Unsupervised Vehicle Re-identification with Progressive Adaptation

Jinjia Peng[1,2] , **Yang Wang**[3,4*] , **Huibing Wang**[1,2] , **Zhao Zhang**[3,4*] , **Xianping Fu**[1,2*] , **Meng Wang**[3,4]

[1]College of Information and Science Technology, Dalian Maritime University, Liaoning, Dalian
[2]Pengcheng Laboratory, Shenzhen, Guangdong
[3]Key Laboratory of Knowledge Engineering with Big Data, Ministry of education, Hefei University of Technology, China
[4]School of Computer Science and Information Engineering, Hefei University of Technology, China
{jinjiapeng, huibing.wang,fxp}@dlmu.edu.cn, yangwang@hfut.edu.cn,
{cszzhang,eric.mengwang}@gmail.com

Abstract

Vehicle re-identification (reID) aims at identifying vehicles across different non-overlapping cameras views. The existing methods heavily relied on well-labeled datasets for ideal performance, which inevitably causes fateful drop due to the severe domain bias between the training domain and the real-world scenes; worse still, these approaches required full annotations, which is labor-consuming. To tackle these challenges, we propose a novel progressive adaptation learning method for vehicle reID, named PAL, which infers from the abundant data without annotations. For PAL, a data adaptation module is employed for source domain, which generates the images with similar data distribution to unlabeled target domain as "pseudo target samples". These pseudo samples are combined with the unlabeled samples that are selected by a dynamic sampling strategy to make training faster. We further proposed a weighted label smoothing (WLS) loss, which considers the similarity between samples with different clusters to balance the confidence of pseudo labels. Comprehensive experimental results validate the advantages of PAL on both VehicleID and VeRi-776 dataset.

1 Introduction

Vehicle-related research has attracted a great deal of attention, ranging from vehicle detection, tracking [Tang *et al.*, 2019] to classification [Hu *et al.*, 2017]. Unlike them, vehicle re-identification (reID) aims to match a specific vehicle across scenes captured from multiple non-overlapping cameras, which is of vital significance to intelligent transport. Most of the existing vehicle reID methods, in particular for deep learning models, usually adopt the supervised approaches [Zhao *et al.*, 2019; Lou *et al.*, 2019; Bai *et al.*, 2018; Wang *et al.*, 2017; Guo *et al.*, 2019] for an ideal performance. However, they suffer from the following limitations.

---

*Corresponding Author

On one hand, due to the domain bias, well-trained vehicle reID models under these supervised methods may suffer from a poor performance when directly deployed to the real-world large-scale camera networks. *On the other hand*, these methods heavily relied on the full annotations, i.e., the identity labels of all the training data from multiple cross-view cameras. However, it is labor expensive to annotate large-scale unlabeled data in the real-world scenes. In particular for the vehicle reID task, it is always required to annotate the same vehicle under all cameras. Hence, how to incrementally optimize the vehicle reID algorithms utilizing the combination of the abundant unlabeled data and existing well-labeled data is practical but challenging.

To these ends, a few unsupervised strategies have been proposed. Specifically, [Bashir *et al.*, 2019] takes the cluster to be pseudo labels and then select the reliable pseudo-labeled samples for training. However, the incorrect annotations assigned by clustering are inevitable. One may try transferring images from the well-labeled domain to the unlabeled domain via style transfer for the unsupervised vehicle reID [Isola *et al.*, 2017; Yi *et al.*, 2017]. The generated images are employed to train the reID model, which preserves the identity information from well-labeled domain, while learn the style of unlabeled domain. However, this solution is limited by the learned style that is different from the unlabeled domain, and may fail to adapt to the real-world scenes without label information.

