Joint Semi-supervised Learning and Re-ranking for Vehicle Re-identication

Fangyu Wu*†, Shiyang Yan*†, Jeremy S. Smith† and Bailing Zhang*
†University of Liverpool
*Xi'an Jiaotong-Liverpool University
Department of Computer Science and Software Engineering

***Abstract*—Despite the promising progress made in recent years, vehicle re-identification (re-ID) remains a challenging task due to the complex variations in vehicle appearances from different camera views. For this challenging problem, most existing algorithms have been developed in the fully-supervised setting, requiring access to a large amount of labeled training data. However, it is impractical to expect the availability of large quantities of labeled data because labeling data is very costly. Besides, when considering vehicle re-ID as a retrieval process, re-ranking is a critical step to improve its accuracy. Yet in the vehicle re-ID community, limited effort has been devoted to re-ranking. To address these problems, in this paper, we propose a semi-supervised pipeline based on the CNN and re-ranking strategy for Vehicle re-ID. Specifically, we obtain more training data by adopting the generative adversarial network (GAN) to generate unlabeled samples, then label smoothing regularization for outliers (LSRO) will assign a uniform label distribution to the unlabeled images, which regularizes the supervised model and improves the baseline. To optimize the re-ID results, an improved re-ranking method is exploited to optimize the initial rank list. Experimental results on publically available datasets VeRi-776 and VehicleID demonstrate that the method significantly outperforms the state-of-the-art.**

I. Introduction

Vehicle is an active object class in computer vision, which is associated with many interesting research topics including vehicle classification [1], segmentation [2] and detection [3]. Vehicle re-identification (Re-ID) is a relatively new frontier which is often neglected by researchers. Vehicle Re-ID is a problem of searching a vehicle from the gallery for images that contain the same vehicle in a cross-camera mode. Fig.1 gives a straightforward description. Vehicle re-ID has pervasive applications in video surveillance [4], intelligent transportation [5] and public security [3], which can quickly discover, locate, and track the target vehicles. The technique is particularly relevant when the re-identification needs to build upon visual characteristics rather than number plates information [6].

A typical procedure in vehicle re-identification is composed of the following two steps [7], [8]. Firstly, discriminative and robust appearance features are extracted from vehicle images. Then a learned metric using training data with correct matching pairs is applied to increase discriminative power of the extracted features. The outcome of the matching is the rank list, which ultimately involves assigning ranks or numbers to images being searched after matching. This rank list is further used to assess the exact matches in vehicle re-identification system. Through many re-identification methods perform well,

Fig. 1. Vehicle re-id: matching with manually cropped vehicle.

their rank accuracy is unsatisfactory. As a result, re-ranking is receiving increasing attention and has gained promising results [9], [10], [11]. After an initial ranking list is obtained, a good practice consists of adding a re-ranking step, with the expectation that the relevant images will receive higher ranks. In particular, [9] proposes a re-ranking method with k-reciprocal encoding, which combines the original distance and Jaccard distance.

Recently, Convolutional Neural Networks (CNN) has been applied to the re-ID research, which unify the feature extraction and the distance metric learning processes into one framework [12], [13], [14]. CNN-based method requires a large amount of annotated data to obtain high performances, which is an impediment to vehicle re-ID where the amount of available training data in vehicle re-identification is often insufficient. Data annotation is costly, because one has to draw a vehicle bounding box and assign an ID label to it. To solve this problem, one solution is to apply a semi-supervised learning paradigm to exploit unlabelled data. There are a few existing methods, e.g.,"pseudo label" [15], "All in one" [16] and label smoothing regularization for outliers (LSRO) method [17]. Among these strategies, the LSRO method makes a less stronger assumption (label smoothing) towards the unlabelled images and has a superior performance in dealing with unlabelled images.

