle data between different cities are typically not shared, in many cases data exchange between domains is infeasible. Consequently, the target domain model cannot directly access source domain data, greatly impacting the model's adaptability performance.

To this end, this paper proposes an unsupervised domain-adaptive vehicle re-identification method based on source-free knowledge transfer; that is, a target domain sample is given, the migration image of the sample is obtained through the generator, and the two images are, respectively, provided to the target model and the source model. The difference between the image pair compensates for the knowledge difference between the domain models, so that the output of the two domain models is similar, and then the generator is trained by constraining the output similarity, and the target domain data can be transformed to have the style of the source domain. Therefore, these "source-like samples" can replace the role played by the source domain data in the model adaptation of the target domain, and since the generated sample content is provided by the target domain, it has more affinity in the process of model adaptation, which helps to solve the problem that the target domain cannot access the data of the source domain. The method can be divided into two stages:

- (1) In the first stage, we construct a source-free knowledge transfer module. It trains a generator to produce "source-like samples" using only the source domain model and the target domain model trained without labeled target domain data as supervision. Importantly, this process does not involve accessing source domain data. The "source-like samples" exhibit a style matching the source domain and content matching the target domain.
- (2) In the second stage, we employ a progressive joint training strategy to gradually train an adaptive model by inputting different proportions of "source-like samples" and target domain data. This process can be viewed as a means of data augmentation. Compared to directly applying target domain data to the source domain model, the "source-like samples" infused with source domain knowledge exhibit greater affinity to the model. Through iterative training, they effectively reduce domain discrepancies, thereby enhancing the model's generalization performance.

The contributions of this paper can be summarized as follows:

- (1) We propose an unsupervised domain-adaptive vehicle re-identification method based on source-free knowledge transfer. Without the need to access source domain data, we utilize domain discrepancy information inherent in the source domain model and the target domain model to constrain a generator in generating "source-like samples." These samples serve as a means of data augmentation to assist in model training for vehicle re-identification tasks.
- (2) We introduce "source-like samples" and a progressive joint training strategy for the target domain. These "source-like samples" are adapted to the same style as the source domain model and matched in content to the target domain data. They serve

as an intermediate bridge between the source domain model and the target domain data, alleviating domain discrepancies and thus enhancing model performance.

2. Method

In this section, we give a detailed description of the proposed method. The schematic diagram of the method is shown in Figure 1. It only needs to be given a source model and a target model, and through the source-free knowledge transfer module constructed in this paper, a generator $G(\cdot)$ can be trained to generate “source-like samples” \tilde{x} . This sample contains source domain knowledge and is more compatible with the source domain model compared to directly applying it to the target domain. It can be utilized as a data augmentation technique to assist in training the target model, thus contributing to improved model performance. Furthermore, we employ a progressive joint training strategy, using this sample in conjunction with target domain data. We control the proportion of “source-like samples” relative to the original target domain samples to avoid potential degradation in model performance due to an excessive proportion of noisy samples. Our approach in this paper eliminates the need to access source domain data, overcoming the limitations imposed by existing unsupervised domain-adaptive methods that require access to the source domain. This eliminates potential security and transmission concerns associated with accessing source domain data.

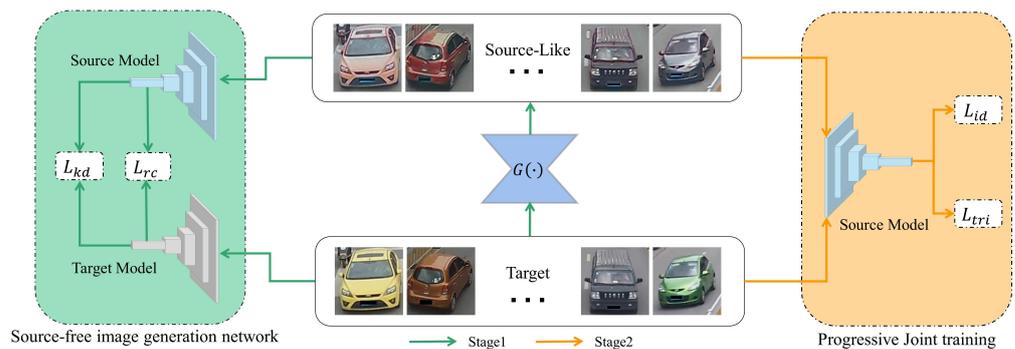


Figure 1. Schematic diagram of unsupervised domain-adaptive vehicle re-identification method based on source-free knowledge transfer.