In this paper, we propose a novel unsupervised method, named PAL, together with Weighted Label Smoothing (WLS) loss to better exploit the unlabeled data, while adapt the target domain to vehicle reID "*progressively*". Unlike the existing unsupervised reID methods, a novel adaptation module is proposed to generate "pseudo target images", which learns the style of unlabeled domain and preserves the identity information of the labeled domain. Besides, dynamic sampling strategy is proposed to select reliable pseudo-labeled data from the clustering result. Furthermore, the fusion data that combines the "pseudo target images" and reliable pseudo-labeled data is employed to train the reID model in the subsequent training. To facilitate the presentation, we illustrate the major framework in Fig. 1.
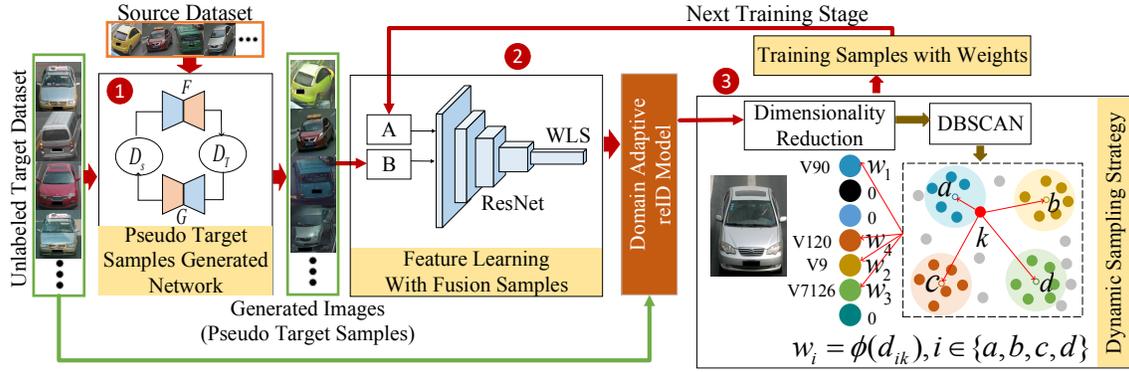
Figure 1: Illustration of the PAL framework. The images are transferred from the source domain to target domain for pseudo target samples generation. During each iteration, we 1) train the reID model by the proposed WLS-based feature learning network that utilizes the fusion data, combining the pseudo target samples with selected samples, and 2) assign pseudo labels with various weights for unlabeled images and select reliable samples according to a dynamic sampling strategy. A means pseudo-labeled images and B is generated images.

Our major contributions are summarized as follows:

- A novel progressive method, named PAL, is proposed for unsupervised vehicle reID to better adapt the unlabeled domain, which iteratively updates the model by WLS based feature learning network, and adopts dynamic sampling strategy to assign labels for selected reliable unlabeled data.

- To make full use of the existing labeled data, PAL employs a data adaptation module based on Generative Adversarial Network (GAN) for generating images as the "pseudo target samples", which are combined with the selected samples from unlabeled domain for model training.

- The feature learning network with WLS loss is proposed, which considers the similarity between the samples and different clusters to balance the confidence of pseudo labels to improve the performance.

Experimental results on benchmark data sets validate the superiority of our method, which is even better than some supervised vehicle reID approaches.

2 Progressive Adaptation Learning for Vehicle ReID

In this section, we formally discuss our proposed technique of progressive adaption learning, named PAL, for vehicle reID. Specifically, as shown in Fig.1, a data adaptation module based on GAN is trained to transfer the well-labeled images to the unlabeled target domain, which aims at smoothing the domain bias and make full use of the existing source domain images. Then the generated images are employed as the "pseudo target samples" and combined with selected unlabeled samples to serve as the input to ResNet50 [He et al., 2015] for feature learning, which adapts the target domain progressively. When the model is trained, WLS loss is proposed to balance the confidence of unlabeled samples and different clusters, which exploits the pseudo labels with different weights, according to the model trained by the last

iteration. Then the output features of the reID model are employed to select reliable samples by dynamic sampling strategy. Finally, the "pseudo target samples" with accurate labels and selected samples from unlabeled domain with pseudo labels are combined to be the training sets for the next iteration. In this way, a more stable adaptive model could be learned progressively.

2.1 Pseudo Target Samples Generated Network

For a target domain, the supervised learning approaches are limited by the unlabeled samples, which can't be utilized to train reID model. Though there are well-labeled datasets, directly applying them to target domain may suffer from a poor performance because of the domain bias mainly caused by diversified illuminations and complicated environment. To remedy this, CycleGAN [Zhu et al., 2017; Lin et al., 2019] is employed to make full use of these well-labeled data, which generates **"pseudo target samples"** to narrow down the domain bias by transferring the style between source domain and target domain. The generated images share the similar style with the target domain while preserving the identity information of the source domain. Specifically, it comprises of two generator-discriminator pairs, (G, D_T) and (F, D_S), which maps a sample from source (target) domain to target (source) domain and generate a sample which is indistinguishable from those in the target(source) domain[Almahairi et al., 2018]. For PAL, besides the traditional adversarial losses and cycle-consistent loss, a content loss [Taigman et al., 2016] is utilized to preserve the label information from the source domain, which is formulated as:

$$L_{id}(G, F, X, Y) = E_{y \sim p_{data}(y)} ||F(y) - y||_1 \\ + E_{x \sim p_{data}(x)} ||G(x) - x||_1, \quad (1)$$

where X and Y represent the source domain and target domain, respectively. $p_{data}(x)$ and $p_{data}(y)$ represent the sample distributions in the source and target domain.

