machines

MDPI

Communication

A Vehicle Comparison and Re-Identification System Based on Residual Network

Weifeng Yin, Yusong Min and Junyong Zhai *

School of Automation, Southeast University, Nanjing 210096, China
* Correspondence: 101010807@seu.edu.cn

Abstract: In the highway intelligent monitoring system, it is difficult to find the target vehicle through millions of pictures because of the presence of fake-licensed vehicles. In order to solve this problem, a vehicle comparison and re-identification (Re-ID) system is built in this paper. By introducing Circle loss and Generalized-Mean(GeM) pooling, vehicle feature extraction and storage, vehicle comparison and vehicle search can be realized. Experimental results show that the proposed algorithm reaches 95.79% of the mean Average Precision (mAP) on the vehicle search task, which meets the requirements of practical applications.

Keywords: vehicle re-identification; deep learning; convolutional neural network (CNN); residual network

Citation: Yin, W.; Min, Y.; Zhai, J. A Vehicle Comparison and Re-Identification System Based on Residual Network. *Machines* **2022**, *10*, 799. https://doi.org/10.3390/machines10090799

Received: 17 August 2022
Accepted: 8 September 2022
Published: 10 September 2022

1. Introduction

With the rapid development of highways, the traditional vehicle recognition methods based on the technology of license plate recognition and electronic toll collection (ETC) no longer meet the needs of the intelligent monitoring system. The vehicles with fake plates and unlicensed vehicles increase the difficulties of localizing the target vehicle from a gallery set of images. To solve these problems, it is necessary to recognize the vehicle accurately by virtue of other features. In this work, we use the front image of a vehicle as the main feature to complete the task of vehicle re-identification (Re-ID). There are three main methods to implement vehicle Re-ID, which are traditional feature extraction algorithms, hash retrieval algorithms and deep learning. Researchers in the field of computer vision have designed many feature extraction algorithms, including global feature extraction and local feature extraction. The global feature abstractly expresses the global information of the image [1,2]. The local feature emphasizes the line change trend of the image, and expresses the information through local changes [3,4]. Traditional feature recognition algorithms are based on manually extracting features, which have good results, and have the advantages of repeatability, distinguishability, and high efficiency. However, compared with the image of the whole vehicle, there are fewer effective features on the front of the vehicle. At the same time, factors such as illumination changes and partial occlusion reduce the effectiveness of artificial features, which means the accuracy of this approach cannot achieve as high as that of deep learning.

On the other hand, the main idea of the hash retrieval algorithm is to map the original data from the high-dimensional space to the low-dimensional space through a series of hash functions. It maintains the similarity of the original data in the high-dimensional space during the mapping process, thereby achieving the purpose of reducing calculation and storage overhead. The existing hash algorithms can be roughly divided into three categories: unsupervised [5], semi-supervised [6] and supervised methods [7]. With its powerful feature learning capabilities, the hash algorithm based on deep learning surpasses the traditional hash method based on hand-designed features. However, the goal of the hash method is to obtain a binary code, so the discrete value constraints are often encountered in the optimization process, so it is generally difficult to use gradient-based

methods to optimize the objective function. Deep learning was first applied to person Re-ID, enabling it to achieve inspiring performance on the widely used benchmarks [8]. Recently, there have been a lot of theoretical studies on the open-world setting, but compared with the close-world one, there are some problems with the effectiveness of learning. At the same time, it is difficult to put it into practical applications since the performance of unsupervised learning cannot keep up with supervised learning. In this paper, a vehicle Re-ID algorithm based on strong-baseline [9] is proposed as a supplement to license plate recognition. The main contributions of this paper are summarized as follows: (i) A vehicle Re-ID dataset which contains 102,187 images of 5924 vehicles is collected at the Expressway toll stations to train the model for practical applications. (ii) The functions of vehicle feature extraction and storage, vehicle comparison, and vehicle search are realized in the system, which can be used in video surveillance and intelligent transportation. Only the global features are trained in the algorithm, which eliminates the system's dependence on other auxiliary annotations such as vehicle colors, vehicle types and license plates. (iii) By introducing Circle loss and Generalized-mean(GeM) pooling, the accuracy of the system is improved. The storage of the vehicle features also decreases the time of inference. The experiments show that the performance of our algorithm performs better than the alternative one and other Re-ID algorithms on the dataset.

