Vehicle re-identification method based on global and local feature fusion

Zhou Bing¹, Wei Wei¹* and Wu Huaihong¹

¹School of Software Engineering, Chengdu University of Information Technology, Chengdu, Sichuan, China

*Corresponding author’s e-mail: weiwei@cuit.edu.cn

Abstract. Vehicle re-identification is still an issue worth discussing due to change of view, light and angle changes. At present, the key of vehicle reidentification research is to solve the problems that the difference between the same class is big and the difference between different classes is small. For the study, a vehicle re-identification method ground on view classification is proposed. The vehicle images are divided into four views, namely, front, rear, top and side, and local features are extracted by the segmented view information. Using the CNN feature extractor, the global feature representation with car ID attribute is learned. We have done experiments with the VERI-776 data sets. The mean average accuracy of 78.13% is obtained, which proves the effectiveness of the method.

1. Introduction

The purpose of vehicle re-identification technology is to retrieve the same vehicle in a specific traffic monitoring area by using the unique characteristic information of the vehicle. The technology uses deep learning technology to collect a large number of view information, and obtains the vehicle's time and space information through the road traffic camera to provide the comprehensive information of the road and its environment. Therefore, vehicle re-identification in traffic safety systems accounts for an important location, especially in the application of unmanned driving or intelligent parking lot.

Because the vehicle is of steel structure, the difference between vehicles with the same brand and model but different labels may be negligible. Because of the different real surveillance camera position of light, perspective, and the difference of resolution, resulting in the same car under different perspective to produce their own differences form a larger vehicle in differences. This makes the vehicle re-identification tasks on the concrete implementation challenges as shown in figure 1.

Figure 1.(a). Intra-class differences for the same vehicle. (b). Similarities between classes of different vehicles.
2. Related Work
With the completion of research data sets such as VERI[1], VehicleID[2], VERI-WILD[3], etc. which are all good vehicle data sets, Vehicle re-identification ways because of deep convolutional neural network also appear one after another. Saghir Alfasly et al.[4] A new vehicle re-identification method based on MLSL is proposed, and an effective model based on deep learning is obtained. Zheng et al.[5] proposed to use camera view, car model, colour and else meaningful attributes as guidance effect on vehicle re-identification. Chen et al.[6] devised a logical triple loss fusion tag smoothing cross entropy to extract fine-grained feature embedding.

Although re-identification has made great progress, the field of vehicle re-identification still lags behind. Therefore, there is furthermore space for development, especially in the aspect of feature extraction accuracy of vehicle re-identification.

3. Materials and Methods
In order to solve the inter-class similarity and intra-class difference between a group of sample pairs in vehicle re-identification, we mainly consider the two core parts of feature extraction and feature distance measurement calculation between different images. In this study, we adopt a deep learning architecture on account of multi view information. Different from the existing framework, we introduce view information to raise the identification accuracy of vehicle re-identification and introduce a new cost function to speed up the training convergence. In addition use double-branch convolutional neural network to classify and metric the global and local characteristics of vehicles. In addition, this paper focuses on the process of re-identification of vehicles using visual information, ignoring additional or background information.

3.1. Problem Formulation
Define a vehicle image V1 to be queried. Our goal is to retrieve the library set G and sort all images to create a sort list of all library images, so that their exterior similarity with the queried image is sorted in descending order. We define an image pair as<lx, ly> and the distance between them is shown in Equation (1). Where, f(·) is the feature extractor.

$$D_{x1, x2} = \| f_{x1} - f_{x2} \|^2$$  \hspace{1cm} (1)

3.2. Vehicle View Classification
We know that the vehicle is a cube structure, which can be classified according to different views. Generally, the camera cannot capture the images of the bottom of the two cars, so we can divide the vehicle into four views: front, back, side and top. This paper mainly considers two aspects in adding view classification information: First, a vehicle image will most likely contain information from three different views, which means that there will be at least two identical views in an image pair. Second, the partitioning of the view, which covers the entire vehicle under a particular view, allows you to capture local subtle differences between different vehicles. We use the vehicle parsing method in VaNet[7] to classify the views and train the view parsing model.
3.3. Feature Extraction
Feature extraction can be regarded as a data dimension reduction process, that is to transform the input image data into feature vectors with certain dimensions, so that the images are comparable. The random partial erasure method is used to reduce the overfitting of the model and raise the robustness of the feature representation. For the extraction of global features, we used MobileNet[8] in CNN as the backbone network to carry out pre-training on ImageNet[9] data set. Then, the 1024-dimensional feature vectors from the last layer of the MobileNet[8] convolution layer are provided to the Softmax layer for classification according to vehicle ID progression. The reason why we choose this network is that it can guarantee high accuracy with less computation time and memory overhead. The feature extraction process is shown in Figure 3.