Given the above considerations, we propose a semi-supervised pipeline based on the CNN and re-ranking strategy for Vehicle Re-ID. In the literature, vehicle images used in training are only provided by the training sets which without being expanded [7], [8], [14], [19]. To directly utilize abundant unlabeled data with the help of labeled data, we firstly train DCGAN [18] on the original re-ID training set and generate new vehicle images to enrich the training set. As shown in Fig. 2, the pipeline feeds the newly generated samples into CNN model. Then, these unlabeled GAN-generated data are

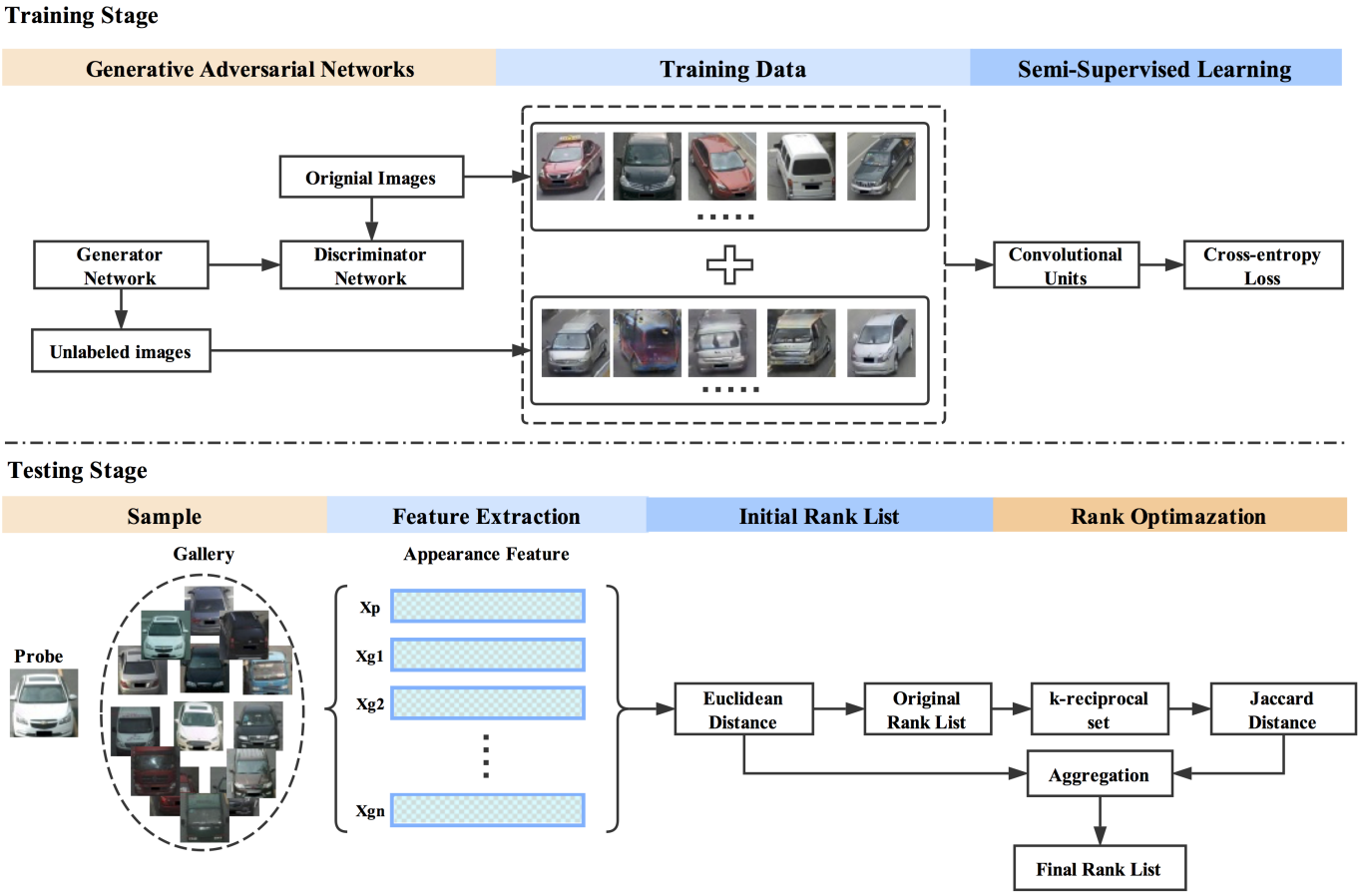

Fig. 2. The workflow of the proposed method. There are two stages including training and testing: (1) Training Stage. A generative adversarial networks[18] for generated new unlabeled samples and a convolutional neural network for semi-supervised learning. "Original Images" represent the labeled data in the given training set; "Training data" combines the "Original Images" and the "Unlabeled Images". We aim to learn more discriminative embeddings with the "Training data"; (2) Testing Stage. Given a probe and gallery set, we used the trained CNN model from training stage to extract the appearance feature for each vehicle. Then the euclidean distance and Jaccard distance are calculated for each pair of the probe vehicle and gallery vehicle. The final distance is computed as the combination of euclidean distance and Jaccard distance, which is used to obtain the proposed ranking list.

