Vehicle Re-identification in Still Images: Application of Semi-supervised Learning and Re-ranking

Fangyu Wu[a,b,*], Shiyang Yan[a,b], Jeremy S. Smith[a], Bailing Zhang[b]

[a] *Department of Electrical Engineering and Electronic, University of Liverpool, Liverpool, United Kingdom*
[b] *Department of Computer Science and Software Engineering, Xi'an Jiaotong-liverpool University, SuZhou, JiangSu Province, China*

Abstract

Vehicle re-identification (re-ID), namely, finding exactly the same vehicle from a large number of vehicle images, remains a great challenge in computer vision. Most existing vehicle re-ID approaches follow a fully-supervised learning methodology, in which sufficient labeled training data is required. However, this limits their scalability to realistic applications, due to the high cost of data labeling. In this paper, we adopted a Generative Adversarial Network (GAN) to generate unlabeled samples and enlarge the training set. A semi-supervised learning scheme with the Convolutional Neural Networks (CNN) was proposed accordingly, which assigns a uniform label distribution to the unlabeled images to regularize the supervised model and improve the performance of the vehicle re-ID system. Besides, an improved re-ranking method based on Jaccard distance and k-reciprocal nearest neighbors is proposed to optimize the initial rank list. Extensive experiments over the benchmark datasets VeRi-776, VehicleID and VehicleReID have demonstrated that the proposed method outperforms the state-of-the-art approaches for vehicle re-ID.

Keywords: Vehicle re-identification, Convolutional Neural Networks, Semi-supervised Learning, Re-ranking

*Corresponding author
Email addresses: Fangyu.Wu@xjtlu.edu.cn (Fangyu Wu), Shiyang.Yan@xjtlu.edu.cn (Shiyang Yan), J.S.Smith@liverpool.ac.uk (Jeremy S. Smith), Bailing.Zhang@xjtlu.edu.cn (Bailing Zhang)

1. Introduction

With the explosive growth of video data captured by various surveillance cameras, there is an increasing demand for improved surveillance video analysis capabilities which require a large number of vehicle related tasks, such as vehicle detection, classification and verification. In this work, we focus on the task of vehicle re-identification (re-ID) in still images, which aims to quickly discover, locate and track the target vehicles across multiple cameras, thus automating the time consuming manual task. Vehicle re-ID has practical applications in surveillance systems and intelligent transportation [1]. In vehicle re-ID systems, a query image, also called a probe image, is compared with the gallery images that contain various vehicles captured by multiple cameras. Normally, a rank list is generated that has several matched images from the gallery set. Fig.1 further explains the vehicle re-ID task.

Traditionally, the combination of sensor data and multiple clues are used to solve the task of vehicle re-ID, such as the transit time [2] and the wireless magnetic sensors [3]. However, these methods are sensitive to the fickle environment (e.g., thunder and lightning) and require the extra cost of additional hardware. In addition, the license plate is an important clue which contains the unique ID of vehicle, thus the technologies related to license plate have been proposed in [4], [5]. Nevertheless, it's easy to occlude, remove, or even forge the license plate, especially in criminal circumstances. To alleviate these limitations, we focus on this task based on on its visual appearance, which is essential for fully-fledged vehicle re-ID system.

To this end, the discriminative features should be extracted to distinguish different vehicles for robust vehicle re-ID [6]. Basically, there exists two challenges. (1) Different lighting and complex environments causes difficulties for appearance-based vehicle re-ID. Also, large variations in appearance will be produced if capture vehicle using different cameras. How to take such large intra-class variance into account for feature representation is crucial. (2) Compared with the person re-ID, vehicle re-ID is more challenging as different vehicles

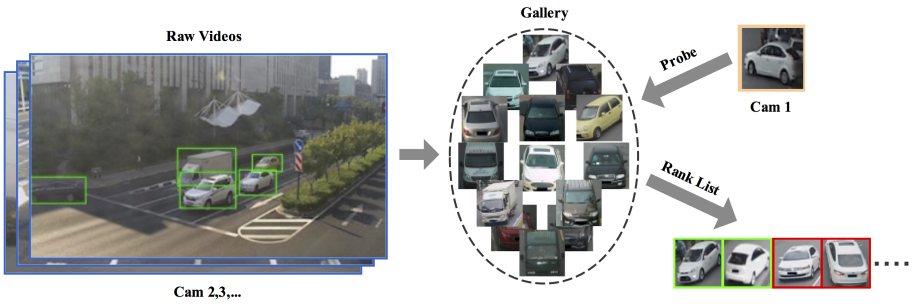

Figure 1: Explanation of the task of vehicle re-ID. Given a snapshot of a vehicle (the probe), a re-ID system retrieves from a database (the gallery) which contains a list of other snapshots of vehicles, usually taken from different cameras at different time, and ranks them by decreasing similarities to the probe.

can be visually very similar to each other, especially when they are from the same category. Fig.2 further explains the situations of intra-class variance and inter-class similarity.

The deep embedding method has shown generalization abilities and promising performance in the re-ID task, which aims at learning compact features embedded in some semantic spaces through a deep convolutional neural network (CNN). The objective of embedding is typically expresses as pulling the features from similar images closer and pushing the features from dissimilar images further away. Among these methods, learning identity-sensitive and view-insensitive features is crucial to ensure the learning effectiveness of the CNN model. Hence rich labelled data from different camera views is required to learn a feature representation that is invariant to the appearance changes. However, relying on manually labelled data for each camera view results in poor scalability. This is due to two reasons: (1) It's a tedious and difficult task for humans to match an identity correctly among hundreds of data from each camera. (2) In real-world applications, there are a large number of cameras in a surveillance network (e.g., those in an airport or shopping mall), it's infeasible to annotate sufficient training samples from all the camera views. Therefore,

(a) Intra-class variance **(b) Inter-class similarity**

Figure 2: Examples explaining the intra-class variance and inter-class similarity. (a) Due to the different viewing angles and illuminations levels of cameras, the images of each row on the left column from the same vehicle produce the significant intra-class variance. (b) The images of each row on the right column belonging to the different vehicles from the same class and produce inter-class similarities. It's challenging to distinguish the vehicles with similar appearance.

these practical issues severely limit the applicability of the existing vehicle re-ID methods.

To alleviate the large demand of training data, the approaches of semi-supervised learning have been proposed recently which uses the unlabeled samples to boost the performance on a specific task. It is driven by the practical value in learning faster, cheaper, and better feature representations. Semi-supervised learning attempts to obtain a deep model that can more accurately predict unseen test data than a deep model learned only from labeled training data. Common semi-supervised learning methods include variants of generative models [7], co-training [8] and graph Laplacian based methods [9]. Above works in semi-supervised learning are based on the fact that sufficient unlabeled data is available. However, if the number of unlabeled sample is scarce or difficult to collect, traditional semi-supervised methods may become useless. In our work, instead of using unlabeled data from the real sample space, we propose a semi-supervised feature embedding method which directly uses a generative ad-