3.4. Loss Functions
Cross entropy loss input is a pair of vehicle diagrams x1 and x2. These two pictures of vehicles may have the same ID, or they may have unlike IDs. y is used to indicate whether the two vehicle images belong to the same ID. If $y = 1$ show the two cars belong to the same label, we call it a positive sample; $y = 0$ show the cars are from different IDs, it is a negative sample. Wherein, the formula of cross entropy loss function is shown in (2). Where $d_{x1,x2}$ is obtained by Equation 1, express the Euclidean
distance calculated for the input images after two feature extraction, that is, their similarity is calculated. margin is the set threshold parameter.

\[ L_{id} = \frac{1}{1+y} \sum_{i=1}^{y} x_{i} d_{1}, x_{2}^{2} + \sum_{i=1}^{1-y} \max (m \arg \min (d_{1}, x_{2}, 0))^{2} \] (2)

The advantage of Triplet Loss is that it can be detailed. That is, when two similar samples are input, Triplet Loss can better model the details, which means that the loss is primarily a measure of the differential characteristics of the input samples. The triple loss network contains three inputs: the anchor point graph (referred to as a), and the positive sample (p for short) pertain to the identical ID as the anchor point, And a negative sample of Negative (n for short) belonging to a different ID from the anchor point. Inside, a and p are samples belonging to the same category, and a and n are samples not belonging to the same category. The triplet loss function is shown in Formula (3):

\[ L_t = (d_{a,p} + a - d_{a,n}) \] (3)

When using triplet loss to learn similarity measure, The purpose is to make the samples of the same label as close as possible, and as far away as possible from other samples that do not belong to the same label. By constant training and learning of the loss network, the vehicle samples with the same ID will eventually form a cluster in the feature space, so as to complete the re-identification task of the vehicle.

Circle loss pointed out the shortcomings of using class tags and positive and negative sample tags for learning at present, Put forward the (s_n-b_p) to a generalization of (α_ns_n-α_pb_p). s_n is to minimize the similarity between classes, s_p to maximize intra-class similarity. So that s_n and b_p can learn in different ways. Specifically, α_n is realized as a linear function of s_n, and α_p is realized as a linear function of b_p, so that the learning speed of each of them can adapt to the optimal state. The weighting factor increases as the distance between the similarity degree and the optimal value increases. The decision boundary obtained through the above optimization is a_ns_n − a_pb_p = m, Can prove that this interface is an arc in (s_n, s_p) space, Therefore, this newly proposed loss function is called Circleloss, as shown in Equation (4):

\[ L_{c} = \log [1 + \sum_{j=1}^{L} \exp (\gamma \alpha_{n}^{j} s_{n}^{j}) \sum_{i=1}^{N} \exp (-\gamma \alpha_{p}^{j} s_{p}^{j})] \] (4)

The overall loss function L_all of the vehicle reidentification network can be expressed as:

\[ L_{all} = L_{id} + aL_{t} + bL_{c} \] (5)

Where a and b are the weights of triple loss and circle loss respectively. In our model, we set the a as well as b to 1.

4. Experiments

4.1. Dataset

The Veri776[9] data set, derived from real traffic scenarios, is one of the more widely used vehicle re-identification data sets. The veri776 dataset include 49360 images from 776 vehicles with a rich array of labeling information, including bounding boxes, vehicle type, color, manufacturer, and more.