fed into the ResNet model [20]. The LSRO method regularizes the learning process by integrating the unlabeled data and, thus, reduces the risk of over-fitting. Finally, after obtaining an initial rank list, we will perform an additional re-ranking step to improve vehicle re-ID accuracy.

To summarize, the contributions of this paper are:

1. We first propose a combination method that explore both of the semi-supervised learning and re-ranking to optimize the performance of vehicle re-ID.

2. We validated our approach on two datasets VeRi-776 [21] and VehicleID [1], and confirmed that the proposed method achieves more favorable performance than state-of-the-art methods.

The rest of this paper is organized as follows: in section II, we provide a detailed description of the proposed methods; implementation details and experimental results are provided in Section III, followed by conclusion in Section IV.

II. The Proposed Approach

In this section, we describe the pipeline of the proposed method, which includes the new samples generated by Generative Adversarial Network (GAN), Label Smoothing Regularization for Outliers (LSRO) and re-ranking method. Fig. 2 describes the details of proposed approach.

A. Generative Adversarial Networks

The generative adversarial network generally consists of two sub-networks: a generator and a discriminator. The adversarial training process is a minimax game: both subnetworks aim to minimize its own cost and maximize the other's cost. This adversarial process leads to a converged status that the generator outputs realistic images and the discriminator extracts deep features. The GAN frameworks we used in this pipeline is DCGAN [18]. We follow the settings in [18] for the generator. We start with a 100-dim random vector and enlarge it to $2\times2\times16$ using a linear function. To enlarge the tensor, five deconvolution functions are used with a kernel size of 5×5 and a stride of 2. Every deconvolution is followed by a rectified linear unit and batch normalization. Additionally, one optional deconvolutional layer with a kernel size of 5×5 and a stride of 1, and one tanh function are added to fine-tune

the result. A sample that is 128×128×3 in size can then be generated.

The input of the discriminator network includes the generated images and the real images in the training set. We use five convolutional layers to classify whether the generated image is fake. Similarly, the size of the convolutional filters is 5×5 and their stride is 2. We add a fully-connected layer to perform the binary classification (real or fake).

B. Label Smoothing Regularization for Outliers

The label smoothing regularization for outliers (LSRO) [17] was proposed to incorporate the unlabeled images in the network. This strategy extends LSR [22] from the supervised domain to leverage unsupervised data generated by the GAN. The basic principles are as follows: let $n \in \{1, 2, ..., N\}$ be the pre-defined classes of the real images in the training data, where N is the number of classes. The cross-entropy loss can be formulated as:

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

where $p(n) \in [0, 1]$ is the predicted probability of the input belonging to class n, and can be outputted by CNN model. It is derived from the softmax function which normalizes the output of the previous fully-connected layer. $q(n)$ is the ground truth distribution, let x be the ground truth class label, $q(n)$ can be defined as:

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

We expect that the maximum class probability of a generated image will be low during the test, which means that the network will fail to predict a particular class with high confidence. Therefore, the class label distribution for the generated samples with no label $q_{LSRO}(n)$ is written as:

$$q_{LSRO}(n) = \frac{1}{n} \tag{3}$$

Eq.3 can be termed as the label smoothing regularization for outliers (LSRO). We can re-write the cross-entropy loss by combining Eq. 1, Eq. 2 and Eq. 3:

$$l_{LSRO} = -(1 - Z)log(p(x)) - \frac{Z}{N}\sum_{n=1}^{N} log(p(n)) \tag{4}$$

For a real training image, Z = 0. For a generated training image, Z = 1. So the real images and generated images actually have different types of loss in the system.