The remainder of this paper is organized as follows. Section 2 introduces the related works, including the convolutional neural network and its application in Re-ID tasks. The framework and the main steps of the algorithm are depicted in Section 3. Section 4 presents the experimental results. Finally, the conclusions are given in Section 5.

2. Related Work

The vehicle comparison and Re-ID systems are performed in a network consisting of multiple surveillance cameras. It has many similarities with deep person Re-ID, which commonly uses a convolutional neural network (CNN) to extract features from images.

2.1. Convolutional Neural Network

CNN is an effective feature extraction method that contains three parts: (i) the convolution operator extracts local features; (ii) the activation function increases the nonlinearity in the output; (iii) the pooling operator provides spatial invariance to the extracted features. The effectiveness of CNN on feature extraction enables it to perform well in vision applications such as object detection, object recognition, object categorization and other tasks. Re-ID is one of the tasks making use of CNN to extract the feature of the image and search the same target according to the similarity between the features.

2.2. Person Re-Identification

A standard deep person Re-ID system contains feature representation learning and deep metric learning [10]. Feature representation learning strategies use CNN to extract representation features and treat the Re-ID task as a classification problem. The representation features can be divided into four categories: (i) global feature [11]; (ii) local feature [12]; (iii) auxiliary feature [13,14]; (iv) video feature [15].

Deep metric learning is widely used in the field of image retrieval, which aims to learn the similarity of two images through a network. The widely used loss functions include contrastive loss [16], triplet loss [17] and quadruplet loss [18].

2.3. Vehicle Re-Identification

Many methods of person Re-ID algorithm can be applied to vehicle Re-ID, while the latter is more complex and challenging than the former [19]. The same vehicle varies greatly in appearance in different surveillance cameras due to multiple viewpoints and blurry photos, and different vehicles may also have similar colors and shapes, especially when they belong to the same manufacturer [20,21]. Some studies focused on license

plate recognition to complement vehicle Re-ID [22]. Spatio-temporal path proposals and viewpoint information have also been proved to make vehicle Re-ID more effective [23,24].

The methods above are designed specifically for certain datasets with auxiliary annotations. It is difficult to apply them to practice as the supplement to license plate recognition and the accuracy of vehicle comparison may be affected by the redundant information. In this paper, CNN is used to acquire camera-independent features based on the strong-baseline [9], which is a baseline of person Re-ID. By training with the dataset collected by the surveillance cameras at the expressway toll stations, the system can be used in general situations of highway surveillance systems. Only the vehicle ID is utilized as the label in the algorithm, which reduces the manual labeling work. The experiments show that the algorithm can achieve better performance than those using auxiliary annotations.

3. System Description

This section first introduces a framework of the algorithm, and then briefly describes the main steps in the framework.

3.1. Framework

The framework of the algorithm contains the following seven stages: pre-processing; backbone; aggregation; head; loss; metric distance; and post-processing shown in Figure 1, which can be divided into two main parts: training and inference. In the training part, the original images are input to the neural network, the network parameters need to be trained, and finally, the images are extracted into 1×2048 dimensional features. In the inference part, the image features in the gallery are sorted by comparing the measured distance with the one in the query, and finally the images most likely to be the same vehicle are selected.