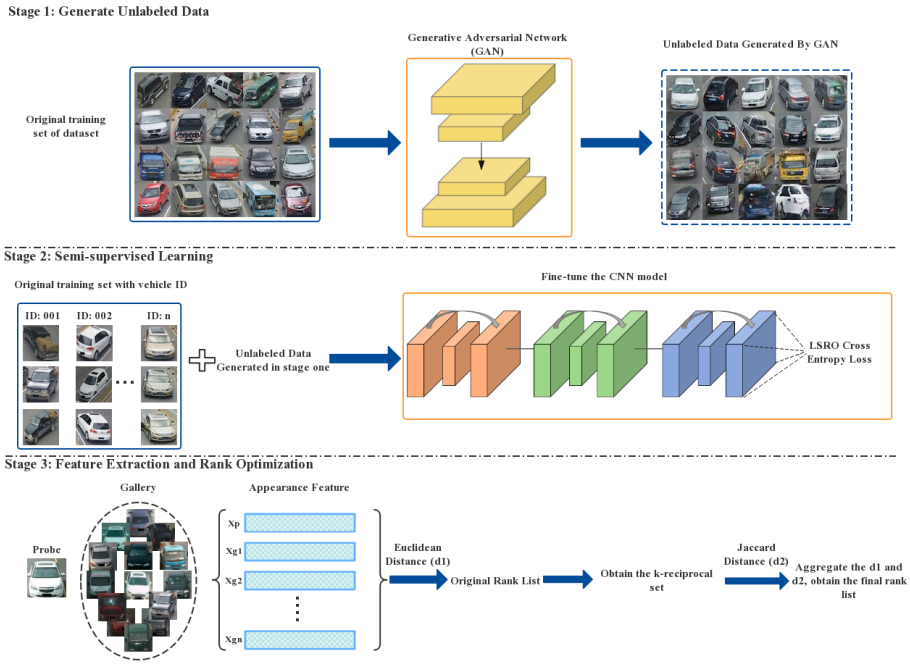

Figure 3: The workflow of the proposed method. There are three stages: (1) Generation of unlabeled data by using the original training set of vehicles to train the generative adversarial network [10]; (2) Semi-supervised learning by combining the labeled training set with vehicle ID and the unlabeled images data to fine-tune the CNN model with LSRO; (3) Feature extraction and rank optimization. We achieve an initial ranking based on the pairwise Euclidean distance of deep feature for the probe and each image in gallery set. To improve the initial ranking list, we finally add the re-ranking step.

versarial network (GAN) to generate unlabeled samples. Goodfellow et al. [11] first proposed the GAN to obtain the optimal discriminator network between real samples and generated samples based on the min-max game between generator and discriminator. Besides, the performance of image generator network will be improved simultaneously. Rather than investigating how to enhance the quality of the generated samples [10], [12], our research will focus on how to use GAN to promote the performance of classifiers. Specifically, we incorporate the generated samples with original training images to train CNN models with semi-supervised learning.

As illustrated in Fig.3, there are three stages in the proposed algorithm. Initially, we obtain the generated vehicle images by using the original images in training set to train DCGAN [10]. In the second stage, we improve the discriminative power of the deep model for the re-ID task by using a larger training set which includes unlabeled images. More precisely, we use the initially labeled target dataset plus the unlabeled data generated in stage one to fine-tune the CNN model. In this manner, the improved ResNet-50 model [13] is trained with all the data simultaneously. This stage is in the setting of semi-supervised learning, as the training dataset includes images with labels and images without labels.

Although significant progress has been achieved from previous researches of appearance based deep learning approaches for vehicle re-ID, their ranking accuracies are often unsatisfactory. To further improve the performance of vehicle re-ID, a technique is presented that uses a distance metric for rank optimization in the third stage. Specifically, we apply the trained CNN model from the second stage to extract the CNN features for probe image and each vehicle in gallery set. The initial ranking list can be achieved by calculating the pairwise Euclidean distances between the probe and the gallery. Then we compute the Euclidean distance and the Jaccard distance by comparing their k-reciprocal nearest neighbor set. We integrate the Euclidean distance and the Jaccard distance to obtain the proposed ranking list. We validate the performance of the proposed technique on three publically available vehicle re-ID datasets, VeRi-

776 [14], VehicleID [15] and VehicleReID [16] dataset, all with promising results. Our contributions can be summarized as follows:

- We propose a semi-supervised deep learning scheme for vehicle re-ID task which makes learning rich feature representations of vehicles from a limited number of labeled data possible.
- We present a re-ranking algorithm for ranking optimization which is firstly introduced for the vehicle re-ID task. Since the sample label is not required, the process of the re-ranking algorithm can be performed in unsupervised learning.
- We conduct extensive experiments and improve state-of-the-art vehicle re-ID performance on two benchmark datasets, VeRi-776 [14] and VehicleID [15] and demonstrate the effectiveness of our proposal. We apply the single shot setting on the VehicleReID [16] dataset for the first time and achieved promising results, providing baseline data for subsequent research.

The remainder of this paper is organized as follows: Section 2 offers a brief overview of the vehicle re-ID literature. We then provide a detailed description of the proposed method in Section 3. The implementation details and experimental results are discussed in Section 4, followed by the conclusion in Section 5.

2. Related work

As an emerging research topic, vehicle re-ID has recently attracted great significant interest [14], [15], [16], [17], [18]. In this section, we review the relevant works from three aspects: semi-supervised learning, re-ranking for person re-ID and vehicle re-ID.

2.1. Semi-supervised Learning

Semi-supervised learning exploits both the labeled data and unlabeled data to perform the learning task and bridges the gap between the fully-supervised

learning and unsupervised learning. Some research exploits weak label annotations for each bounding box [19], or image [20] to enrich the training data. Compared with strong annotations, i.e., pixel-wise segmentations, weak annotations for bounding boxes and images cost less time. Therefore, they generally assume that there are a large number of weak annotations available for training, while the amount of training images with strong annotations are limited. In this setting, weakly annotated samples are used to update the supervised deep model by iteratively inferring and refining hypothetical segmentation labels.

A framework of semi-supervised feature selection has been introduced in [21], both labeled and unlabeled training data are exploited to analyse the feature space. The researches in [22], [23], [24] explore the idea of assigning virtual labels to the generated samples in the setting of semi-supervised learning. Salimans et al. [22] and Odena et al. [23] proposed an all-in-one method which simply take all the generated images as a new class. In practice, N defines the number of classes in the real training sets, then $N+1$ is assigned to each generated sample. However, the generated samples tend to belong to the classes in N rather than the $N+1$ class due to the fact that they are generated from distribution of the real samples. Without using an extra class, the method of assigning virtual label to generated samples has been proposed in [24], which exploits the maximum predicted probability generated for unlabeled image. After fedding an unlabeled sample into network, it will be fitted to a certain pre-defined class after several training epochs. A virtual label smoothing regularization for outliers (LSRO) was introduced by Zheng et al. [25] to address the over-fitting problem in [24]. LSRO assigns a uniform label distribution on generated samples to regularize the training process of deep network.

2.2. Re-ranking for Person re-ID

Recently, several re-ranking methods are proposed to improve the performance of person re-ID by optimizing the original ranking list [26], [27]. In [28], a re-ranking model is developed by analyzing the correlation of nearest neighbors of each pair images. Garcia et al. [29] introduced a re-ranking method for

person re-id, in which the content and content information are both considered to remove ambiguous samples. A bidirectional ranking method has been proposed in [30], which joins the contextual similarities with content similarities to revise the initial ranking list.

Some researchers have exploited the nearest neighbors of the multiple baseline methods to the re-ranking task [31], [32]. In [31], the common nearest neighbors of local and global features are combined as new queries, then aggregate the global and local feature to optimize the initial ranking list. Ye et al. [32] calculated both the similarity and dissimilarity of the k-nearest neighbor set from different baseline methods to optimize the initial ranking list. These re-ranking methods have made contributions to discover the potential information from the k-nearest neighbors.

However, the overall performance from the above works may be restricted if the k-nearest neighbors are used to achieve the task of re-ranking directly, because false matches are often included. In the literature, the k-reciprocal nearest neighbor [33], [34] is effective to increase the amount of true matches on the top-k images. We regard the two images as k-reciprocal nearest neighbors [34] if they are both ranked between top-k in the ranking list when the other image is used as the probe. In this paper, we propose an effective re-ranking method for vehicle re-ID and study the importance of the k-reciprocal neighbors.