Using LSRO, we can deal with more outliers that are located near the real training images in the space by introducing more variances of color and type to regularize the model. For example, if the training set has only one red-color vehicle, the discriminative power of the model will be limited because the network may be misdirected considering that the color red is a discriminative feature. By adding generated training samples, such as an unlabeled red-color vehicle, the classifier will be penalized if it makes the wrong prediction towards the labeled red-color vehicle. In this way, the network will be encouraged to find more underlying causes and to be less prone to overfitting.

C. Re-ranking Method

Problem Definition. For a given probe image, an initial rank list $L(p, G) = \{g_1, g_2, ..., g_N\}$ can be obtained according to the pairwise euclidean distance between probe p and gallery g_i. Our goal is to re-rank $L(p, G)$, so that more positive samples rank top in the list, and thus to improve the performance of vehicle re-ID.

Similarity Ranking Optimization. Following [23], the k-nearest neighbors (i.e. the top-k samples of the ranking list) of a probe p can be defined as $N(p, k)$:

$$N(p, k) = \{g_1, g_2, ..., g_N\} \tag{5}$$

where k represents the number of candidates in the set. The k-reciprocal nearest neighbors $R(p, k)$ can be defined as,

$$R(p, k) = \{g_i | (g_i) \in N(p, k)) \wedge (p \in N(g_i, k))\} \tag{6}$$

According to the previous description, the k-reciprocal nearest neighbors are more related to probe p than k-nearest neighbors. However, due to variations in views, illuminations and occlusions, the k-nearest neighbors and k-reciprocal nearest may not include the positive images. To address this problem, we incrementally add the $\frac{1}{4}k$-reciprocal nearest neighbors of each candidate in $R(p, k)$ into a more robust set $R^*(p, k)$ according to the following condition:

$$R^*(p, k) \leftarrow \{g_i | (g_i) \in N(p, k)) \wedge (p \in N(g_i, k))\} \tag{7}$$

$R^*(p, k)$ has been added more positive samples which are more similar to the candidates in $R(p, k)$ by this operation. We consider $R^*(p, k)$ as contextual knowledge to re-calculate the distance between the appearance feature of probe and gallery. The pairwise distance between the probe p and the gallery g_i will be re-calculated by comparing their k-reciprocal nearest neighbor sets. As described earlier [10], [24], we believe that if two images are similar, their k-reciprocal nearest neighbor sets overlap, i.e., there are some duplicate samples in the sets. And the more duplicate samples, the more similar the two images are. The new distance between p and q_i can be calculated by the Jaccard metric of their k-reciprocal sets as:

$$d_j(p, g_i) = 1 - \frac{|R^*(p, k) \cap R^*(g_i, k)|}{|R^*(p, k) \cup R^*(g_i, k)|} \tag{8}$$

Inspired by [9], we address the importance of original distance in re-ranking by jointly aggregate the original distance and Jaccard distance to revise the initial ranking list, the final distance d_f is defined as

$$d_f(p, g_i) = (1 - \lambda)d_j(p, g_i) + \lambda(p, g_i) \tag{9}$$

where λ denotes the penalty factor, it penalizes galleries far away from the probe p. When $\lambda = 0$, only the Jaccard distance is considered. On the contrary, when $\lambda = 1$, only the original distance is considered. Finally, the revised ranking list $L_{new}(p, G)$ can be obtained by ascending sort of the final distance.

III. Experiments

In the following section, we will first introduce the datasets used in our experiments and then the experimental set up will be briefly outlined including performance comparison, followed by the details of experiments on datasets for vehicle re-ID.

A. Vehicle Re-id Datasets

VeRi-776 [21] contains over 50,000 images of 776 vehicles with identity annotations, image timestamps, camera geo-locations, license plates, car types and colors information. Each vehicle is captured by 2 to 18 cameras in a road network during a 24-hour time period. The dataset is split into a train set consisting of 37,781 images of 576 vehicles, and a testing set of 11,579 images belonging to 200 vehicles. A subset of 1,678 probe images in the testing set are used as to retrieve corresponding images from all other test images.

VehicleID [1] is currently the largest publicly available vehicle re-id dataset contains data captured during daytime by multiple real-world surveillance cameras distributed in a small city in China. There are 221763 images of 26267 vehicles in total, which is split into two parts for model training and testing. The first part contains 110178 images of 13134 vehicles and the second part contains 111585 images of 13133 vehicles. The testing data provides three subsets(i.e. small, medium and large) ordered by their size from the original testing data for vehicle re-ID task.