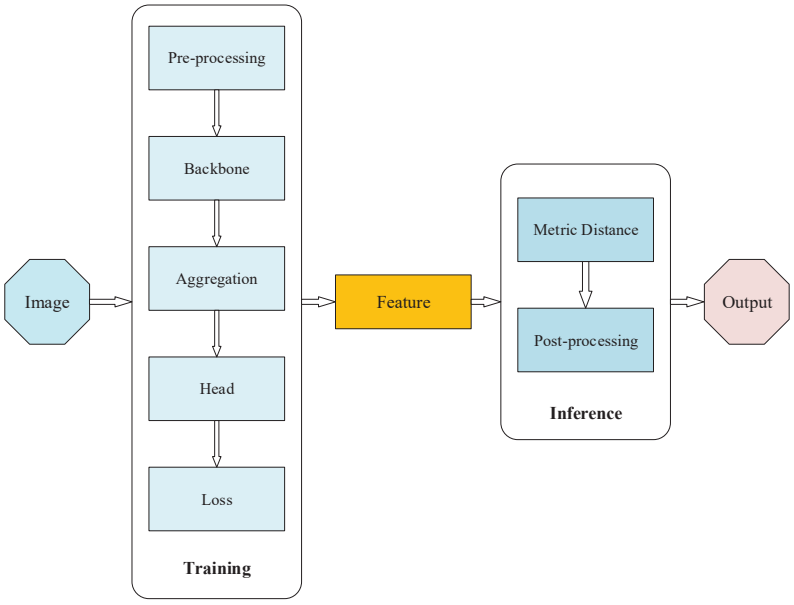


Figure 1. A framework of the vehicle Re-ID system.

3.2. Pre-Processing

Before entering the network, the original image needs to be resized and normalized to unify the image size. In addition, data enhancement needs to be applied to enhance the robustness of the model, including horizontal flipping, random cropping, filling, and color dithering. Here, we apply the auto-augment method proposed in [25] to achieve effective data enhancement. Automatic enhancement uses a search algorithm based on reinforcement learning to find the best choice and sequence of enhancement operations so that the neural network can obtain better verification set accuracy.

3.3. Backbone

ResNet is the abbreviation of the residual network, which is a CNN architecture. It introduces a residual network and uses shortcut connections to solve the problem of gradient disappearance in deep neural networks. The backbone aims to infer the image as a feature map. In order to obtain the feature map with increased spatial size, the last step of ResNet is changed from 2 to 1, which brings a significant improvement by slightly increasing the computational cost. On the basis of ResNet, the instance batch normalization (IBN-Net) [26] module is added to the backbone to improve cross-domain capabilities. IBN-Net unifies instance normalization (IN) and batch normalization (BN) in an appropriate way. The IN and BN features remain in the shallow layer, while the BN features remain in the deeper layer. It turns out that IBN-Net improves the performance of Re-ID problems.

3.4. Aggregation

After obtaining the feature maps from the backbone, the network uses a pooling layer to aggregate them into a global feature. The pooling layer takes a 3D tensor X of $W \times H \times C$ dimensions as an input, and produces a vector f of $1 \times 1 \times C$ dimensions as an output. C is the number of feature maps while W and H are the width and the height of a feature map. Different from the two common operations in Re-ID, global average pooling and maximum pooling, generalized mean (GeM) pooling in [27] is applied after ResNet to obtain better results in the target dataset. The functions of global vector $f = [f_1, \ldots, f_c, \ldots, f_C]$ and the set of feature maps $X = [X_1, \ldots, X_c, \ldots, X_C]$ are as follows:

$$\text{Max Pooling: } f_c = \max_{x \in X_c}\{x\} \tag{1}$$

$$\text{Average Pooling: } f_c = \frac{1}{|X_c|}\sum_{x \in X_c} x \tag{2}$$

$$\text{GeM Pooling: } f_c = \left(\frac{1}{|X_c|}\sum_{x \in X_c} x^p\right)^{\frac{1}{p}} \tag{3}$$

where p is a pooling parameter to be learned. GeM pooling can turn into max pooling or average pooling by changing p. By increasing p, the feature map responses will become more localized. The experimental results show that the GeM pooling performs better on the dataset.

3.5. Head

Inspired by the BN-Neck, which is introduced by the strong-baseline, we use classification learning and pairwise learning to train the model, respectively. The framework of the head stage is shown in Figure 2. Getting through the pooling layer, the feature of the image was descended into 1×2048 dimension, which is marked as f_t. It is trained with pairwise learning. After a Batch Normalization (BN) layer, feature f_t is transformed into f_i to be trained using classification learning. The BN layer helps to smoothen the feature distribution, which improves the performance of classification learning.