2.3. Vehicle re-ID

In recent years, the researches on various computer vision tasks have achieved significant progresses, including object matching [35], [36], traffic scene recognition [37], action recognition [38], [39] and vehicle related works [18], [40]. Several researchers have proposed to apply the visual characteristics and the semantic attributes for vehicle retrieval. A vehicle retrieval and detection system was presented in [41], in which the task of attribute recognition and vehicle retrieval were both achieved. Liu et al. [14] exploited the real-world spatial-temporal environment to achieve a content assisted search for vehicle re-ID. There are some works focused on applying the linear discriminant analysis (LDA) [42], [43] to

optimize distance metrics in re-ID tasks. LDA learns a transformation matrix for feature space from high-dimensional to low-dimensional while preserving the class discrimination information as much as possible [44]. In [45], Local Fisher Discriminant Analysis was employed to learn a distance metic. Wu et al. [46] approximated the variations of intra-class and inter-class by training a hybrid deep architecture with an LDA criterion.

Additionally, hybrid features have been proposed to enhance the recognition of vehicle characteristics in some published works. For example, Cormier et al. [18] proposed a mixed descriptor for low resolution vehicle re-ID, in which the local variance and local binary patterns (LBP) were combined. Liu et al. [40] presented a vehicle re-ID method that incorporated the feature of metric learning and vehicle model into one network. Despite these progresses on vehicle re-ID, how to exploit unlabeled samples and the re-ranking algorithm have not been well investigated in detail, which can significantly influence vehicle recognition performance. In this work, we propose to use GAN generated samples and re-ranking to boost the vehicle re-ID performance of off-the-shelf CNN.

3. Proposed Approach

3.1. Generative Adversarial Networks

A generator and a discriminator are two sub-networks in the generative adversarial network (GAN)[11]. A generator produces a model distribution by transforming a random noise seed. A discriminator then tries to distinguish between samples between that model distribution and the target distribution. The training process of adversarial can be regarded as a minimax game: both the generator and discriminator oppose each other's objective and minimize its own cost, which leads a converged status that minimize the distance between the distribution of real samples and generated samples.

We use the basic framework of DGCAN [10] in our research. Many other variants of GAN have been proposed, such as conditional GAN [47] and stackedGAN [48]. While most of previous researches are focus on studying the methods of

generating more complex sample by training with high-quality images of objects, our aim is to modify the basic GAN model [10] and use it to generate unlabeled samples from the low-quality surveillance image of vehicles, thus helping improve the discriminative learning.

Five deconvolution functions are used to expand the tensor, which is defined as a data container with an N-dimensional array. The stride of the deconvolution filters 2 and their size is 5×5. Following with a tanh activation function, we add one deconvolutional layer with a stride of 1 and kernel size 5×5 to fine-tune the result. An image can then be drawn from the generator net after training. We combine the original training set with the generated images and then fed them into the discriminator network. Five convolutional layers with a stride of 2 and kernel size 5×5 are used to identify whether the generated images are fake.

3.2. Label Smoothing Regularization for Outliers

Our model computes the probability of each class $n \in \{1, 2, ..., N\}$: $p(n|x) = \frac{exp(z_n)}{\sum_{n=1}^{N} exp(z_n)}$ for each training image x. Here, N is the number of pre-defined classes in the training set and z_n represents the logits or unnormalized log-probabilities. We normalize the ground-truth distribution over labels $q(n|x)$ for image x so that $\sum_n q(n|x)$=1. We define the he cross-entropy loss as Eq.1, which omits the dependence of p and q on example x.

$$l = -\sum_{n=1}^{N} log(p(n))q(n) \tag{1}$$

Minimizing the cross-entropy loss is equal to maximize the expected log-likelihood of a label, which is selected according to its ground-truth distribution $q(n)$. Cross-entropy loss is widely appied for gradient training of deep models. The gradient can formulated as $\frac{\partial l}{\partial z_n} = p(n) - q(n)$, the bounded range for it defined as $[-1, 1]$. Suppose there exsits a single ground truth label y, we can express the $q(n)$ as:

$$q(n) = \begin{cases} 0 & n \neq y \\ 1 & n = y \end{cases} \tag{2}$$

In this case, the objective of minimizing the cross-entropy loss is equal to maximize the predicted probability of the expected log-likelihood of the ground truth label. For a particular image x with ground truth y, the log-likelihood is maximized for $q(n)$, which equals to 1 for $n = y$. This maximum is not achievable for finite z_n but is approached if $z_y > z_n$ for all $n \neq y$, which means the logit of ground-truth label is larger than other logits. However, two problems can be caused. First, it may result in overfitting: the generalize can not be guaranteed if the model assigns full probability to the ground-truth label for each training example. Second, the model is overconfident about its predictions, resulting in a larger difference between the maximum logit and all other logits.

To address the second problem, the label smoothing regularization (LSR) has been introduced in [49] to encourage the model to be less confident. While it not consistent with the goal of maximizing training tags, it does regularize the model and make it more adaptable. In [49], the label distribution $q_{LSR}(n)$ is written as:

$$q_{LSR}(n) = \begin{cases} \frac{\epsilon}{N} & n \neq y \\ 1 - \epsilon + \frac{\epsilon}{N} & n = y \end{cases} \tag{3}$$

where $\epsilon \in [0, 1]$ is a smoothing parameter. If set ϵ to zero, Eq.3 will reduce to Eq.2. On the contrary, the model may not be able to predict ground truth label if ϵ is too large. Therefore, the value of ϵ equals 0.1 in most cases. The cross-entropy loss evolves to Eq.4 by considering Eq.1 and Eq.3:

$$l_{LSR} = -(1 - \epsilon) log(p(y)) - \frac{\epsilon}{N} \sum_{n=1}^{N} log(p(n)) \tag{4}$$

In order to use the generated images in the process of deep feature learning, Zheng et al. [49] propose the label smoothing regularization for outliers (LSRO) method, which extends LSR [49] from the fully-supervised learning to the semi-supervised learning. It assumes the generated samples do not belong to any pre-defined class and sets the virtual label distribution to be uniform over all classes. Therefore, the maximum probability that is produced for the generated samples will be very low, which makes the network cannot make prediction for them. So the class label distribution for the unlabeled samples $q_{LSRO}(n)$ is

defined as:

$$q_{LSRO}(n) = \frac{1}{N} \tag{5}$$

We combine Eq. 1, Eq. 2 and Eq. 5 to re-write the cross-entropy loss:

$$l_{new} = -(1 - Z)log(p(y)) - \frac{Z}{N}\sum_{n=1}^{N} log(p(n)) \tag{6}$$

For a real training image, $Z = 0$. For a generated training image, $Z = 1$. Therefore, the loss for the real images and generated images are different in the system. During the training process, we define the loss of LSRO on a generated sample as follows:

$$l_{LSRO} = \frac{1}{N}\sum_{n=1}^{N} log(p(n)) \tag{7}$$

With the help of LSRO, we can regularize the model by processing more training images (outliers) that are located near the real training images in the sample space, which introduces more variances such as lighting and color. For example, if only one black-color vehicle exists in the training set, the discriminative power of the model will be limited because the model may be misled and regarded the black-color as discriminative feature. By adding generated images, such unlabeled black-color vehicle, the classifier will be punished if it misjudges the labeled black-color vehicle. In this manner, the network will be encouraged to look for more underlying causes and to be less prone to over-fitting.

3.3. Re-ranking Method

Problem Definition. Given a gallery set $G = \{g_i | i = 1, ..., T\}$ and a probe vehicle image b, where i defines the index of each image and T is the size of the gallery. After comparing the Euclidean distance between probe b and each image in gallery g_i, we reorder the indices of images in G so that $\{g_1, g_2, ..., g_T\}$ correspond to $L(b, G)$. The similarities between b and g_i satisfy $S(b, g_1) > S(b, g_2) > S(b, g_3) > \cdots > S(b, g_T)$. The objective of re-ranking method is to make more true matches rank top in the ranking list, thus improve the performance of the vehicle re-ID.