One may wonder why the generated network could make full use of the well-labeled data, we answer this question from the following two aspects:

- Through CycleGAN, the generated "pseudo target samples" have similar distribution for the target domain, which reduces the bias between source and target domain.

- Furthermore, the identity information of source domain is also preserved by turning the content loss during the transferring phrase, implying that the well labeled annotations could be re-utilized in the subsequence.

2.2 Feature Learning Network with WLS Loss

Feature learning plays a vital role of the PAL, which trains the model by combining the generated "pseudo target images" with the selected pseudo labeled samples. For the "pseudo target images", it's easy to obtain the label information. However, how to assign labels for the pseudo labeled samples reasonably is a big challenge, due to the following facts:

- If the clustering centroids serve as the pseudo labels, it may cause ambiguous prediction during the training phase due to the inaccurate clustering results.

- Moreover, it is not reasonable to assign the same labels to all the samples regardless of the distance to the clustering centroids.

Hence, WLS loss is proposed to set the pseudo label distribution as a weighted distribution over all clusters, which effectively regularizes the feature learning network to the target training data distribution.

Specifically, we model the virtual label distribution as a weighted distribution over all clusters for unlabeled data according to the distance between the features and each centroid of clusters. Thus, the weights over all clusters are different in WLS loss. In this way, a dictionary α is constructed to record the weights. For an image g, the weights of the label can be calculated as:

$$w_k^g = \frac{1}{K}\alpha_k^g, k \in [1, K], \tag{2}$$

where α_k^g represents the weight of the image g over the k-th cluster. To obtain α_k^g, unlabeled samples are clustered to obtain centroids set $C = \{c_1, c_2, ..., c_k\}, k \in [1, K]$, which is introduced in section 2.3. K is the number of clusters, while the similarity between g and c_k can be calculated as $d_k^g = ||f_g - f_{c_k}||_2$, where f represents the feature of images or centroids. The set of distance of image g over K centroids could be described as: $d^g = \{d_1^g, d_2^g, ..., d_k^g\}, k \in [1, K]$. Inspired with [Huang *et al.*, 2019], then all elements in d^g are sorted with descending order, and saved to ds^g. α_k is obtained by taking the corresponding index of ds_k^g in the set of ds^g:

$$\alpha_k^g = (1 - \frac{d_k^g}{max(d^g)}) \cdot \psi_{ds^g}(d_k^g), \tag{3}$$

where $\psi_{ds^g}(\cdot)$ is the index of d_k^g in ds^g. Thus, the corresponding relationship between images and cluster centroids is constructed with different weights. In order to filter noise, top-m in w are selected as reliable weights, with others set to 0. To this end, we have:

$$w_k^g = \begin{cases} 0, & \text{if } w_k^g < \theta \\ w_k^g, & \text{if } w_k^g >= \theta, \end{cases} \tag{4}$$

where θ is a threshold. Thus, the WLS loss of unlabeled data ℓ_{wls} can be formulated as:

$$\ell_{wls} = -\sum_{k=1}^{K} w_k log(p(k)) \tag{5}$$

Besides the real unlabeled samples, there are some generated images by CycleGAN that are combined to train the reID model. The training loss is defined as follows:

$$\ell = -(1 - \sigma) \cdot log(p(y)) - \sigma \cdot \lambda \cdot \sum_{k=1}^{K} w_k log(p(k)) \tag{6}$$

For a generated image $\lambda = 0$, the loss is equivalent to cross-entropy loss and y is the label of the generated image. When $\lambda = 1$ means the image is from the unlabeled data and y is the cluster it belongs to. Beyond that, for the unlabeled data, σ is a smoothing factor between cross-entropy loss and WLS loss.