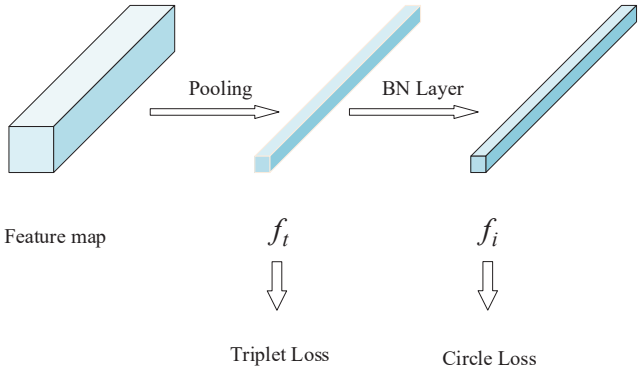


Figure 2. The framework of the Head.

3.6. Loss Function

The total loss function includes triplet loss [28] and Circle loss [29] as follows:

$$L_{total} = L_{tri} + L_{circle} \tag{4}$$

The optimization objective of the triplet loss function is to minimize the distance between the anchor image and the positive image (the images of the same vehicle), and to maximize the distance between the anchor image and the negative image of a different vehicle. The formula is as follows:

$$L_{tri} = [d_p - d_n + \alpha]_+ \tag{5}$$

where d_p represents the distance between the anchor image and the positive image; d_n represents the distance between the anchor image and the negative image. α is a margin and $[x]_+$ means $\max(x, 0)$.

Because of the large dataset, there are numerous possible combinations of triplets. To improve the generalization of the network, the hardest samples are chosen to train the network. For each anchor image, a hardest positive sample and a hardest negative sample will be selected to form a triplet. Softmax cross-entropy loss is commonly used in the Re-ID models, including the strong-baseline. Meanwhile, the Cosface loss normalizes the feature and weight vector to eliminate radial changes, and reformulates the traditional softmax loss as a cosine loss. Their functions are given below:

$$L_{Softmax+CE} = -\log \frac{e^{\gamma z_y}}{e^{\gamma z_y} + \sum_{j \neq y} e^{\gamma z_j}} = \log[1 + \sum_{j \neq y} e^{\gamma (z_j - z_y)}] \tag{6}$$

$$L_{Cosface} = -\log \frac{e^{\gamma (s_y - m)}}{e^{\gamma (s_y - m))} + \sum_{j \neq y} e^{\gamma s_j}} = \log[1 + \sum_{j \neq y} e^{\gamma (s_j - s_y + m)}] \tag{7}$$

where γ is a scale factor. Softmax cross-entropy loss uses the inner product between feature and weight z_j to denote the activation of a fully connected layer, while Cosface loss uses the cosine of the angle between feature and weight s_j. z_y and s_y mean the within-class similarity score of the feature and its ground truth. m is a cosine margin term introduced by Cosface loss to further maximize the decision margin in the angular space.

In this paper, Circle loss is used in the network to replace softmax cross-entropy loss since it has better performance in Re-ID tasks. Similar to Cosface loss, Circle loss uses cosine similarity s_p and s_n to represent within-class similarity and between-class similarity, respectively. It is assumed that there are K within-class similarity scores and L between-class similarity scores associated with the target feature. Each similarity is able to learn at its own speed by introducing non-negative weighting factors α_p and α_n, which bring more flexible optimization and more definite convergence to the model. Δ_p and Δ_n are the between-class

margin and inter-class margin, respectively. Our task is to use classification training to train the feature f_i. Suppose there are N training classes. There is only one within-class similarity score and $N-1$ between-class similarity scores. The function conversion of Circle loss is as follows:

$$\begin{aligned} L_{circle} &= -\log \frac{e^{\gamma(a_p(s_p-\Delta_p))}}{e^{\gamma(a_p(s_p-\Delta_p))} + \sum_{j=1}^{N-1} e^{\gamma(a_n^j(s_n^j-\Delta_n^j))}} \\ &= \log[1 + \sum_{j=1}^{N-1} e^{\gamma\alpha_n^j(s_n^j-\Delta_n)} e^{-\gamma\alpha_p(s_p-\Delta_p)}] \end{aligned} \tag{8}$$

3.7. Distance Metrics

The feature $f_i \in \mathbb{R}^{1\times 2048}$ is used to infer the closer image of the target image in the query. Cosine similarity is implemented to compare the metric distance between two features. The similarity between two features f_1 and f_2 is calculated as:

$$Similarity(f_1, f_2) = \frac{f_1 f_2^T}{\|f_1\|_2 \cdot \|f_2\|_2} \tag{9}$$

To ensure that the parameters in the training process and the inference process are consistent, the cosine similarity is also used in the triple function to replace the Euclidean distance to represent the distance between the features.