K-reciprocal Nearest Neighbors. Following [34], we define the k-nearest neighbors as the top-k samples of the ranking list of a probe b, it can be expressed as $R(b,k)$:

$$R(b,k) = \{g_1, g_2, ..., g_k\} \tag{8}$$

A potential assumption is that the returned image can be used for the subsequent re-ranking when it ranks within the k-nearest neighbors of the probe. However, some traditional methods which directly using the top-k images in the ranking list to perform re-ranking may introduce noise into the system and affect the final result. Therefore, we apply the k-reciprocal nearest neighbor $R^*(b,k)$ [33], [34] to solve this problem. It can be defined as:

$$R^*(b,k) = \{g_i | (g_i \in N(b,k)) \wedge (b \in N(g_i,k))\} \tag{9}$$

Rank Aggregation. Compared with the k-nearest neighbors, the k-reciprocal nearest neighbors are more relevant to probe b. However, the true matches may not appear in the $R^*(b,k)$ due to the variations in occlusions, illuminations, poses and views. To solve this problem, for each sample q in $R^*(b,k)$, we add the half of the samples in its k-reciprocal nearest neighbors set into another set $R_{new}(b,k)$ as the following step:

$$R_{new}(b,k) \leftarrow R^*(b,k) \cup R^*(q, \frac{1}{2}k) \tag{10}$$

Therefore, $R_{new}(b,k)$ includes more images that are more relevant to the samples in $R^*(b,k)$. Then we consider the $R_{new}(b,k)$ as contextual knowledge and re-calculate the distance between the deep features of the probe and the images in gallery set. As described in [32], the similarity of two images is higher if more duplicate samples in their k-reciprocal nearest neighbor sets. We calculate the new distance between the k-reciprocal sets of b and gallery g_i according to the Jaccard metric:

$$d_j(b,g_i) = 1 - \frac{|R_{new}(b,k) \cap R_{new}(g_i,k)|}{|R_{new}(b,k) \cup R_{new}(g_i,k)|} \tag{11}$$

Algorithm 1 Rank Aggregation Algorithm

Input: A probe image b and a gallery set $G = \{g_i | i = 1, ..., T\}$

Output: A rank list for the probe image

Offline:

Compute the pairwise Euclidean distance between the probe vehicle b and images in gallery set.

Reorder the indices of images in G by sorting the pairwise Euclidean distance.

Correspond the set $\{g_1, ...g_T\}$ to the initial ranking list $L(b, G)$, and obtain the top-k galleries $R(b, k)$ from $L(b, G)$ of the probe image.

Query each image g_i in the gallery G.

Obtain the top-k galleries $R(g_i, k)$ of each image g_i.

Online:

for $i = 1$ to $|L(b, G)|$ **do**

  g_i is the i-th item in $L(b, G)$

  Get the k-reciprocal nearest neighbors of probe b by Eq. (9)

  Add more positive samples into $R_{new}(b, k)$ by Eq. (10)

end for

for $i = 1$ to $|R^*(b, k)|$ **do**

  g_i is the i-th item in $R(b, k)$

  Compute the new distance $d_j(b, g_i)$ between b and g_i by the Jaccard metric of their k-reciprocal sets as Eq.(11)

  Compute the final distance d_f between b and g_i as Eq.(12)

end for

Use the final distance to obtain the new rank list revised ranking list $L_{new}(b, G)$

.

Inspired by [50], the original distance and the Jaccard distance are aggregated to emphasize the importance of the original distance and improve the initial ranking list. We define the final distance d_f as:

$$d_f(b, g_i) = (1 - \lambda)d_j(b, g_i) + \lambda d(b, g_i) \tag{12}$$

where λ represents the weight of original distance in the final distance, and d represents the Euclidean distance. Finally, we obtain the new ranking list for probe b $L_{new}(b, G)$ by sorting the final distance d_f. We denote the size of $R_{new}(b, k)$ and $R(b, k)$ as $k1$ and $k2$, respectively. Our rank aggregation algorithm is summarized in Algorithm 1.

3.4. Complexity Analysis

In the proposed re-ranking method, calculating the pairwise distance of all image pairs requires a large amount of computational cost. We define the gallery size as t, $O(t^2)$ and $O(t^2 \, log \, t)$ represent the computation complexity of distance measure and the ranking process, respectively. Since the work of calculating the pairwise distance and obtaining the initial ranking list for the probe can be done in advance offline, the computation costs will be reduced in practical applications. Therefore, the computation costs include only $O(t)$ and $O(t \, log \, t)$, the former representing the calculation of pairwise distance between probe and gallery, the latter representing the complexity of ranking all final distances.

4. Experiments Results and Discussion

4.1. Datasets Introduction

Extensive experiments are conducted on three vehicle re-ID benchmark datasets: VeRi-776 [14], VehicleID [15] and VehicleReID dataset [16].

VeRi-776 [14] consists of 50,000 labeled images of 776 vehicles which collected by 20 cameras in a road network in 24 hours. The specific information of vehicles are also provided, such as car model, camera locations and license plates. The dataset has been divided into two parts, a training set and a testing

set. The training set contains 37,778 images of 576 vehicles, and the testing set consists of 9,919 images belong to 200 vehicles. For the vehicle re-ID task, the 1,678 probe vehicle images in testing set are selected randomly to search the other images in testing set.

VehicleID [15] is currently the largest publicly available vehicle re-ID dataset. It contains 222,628 images belong to 26,328 vehicles collected from the traffic surveillance system. There are two parts in the dataset: a training set and a testing set. The training set contains 113,346 images belong to 13,164 vehicles and the testing set contains 109,282 images captured from 13,164 vehicles. The testing data provides three subsets including small, medium and large scale for the vehicle re-ID task.

VehicleReID [16] contains 1,232 vehicle image pairs obtained from two surveillance camera. The appearance of the same vehicle is changed by variations of viewpoints, illuminations and the locations of cameras.There are 553 vehicles from camera view A and 530 from camera view B, with 423 common vehicles in both views.

4.2. Implementation Details

CNN Baseline. The ResNet-50 [13] model which pre-trained on the ImageNet dataset is slightly improved and used in our experiments as the basic CNN network. We fine-tune the model using the training set to classify the training identities. ResNet-50 is a state-of-the-art architecture that exhibits top performance in several tasks in the field of computer vision, such as face identification, object classification and action recognition. It is composed of multiple basic blocks that are serially connected to each other and introduces shortcut connections summed after every few layers, so as to represent residual functions. In such way, it allows for a very deep architecture without hindering the learning process and at the same time shows less complexity in comparison to other networks of even smaller depth. Although there exists deeper versions of ResNet, we choose the 50-layer variant as the baseline model, as computation time is still crucial for this task.

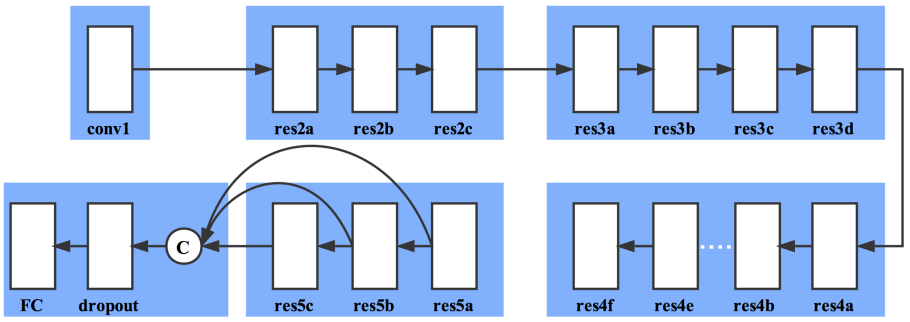

Figure 4: The structure of the improved ResNet-50 model. There are two modifications in the ResNet-50 model structure,1) we add a dropout layer before FC layer to reduce the possibilities of overfitting; 2) we concat the $5a$, $5b$ and $5c$ convolutional layers to obtain mid-level identity-sensitive information.