4.2. Evaluation Metrics

4.2.1. mAP

The re-identification method uses mean Average Precision as the evaluation standard of its overall performance and means the average of the accuracy of all retrieved results. Divide the test set into two parts, the query set and the image library. First, the average precision AP is calculated for the retrieval results of image q in each query set. As shown in Equation (6), The serial number of the image set is denoted as k; n is the sum of library images, and N represents the total number of images of the target...
vehicle. P(k) means the accuracy up to the k bit in the retrieval sequence, gt(k) used to judge whether the image No. K is the target vehicle, and Q shows the total number of queried picture. Finally, the average value of the average precision of the retrieval results of all query sets is calculated, Namely mAP, as shown in Equation (7).

\[
AP = \frac{\sum_{k=1}^{Q} P(k) \times gt(k)}{N}
\]

\[
mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q}
\]

4.2.2.CMC

The Cumulative Matching Characteristic (CMC) Indicates the probability of matching as the correct result in the first k search results. The abscissa represents K, and the ordinate represents the accuracy. CMC before the k position is shown in Equation (8), When the correct matching target of image q appears before the k bit of the retrieval sequence, gt(q,k) is equal to 1.

\[
Rank(k) = CMC(k) = \frac{\sum_{q=1}^{Q} gt(q,k)}{Q}
\]

4.3. Implementation Details

We use MobileNet[8] as our backbone network to conduct parameter initialization training on the dataset ImageNet[9]. All images are adjusted to a size of 255×255 and the initial learning rate is 0.00035. Adam is used as the parameter optimizer. Use the default parameter Settings. We conducted 110 epoch of training with the batch size set to 64, and we used random erasure for data enhancement. Our approach is implemented on PyTorch, a deep learning framework.

4.4. Results on VeRi-776 Dataset

![Loss curve during training](image1)

![The change curve of accuracy in the training process](image2)

Figure 4. Changes in loss and accuracy during training.

| Model                     | mAP(%) | Rank-1(%) | Rank-5(%) |
|---------------------------|--------|-----------|-----------|
| FACT[1]                   | 18.49  | 50.95     | 73.48     |
| FACT+Plate-SNN+STR[1]     | 27.77  | 61.44     | 78.78     |
| FDA-Net[3]                | 55.49  | 84.27     | 92.43     |
| DF-CVT[5]                 | 61.06  | 91.36     | 95.77     |
| MLSL[4]                   | 61.13  | 90.04     | 96.00     |

Table 1. Comparative experiment on VeRi-776 dataset.
5. Conclusion

In order to solve the vehicle differences caused by the changes of different perspectives, we put forward an end-to-end deep convolutional network architecture that combines multiple views information and use double-branch convolutional neural network to classify and measure the global and local features of the vehicle. Experiments on public datasets VERI-776 show that the proposed algorithm can achieve better performance than most existing re-identification algorithms.

References

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

[2] Liu, H., Tian, Y., Yang, Y., Pang, L., & Huang, T. (2016). Deep relative distance learning: Tell the difference between similar vehicles. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2167-2175.

[3] Lou, Y., Bai, Y., Liu, J., Wang, S., & Duan, L. (2019). Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3235-3243.

[4] Alfasly, S., Hu, Y., Li, H., Liang, T., Jin, X., Liu, B., & Zhao, Q. (2019). Multi-label-based similarity learning for vehicle re-identification. IEEE Access, 7, 162605-162616.

[5] Zheng, A., Lin, X., Li, C., He, R., & Tang, J. (2019). Attributes guided feature learning for vehicle re-identification. arXiv preprint arXiv:1905.08997.

[6] Chen, X., Sui, H., Fang, J., Feng, W., & Zhou, M. (2020). Vehicle Re-Identification Using Distance-Based Global and Partial Multi-Regional Feature Learning. IEEE Transactions on Intelligent Transportation Systems, 22(2), 1276-1286.

[7] Chu, R., Sun, Y., Li, Y., Liu, Z., Zhang, C., & Wei, Y. (2019). Vehicle re-identification with viewpoint-aware metric learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8282-8291.

[8] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. pp. 770–778.

[9] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June).Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. pp. 248-255.

[10] Liu, X., Liu, W., Mei, T., & Ma, H. (2016, October). A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In European conference on computer vision. Springer, Cham. pp. 869-884.

[11] Liu, C., Huynh, D. Q., & Reynolds, M. (2019, July). Urban area vehicle re-identification with self-attention stair feature fusion and temporal bayesian re-ranking. In 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1-8.