B. Implementation Details

CNN Baseline. The Matconvnet [25] package is used for ResNet-50 model training. During training, we modify the fully-connected layer to have 571 and 13166 neurons for VeRi-776, VehicleID, respectively. We insert a dropout layer before the final convolutional layer and set the dropout rate to 0.8 for VeRi-776 and 0.6 for VehicleID. During testing, we extract the 2,048-dim CNN feature in the last convolutional layer for each image. Then, the initial rank list for a given probe image will be generated by euclidean distance.

GAN training and testing. We use Tensorflow [26] and the DCGAN package to train the GAN model using the provided data in the original training set without preprocessing. The outputted image is resized to 256×256 and then used in CNN model training (with LSRO).

Evaluation metrics. To evaluate the performance of the proposed algorithms in Section II, we use the Cumulative Matching Characteristic (CMC) curve which is widely used in Re-Id. It depicts the recognition performance as a function of the rank score and represents the expectation of finding the correct match in the rank-k ones. Moreover, we also use mean average precision (mAP), rank-1 and rank-5 to evaluate the overall performance.

C. Experiments on VeRi-776

We first evaluate our method on the VeRi-776 dataset [21] which is the only existing vehicle re-identification dataset providing spatial and temporal annotations. In our implementation, the CNN baseline has a rank-1 recognition accuracy of 85.64%. We compared the baseline result with other existing vehicle re-ID methods in Table II, the baseline alone has slightly exceeds many previous work [8], [21]. We train DCGAN on the 37,781 images in the VeRi-776 training set, and combine the real images with the generated images (see Fig. 3) to train the CNN model. As shown in Table I, when we add 21,000 GAN images to the CNN training, our method significantly improves the re-ID performance on VeRi-776. We observe improvement of +2.06% (from 85.64% to 87.70%), +2.63% (from 90.04% to 92.67%) and +4.43%(from 53.80% to 58.23%) in rank-1, rank-5 and mAP, respectively. These results indicate that the unlabeled images generated by the GAN effectively yield improvements over the baseline using the LSRO method. More GAN images are shown in Fig. 3.

With the help of re-ranking, the rank-1 accuracy and mAP are further improved to 88.62% and 64.78%. Three example results are shown in Fig.4. The proposed method, GAN+LSRO

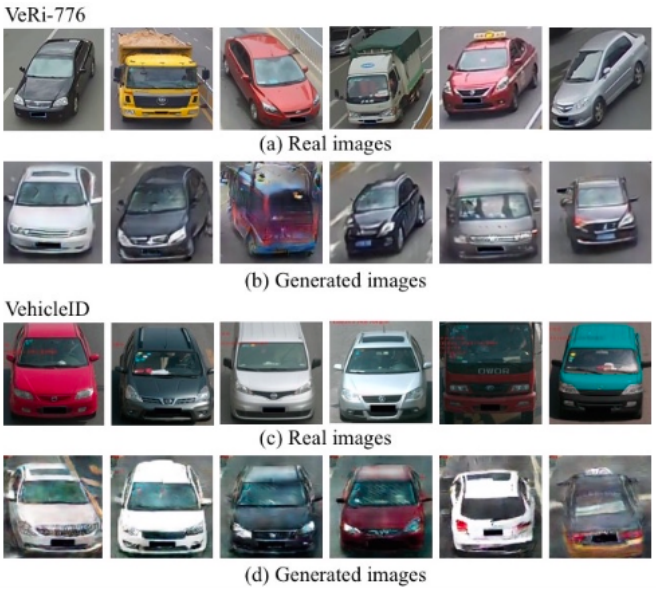

Fig. 3. Examples of real images and GAN generated images for VeRi-776 and VehicleID. (a)(c) The top one and three row show the real images in VeRi-776 training set; (b)(d) The second and bottom row show the unlabeled images generated by DCGAN[18] trained on the VeRi-776 training set and VehicleID training set. Although the generated images in (b) and (d) can be easily recognized as fake images by a human, they are added to the training sets of VeRi-776 and VehicleID to regularize the CNN model in our experiment.