3.8. Post-Processing

In order to reduce the search time, the images in the gallery are first extracted as 1×2048 dimensional features and saved. When the target image is input into the algorithm, its features can be directly used to calculate the distance to the saved feature. At the same time, excluding photos taken by the same camera can improve retrieval efficiency. In order to visually view the results, a visualization tool is provided. The images in the gallery are sorted by distance, with the nearest at the top. The visualization tool displays the predicted first 10 images.

4. Experiments

In this section, the details of the experiments are described. The algorithms and models are based on Python 3.8 environments and the platform framework is based on PyTorch 1.6, CUDA 10.1. The experiments are trained on 4 TITAN Xp GPUs. Firstly, the collection of the dataset and the evaluation indicators are introduced. Then the thresholds of metric distance are discussed for vehicle comparison. Finally, the comparative experiments and the ablation experiments of vehicle Re-ID are conducted on the public dataset VeRi [30] and our own dataset.

4.1. Dataset

Traditional vehicle Re-ID datasets such as VeRi [30] and VehicleID [31] are not well suited for realistic highway scenarios due to their collection locations and recording duration. To use the algorithm in the expressway monitoring system, a dataset is collected at the expressway toll stations. The images are captured by the surveillance cameras when the vehicles pass through the station. The license plate numbers are saved as labels, which reduces the time of annotation. The dataset contains 102,187 images of 11,854 vehicles. We put one image of a vehicle into the query set and the rest into the gallery set. They are split in half into a test set and a training set, each containing the frontal face images of 5924 vehicles. The images are resized to 128×256.

Meanwhile, to compare our algorithm with those methods using auxiliary annotations, the public dataset VeRi is used to evaluate the proposed method. The dataset contains over 50,000 images of 776 vehicles annotated with vehicle ID and vehicle attribution.

The 576 vehicles with 37,781 images are annotated as the training set and the rest 200 vehicles with 11,579 images are set as the test set.

4.2. Evaluation Indicators

Precision and recall are two performance measures commonly used in the evaluation of information retrieval. Precision is the fraction of retrieved documents that are relevant to the query, while recall is the fraction of the relevant documents that are successfully retrieved.

For the Re-ID system, there are three widely used evaluation indexes, including Rank-n, mean average precision(mAP) and mean inverse negative penalty (mINP). Rank-n means the probability that the first n images in the search result are the correct results. Generally, the average value is taken through multiple experiments. mAP is widely used to measure the retrieval abilities of the system. The function of mAP is shown as follows:

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{10}$$

$$AP_i = \frac{1}{m_i}\sum_{j=1}^{m_i} Precision_i^j \tag{11}$$

where n is the number of the images in the query. m_i means the number of relevant images in the gallery for the target vehicle. $Precision_i^j$ represents the precision for each additional correct image.

mINP is used to evaluate the ability of a model to search for the most difficult samples. The function of mINP is shown as:

$$mINP = \frac{1}{n}\sum_{i=1}^{n} (1 - NP_i) \tag{12}$$

$$NP_i = \frac{R_i^{hard} - G_i}{R_i^{hard}} \tag{13}$$

where NP_i is a computationally efficient metric that measures the plenty to find the hardest correct match for query i. G_i indicates the total number of correct matches, and R_i^{hard} represents the rank position of the hardest match.

4.3. Vehicle Comparison

The task of vehicle comparison is to determine whether the two input images belong to the same vehicle. After training, the network can extract the corresponding feature from each image, and its dimension is 1×2048. By measuring the metric distance between two image features in the cosine space, their similarity can be obtained. Table 1 shows the similarities between different images.

Table 1. The similarity between different images.