2.3 Dynamic Sampling Strategy

It is crucial to obtain appropriately selected candidates to exploit the unlabeled data. When the model is weak, small reliable measure is set, which is nearby to the cluster centroids in the feature space. As the model becomes stronger in subsequent iterations, various instances should be adaptively selected as the training examples. Hence, a dynamic sampling strategy is proposed to ensure the reliability of selected pseudo-labeled samples. As shown in Fig.1, images in the target domain is processed by the well-trained reID model to output the features with high dimensions. Most methods select the K-Means to generate clusters, which are required to be initialized by the cluster centroids. However, it is uncertain on how many categories are required in the target domain. Hence, DBSCAN is selected as the clustering method. Specifically, instead of employing the fixed clustering radius, the paper employs a dynamic cluster radius rad that is calculated by K-Nearest Neighbors (KNN). After DBSCAN, in order to filter noise, some of top reliable samples are selected to be assigned with soft labels, according the distance between features of the samples and cluster centroids. For our method, samples with $||f_g, c_{f_g}||_2 < \gamma$ are satisfied for the next iteration for training model, where f_g is the feature of g-th image and c_{f_g} is the feature of the cluster centroid where the f_g belongs to. Our method is summarized in Algorithm 1.

3 Experiments

3.1 Datasets and Evaluation Metrics

The experiments are conducted over the following two typical data sets for Vehicle reID: VeRi-776 and VehicleID. **VeRi-776** [Liu *et al.*, 2018] is a large-scale urban surveillance vehicle dataset for reID, which contains over 50,000 images of 776 vehicles, where 37,781 images of 576 vehicles are employed as training set, while 11,579 images of 200 vehicles are employed as a test set. A subset of 1,678 images in the test set generates the query set. **VehicleID** [Liu *et al.*, 2016a] is a surveillance dataset from the real-world scenario, which

Algorithm 1 PAL for Unsupervised Vehicle reID

Require: Numbers of images on the target domain N, labeled Source domain S, unlabeled target domain T, iteration number M, cluster number K, reliability threshold γ, new Training Set D

Ensure: An encoder E for target domain

1: Transfer style from S to T by GAN to generate pseudo target images ST
2: Initialization $E^{(0)}$ with ST, D: $D = ST$
3: **for** i=1 to M **do**
4:    Train $E^{(i)}$ with D utilizing WLS-based feature learning network, compute $ft = E^{(i)}(D)$
5:    Reduce dimension by manifold learning $f = mad(ft)$, calculate number and centroids of clusters: $(K, C) = DBSCAN(f)$
6:    Select features of centroids: $\{c_k\}_{k=1}^{K} \rightarrow \{f_{c_k}\}_{k=1}^{K}$
7:    $D = ST$
8:    **for** k=1 to K **do**
9:       **for** $g = 1$ to N **do**
10:          **if** $||f_g, f_{c_k}||_2 > \gamma$ **then**
11:             $D = D \cup T_g$
12:             Calculate weights w_k^g by Eq.(2)(3)(4)
13:          **end if**
14:       **end for**
15:    **end for**
16: **end for**

---

contains 221,763 images corresponding to 26,267 vehicles in total. From the original testing data, four subsets, which contain 800, 1,600, 2,400 and 3,200 vehicles, are extracted for vehicle search for multi-scales. CMC curve and mAP are employed to evaluate the overall performance for all test images. For each query, its average precision (AP) is computed from the precision-recall curve.

3.2 Implementation Details

For CycleGAN, the model is trained in the tensorflow [Abadi *et al.*, 2016]. It is worth mentioning that any label notation aren't utilized during the learning procedure. In the stage of the feature learning, the ResNet50 [He *et al.*, 2015] is employed as the backbone network. For PAL, the images are transferred by CycleGAN from source domain to target domain, which are as "Pseudo target samples" for training the feature learning model. Considering the limit of device, when training reID model on the VeRi-776, 10,000 transferred images from VehicleID are utilized as the "pseudo target images". The same implementations are conducted when the reID model is trained on the VehicleID. Besides that, when training the unsupervised model on VehicleID, only 35,000 images from the VehicleID are selected as the training set. Moreover, any annotations of target domain aren't employed in our framework.