TABLE I
An Example of a Table

GAN Images	rank-1	rank-5	Mean AP(%)
0	85.64	92.67	53.80
12000	87.60	93.42	57.53
21000	**87.70**	**93.92**	**58.00**
30000	86.82	93.27	54.28
48000	86.83	93.35	56.85

TABLE II
RANK-1 AND MAP ACCURACIES BY COMPARED METHODS ON THE VERI-776 DATASET [21]

Method	rank-1	rank-5	Mean AP
FACT [8]	50.95	73.48	18.49
FACT+Plate-SNN+STR [21]	61.44	78.78	27.77
Siamese-CNN+Path-LSTM [19]	83.49	90.04	58.27
Baseline (Ours)	85.64	92.67	53.80
GAN+LSRO (Ours)	87.70	93.92	58.23
GAN+LSRO+ re-ranking (Ours)	**88.62**	**94.52**	**64.78**

Fig. 4. Example vehicle re-identification results (rank-5) of two probes on the VeRi-776 dataset. For each probe, the first row correspond to the ranking results produced by our baseline, the second row describes the improved ranking results produced by our proposed method (GAN+LSRO+re-ranking). Vehicle surrounded by red box denotes the same vehicle as the probe, otherwise blue.

TABLE III
GALLERY AND PROBE SPLIT FOR VEHICLEID DATASET

Number of images	Small	Medium	Large
Gallery size	800	1600	2400
Probe size	6532	11395	17638

+ re-ranking, effectively ranks more true vehicles in the top of ranking list which are missed in the ranking list of our baseline. Fig.5. further shows the CMC curves of the proposed methods.

D. Experiments on VehicleID

To further test the effectiveness of our method, we provide the result of the largest image-based vehicle Re-ID dataset [1] in Table IV and Table V. Following the setting in [1], we randomly select one image of each vehicle and put it into the gallery set, the others images are all probe queries. The detailed information of the gallery set and the probe set in each test subset shows in III.

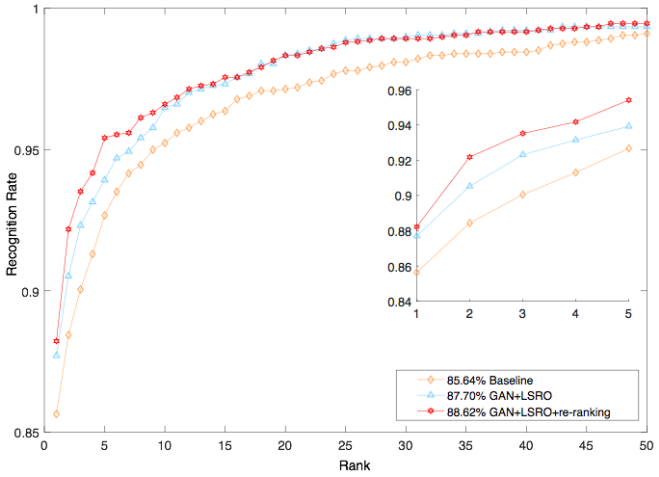

Fig. 5. The CMC curves of proposed methods for VeRi-776.

TABLE IV
RANK-1 ACCURACY BY COMPARED METHODS ON THE VEHICLEID DATASET [1]

Method		Small	Medium	Large
VGG+Triplet Loss[27]	rank-1	0.404	0.354	0.319
VGG+CCL [1]		0.436	0.370	0.329
Mixed Diff+CCL[1]		0.490	0.428	0.382
Baseline (Ours)		83.42	80.31	76.10
GAN+LSRO (Ours)		85.31	81.97	77.87
GAN+LSRO+ re-ranking (Ours)		**88.95**	**83.56**	**79.46**

TABLE V
MAP ACCURACY ON THE VEHICLEID DATASET [1]

Method		Small	Medium	Large
Baseline (Ours)	mAP	72.45	71.37	70.21
GAN+LSRO (Ours)		74.12	73.12	71.56
GAN+LSRO+ re-ranking (Ours)		**79.34**	**78.56**	**77.32**

Following the common method when evaluating model prediction accuracy, we repeat it 20 times in testing phase to get the final CMC curve. The detailed match rate from rank-1 to rank-50 of the proposed methods evaluated on the three scale test data are illustrated in Fig 6. In our experiments, we use the original training set and 50000 unlabeled images generated by GAN to train the network ResNet-50. Table IV and Table V illustrate the rank-1 match rate and mAP of our proposed method on all three test data splits, from the results, we can find that after applying our proposed semi-supervised learning and re-ranking structure, the match rate further increased about 30%−40% than previous methods, which reveal the significant advantages of our method again.