2.1. Pre-Trained Source Model and Target Model

Among the existing unsupervised domain-adaptive methods [16], the usual practice is to train the model $F(\cdot | \theta)$ on the source domain first, where θ represents the parameters of the current model, and then transfer the model to the target domain for learning.

The method in this paper does not need to access the source data, and its source domain model is expressed as shown in Equation (1):

$$\mathbb{D}_s = \{x_i^s, y_i^s\} |_{i=1}^{N_s} \tag{1}$$

where x_i^s represents the i -th training sample, y_i^s represents its identity label, and N_s represents the total number of samples. The target domain model is expressed as shown in Equation (2):

$$\mathbb{D}_t = \{x_i^t\} |_{i=1}^{N_t} \tag{2}$$

which does not have any associated labels. The experimental steps are as follows:

First, the source domain is accessed to train the source model as well as a learnable source domain classifier $C^s : f^s \rightarrow \{1, 2, \dots, M_s\}$, where M_s represents the number of sample identities.

Second, it is optimized using the identity classification loss $\mathcal{L}_{id}^s(\theta)$ and triplet loss $\mathcal{L}_{tri}^s(\theta)$ [17] composed of the cross-entropy loss function L_{ce} , as shown in Equations (3) and (4):

$$\mathcal{L}_{id}^s(\theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathcal{L}_{ce}(C^s(F^s(x_i^s | \theta)), y_i^s) \tag{3}$$

$$\mathcal{L}_{tri}^s(\theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} \max(0, \|F^s(x_i^s | \theta) - F^s(x_{(i,p)}^s | \theta)\| + m - \|F^s(x_i^s | \theta) - F^s(x_{(i,n)}^s | \theta)\|) \tag{4}$$

where $\|\cdot\|$ represents the L_2 -norm; the subscripts (i, p) and (i, n) represents the positive sample and the negative sample of the i -th sample, respectively; and $m = 0.5$ represents the distance margin of triplet loss. The overall loss $\mathcal{L}^s(\theta)$ of the pre-trained source domain model is therefore calculated as shown in Equation (5):

$$\mathcal{L}^s(\theta) = \mathcal{L}_{id}^s(\theta) + \lambda^s \mathcal{L}_{Ptri}^s(\theta) \tag{5}$$

where λ^s represents the weight of two losses. After obtaining the source model, we train a target model by loading the source model parameters, clustering the target domain, and then predicting the pseudo labels \tilde{y}_i^t . The overall loss $\mathcal{L}^t(\theta)$ of the target domain model is therefore calculated as shown in Equation (6):

$$\mathcal{L}^t(\theta) = \mathcal{L}_{id}^t(\theta) + \lambda^t \mathcal{L}_{tri}^t(\theta) \tag{6}$$

2.2. Source-Free Image Generation Module

The existing unsupervised domain-adaptive vehicle re-identification methods usually need to access the source domain data, and transfer the well-marked source domain data to the unmarked target domain style through style-based transfer [18,19] or a generative confrontation network [20,21], so as to smooth domain bias to better apply source domain models to target domain data. However, in cases where certain data are subject to security and privacy restrictions, accessing source domain data can be extremely challenging. To address this issue, this paper constructs a source-free image generation module. Its purpose is to leverage the implicit domain information within the model to compel the generator to produce “source-like samples” in the target domain data, mimicking the style of the source domain. This generation of samples aims to bridge the knowledge gap between models.