3.3 Comparison with the State-of-the-art Methods

In this section, the results of the comparison between PAL and other state-of-the-art methods are reported in Tables 1, 2 and Figures 2, 3, which includes: (1) FACT [Liu *et al.*, 2016b]; (2) FACT+Plate-SNN+STR [Liu *et al.*, 2016b]; (3)
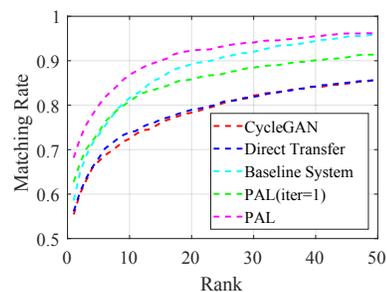

Figure 2: CMC results of several typical methods on VeRi-776. The proposed PAL outperforms other compared methods, especially the "CycleGAN" and the "Direct Transfer". Besides that, compared with the "PAL (iter=1)", "PAL" has a high improvement in mAP, which demonstrates the progressive learning could increase the adaptive ability for the reID model in the unlabeled target domain.

Method	mAP(%)	Rank1(%)	Rank5(%)
FACT	18.75	52.21	72.88
FACT+Plate-SNN+STR	27.77	61.44	78.78
VR-PROUD	22.75	55.78	70.02
PUL	17.06	55.24	66.27
CycleGAN	21.82	55.42	67.34
Direct Transfer	19.39	56.14	68.00
Baseline System	31.94	58.58	73.24
PAL	**42.04**	**68.17**	**79.91**

Table 1: Performance of different methods on VeRi-776. The best results are shown in bold face. PAL can achieve best performance.

Mixed Diff+CCL [Liu *et al.*, 2016a]; (4)VR-PROUD [Bashir *et al.*, 2019]; (5) CycleGAN [Zhu *et al.*, 2017]. This is method of style transfer, which is employed for the domain adaptation; (6) Direct Transfer: It directly employed the well-trained reID model by the [Zheng *et al.*, 2018] on source domain to the target domain; 7)Baseline System. Compared with the framework of PAL, it utilizes the original samples from source domain instead of generated data and the reID model is only trained with cross-entropy (CE) loss; 8)PUL [Fan *et al.*, 2018]. The methods of (1), (2) and (3) are supervised vehicle reID approaches. And others are unsupervised methods. Specially, the PUL is an unsupervised adaptation method of person reID. Since only a few works focused on the unsupervised vehicle reID, PUL is compared with the proposed PAL in this paper. There are some other methods that similar with PUL. However, most of them require special annotations, such as labels for segmenting or detecting keypoints, which are not annotated in the existing vehicle reID datasets. From the Tables 1, 2, we note that the proposed method achieves the best performance among the compared with methods with Rank-1 = 68.17%, mAP = 42.04% on VeRi-776, Rank-1 = 50.25%, 44.25%, 41.08%, 38.19%, mAP = 53.50%, 48.05%, 45.14%, 42.13% on VehicleID with the test size of 800, 1600, 2400, 3200, respectively.

Compared with PUL [Fan *et al.*, 2018] and VR-PROUD [Bashir *et al.*, 2019], PAL has 24.98% and 19.29% gains on VeRi-776, respectively. Our model also outperforms the PUL and VR-PROUD in Rank-1, Rank-5 and mAP on VehicleI-

Method	Test size = 800 (%)			Test size = 1600 (%)			Test size = 2400 (%)			Test size = 3200 (%)		
	mAP	Rank1	Rank5	mAP	Rank1	Rank5	mAP	Rank1	Rank5	mAP	Rank1	Rank5
FACT	–	49.53	67.96	–	**44.63**	64.19	–	39.91	60.49	–	–	–
Mixed Diff+CCL	–	49.00	**73.50**	–	42.80	**66.80**	–	38.20	**61.60**	–	–	–
PUL	43.90	40.03	56.03	37.68	33.83	49.72	34.71	30.90	47.18	32.44	28.86	43.41
CycleGAN	42.32	37.29	58.56	34.92	30.00	49.96	31.89	27.15	46.52	29.17	24.83	42.17
Direct Transfer	40.58	35.48	57.26	33.59	28.86	48.34	30.50	26.08	44.02	27.90	23.85	39.76
Baseline System	42.96	39.11	55.24	38.03	34.04	50.91	34.04	30.10	48.41	31.98	28.24	43.77
PAL	**53.50**	**50.25**	64.91	**48.05**	44.25	60.95	**45.14**	**41.08**	59.12	**42.13**	**38.19**	**55.32**

Table 2: Performance of various methods over different reID methods on VehicleID. It is notable that the best results are shown in bold face. PAL can achieve best performance in most situations. Mixed Diff+CCL can also achieve good performance.