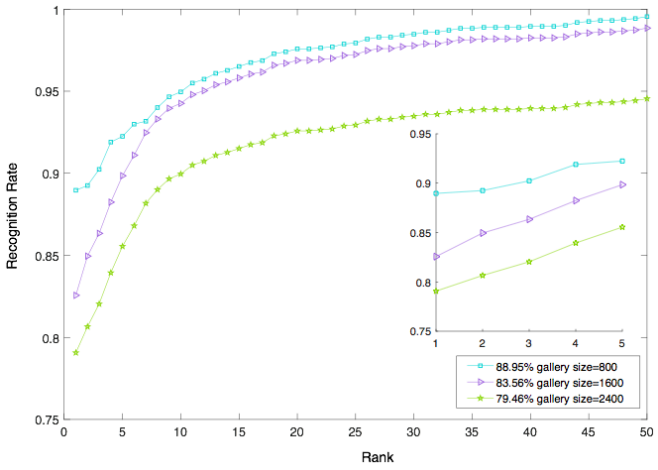

Fig. 6. The CMC curves of proposed method for VehicleID.

IV. Conclusion

In this paper, we propose an efficient approach to jointly apply the semi-supervised learning and re-ranking for vehicle re-id problem. Using a baseline DCGAN model [27], we show that the imperfect GAN images effectively demonstrate their regularization ability when trained with a ResNet baseline model. Through the LSRO method, we mix the unlabeled GAN images with the labeled real training images for simultaneous semi-supervised learning. We have also addressed the re-ranking problem by improving the k-reciprocal method. the proposed approach outperforms state-of-the-arts methods on the VeRi-776 dataset. Our experimental results indicate that the proposed approach outperforms state-of-the-arts methods on the VeRi-776 and VehicleID dataset.

References

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

[2] S. Mahendran and R. Vidal, "Car segmentation and pose estimation using 3d object models," *arXiv preprint arXiv:1512.06790*, 2015.

[3] B. C. Matei, H. S. Sawhney, and S. Samarasekera, "Vehicle tracking across nonoverlapping cameras using joint kinematic and appearance features," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3465–3472.

[4] M. Valera and S. A. Velastin, "Intelligent distributed surveillance systems: a review," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 152, no. 2, pp. 192–204, 2005.

[5] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.

[6] M. A. Saghafi, A. Hussain, M. H. M. Saad, N. M. Tahir, H. B. Zaman, and M. Hannan, "Appearance-based methods in re-identification: a brief review," in *Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on*. IEEE, 2012, pp. 404–408.

[7] D. Zapletal and 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.

[8] X. Liu, W. Liu, H. Ma, and 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.

[9] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," *arXiv preprint arXiv:1701.08398*, 2017.

[10] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, and R. Hu, "Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2553–2566, 2016.

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

[12] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3908–3916.

[13] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1249–1258.

[14] H. Liu, Y. Tian, Y. Yang, L. Pang, and 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.

[15] 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.

[16] A. Odena, "Semi-supervised learning with generative adversarial networks," *arXiv preprint arXiv:1606.01583*, 2016.

[17] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," *arXiv preprint arXiv:1701.07717*, 2017.

[18] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[19] Y. Shen, T. Xiao, H. Li, S. Yi, and X. Wang, "Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals," *arXiv preprint arXiv:1708.03918*, 2017.

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

[21] X. Liu, W. Liu, T. Mei, and 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.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and 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.

[23] D. Qin, S. Gammeter, L. Bossard, T. Quack, and 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.

[24] S. Bai and X. Bai, "Sparse contextual activation for efficient visual re-ranking," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1056–1069, 2016.

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

[26] 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.

[27] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognition*, vol. 48, no. 10, pp. 2993–3003, 2015.