As shown in Figure 2, the first row consists of target sample images, while the second row shows the “source-like samples” generated through the source-free image generation module. The most significant characteristic of these samples is that their content matches that of the target domain data, while their style can match that of the source domain model. They serve as a bridge between the source domain and the target domain. In the subsequent model optimization process, training these samples in conjunction with target domain samples effectively enhances the generalization performance of the vehicle re-identification model.

The source-free image generation module is designed to train an image generator $G(\cdot)$ using the implicit domain information within the model, thereby generating “source-like samples” with the style of the source domain. These samples replace source domain data to match the source model. To describe the knowledge adapted in the “source-like samples,” in addition to the traditional knowledge distillation loss \mathcal{L}_{kd} , a channel-level relationship consistency loss \mathcal{L}_{rc} is introduced. This loss focuses on the relative channel relationships between feature maps in the target domain and the “source-like samples”. Therefore, the total loss \mathcal{L}_{SFIG} is calculated as shown in Equation (7):

$$\mathcal{L}_{SFIG} = \mathcal{L}_{kd} + \mathcal{L}_{rc} \tag{7}$$

In the following sections, the details of the two losses will be introduced in detail.



Figure 2. Schematic diagram of comparison between source-like samples and target samples. (a) target. (b) source-like.

2.2.1. Knowledge Distillation Loss

In our proposed source-free image generation network, we utilize a combination $F_s(G(x_i^t))$ of the source model and generator to describe the knowledge adapted in the target model. This approach can be considered as a special application of knowledge distillation. Our aim is to extract knowledge differences between two domains into the generator. In this case, we compose the knowledge distillation loss \mathcal{L}_{kd} by using the output $\rho(F^s(\tilde{x}))$ obtained by feeding “source-like samples” into the source model and the output $\rho(F^t(x_i^t))$ obtained by feeding target domain samples into the target model, as shown in Equation (8):

$$\mathcal{L}_{kd} = \mathcal{D}_{kl}(\rho(F^s(\tilde{x})), \rho(F^t(x_i^t))) \tag{8}$$

where $\mathcal{D}_{kl}(\cdot)$ represents the Kullback–Leibler divergence(KL).

2.2.2. Channel-Level Relational Consistency Loss

In unsupervised domain-adaptive tasks, it is usually assumed that there is a fixed classifier, so it can be considered that the global features obtained by the target domain through the target model should be similar to the global features obtained by “source-like samples” through the source model. To promote similar channel-level relationships between feature maps $f^s \rightarrow F^s(\tilde{x})$ and $f^t \rightarrow F^t(x_i^t)$, a relation consistency loss is used to constrain the model.

Previous knowledge distillation work is usually constrained by maintaining batch-level or pixel-level relationships [22,23]. However, this constrained approach is not suitable for the current task. First of all, the batch-level relationship cannot supervise the generation task of each image well, which will cause damage to the generated effect. Second, the effectiveness of pixel-level relationships will be greatly reduced after global pooling. Compared with the two, the channel-level relationship [24] is computed on a per-image basis and is not affected by global pooling. Therefore, the channel-level relationship is more suitable for computing \mathcal{L}_{rc} .

Given the feature map f^s of “source-like samples” and the feature map f^t of the target domain, we resize them into feature vectors \mathcal{F}^s and \mathcal{F}^t , as shown in Equations (9) and (10):

$$f^s \in \mathbb{R}^{D*H*W} \rightarrow \mathcal{F}^s \in \mathbb{R}^{D*HW} \tag{9}$$

$$f^t \in \mathbb{R}^{D*H*W} \rightarrow \mathcal{F}^t \in \mathbb{R}^{D*HW} \tag{10}$$

where $D, H,$ and $W,$ represent the feature map’s depth (number of channels), height, and width, respectively. Next, we compute their channel-level self-correlation, the gram matrix, as shown in Equation (11):