Similarity	01_c1	01_c2	02_c1	02_c2	03_c1
01_c1	1.0000	0.7476	−0.0209	−0.0482	0.0562
01_c2	0.7476	1.0000	−0.0075	−0.0247	0.1228
02_c1	−0.0209	−0.0075	1.0000	0.8781	−0.0241
02_c2	−0.0482	−0.0247	0.8781	1.0000	−0.0648
03_c1	−0.0562	0.1228	−0.0241	−0.0648	1.0000

The number in the front represents the vehicle number, and the number starting with c in the back represents the camera number. The similarity ranges from -1 to 1. The closer the similarity is to 1, the closer the two images are. Therefore, it is necessary to select a threshold to determine whether the two images belong to one vehicle. It is difficult to

increase the precision rate and recall rate at the same time. Table 2 shows the average precision and average recall rate of all vehicles under different thresholds.

Table 2. The average precision and recall of all vehicles under different thresholds.

Measure	0.60	0.62	0.64	0.66	0.68	0.70
Precision	90.23	92.70	94.46	95.75	96.75	97.45
Recall	97.11	96.34	95.32	94.06	92.50	90.50

4.4. Vehicle Re-ID

The task of Vehicle Re-ID is to search the target vehicle in a gallery with amounts of images captured by surveillance cameras. It is important to improve the speed and accuracy of the search.

4.4.1. Comparison with Attribute Based Methods on VeRi Dataset

Table 3 lists the mAP and the matching rates of our proposed method and other vehicle Re-ID algorithms using auxiliary annotations. Our method achieves a mAp of 81.6% without using other annotations such as spatio-temporal path information [23,24] and vehicle types [20,21].

Table 3. Comparison with attribute based methods on VeRi dataset.

Method	Rank-1	Rank-5	mAP
AGNet [20]	90.90	96.20	66.32
SAN [21]	93.30	97.10	72.50
PROVID [23]	81.56	95.11	53.42
Siamese-CNN+Path-LSTM [24]	83.49	90.04	58.27
Ours	96.6	97.0	81.6

4.4.2. Ablation Studies

To verify the improvement strategies for the method proposed in this paper, ablation experiments are conducted based on the test set. The result of the experiment is shown in Table 4.

Table 4. The ablation experiments.

GeM Pooling	Circle Loss	Feature Stored	Rank-1	mAP	Inference-Time
			97.47	91.25	8.52 ms
✓			98.12	92.47	8.61 ms
	✓		98.42	93.96	8.53 ms
		✓	97.47	91.25	2.13 ms
✓	✓	✓	98.62	95.79	2.34 ms

The introduction of Circle loss and GeM pooling improves the accuracy of the algorithm with little cost to inference speed. Meanwhile, the features of the images in the gallery are stored in advance, which saves the time of extracting them from the images before computing the metric distances between them and the target image.

4.4.3. Comparison with State-of-the-Art Methods on Our Dataset

The proposed method is compared with state-of-the-art Re-ID methods on our dataset. AGW [10] is a baseline for person Re-ID. Designed on top of the strong-baseline [9], it achieves competitive performance on both image and video Re-ID tasks. MGN [32] is a multi-branch deep network architecture consisting of global feature representations and local feature representations. Transreid [33] introduces the transformer model as the backbone to extract vehicle features. The results are shown in Table 5. By improving

the strong-baseline and introducing methods such as GeM pooling and Circle loss, our algorithm achieves the best performance among the methods, with mAP of 95.79% and mINP of 85.52%.

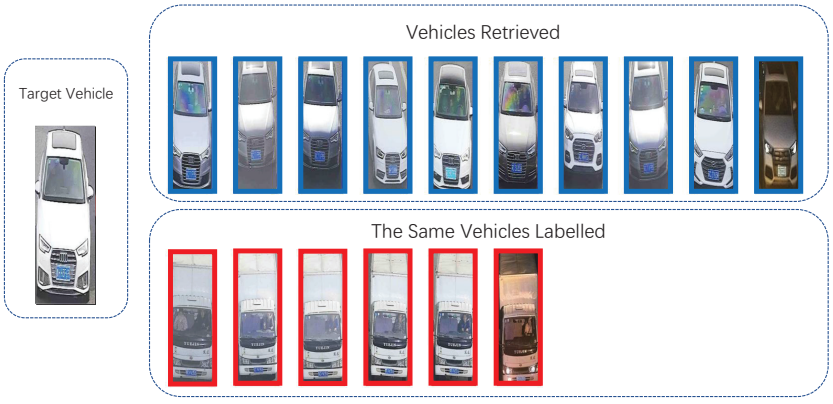
Table 5. Comparison with State-of-the-Art Re-ID methods on our dataset.