We use the Matconvnet [51] package to implement the network training and resize all the images to 256 $\times$ 256. During training, random horizontal flipping is applied to crop the images to 224 $\times$ 224 randomly. A dropout layer has been inserted before the final convolutional layer to reduce the possibility of overfitting. Assume the original training set has K vehicle identities, we add K neurons in the last fully-connected layer to predict the K-classes. In most existing deep re-ID models, the final convolution layer will compute the feature vector. Inspired by [52], which demonstrates that the useful information of mid-level identity-sensitive can be obtained before the last fully-connected layer in a DNN, as shown in Fig.4, we thus concat the $5a$, $5b$ and $5c$ convolutional layers of the ResNet-50 structures into a 2048-dim feature vector after the last fully-connected layer.

The GAN model. We used the Tensorflow [53] and DCGAN package to train the GAN model. Before training, we resize all the images in the training set to 128$\times$128 and perform randomly flipped on them. The model is trained with mini-batch stochastic gradient descent (SGD) with a mini-batch size of 64. We use a zero-centered Normal distribution to initialize the weights and set the

Figure 5: Examples of original images in training set and images generated by GAN from (a) VeRi-776 [14], (b) VehicleID [15], (c) VehicleReID [16].

standard deviation as 0.02. We apply the Adam stochastic optimization with parameters $\beta 1$ and $\beta 2$ which are used to define a memory for Adam, and average the gradient and squared gradient, respectively. Following the practice in [54], the good default settings for the tested machine learning problems are $\beta 1 = 0.9$, $\beta 2 = 0.99$. During testing, we fed a 50-*dim* random vector with Gaussian noise distribution into the GAN to generate vehicle images. Finally, all the generated samples are resized to 256×256 and are used in training the CNN with the LSRO. Fig.5 illustrates the generated and real samples on these three datasets. Although human can easily recognize the generated samples as fake, they are still effective in promoting the performance by adding the LSRO as virtual labels in our experiment.

Evaluation Metrics. We use Mean Average Precision (mAP) and Cumulative Match Curve (CMC) to measure the re-ID quantitatively.

Mean Average Precision: The mAP metric evaluates the overall per-

formance of re-ID. For each probe image b, average precision is calculated as follows:

$$\rho = \frac{\sum_{k=1}^{n} P(k) \times rel(k)}{N_{gt}} \tag{13}$$

where k defines the rank in the list of retrieved vehicles, n denotes the number of retrieved vehicles, N_{gt} is the number of ground truth retrievals for the probe. $P(k)$ denotes the precision at cut-off k, $rel(k)$ indicates whether the k-th recall image is right match or not. So we define the mAP as follows:

$$mAP = \frac{\sum_{b=1}^{Q} \rho}{Q} \tag{14}$$

where Q denotes the number of probe images.

Cumulative Match Characteristics: The CMC curve describes the expectation of positive samples within the first k ranks, we calculate the CMC value for top k ranks as follows:

$$CMC@k = \frac{\sum_{i=1}^{Q} f(b_i, k)}{Q} \tag{15}$$

where b_i is i-th probe vehicle, $f(b_i, k)$ is an indicator function which equals to 1 when the positive samples are within the top k ranks, otherwise, it equals to 0.

4.3. Semi-supervised Learning Results

Performance Comparisons on VeRi-776 Dataset. The proposed method was evaluated on the VeRi-776 dataset [14] firstly which is the only existing vehicle re-ID dataset providing spatial and temporal annotations. We used the previously explained semi-supervised learning of the CNN model, and applied the re-ranking for the final identification. The Cumulative Match Curve (CMC) metric and mean Average Precision (mAP) are adopted for the evaluation. We describe the details of experiment procedure and three comparative settings as follows:

(1) The CNN baseline.

Following the procedure of training and testing described in Section 4.2, the final results of the VeRi-776 dataset are reported in Table 1, Table 2

Table 1: Match rate (CMC@Rank-R, %) and mAP (%) under different dropout rate on the VeRi-776 dataset [14]

Methods	Rank-1	Rank-5	Mean AP
CNN baseline (Without dropout layer)	82.54	90.52	48.90
CNN baseline (Dropout rate 0.5)	84.74	92.49	54.36
CNN baseline (Dropout rate 0.6)	86.23	92.37	53.95
CNN baseline (Dropout rate 0.7)	85.52	92.13	53.17
CNN baseline (Dropout rate 0.8)	85.76	92.67	54.47
CNN baseline (Dropout rate 0.9)	**85.88**	**92.85**	**54.59**

and Table 3. To evaluate the stand-alone performance of ResNet-50, we extracted the CNN feature from the first fully connected layers (FC6) for each vehicle image and directly apply it for vehicle re-ID as a comparative baseline. As shown in Table 1, the CNN model with dropout layer gains about 5.46 points increase in mAP, from 48.90% to 54.36%. To select the best dropout rate, the extensive comparative experiments were further performed. As can be seen in Table 1, the best performance was achieved when the dropout rate is 0.9. Therefore, in our implementation, the final result of CNN baseline has a Rank-1 match rate of 85.88% and 54.59% mAP. We also compared the result of CNN baseline with other published vehicle re-ID results, from Table 2, the CNN baseline achieves better performance than previous works [14], [17]. There are no unlabeled samples in this scenario and the re-ranking methods have not been taken into account. We report the results of semi-supervised learning with different numbers of generated images in Table 3. The performance of vehicle re-ID has been improved when we fed different numbers of unlabeled data into the process of CNN training, which implies that CNN features alone are insufficient compared with semi-supervised learning.

(2) Semi-supervised learning with different numbers of generated images.

Table 2: Match rate (CMC@Rank-R, %) and mAP (%) for different methods on the VeRi-776 dataset [14]

Methods	Rank-1	Rank-5	Mean AP
FACT [17]	50.95	73.48	18.49
FACT+Plate-SNN+STR [14]	61.44	78.78	27.77
Siamese-CNN+Path-LSTM [55]	83.49	90.04	58.27
VGG+C+T+S [56]	86.59	92.85	57.40
CNN Baseline (Ours)	85.88	92.85	54.59
SSL (Ours)	88.57	93.56	61.07
SSL+re-ranking (Ours)	**89.69**	**95.41**	**69.90**

We trained DCGAN on the VeRi-776 training set, and combined the original training set with the generated images to fine-tune the CNN model. We evaluated the effect of the number of generated images on re-ID performance. Since unlabeled data is easy to obtain, we hope that as the number of unlabeled images increases, the model will obtain more general information. We compare the number of real training images (37,778) with the number of generated images fed into network, then two conclusions are obtained after analyzing the results in Table 3. First, the baseline has been consistently improved by adding different numbers of generated images. Adding approximately 2 times generated images (70,000) that of the real training set still obtain +1.44 points improvement to rank-1 match rate. Second, the peak performance is achieved when 0.3 times generated images (10,000) that of the real training set are added. From Table 3, when 10,000 generated images are added to the semi-supervised learning, the re-ID performance on VeRi-776 has been significantly improved. We observed the improvement of 3.09 points (from 85.88% to 88.97%), 0.71 points (from 92.85% to 93.56%) and 6.48 points (from 54.59% to 61.07%) in the Rank-1, Rank-5 match rates and mAP, respectively. Too many or too few images generated images incorporated into the semi-supervised learning will pro-

Table 3: Match rate (CMC@Rank-R, %) and mAP (%) after using different numbers of generated images on the VeRi-776 dataset [14]

The number of generated images	Rank-1	Rank-5	Mean AP
0 (basel.)	85.88	92.85	54.59
2,000	86.12	92.96	55.43
5,000	86.78	93.21	57.68
8,000	88.31	93.35	59.34
10,000	**88.97**	**93.56**	**61.07**
30,000	88.19	93.54	59.00
50,000	87.90	92.90	59.10
70,000	87.34	92.61	58.87

duce negative impacts on the model.