$$\mathcal{G}^s = \mathcal{F}^s \cdot (\mathcal{F}^s)^T, \mathcal{G}^t = \mathcal{F}^t \cdot (\mathcal{F}^t)^T \tag{11}$$

where $\mathcal{G}^s, \mathcal{G}^t \in \mathbb{R}^{D \times D}$. Like other similarity-preserving losses for knowledge distillation, we apply row-level L_2 -norm, as shown in Equation (12):

$$\tilde{\mathcal{G}}_{[i,:]}^s = \frac{\mathcal{G}_{[i,:]}^s}{\|\mathcal{G}_{[i,:]}^s\|_2}, \tilde{\mathcal{G}}_{[i,:]}^t = \frac{\mathcal{G}_{[i,:]}^t}{\|\mathcal{G}_{[i,:]}^t\|_2} \quad (12)$$

where $[i, :]$ represents row i in the matrix. Finally, the channel-level relational consistency loss \mathcal{L}_{rc} is the mean squared error (MSE) between the normalized Grann matrices, as shown in Equation (13):

$$\mathcal{L}_{rc} = \frac{1}{D} \|\tilde{\mathcal{G}}^s - \tilde{\mathcal{G}}^t\|_F^2 \quad (13)$$

2.3. Progressive Joint Training Strategy

In the previous section, the target domain data were used to generate “source-like samples” through the source-free knowledge transfer module. Since this sample combines the style of the source domain and the vehicle content of the target domain, it can be used for training to improve the target. The performance degradation caused by deploying domain data to the source domain model improves the generalization performance of the model. In addition, during the process of model adaptation, due to issues such as image style and image quality, “source-like samples” may be regarded as noise by the model. Therefore, a progressive joint training strategy is introduced to control the feeding ratio of “source-like samples” and target domain data, effectively preventing the model from being adversely affected by excessive noise caused by a one-time input of “source-like samples.” Additionally, as the proportion of “source-like samples” increases, the model’s adaptability to the fusion of both sample types also improves, allowing it to learn more discriminative features in the target domain. The maximum proportion of “source-like samples” to the target domain is set at 1:1.

During the training process, the “source-like samples” and target domain data are processed by using the pre-trained source model of vehicle re-identification with good performance to output high-dimensional features. Most of the previous methods chose K-means to generate clusters, which need to be initialized by cluster centroids. However, it is uncertain how many categories are required in the target domain. Therefore, DBSCAN [25] is chosen as the clustering method. Specifically, instead of using a fixed clustering radius, this paper adopts a dynamic clustering radius calculated using K-Nearest Neighbors (KNN). After DBSCAN, in order to filter noise, some of the most reliable samples are selected for soft label assignment according to the distance between sample features and cluster centroids. For the proposed method, samples satisfying $\|f_i, c_{f_i}\| < \gamma$ are used for the next iteration, where f_i is the feature of the i -th image, c_{f_i} is the feature of the centroid of the cluster to which f_i belongs, and γ represents the metric radius belonging to the same category.

3. Experiment

3.1. Experimental Environment Settings

Currently commonly used open-source deep learning frameworks include Caffe, Tensorflow, PyTorch, etc. The method proposed in this paper is trained in the PyTorch framework, including the ablation experiment, which is also conducted in PyTorch. Compared with other frameworks, Python, the development language of PyTorch, has the biggest advantage of supporting dynamic neural networks [26], and the framework is more intuitive and concise as a whole. All models were trained on the NVIDIA GeForce RTX 3060 graphics card, the initial learning rate was set to 0.0005, and the weight decay was set to 0.0005. The total number of iterations was set to 50 and optimized using the Adam optimizer. In order to facilitate the experiment, the model training framework proposed in [6] was used here to obtain the initial source domain model and target domain model. The source domain data were accessed here only to provide the initial source domain model,

and the source domain data were no longer accessed in subsequent experiments. The main architecture of the model adopts ResNet50 [27]. Due to memory issues, the generator in the source-free knowledge transfer module was built using a modified CycleGAN architecture with three residual blocks. The source and target domain models were fixed during training. During the training process, “source-like samples” generated by the source-free knowledge transfer module were used for training in various ratios with target domain data. The ratios between “source-like samples” and target domain data were set as 1:5, 1:4, 1:3, 1:2, and 1:1. The clustering method employed is DBSCAN.