Model	Rank-1	Rank-5	mAP	mINP
Strong-baseline	97.47	98.94	91.25	74.39
Transreid	93.88	97.80	80.99	55.01
MGN	94.34	98.16	84.61	62.85
AGW	97.79	99.16	92.98	78.36
Ours	98.62	99.04	95.79	85.52

The experimental results show that the mAP of our algorithm on the dataset can reach 95.79% and the inference time per 128 images can reach 0.293 s. The search results are presented by visualization as shown in Figure 3.

The image on the left is the target image that needs to be retrieved in the query. Ten images retrieved from the gallery are displayed in the upper right corner of the figure. The vehicle images below are photos of this target vehicle taken at different locations. The red frame means the same vehicle while the blue one means not. At the same time, by visualizing the 100 vehicles with the lowest accuracy in the query, it was found that 62 of them are caused by data labeling errors. An example of incorrect labeling results is shown in Figure 3b.

(**a**) The example of the right results



(**b**) The example of the wrong results

Figure 3. The example of the search results.

Despite the wrong label, the predicted vehicle is similar to the target vehicle. In other wrong predictions, the target images are severely distorted due to exposure effects or darkness, and the predicted images are also distorted. In summary, the trained model performs well on the vehicle frontal dataset and meets the requirements of the Re-ID task in the original scenes.

5. Conclusions

In this article, a dataset is collected at the highway toll stations to train a vehicle comparison and Re-id system for practical application in the highway intelligent monitoring system. With a lack of auxiliary annotations, the global features of the vehicle images are extracted to accomplish the vehicle comparison and Re-ID task. The Circle loss and Gem pooling are introduced based on the strong-baseline to improve the accuracy. Compared with other attributed-based algorithms, our method achieves higher accuracy with 81.6% of mAP on the public dataset VeRi. The experiments also indicate that our method performs well with 95.79% of mAP among the state-of-the-art Re-Id networks on our dataset. By saving the vehicle features in advance, the inference process is simplified by sorting the features in the gallery by the cosine similarities with the feature extracted from the target image. The inference time for each vehicle was reduced to 2.34 ms. In the future, we will focus on studying how to re-identify vehicles from different viewpoints in different weather. We will try to use more efficient and intelligent algorithms to train the feature extraction network.

Author Contributions: Conceptualization, W.Y.; methodology, J.Z.; software, Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 61873061 and Natural Science Foundation of Jiangsu Province (BK20211162).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

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

References

1. Baek, N.; Park, S.M.; Kim, K.J.; Park, S.B. Vehicle color classification based on the support vector machine method. In *International Conference on Intelligent Computing*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 1133–1139.
2. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
3. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
4. Ojala, T.; Pietikainen, M.; Harwood, D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1994; Volume 1, pp. 582–585.
5. Datar, M.; Immorlica, N.; Indyk, P.; Mirrokni, V.S. Locality-sensitive hashing scheme based on p-stable distributions. In Proceedings of the 20th Annual Symposium on Computational Geometry, Buffalo, NY, USA, 7–11 June 2021; pp. 253–262.
6. Wang, J.; Kumar, S.; Chang, S.F. Semi-supervised hashing for large-scale search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2393–2406. [CrossRef] [PubMed]
7. Xia, R.; Pan, Y.; Lai, H.; Liu, C.; Yan, S. Supervised hashing for image retrieval via image representation learning. In Proceedings of the National Conference on Artificial Intelligence, Madeira, Portugal, 24–26 September 2014; Volume 3, pp. 2156–2162.
8. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1116–1124.
9. Luo, H.; Jiang, W.; Gu, Y.; Liu, F.; Liao, X.; Lai, S.; Gu, J. A strong baseline and batch normalization neck for deep person re-identification. *IEEE Trans. Multimed.* **2020**, *22*, 2597–2609. [CrossRef]
10. Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Shao, L.; Hoi, S.C. Deep learning for person re-identification: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 2872–2893. [CrossRef] [PubMed]