In semi-supervised learning with LSRO, generated images are used to learn more discriminative features and reduce the possible of over-fitting by assigning a uniform label distribution to the generated images to regularize the CNN model. When we incorporate too few GAN samples, the regularization ability of the LSRO is inadequate. In contrast, if we add too many GAN samples to fine-tune the network, the CNN model will tend to converge towards assigning a uniform label distribution to all the training images, which lead to overfitting and affect the discriminative learning from real images. Therefore, we recommend to make a trade-off of GAN samples to avoid poor regularization and overfitting.

(3) Ranking Optimization with different metrics. We set the parameter $k1$ =50, $k2 = 10$, and λ=0.3 which have the best performance in the test. After adding the step of re-ranking, the Rank-1, Rank-5 match rates and mAP are further improved to 89.69%, 95.41% and 69.90%. Table 2 compares the performance of our best approach and semi-supervised learning with re-ranking, against other state-of-the-art methods.

We compare our results with the methods in [14],[17], in which the hand

Table 4: Match rate (CMC@Rank-R, %) and mAP (%) for the compared methods on the VeRi-776 dataset [14]

Methods	Rank-1	Rank-5	Mean AP
SSL+KISSME	86.84	92.37	60.12
SSL+KISSME+ re-ranking	**88.66**	**94.62**	**64.71**
SSL+XQDA	87.49	93.80	60.11
SSL+XQDA+ re-ranking	**88.72**	**94.92**	**67.48**

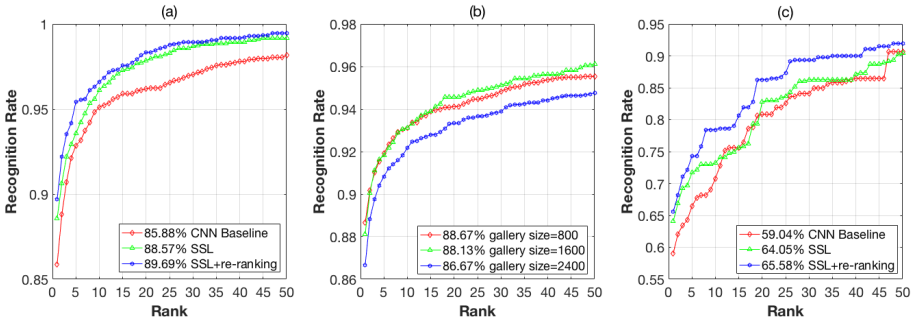

Figure 6: The CMC curves of the proposed methods on VeRi-776 (a), VehicleID (b), VehicleReID (c). The recognition rate shows the percentage of the probes that are correctly recognized within the top k matches in the gallery. The numbers in the legend of curves are the top 1 value of CMC.

crafted features were adopted for vehicle re-ID. It can be observed that our method achieves significant improvement over them, proving the advantage of deep feature. In [14], the license plate information (Plate-SNN) and spatio-temporal information were additionally used to improve the performance of vehicle re-ID. Compared with [14], our method based on vehicle appearance further yields an improvement of 28.25 points (from 61.44% to 89.69%) in Rank-1, 16.63 points (78.78% to 95.41%) in Rank-5 and 42.13 points (from 27.77% to 69.90%) in mAP. We also compare our method with the appearance-based deep learning approach [56], which improved triplet-wise training of CNN for vehicle re-ID. As shown in Table 2, the proposed

method with both semi-supervised learning and re-ranking leads to significant improvements compared with the best method (VGG+C+T+S) in [56]. The CMC curves of the proposed methods are shown in Fig.6 (a) Moreover, experiments conducted with two popular metric learning methods, KISSME [57] and Cross-view Quadratic Discriminant Analysis (XQDA) [58] verify the effectiveness of our ranking optimization method on different distance metrics as shown in Table 4. In [57], the Mahalanobis distance is learned by considering the log likelihood ratio test of two Gaussian distributions. Based on the idea of KISSME [57], the XQDA further learns a discriminant subspaces with more efficient metrics.

Performance Comparisons on VehicleID Dataset. We provide our results from the largest vehicle re-ID dataset [15] in Table 5 to further demonstrate the effectiveness of the proposed method. Following the dataset setting in [15], we randomly select one image from each vehicle and put it into gallery set, then the remaining images are all used as probe images. The details of the three testing subsets are listed in Table 6. We preform the testing process with different values of $k1$, $k2$ and λ, and obtain the best performance when $k1$=10, $k2$=6 and λ=0.3. The evaluation procedure was repeated for 10 times to evaluate model prediction accuracy and obtain the final CMC curve.

The detailed match rates from Rank-1 to Rank-50 of the proposed methods evaluated on the three scale test subset are presented in Fig.6 (b). For VehicleID dataset, we fine-tuned the improved ResNet-50 model by using the combination training set of original training set and 40,000 generated images. The vehicle re-ID results of our proposed method on three scale test subsets are shown in Table 5. Compared with the best state of the art method [56], the proposed method improves the Rank-1 and Rank-5 match rates for large subset by 2.44 points (from 84.23% to 86.67%) and 2.16 points (from 88.67% to 90.83%), respectively, which proves once again that our method has significant advantages. Four examples are shown in Fig.7. The proposed method, semi-supervised learning+re-ranking, effectively ranks more positive samples at the

Table 5: Match rate (CMC@Rank-R, %) and mAP (%) of the comparison methods on the VehicleID dataset [15]

Methods		Small	Medium	Large
VGG+Triplet Loss [59]	Rank-1	40.40	35.40	31.90
VGG+CCL [15]		43.60	37.00	32.90
Mixed Diff+CCL [15]		49.00	42.80	38.20
VGG+C+T+S [56]		69.90	66.20	63.20
Baseline (Ours)		81.93	81.44	81.37
GAN+LSRO (Ours)		85.72	85.12	84.23
GAN+LSRO+ re-ranking (Ours)		**88.67**	**88.13**	**86.67**
VGG+Triplet Loss [59]	Rank-5	61.70	54.60	50.30
VGG+CCL [15]		64.20	57.10	53.30
Mixed Diff+CCL [15]		73.50	66.80	61.60
VGG+C+T+S [56]		87.30	82.30	79.40
CNN Baseline (Ours)		86.93	86.44	86.67
SSL (Ours)		89.12	88.12	88.67
SSL+re-ranking (Ours)		**91.92**	**91.81**	**90.83**
CNN Baseline (Ours)	mAP	70.13	66.67	65.47
SSL (Ours)		74.13	69.84	68.74
SSL+re-ranking (Ours)		**76.42**	**71.39**	**70.59**

Table 6: The three subset of testing set for the VehicleID Dataset [15]

Number of images	Small	Medium	Large
Gallery size	6,493	11,777	17,377
Probe size	800	1,600	2,400

top of the ranking list which are not included in the ranking list of our baseline.

Performance Comparisons on the VehicleReID Dataset. Furthermore, we study the effectiveness of our method on the VehicleReID dataset by using the single shot setting. There are 423 vehicles from both camera view A and camera view

Figure 7: Four examples of vehicle re-ID results (Rank-5) on the VehicleID dataset. For each probe, the ranking results produced by our baseline are presented in the first row, the second row corresponds to our proposed method (Semi-supervised learning+re-ranking) which improves the baseline ranking results. The green box indicates a true matches, the red box identifies the false matches.