3.2. Datasets Setting and Evaluation Index Level

All experiments in this paper were conducted on two significant publicly available datasets in the vehicle re-identification domain: VeRi776 [28] and VehicleID [29]. For each vehicle in the test datasets, a query image was selected from the perspective of each camera where it was located, the input images in the experiment were modified to 240×240 , and the number of input images in each batch was 32. For training, the input images were preprocessed by random horizontal flipping and random erasing. The evaluation metrics used in the experiments include Rank-1, Rank-5, and mean Average Precision (mAP) computed from the cumulative matching characteristics. Rank-N represents the recognition accuracy of identifying vehicles belonging to the same identity as the query image in the top N sorted results. mAP is another crucial evaluation metric in vehicle re-identification tasks, measuring the comprehensive performance of a model in recognizing objects in images with high accuracy and locating them accurately. A higher mAP score indicates better model performance.

3.3. Experimental Results and Analysis

This subsection compares the proposed method with existing unsupervised methods. Tables 1–3 show the comparison results. In Table 1, the initial source domain model is trained on the VehicleID datasets, and the target domain is VeRi776. In Tables 2 and 3, the initial source domain model is trained on the VeRi776 datasets, and the target domain is VehicleID.

Table 1. Comparative experimental results of the method in this paper on VeRi776.

Methods	VeRi776		
	Rank-1 (%)	Rank-5 (%)	mAP (%)
PUL [30]	55.24	66.27	17.06
SPGAN [31]	57.4	70.0	16.4
HHL [32]	56.20	67.61	17.52
ECN [33]	60.8	70.9	27.7
UDAP [6]	73.9	81.5	35.8
PAL [13]	68.17	79.91	42.04
Direct Transfer	62.1	73.9	27.6
ours	74.4	82.1	37.9

Table 2. Comparative experimental results of the method in this paper on VehicleID (part 1).

Methods	Test Size = 800			Test Size = 1600		
	Rank-1 (%)	Rank-5 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	mAP (%)
PUL [30]	40.03	46.03	43.9	33.83	49.72	37.68
CycleGAN [20]	37.29	58.56	42.32	30.00	49.96	34.92
PAL [13]	50.25	64.91	53.50	44.25	60.95	48.05
Direct Transfer	39.56	56.03	43.01	35.01	50.84	39.17
ours	52.76	67.29	58.33	47.65	63.83	53.72

Table 3. Comparative experimental results of the method in this paper on VehicleID (part 2).

Methods	Test Size = 2400			Test Size = 3200		
	Rank-1 (%)	Rank-5 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	mAP (%)
PUL [30]	30.90	47.18	34.71	28.86	43.41	32.44
CycleGAN [20]	27.15	46.52	31.86	24.83	42.17	29.17
PAL [13]	41.08	59.12	45.14	38.19	55.32	42.13
Direct Transfer	31.05	48.52	34.72	28.12	42.98	31.99
ours	43.87	62.43	50.42	41.77	60.42	47.29

3.3.1. Experimental Results and Analysis on VehicleID→VeRi776

The experimental results of this method on VehicleID→VeRi776 are shown in Table 1. The proposed method has clearly achieved good experimental results on the VeRi776 datasets, with Rank-1, Rank-5, and mAP being 74.4%, 82.1%, and 37.9%, respectively. In particular, PUL [30], HHL [32], and other methods are based on human-based unsupervised re-identification methods. Since most of the current vehicle re-identification is based on supervised methods, they can be used for unsupervised vehicle re-identification for comparison. Methods are scarce, so these pedestrian-based unsupervised re-ID methods are applied to vehicle re-ID to compare with the proposed method.

