MULTI-VIEW FEATURE FUSION NETWORK FOR VEHICLE RE-IDENTIFICATION

Haoran Wu, Dong Li, Yucan Zhou, and Qinghua Hu

School of Computer Science and Technology,
Tianjin University, Tianjin, China

ABSTRACT

Identifying whether two vehicles in different images are same or not is called vehicle re-identification. In cities, there are lots of cameras, but cameras cannot cover all the areas. If we can re-identify a car disappearing from one camera and appearing in another in two adjacent regions, we can easily track the vehicle and use the information help with traffic management. In this paper, we propose a two-branch deep learning model. This model extracts two kinds of features for each vehicle. The first one is license plate feature and the other is the global feature of the vehicle. Then the two kinds of features are fused together with a weight learned by the network. After, the Euclidean distance is used to calculate the distance between features of different inputs. Finally, we can re-identify vehicles according to their distance. We conduct some experiments to validate the effectiveness of the proposed model.

KEYWORDS

vehicle re-identification, deep learning model, feature fusion

1. INTRODUCTION

With the development of economy, urban traffic management becomes extremely important. Usually, large cities have many monitoring cameras, and these cameras can be used for vehicles tracking. However, for economic and privacy reasons, it is difficult for these monitoring cameras to cover all areas. In this situation, re-identifying a vehicle when it disappears from one camera and appears in another is necessary for tracking. In addition to that, vehicle re-identification also benefits to criminal investigation, urban construction and so on.

Usually, there are two problems in re-identification: finding a feature used for re-identification and finding a suitable method for distance metrics. The feature should be robust in different situations. The distance metrics method aims to learn a distance which is able to distinguish inner class and inter class. In vehicle re-identification based on monitor video, we can use license plates to identify whether two cars are the same or not, as the license plate is the only identification of a vehicle. Certainly, there has been a lot of effective license plate recognition methods proposed in recent years. Some methods use traditional image processing techniques [1, 2] and some of them take use of the neural network [3]. However, in real situation, the license plates are not always identified. When identifying license plate, some plates are too fuzzy to be recognized as shown in Fig. 1 (a). There are also some vehicles without license plates like Fig. 1 (b). In condition of

Natarajan Meghanathan et al. (Eds) : ICAITA, Signal, SAI – 2017
pp. 39– 47, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71304

traffic congestion, some vehicles are shielded by vehicles in front of them like Fig. 1 (c). In these conditions, license plates are useless for vehicles re-identification. Thus, a more common feature is required for vehicle re-identification. Since Geoffrey and Alex put forward the AlexNet in 2012, it is popular to use deep neural network for extracting features [4, 5, 6]. For the vehicle images in which the license plates are hard to be recognized, we can use a deep neural network to extract global vehicle features. However, different from person re-identification, the vehicles of the same type usually have little difference. In this situation, the deep neural features maybe useless for vehicle re-identification.

(a) fuzzy (b) unlabeled (c) shield

Fig. 1. Sample images. (a) is a car whose plate is fuzzy. (b) is a vehicle without plate. (c) is a car whose plate is shielded by the car in front of it.

In order to solve the problems above, we propose a deep neural network, called Multi-View Feature Fusion Network (MVFFN). MVFFN is a two-branch network. This net aims to adapt to the following situations: license plates are hard to be recognized and vehicle types are similar. It fuses license plate features and vehicle features together according to different situations. If the effect of license plate features is small, the weights of license plates features will be small and the weights of vehicle features will be heavier. On the contrary, when re-identifying two similar cars, the weights of vehicle features will be small and weights of license plate features will be big. Experiments are conducted on dataset collected by ourselves and the result proves our method to be very effective.

The rest of the paper is organized as follows: Section 2 will give a review of related work. Section 3 will introduce the model and the specifically designed fusion method. Section 4 describes the experiments and results. The conclusion will be drawn in Section 5.

2. Related Work

2.1. Feature Fusion Network

The previous re-identification works focus on person re-identification, and the purpose of person re-identification is similar to vehicle re-identification. Therefore, the researches of person re-identification are helpful to our work.

In the past, the person re-identification techniques usually rely on some handcrafted features, like color and texture histogram [10, 11]. With the development of deep neural network, there are also some work using deep neural networks [8, 12]. In the work, the Feature Fusion Network (FFN) [8] combines handcrafted features and deep neural features together, which gives us inspiration.

FFN extracts features by a deep neural network and hand-crafted method ELF [9], respectively. Then the two kinds of features are fused in the fusion layer. The output of the fusion layer is calculated by:

$$Z(x) = h(W^T x + b) \tag{1}$$

where x is the concatenation of two kinds of features and h is the ReLu function. W and b are weights and bias, respectively.