B, for solving the vehicle re-ID task, we chosen this subset from the original sets. We randomly split the vehicles in both camera A and camera B into two almost equal subsets, where 211 vehicles for training and 212 vehicles for testing. Among the 212 vehicles for testing, we treat the images from camera A as the probe set and use the images from camera B as the gallery set. During the testing process, we search the 212 test vehicles in all vehicles from camera B.

We followed the semi-supervised learning method to fine-tune the CNN model as previously explained, and applied the ranking optimization algorithm for the final prediction. Specifically, the DCGAN was trained to generate unlabeled vehicle images, then we combined the generated images with original training set to fine-tune the improved ResNet-50 model. The ranking optimization was accomplished after the initial list generated by the Euclidean distance. We set the appropriate value to $k1$=6, $k2$=3 and λ=0.8. The testing phase is repeated for 10 times with the average results reported in Table 7. Our semi-supervised learning method gains 5.01 points improvement in Rank-1 match rate and significant 4.11 points improvement in mAP for CNN baseline. After applying the re-ranking algorithm, our method further gains an improvement of 1.53 points in Rank-1 match rate and 3.48 points in mAP. Experimental results

Table 7: Match rate (CMC@Rank-R, %) and mAP (%) for the compared methods on the VehicleReID dataset [16]

Methods	Rank-1	Rank-5	Mean AP
CNN Baseline (Ours)	59.04	66.45	62.53
SSL (Ours)	64.05	72.56	66.64
SSL + re-ranking (Ours)	**65.58**	**74.29**	**70.12**

demonstrate that our method is also effective on the re-ID problem of single-shot setting. Fig.6 (c) shows the CMC curve on the VehicleReID dataset.

5. Further Evaluation

5.1. The impact of the scale of random vector fed to the GAN.

The generator, G, used in GAN input a random noise vector z which passed through each layer in the network and generates a fake sample $G(z)$ from the final layer. We evaluate whether the scale of the random vector z fed to the GAN impacts the performance of vehicle re-ID. To investigate the effect, we tried three different ranges of the random vector, i.e., [-0.5,0.5], [-1,1], and [-1.5,1.5], with a normal distribution. The results of vehicle re-ID on the VeRi-776 dataset are presented in Table 8. We find that the [-0.5,0.5] yields higher re-ID performance than the other two ranges. The visual examples are shown in Fig.8. We find that visual examples of [-1.5, 1.5] show obvious differences among the three ranges, with some strange shapes of vehicles. Typically, a larger range may contain some strange variations and affect the quality of generated images.

5.2. Analysis of the parameters of ranking optimization method

The parameters of ranking optimization method are evaluated in this subsection. We observe the influence of $k1$, $k2$ and λ on the VeRi-776 dataset. Fig.9 (a)(b) show the impact of the size of k-reciprocal neighbors set on Rank-1 match rate and mAP. As $k1$ grows, the Rank-1 match rate first increases with fluctuations, and then starts a slow decrease after $k1$ passes the optimal point

Figure 8: The GAN generated images with different scales of the random vector, i.e. [-0.5,0.5], [-1.0,1.0], [-1.5,1.5]. We hardly find any significant visual differences between them.

Table 8: Match rate (CMC@Rank-R, %) and mAP (%) after using the GAN generated images with different scales of the random vector on the VeRi-776 dataset [14]

Random Range	Rank-1	Rank-5	Mean AP
[-0.5,0.5]	**89.65**	**95.41**	**68.97**
[-1,1]	89.46	95.12	68.46
[-1.5,1.5]	89.13	94.97	68.40

at around 50. Similarly, the mAP increases with the growth of $k1$, and it starts to slowly decline after $k1$ passes the optimal point. If $k1$ is too large, more false matches will be included in the k-reciprocal set and cause performance degradation.

The impact of $k2$ is shown in Fig.9 (c)(d). Obviously, the performance will increase as $k2$ grows within a reasonable range (e.g, smaller than 10). However, the performance declines when the value of $k2$ is too large due to the set includes more false matches. In fact, it is very important to set an appropriate value to $k2$ and thus further enhance the performance.

Fig.9 (e)(f) show the impact of the parameter λ. The Jaccard distance is only considered when λ equals zero, in contrast, the Jaccard distance is left out when λ equals one, and the result is obtained using only the original distance. It can be observed that our method consistently outperforms the CNN baseline when the Jaccard distance is only considered, which indicates that the proposed

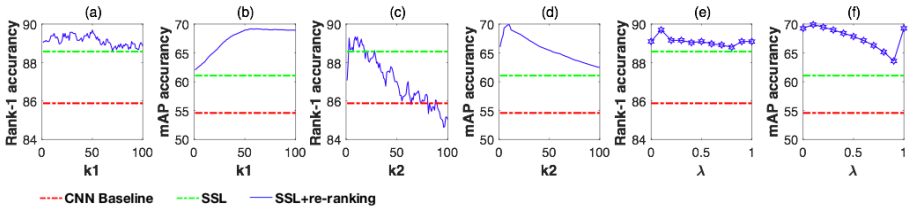

Figure 9: (a)(b): The impact of the parameter $k1$ on the performance of the VeRi-776 dataset. The $k2$ was fixed at 10 and λ set to 0.2; (c)(d): The impact of the parameter $k2$ on the performance of VeRi-776 dataset. The $k1$ was fixed at 50 and λ set to 0.2; (e)(f): The impact of the parameter λ on the performance of VeRi-776 dataset. The $k1$ was fixed at 50 and $k2$ at 10.

Jaccard distance is effective for re-ranking. Moreover, the performance is further improved when we consider the importance of original distance and set the value of λ arounds 0.2.

6. Conclusion

In this paper, we proposed an effective semi-supervised learning approach augmented with ranking optimization for the vehicle re-ID problem. Specifically, a DCGAN model is exploited to generate the unlabelled images and effectively demonstrate their regularization ability when trained with an improved ResNet-50 baseline model. The unlabeled generated images are used to assist the labeled training images for simultaneous semi-supervised learning. We also addressed the re-ranking task by improving the k-reciprocal Nearest Neighbors method. The final distance based on the aggregation of the original distance and Jaccard distance produces effective improvement of the re-ID performance on VeRi-776, VehicleID and VehicleReID datasets. Our experimental results indicate that the proposed methods significantly outperforms state-of-the-arts methods on the VeRi-776 and VehicleID dataset.

References

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, C. Chen, Data-driven intelligent transportation systems: A survey, IEEE Transactions on Intelligent Transportation Systems 12 (4) (2011) 1624–1639.

[2] W.-H. Lin, D. Tong, Vehicle re-identification with dynamic time windows for vehicle passage time estimation, IEEE Transactions on Intelligent Transportation Systems 12 (4) (2011) 1057–1063.

[3] K. Kwong, R. Kavaler, R. Rajagopal, P. Varaiya, Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors, Transportation Research Part C: Emerging Technologies 17 (6) (2009) 586–606.

[4] S. Ribaric, A. Ariyaeeinia, N. Pavesic, De-identification for privacy protection in multimedia content: A survey, Signal Processing: Image Communication 47 (2016) 131–151.

[5] C. Gou, K. Wang, Y. Yao, Z. Li, Vehicle license plate recognition based on extremal regions and restricted boltzmann machines, IEEE Transactions on Intelligent Transportation Systems 17 (4) (2016) 1096–1107.

[6] Y. Bai, Y. Lou, F. Gao, S. Wang, Y. Wu, L. Duan, Group sensitive triplet embedding for vehicle re-identification, IEEE Transactions on Multimedia.