It is worth mentioning that the unsupervised vehicle re-identification method PAL [13] was also found through research. It is a method of learning cross-domain multi-semantic knowledge for unsupervised vehicle re-identification. Its architecture is quite similar to the method proposed in this paper. In comparison, our proposed method achieves an improvement of 6.3% in Rank-1 and 2.2% in Rank-5, with a slightly lower mAP compared to PAL. Experimental results demonstrate that, compared to PAL methods that still require access to the source domain during the transfer process, the proposed source-free knowledge transfer module can, without accessing source domain data, generate reliable “source-like samples”. Additionally, through the use of a joint training strategy, it effectively mitigates domain discrepancies between the target and source domains, resulting in performance improvement.

3.3.2. Experimental Results and Analysis on VeRi776→VehicleID

The experimental results of this method on VeRi776→VehicleID are shown in Tables 2 and 3. It can be seen that the proposed method also achieved the best experimental results on the VehicleID datasets. When the test size is 800, 1600, 2400, and 3200, the values of Rank-1 are 52.76%, 47.65%, 43.87%, and 41.77%, respectively, the values of Rank-5 are 67.29%, 63.83%, 62.43%, and 60.42%, and the values of mAP are 58.33%, 53.72%, 50.42%, and 47.29%. It is worth noting that the method proposed in this paper has a huge improvement compared with the direct application of the target domain on the source model (direct transfer). Taking test size = 800 as an example, Rank-1 and mAP are increased by 13.2% and 15.32%, respectively. In addition, since the scale of VeRi776 is much smaller than that of VehicleID, there are very few generalization studies on the realization of vehicle re-identification tasks from small datasets to large datasets, so we can only select some from a small number of research works. Relatively prominent methods are compared with the model in this paper. It is not difficult to see from Tables 2 and 3 that compared with PAL, which is most similar to the method in this paper, it shows a stable improvement in different test sizes.

To this end, from the experimental results in the Tables 1–3, it is evident that our proposed method demonstrated remarkable performance in both the VehicleID → VeRi776 and VeRi776 → VehicleID experiments. This validates the effectiveness of the proposed approach, indicating its ability to enhance the performance of unsupervised domain-adaptive vehicle re-identification methods even in a source-free domain setting. However, when

compared to some supervised learning methods, there is still a noticeable performance gap in our experimental results. This implies the need for further exploration and refinement to bridge this gap, particularly in the domain adaptation methods.

3.4. Ablation Experiment

In this section, in order to verify the effectiveness of each module of the proposed method, an ablation experiment was conducted on the VeRi776 datasets.

3.4.1. Validation of the Loss Function of the Source-Free Image Generation Module

In the source-free knowledge transfer module, the generator is constrained by the distillation loss and the channel-level relational consistency loss, forcing the target domain image to generate “source-like samples” with the style of the source domain under the guidance of the source domain model and the target domain model. The impact of the loss function of the source-free domain image generation module on performance is shown in Table 4. It can be clearly seen that when the two losses work together, the generated “source-like samples” are the best for training in image augmentation. Compared with \mathcal{L}_{rc} alone, Rank-1 increased by 4.2% and mAP increased by 3.5%; compared with \mathcal{L}_{kd} alone, Rank-1 increased by 6.8% and mAP increased by 5.8%. In addition, the effect of using only \mathcal{L}_{rc} is also slightly higher than using only \mathcal{L}_{kd} . This is due to the fact that focusing on relative channel relationships can better preserve foreground objects (less blurry and more prominent) while transferring the overall image style, resulting in higher recognition accuracy.

Table 4. The loss function of the source-free image generation module’s impact on performance.