Method	Generated Images	Original Images	WLS	CE
BS	×	√	×	√
CEL	√	×	×	√
OIMG	×	√	√	×
PAL	√	×	√	×

Table 3: The settings for different ablation models.

D. For these methods, the K-Means is employed to assign pseudo-labels for unlabeled samples. Due to the uncertainty on how many categories, the K-Means is not appropriate to be utilized in the reID task. In addition, compared with "Direct Transfer", it is obvious that our proposed PAL achieves 22.65% and 12.03% gains in mAP and Rank-1 on VeRi-776. It also has similar improvements on VehicleID. Furthermore, compared with the supervised approaches, such as FACT [Liu *et al.*, 2016b], Mixed Diff+CCL [Liu *et al.*, 2016a] and FACT+Plate-SNN+STR [Liu *et al.*, 2016b], PAL achieves improvements on VeRi-776 and VehicleID, validating that PAL is more adaptive to different domains.

Compared with the CycleGAN [Zhu *et al.*, 2017] that adapts the domain bias by style transfer, our method has large improvements on both VeRi-776 and VehicleID. The proposed PAL achieves 20.22% and 12.75% improvements in mAP and Rank-1 on VeRi-776, respectively. Similarly, our method has 12.96%, 14.25%, 13.93% and 13.36% gains in Rank-1 on VehicleID with the test sets of 800, 1600, 2400 and 3200. The significant improvements are mainly due to the fact that PAL exploits the similarity among unlabeled samples through iteration for unsupervised vehicle reID. Though the generated images have the style of target domain, they are just served as the pseudo samples. The real samples in the target domain could be more reliable to generate the discriminative features during the stage of training. These results suggest that reliable samples in target domain is an important component for the unsupervised reID task, which indicate that PAL could make full use of the unlabeled samples in target domain.

Compared with "Baseline System", PAL has large improvements both on VeRi-776 and VehicleID. The PAL achieves 10.1% increase in mAP on VeRi-776, and 10.54%, 10.02%, 11.1%, 10.15% improvements in mAP on VehicleID with different test sets, respectively. These indicate that the "pseudo target images" and "weighted label smoothing" are two core components in PAL which lead the reID model

Iteration	CEL (%)			BS (%)		
	mAP	Rank1	Rank5	mAP	Rank1	Rank5
iter1	27.71	59.89	72.10	25.04	57.33	71.33
iter2	33.76	64.12	78.06	30.19	58.40	73.53
iter3	35.73	65.55	78.18	32.49	59.41	73.06
iter4	36.01	63.28	77.47	32.63	59.77	74.07
iter5	33.86	60.90	77.11	32.86	60.96	74.91
iter6	34.03	62.09	75.38	31.94	58.58	73.24

Table 4: Performance of comparison between CEL and BS on VeRi-776.

Iteration	CEL (%)			BS (%)		
	mAP	Rank1	Rank5	mAP	Rank1	Rank5
iter1	35.69	31.43	49.54	34.93	30.71	48.41
iter2	35.89	31.44	49.65	34.00	29.90	46.98
iter3	36.49	32.39	50.84	34.33	30.29	47.14
iter4	36.62	32.73	51.01	33.62	29.67	46.38
iter5	37.33	33.69	51.25	34.08	30.15	46.94
iter6	37.12	33.45	51.36	34.04	30.10	46.94

Table 5: Performance of comparison between CEL and BS on VehicleID(2400).

trained by our method to be more robust to different domains. We discuss more details in the next section.

3.4 Ablation Studies

We conduct ablation studies on two major components of PAL, i.e., the data adaptation module and WLS, which are shown in Fig.4. The settings are depicted in Table 3. All of them share the similar structure with PAL. "Generated Images" means employing the transferred images from source domain and image of target domain to train the models, while "Original Images" means to utilize the original images of source domain and samples from target domain for unsupervised vehicle reID. WLS, CE represent that employ the WLS and cross-entropy loss to train reID models, respectively. Fig.4 shows that PAL achieves the best performance on two datasets, demonstrating that the data adaptation module and WLS are effective to adapt to unlabeled domain.