Then the output is sent to a softmax loss layer. Finally, the features extracted from the fusion layer are used to calculate the distance between different inputs. The distance metric method is Mirror Kernel Marginal Fisher Analysis (Mirror KMFA) [9].

2.2. Deep Relative Distance Learning

There are some previous researches on vehicles re-identification. The PROVID [17] also uses the vehicle features and license plate features, but the vehicle features are extracted by conventional methods. Besides, the PROVID is a layered structure. The most related work of vehicle re-identification we found is Deep Relative Distance Learning [13]. It proposes a Deep Relative Distance Learning (DRDL) model. The idea of this method is: two vehicles are different if the models of them are different, if they are vehicles of the same model then other features are applied. This model is based on VGG_CNN_M_1024. It has two branches, the first branch is VGG followed by an attribute recognition loss. This branch is designed to identify whether the inputs are vehicles of the same model or not. The other branch is VGG and coupled cluster loss specially designed for this problem. This branch is used to distinguish vehicles belong to the same model. Two kinds of features are concatenated and sent to another coupled cluster loss.

The inputs of this network are two image sets: a positive set and a negative set. The coupled cluster loss are specially designed for the relative distance learning problem. The loss function is

$$L(W, X^P, X^N) = \sum_i^{N^P} \frac{1}{2} \max\{0, \left\| f(x_i^P) - c^P \right\|_2^2 + \alpha - \left\| f(x_*^n) - c^P \right\|_2^2\} \tag{2}$$

where c^P is the center of the positive set, which is the mean value of all positive samples. And x_*^n is the nearest negative sample to the positive set center point. α is a predefined constant parameter. When the distances of all the positive samples to the center point adds α are less than the distance of the nearest negative sample to the center point, the loss will be zero. On the other hand, the loss will be

$$\sum_i^P \frac{1}{2} (\left\| f(x_i^P - c^P) \right\|_2^2 + \alpha - \left\| f(x_*^n - c^P) \right\|_2^2) . \tag{3}$$

Inspired by their work, we also designed a two branch network to extract different features and fuse them together to get a more robust feature for vehicle re-identification.

3. METHODOLOGY

3.1. Network structure

The structure of the network is shown in Fig. 2. We formulate the problem as a binary classification problem just as the work by E. Ahmed et al. [14]. Therefore, we designed a network with a pair of inputs. The label of the paired inputs is 1 or 0 which represents whether they are the same or not. Each input in the inputs pair contains two images: one is the vehicle image and the other is the license plate image extracted from vehicle image by Faster R-CNN [7]. These four images are sent to VGG16 and get four 4096-dimension vectors. Then the features of the two license plates are sent to FC layer in which they are used to calculate the weight. After that, the vehicle feature and license plate feature are sent to the concatenate layer with the weight. In this layer, vehicle features and license plate features are connected together with the weight. Output of concatenate layer is an 8192-dimension vector. Finally, each branch will get an 8192-dimension feature fusion vector, and the two feature fusion vectors are used to calculate Euclidean distance. The Euclidean distance is used to judge whether the pair of inputs are from the same vehicle or not.

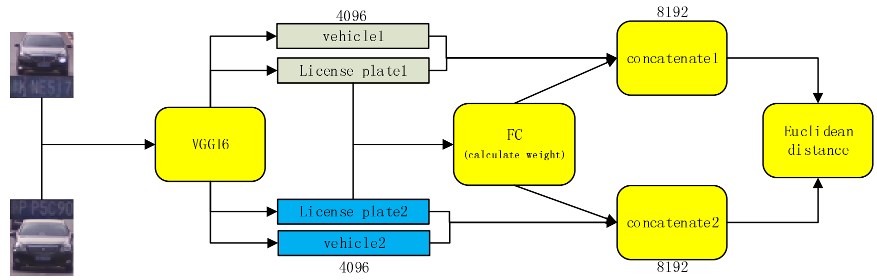


Fig. 2. Multi-View Feature Fusion Network for vehicles re-identification

3.2. Feature Fusion

The features of the input images are extracted by VGG16, the outputs of VGG16 are 4096-dimension vectors. The vehicle features and license plate features of the first input are $\{a_1, b_1\}$. The vehicle features and license plate features of the second input are $\{a_2, b_2\}$. For each input, we use a weight to integrate the vehicle feature and the license plate feature. Therefore, we design a FC layer to calculate the weight.

The input of this layer is

$$x = b_1 - b_2, \tag{4}$$

x is different between the two license plate features. The output of this layer is

$$w = f(W^T x) \tag{5}$$

where W is a 4096-dimension vector, f is the sigmoid function, and w is the weight calculated by this layer.

After we obtain the weight w, the vehicle feature and the license plate feature are fused in the concatenation layer. They are fused as

$$feature = [(1-w)*a, w*b] \tag{6}$$

The output of the concatenation layer is an 8192-dimension vector.