Loss Function	VeRi776		
	Rank-1 (%)	Rank-5 (%)	mAP (%)
\mathcal{L}_{kd}	67.6	78.6	32.1
\mathcal{L}_{rc}	70.2	80.3	34.4
$\mathcal{L}_{kd} + \mathcal{L}_{rc}$	74.4	82.1	37.9

3.4.2. Validation of the Effectiveness of “Source-like Samples”

The impact of “source-like samples” on performance is shown in Table 5. Through data comparison, it is evident that in the “direct transfer” scenario, where target domain data are directly applied to the source domain model, performance degradation occurs due to domain discrepancies. Rank-1 dropped by 29.3% and mAP dropped by 42.9%. After only using “source-like samples” for training was the performance of the model significantly improved, with a 10.4% increase in Rank-1 and a 7.5% increase in mAP. It is worth mentioning that after the joint training of “source-like samples” and target domain data, the model was further improved. This fully demonstrates that using the characteristics of “source-like samples” to match the content of the target domain and the style to match the source domain can help smooth interdomain deviations during training, thereby improving the generalization performance of the model.

Table 5. The impact of “source-like samples” on performance.

Type	VeRi776		
	Rank-1 (%)	Rank-5 (%)	mAP (%)
Supervised Learning	91.4	96.2	70.5
Direct Transfer	62.1	73.9	27.6
Source-Like Samples	72.5	81.5	35.1
Joint Training	74.4	82.1	37.9

3.4.3. Validation of Progressive Joint Training Strategy

The performance impact results of the progressive joint training strategy are shown in Table 6, which are studied by inputting different ratios of “source-like samples” to the target domain data. Comparing the data, it can be found that when the initial proportion of “source-like samples” is small, the performance improvement of the model is also very limited. Rank-1 increased by 5.3%, and mAP increased by 3.7%. When the input ratio was 1:3, the performance change in the model tended to be flat, and it reached the best performance when the input ratio was 1:2; Rank-1 increased by 12.3%, and the mAP increased by 9.7%. This demonstrates that employing a joint training strategy can gradually alleviate domain discrepancies through iterative training, thereby enhancing model performance.

Table 6. The impact of progressive joint training strategy on performance.

Feed Ratio	VeRi776		
	Rank-1 (%)	Rank-5 (%)	mAP (%)
1:5	67.4	76.5	31.3
1:4	69.5	78.9	33.2
1:3	72.6	80.9	36.1
1:2	74.4	82.1	37.9
1:1	73.6	81.7	37.4

4. Conclusions

In this paper, we propose an unsupervised vehicle re-identification method that leverages source-free knowledge transfer for data augmentation, aimed at enhancing the performance of unsupervised vehicle re-identification. We achieve this by employing a source-free knowledge transfer module to generate “source-like samples” and utilizing a joint training strategy to facilitate target domain adaptation, ensuring robust model generalization across diverse environmental conditions. Our approach undergoes comprehensive experimental validation, conducted on two crucial benchmark datasets in the vehicle re-identification domain, namely VeRi776 and VehicleID. When compared to existing unsupervised vehicle re-identification methods, our method exhibits superior performance, indicating its capability to accurately identify vehicles across different environmental settings. Ablation experiments further corroborate the effectiveness of each proposed component. Looking forward, we will continue to explore cross-domain challenges in unsupervised vehicle re-identification to further enhance its performance and applicability.