The Effectiveness of Generated Samples. To demonstrate the effectiveness of the generated samples, BS and CEL are compared, the results are reported in Tables 4 and 5. For CEL, we utilize CycleGAN to translate labeled images from the source domain to the target domain, and regard generated images as the "pseudo target samples". Then the "pseudo tar-
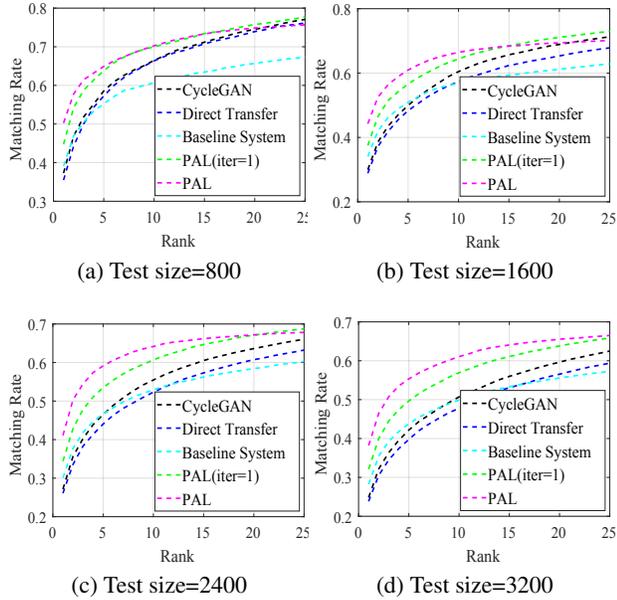
Figure 3: CMC curves of several typical methods on VehicleID. From the curves, it is obvious that better results could be achieved when the model is tested on the different datasets, which demonstrates PAL is effective for different test sets.

Iteration	OIMG (%)			BS (%)		
	mAP	Rank1	Rank5	mAP	Rank1	Rank5
iter1	28.61	60.90	74.19	25.04	57.33	71.33
iter2	30.11	62.09	75.32	30.19	58.40	73.53
iter3	30.52	61.02	74.43	32.49	59.41	73.06
iter4	32.51	63.70	76.16	32.63	59.77	74.07
iter5	33.90	65.19	76.34	32.86	60.96	74.91
iter6	37.33	67.69	79.02	31.94	58.58	73.24

Table 6: Performance of comparison between OIMG and BS on VeRi-776.

get samples" are combined with the images in the target domain to train the reID model. Both CEL and BS are trained by cross-entropy loss. According to the last iteration, compared with BS, the mAP of CEL increases by 2.09% on VeRi-776. Besides that, it also rises to 37.12% and 33.45% in mAP and Rank-1 on VehicleID, demonstrating that the generated images learned the important style information from the target domain, which narrow down the domain gap.

The Effectiveness of WLS. We compare BS with OIMG to validate the effectiveness of the WLS. Tables 6 and 7 show the comparisons on VeRi-776 and VehicleID, where the proposed WLS achieves better performance than cross-entropy loss. According to the last iteration, compared with the BS, the mAP and Rank-1 accuracy increased by 5.39% and 9.11% on VeRi-776 for OIMG, respectively. The similar conclusions hold on VehicleID, which indicates that the WLS loss has better generation ability to achieve discriminative representation during the stage of training.
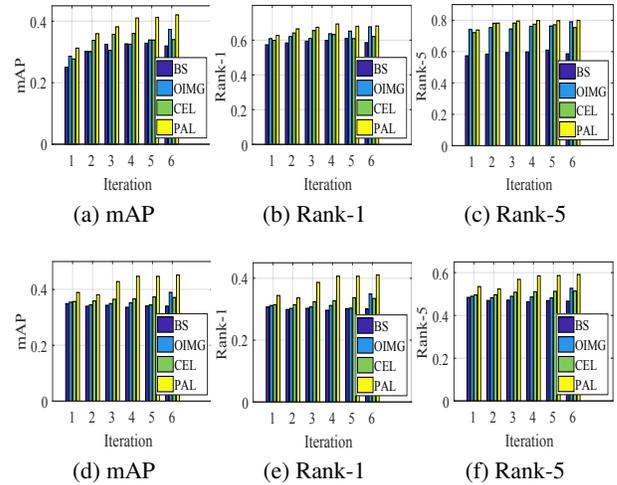

Figure 4: The Comparison Results. (a), (b), (c) are the mAP, Rank-1 and Rank-5 of four comparison methods on VeRi-776. (d), (e), (f) are the results of four comparison methods in every iteration on VehicleID, respectively.