[7] D. P. Kingma, S. Mohamed, D. J. Rezende, M. Welling, Semi-supervised learning with deep generative models, in: Advances in Neural Information Processing Systems, 2014, pp. 3581–3589.

[8] M. Zhang, J. Tang, X. Zhang, X. Xue, Addressing cold start in recommender systems: A semi-supervised co-training algorithm, in: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, ACM, 2014, pp. 73–82.

[9] G. Doquire, M. Verleysen, A graph laplacian based approach to semi-supervised feature selection for regression problems, Neurocomputing 121 (2013) 5–13.

[10] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.

[12] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, arXiv preprint arXiv:1701.07875.

[13] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[14] X. Liu, W. Liu, T. Mei, H. Ma, A deep learning-based approach to progressive vehicle re-identification for urban surveillance, in: European Conference on Computer Vision, Springer, 2016, pp. 869–884.

[15] H. Liu, Y. Tian, Y. Yang, L. Pang, T. Huang, Deep relative distance learning: Tell the difference between similar vehicles, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[16] D. Zapletal, A. Herout, Vehicle re-identification for automatic video traffic surveillance, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 25–31.

[17] X. Liu, W. Liu, H. Ma, H. Fu, Large-scale vehicle re-identification in urban surveillance videos, in: Multimedia and Expo (ICME), 2016 IEEE International Conference on, IEEE, 2016, pp. 1–6.

[18] M. Cormier, L. W. Sommer, M. Teutsch, Low resolution vehicle re-identification based on appearance features for wide area motion imagery, in: Applications of Computer Vision Workshops (WACVW), 2016 IEEE Winter, IEEE, 2016, pp. 1–7.

[19] G. Papandreou, L.-C. Chen, K. Murphy, A. L. Yuille, Weakly-and semi-supervised learning of a dcnn for semantic image segmentation, arXiv preprint arXiv:1502.02734.

[20] P. O. Pinheiro, R. Collobert, Weakly supervised semantic segmentation with convolutional networks, in: CVPR, Vol. 2, Citeseer, 2015, p. 6.

[21] X. Chang, Y. Yang, Semisupervised feature analysis by mining correlations among multiple tasks, IEEE transactions on neural networks and learning systems 28 (10) (2017) 2294–2305.

[22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: Advances in Neural Information Processing Systems, 2016, pp. 2234–2242.

[23] A. Odena, Semi-supervised learning with generative adversarial networks, arXiv preprint arXiv:1606.01583.

[24] D.-H. Lee, Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on Challenges in Representation Learning, ICML, Vol. 3, 2013, p. 2.

[25] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, arXiv preprint arXiv:1701.07717.

[26] A. J. Ma, P. Li, Query based adaptive re-ranking for person re-identification, in: Asian Conference on Computer Vision, Springer, 2014, pp. 397–412.

[27] N. Martinel, A. Das, C. Micheloni, A. K. Roy-Chowdhury, Temporal model adaptation for person re-identification, in: European Conference on Computer Vision, Springer, 2016, pp. 858–877.

[28] W. Li, Y. Wu, M. Mukunoki, M. Minoh, Common-near-neighbor analysis for person re-identification, in: Image Processing (ICIP), 2012 19th IEEE International Conference on, IEEE, 2012, pp. 1621–1624.

[29] J. Garcia, N. Martinel, C. Micheloni, A. Gardel, Person re-identification ranking optimisation by discriminant context information analysis, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1305–1313.

[30] Q. Leng, R. Hu, C. Liang, Y. Wang, J. Chen, Person re-identification with content and context re-ranking, Multimedia Tools and Applications 74 (17) (2015) 6989–7014.

[31] M. Ye, J. Chen, Q. Leng, C. Liang, Z. Wang, K. Sun, Coupled-view based ranking optimization for person re-identification, in: International Conference on Multimedia Modeling, Springer, 2015, pp. 105–117.

[32] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, R. Hu, Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing, IEEE Transactions on Multimedia 18 (12) (2016) 2553–2566.

[33] H. Jegou, H. Harzallah, C. Schmid, A contextual dissimilarity measure for accurate and efficient image search, in: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE, 2007, pp. 1–8.

[34] D. Qin, S. Gammeter, L. Bossard, T. Quack, L. Van Gool, Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 777–784.

[35] S. F. Tahir, A. Cavallaro, Low-cost multi-camera object matching, in: IEEE International Conference on Acoustics, 2014.

[36] M. Ferecatu, H. Sahbi, Multi-view object matching and tracking using canonical correlation analysis, in: 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, 2009, pp. 2109–2112.

[37] F.-Y. Wu, S.-Y. Yan, J. S. Smith, B.-L. Zhang, Traffic scene recognition based on deep cnn and vlad spatial pyramids, in: Machine Learning and Cybernetics (ICMLC), 2017 International Conference on, Vol. 1, IEEE, 2017, pp. 156–161.

[38] L. Wang, H. Sahbi, Nonlinear cross-view sample enrichment for action recognition, in: European Conference on Computer Vision, Springer, 2014, pp. 47–62.

[39] S. Yan, J. S. Smith, B. Zhang, Action recognition from still images based on deep vlad spatial pyramids, Signal Processing Image Communication 54 (2017) 118–129.

[40] H. Liu, Y. Tian, Y. Yang, L. Pang, T. Huang, Deep relative distance learning: Tell the difference between similar vehicles, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2167–2175.

[41] R. S. Feris, B. Siddiquie, J. Petterson, Y. Zhai, A. Datta, L. M. Brown, S. Pankanti, Large-scale vehicle detection, indexing, and search in urban surveillance videos, IEEE Transactions on Multimedia 14 (1) (2012) 28–42.

[42] C. Zhao, X. Wang, D. Miao, H. Wang, W. Zheng, Y. Xu, D. Zhang, Maximal granularity structure and generalized multi-view discriminant analysis for person re-identification, Pattern Recognition 79 (2018) 79–96.

[43] M. Kan, S. Shan, H. Zhang, S. Lao, X. Chen, Multi-view discriminant analysis, IEEE transactions on pattern analysis and machine intelligence 38 (1) (2016) 188–194.

[44] W. Sui, X. Wu, Y. Feng, Y. Jia, Heterogeneous discriminant analysis for cross-view action recognition, Neurocomputing 191 (2016) 286–295.

[45] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3318–3325.

[46] L. Wu, C. Shen, A. van den Hengel, Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification, Pattern Recognition 65 (2017) 238–250.

[47] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, arXiv preprint.

[48] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, D. Metaxas, Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, in: IEEE Int. Conf. Comput. Vision (ICCV), 2017, pp. 5907–5915.

[49] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[50] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, arXiv preprint arXiv:1701.08398.

[51] A. Vedaldi, K. Lenc, Matconvnet: Convolutional neural networks for matlab, in: Proceedings of the 23rd ACM international conference on Multimedia, ACM, 2015, pp. 689–692.

[52] X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, X. Wang, Hydraplus-net: Attentive deep features for pedestrian analysis, arXiv preprint arXiv:1709.09930.

[53] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning., in: OSDI, Vol. 16, 2016, pp. 265–283.

[54] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

[55] Y. Shen, T. Xiao, H. Li, S. Yi, X. Wang, Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals, arXiv preprint arXiv:1708.03918.

[56] Y. Zhang, D. Liu, Z.-J. Zha, Improving triplet-wise training of convolutional neural network for vehicle re-identification, in: Multimedia and Expo (ICME), 2017 IEEE International Conference on, IEEE, 2017, pp. 1386–1391.

[57] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2288–2295.

[58] S. Liao, Y. Hu, X. Zhu, S. Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.

[59] S. Ding, L. Lin, G. Wang, H. Chao, Deep feature learning with relative distance comparison for person re-identification, Pattern Recognition 48 (10) (2015) 2993–3003.