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

Funding: This research was funded by Fujian Science and Technology Plan: the special project of central government guiding local science and technology development (No. 2022L3003). Fujian Provincial Department of Education project (JAT200039).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Huang, L.; Yang, Q.; Wu, J.; Huang, Y.; Wu, Q.; Xu, J. Generated data with sparse regularized multi-pseudo label for person re-identification. *IEEE Signal Process. Lett.* **2020**, *27*, 391–395. [\[CrossRef\]](#)
2. Bai, Y.; Lou, Y.; Gao, F.; Wang, S.; Wu, Y.; Duan, L.Y. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Trans. Multimed.* **2018**, *20*, 2385–2399. [\[CrossRef\]](#)
3. Zhao, Y.; Shen, C.; Wang, H.; Chen, S. Structural analysis of attributes for vehicle re-identification and retrieval. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 723–734. [\[CrossRef\]](#)
4. Guo, H.; Zhu, K.; Tang, M.; Wang, J. Two-level attention network with multi-grain ranking loss for vehicle re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 4328–4338. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Ma, H.; Li, X.; Yuan, X.; Zhao, C. Two-phase self-supervised pretraining for object re-identification. *Knowl.-Based Syst.* **2023**, *261*, 110220. [\[CrossRef\]](#)
6. Song, L.; Wang, C.; Zhang, L.; Du, B.; Zhang, Q.; Huang, C.; Wang, X. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognit.* **2020**, *102*, 107173. [\[CrossRef\]](#)
7. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
8. Lu, Z.; Lin, R.; He, Q.; Hu, H. Mask-aware pseudo label denoising for unsupervised vehicle re-identification. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4333–4347. [\[CrossRef\]](#)
9. 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.* **2022**, *24*, 2935–2946. [\[CrossRef\]](#)
10. Zhu, W.; Peng, B. Manifold-based aggregation clustering for unsupervised vehicle re-identification. *Knowl.-Based Syst.* **2022**, *235*, 107624. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#)
12. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Venice, Italy, 22–29 October 2017; pp. 1125–1134.
13. Peng, J.; Wang, Y.; Wang, H.; Zhang, Z.; Fu, X.; Wang, M. Unsupervised vehicle re-identification with progressive adaptation. *arXiv* **2020**, arXiv:2006.11486.
14. Zheng, A.; Sun, X.; Li, C.; Tang, J. Aware progressive clustering for unsupervised vehicle re-identification. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 11422–11435. [\[CrossRef\]](#)
15. 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. [\[CrossRef\]](#)
16. Ge, Y.; Chen, D.; Li, H. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. *arXiv* **2020**, arXiv:2001.01526.
17. Hermans, A.; Beyer, L.; Leibe, B. In defense of the triplet loss for person re-identification. *arXiv* **2017**, arXiv:1703.07737.
18. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
19. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1501–1510.
20. 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, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
21. Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.Y.; Isola, P.; Saenko, K.; Efros, A.; Darrell, T. Cycada: Cycle-consistent adversarial domain adaptation. In Proceedings of the International Conference on Machine Learning, Stockholm Sweden, 10–15 July 2018; pp. 1989–1998.
22. Tung, F.; Mori, G. Similarity-preserving knowledge distillation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1365–1374.
23. Li, Z.; Jiang, R.; Aarabi, P. Semantic relation preserving knowledge distillation for image-to-image translation. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 648–663.
24. Hou, Y.; Zheng, L. Visualizing adapted knowledge in domain transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13824–13833.
25. Khan, K.; Rehman, S.U.; Aziz, K.; Fong, S.; Sarasvady, S. DBSCAN: Past, present and future. In Proceedings of the Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014), IEEE, Chennai, India, 17–19 February 2014; pp. 232–238.
26. Han, Y.; Huang, G.; Song, S.; Yang, L.; Wang, H.; Wang, Y. Dynamic neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7436–7456. [\[CrossRef\]](#) [\[PubMed\]](#)
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Liu, X.; Liu, W.; Mei, T.; Ma, H. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Trans. Multimed.* **2017**, *20*, 645–658. [\[CrossRef\]](#)

29. Liu, H.; Tian, Y.; Yang, 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 2167–2175.
30. Fan, H.; Zheng, L.; Yan, C.; Yang, Y. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Trans. Multimed. Comput. Commun. Appl. TOMM* **2018**, *14*, 1–18. [[CrossRef](#)]
31. Deng, W.; Zheng, L.; Ye, Q.; Kang, G.; Yang, Y.; Jiao, J. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 994–1003.
32. Zhong, Z.; Zheng, L.; Li, S.; Yang, Y. Generalizing a person retrieval model hetero-and homogeneously. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 172–188.
33. Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; Yang, Y. Invariance matters: Exemplar memory for domain adaptive person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 598–607.

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.