11. Sun, Y.; Zheng, L.; Deng, W.; Wang, S. SVDNet for pedestrian retrieval. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October2017; pp. 3800–3808.
12. Varior, R.R.; Shuai, B.; Lu, J.; Xu, D.; Wang, G. A siamese long short-term memory architecture for human re-identification. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 135–153.
13. Su, C.; Zhang, S.; Xing, J.; Gao, W.; Tian, Q. Deep attributes driven multi-camera person re-identification. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 475–491.
14. Sun, R.; Huang, Q.H.; Fang, W.; Zhang, X. Attributes-based person re-identification via CNNs with coupled clusters loss. *J. Syst. Eng. Electron.* **2020**, *31*, 45–55. [CrossRef]
15. Liu, H.; Jie, Z.; Jayashree, K.; Qi, M.; Jiang, J.; Yan, S.; Feng, J. Video-based person re-identification with accumulative motion context. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 2788–2802. [CrossRef]
16. Varior, R.R.; Haloi, M.; Wang, G. Gated siamese convolutional neural network architecture for human re-identification. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 791–808.
17. Cheng, D.; Gong, Y.; Zhou, S.; Wang, J.; Zheng, N. Person re-identification by multichannel parts-based CNN with improved triplet loss function. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Amsterdam, The Netherlands, 11–14 October 2016; pp. 1335–1344.
18. Chen, W.; Chen, X.; Zhang, J.; Huang, K. Beyond triplet loss: A deep quadruplet network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1320–1329.
19. Deng, J.; Hao, Y.; Khokhar, M.S.; Kumar, R.; Cai, J.; Kumar, J.; Aftab, M.U. Trends in vehicle re-identification past, present and future: A comprehensive review. *Mathematics* **2021**, *9*, 1–35.
20. Wang, H.; Peng, J.; Chen, D.; Jiang, G.; Zhao, T.; Fu, X. Attribute-Guided Feature Learning Network for Vehicle Reidentification. *IEEE Multimed.* **2020**, *27*, 112–121. [CrossRef]
21. Qian, J.; Jiang, W.; Luo, H.; Yu, H. Stripe-based and attribute-aware network: A two-branch deep model for vehicle re-identification. *Meas. Sci. Technol.* **2020**, *31*, 095401. [CrossRef]
22. Chen, G.W.; Yang, C.M.; İk, T.U. Real-time license plate recognition and vehicle tracking system based on deep learning. In Proceedings of the 22nd Asia-Pacific Network Operations and Management Symposium, Tainan, Taiwan, 8–10 September 2021; pp. 378–381.
23. Liu, X.; Liu, W.; Mei, T.; Ma, H. PROVID: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Trans. Multimed.* **2018**, *20*, 645–658. [CrossRef]
24. Shen, Y.; Xiao, T.; Li, H.; Yi, S.; Wang, X. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1918–1927.
25. Cubuk, E.D.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, Q.V. AutoAugment: Learning augmentation policies from data. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 113–123.
26. Pan, X.; Luo, P.; Shi, J.; Tang, X. Two at once: Enhancing learning and generalization capacities via IBN-Net. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 484–500.
27. Radenovic, F.; Tolias, G.; Chum, O. Fine-tuning CNN image retrieval with no human annotation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1655–1668. [CrossRef] [PubMed]
28. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
29. Sun, Y.; Cheng, C.; Zhang, Y.; Zhang, C.; Zheng, L.; Wang, Z.; Wei, Y. Circle Loss: A unified perspective of pair similarity optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6397–6406.
30. Liu, X.; Liu, W.; Mei, T.; Ma, H. A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9906, pp. 869–884.
31. Liu, H.; Tian, Y.; Wang, Y.; Pang, L.; Huang, T. Deep Relative Distance Learning: Tell the Difference between Similar Vehicles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Amsterdam, The Netherlands, 11–14 October 2016; pp. 2167–2175.
32. Wang, G.; Yuan, Y.; Chen, X.; Li, J.; Zhou, X. Learning discriminative features with multiple granularities for person re-identification. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 274–282.
33. He, S.; Luo, H.; Wang, P.; Wang, F.; Li, H.; Jiang, W. Transreid: Transformer-based object re-identification. In Proceedings of the International Conference on Computer Vision, Online, 11–17 October 2021; pp. 14993–15002.