Iteration	OIMG (%)			BS (%)		
	mAP	Rank1	Rank5	mAP	Rank1	Rank5
iter1	35.45	31.19	48.95	34.93	30.71	48.41
iter2	34.48	30.26	48.25	34.00	29.90	46.98
iter3	34.94	30.74	48.93	34.33	30.29	47.14
iter4	35.22	31.20	48.66	33.62	29.67	46.38
iter5	34.48	30.35	48.21	34.08	30.15	46.94
iter6	38.95	34.90	52.69	34.04	30.10	46.94

Table 7: Performance of comparison between OIMG and BS on VehicleID (2400).

4 Conclusion

In this paper, we propose an unsupervised vehicle reID framework, named PAL, which iteratively updates the feature learning model and estimates pseudo labels for unlabeled data for target domain adaptation. The extensive experiments of the developed algorithm has been carried out over benchmark datasets for Vehicle Re-id. It can be observed from the results that compared with other existing unsupervised methods, PAL could achieve superior performance, and even achieve better performance than some typical supervised models.

Acknowledgments

References

[Abadi *et al.*, 2016] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[Almahairi *et al.*, 2018] Amjad Almahairi, Sai Rajeswar, Alessandro Sordoni, Philip Bachman, and Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. *arXiv preprint arXiv:1802.10151*, 2018.

[Bai *et al.*, 2018] Yan Bai, Yihang Lou, Feng Gao, Shiqi Wang, Yuwei Wu, and Ling-Yu Duan. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Transactions on Multimedia*, 20(9):2385–2399, 2018.

[Bashir *et al.*, 2019] Raja Muhammad Saad Bashir, Muhammad Shahzad, and MM Fraz. Vr-proud: Vehicle re-identification using progressive unsupervised deep architecture. *Pattern Recognition*, 90:52–65, 2019.

[Fan *et al.*, 2018] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):83, 2018.

[Guo *et al.*, 2019] Haiyun Guo, Kuan Zhu, Ming Tang, and Jinqiao Wang. Two-level attention network with multi-grain ranking loss for vehicle re-identification. *IEEE Transactions on Image Processing*, 2019.

[He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[Hu *et al.*, 2017] Qichang Hu, Huibing Wang, Teng Li, and Chunhua Shen. Deep cnns with spatially weighted pooling for fine-grained car recognition. *IEEE Transactions on Intelligent Transportation Systems*, 18(11):3147–3156, 2017.

[Huang *et al.*, 2019] Yan Huang, Jingsong Xu, Qiang Wu, Zhedong Zheng, Zhaoxiang Zhang, and Jian Zhang. Multi-pseudo regularized label for generated data in person re-identification. *IEEE Transactions on Image Processing*, 28(3):1391–1403, 2019.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[Lin *et al.*, 2019] Wu Lin, Wang Yang, and Shao Ling. Cycle-consistent deep generative hashing for cross-modal retrieval. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 28(4):1602, 2019.

[Liu *et al.*, 2016a] Hongye Liu, Yonghong Tian, Yaowei Yang, Lu Pang, and Tiejun Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2167–2175, 2016.

[Liu *et al.*, 2016b] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *European Conference on Computer Vision*, pages 869–884. Springer, 2016.

[Liu *et al.*, 2018] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Transactions on Multimedia*, 20(3):645–658, 2018.

[Lou *et al.*, 2019] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan. Embedding adversarial learning for vehicle re-identification. *IEEE Transactions on Image Processing*, 2019.

[Taigman *et al.*, 2016] Yaniv Taigman, Adam Polyak, and Lior Wolf. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*, 2016.

[Tang *et al.*, 2019] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David Anastasiu, and Jenq-Neng Hwang. Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8797–8806, 2019.

[Wang *et al.*, 2017] Zhongdao Wang, Luming Tang, Xihui Liu, Zhuliang Yao, Shuai Yi, Jing Shao, Junjie Yan, Shengjin Wang, Hongsheng Li, and Xiaogang Wang. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2017.

[Yi *et al.*, 2017] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857, 2017.

[Zhao *et al.*, 2019] Yanzhu Zhao, Chunhua Shen, Huibing Wang, and Shengyong Chen. Structural analysis of attributes for vehicle re-identification and retrieval. *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[Zheng *et al.*, 2018] Zhedong Zheng, Liang Zheng, and Yi Yang. A discriminatively learned cnn embedding for person reidentification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(1):13, 2018.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.