3.3. Contrastive loss

The loss function of this model is the contrastive loss. The contrastive loss is

$$L = \frac{1}{2N}\sum_{n=1}^{N} yd^2 + (1-y)\max(margin - d, 0)^2 \tag{7}$$

where y is the label of the input pair and margin is the threshold. The distance d between two fused features is calculated by

$$d = \left\| feature_1 - feature_2 \right\|_2^2 \tag{8}$$

where $feature_1$ and $feature_2$ represent fused features of two inputs, respectively.

4. EXPERIMENT

4.1. Data sets

There are few open datasets in this area. The only dataset we found is VehicleID [13], there is not license plate information, it is unsuitable in our experiment. We collect a new dataset for vehicle re-identification. The dataset is collected from real-world video, which is obtained from a busy street in Tianjin, China. Some images of this dataset are shown in Fig. 3.

In this dataset, there are several images of one vehicle, so it is suitable for the vehicle re-identification task. The split of this dataset is given in Table 1.

Fig. 3. Images in dataset. Each line is the images of one vehicle.

Table 1. Data split for our dataset

Class number	Train images number	Probe images number	Gallery images number
389	4276	1222	2334

4.2. Training strategies

The training pairs of images are chosen from the training images. The positive pairs and negative pairs ratio is 1 to 3. Meanwhile, 10 percent of the training data are used for validation.

We use mini-batch stochastic gradient descent (SGD) [15] in our experiment. The initial learning rate is set as $\gamma = 0.001$, and the learning rate decreases by $\gamma_{new} = \gamma / number\_epoach$. The initialized method of weights of FC layer is random initialization.

4.3. Weights distribution in MVFFN

The weights used to fuse two kinds of features are variable with the change of input pairs of images. That is, according to different situation, the network can learn a suitable weight to fuse the two kinds of features.

Some weights are shown in Table 2. and the corresponding gallery license plate images are shown in Fig. 4.

The first probe license plate is the plate of the first car, the 3rd gallery license plate belongs to the same car and it is clear, so the weight of the 3rd gallery is very high. The 13th gallery license plate is relatively obscure, in this condition, the weight is small. The 91st gallery license plate belongs to an unlicensed vehicles, so the weight will also be small. The 162nd gallery license plate is clear, so the weight is large.

Table 2. Some weights of the first license plate images

Gallery plate number	3	13	91	162
weight	0.8506	0.4324	0.5158	0.8501

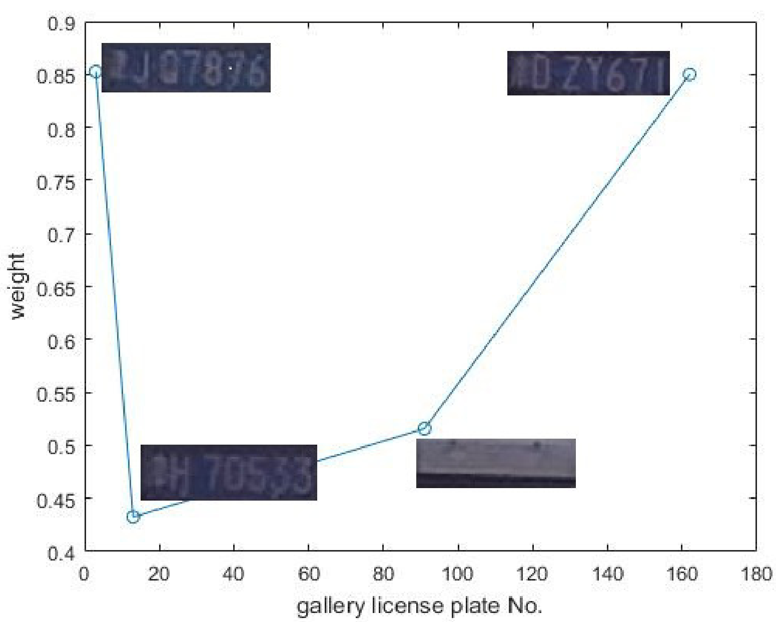


Fig. 4. Weights and corresponding gallery license plate images

4.4. Experiment results and analysis

To verify the effectiveness of our method, we design three experiments. We use the vehicle features and license plate features to re-identify vehicles, respectively. Another method is to concatenate the vehicle features and license plate features together. We use the cumulative match curve (CMC) [16] to evaluate the experimental results. The CMC of different methods is shown in Table 3. And the match rates from top-1 to top-10 are illustrated in Fig. 5.

From the results we can see that it is effective to use license plate information to re-identify vehicles. The vehicle information is not as powerful as license plate information but it is also useful especially when the license plate information is missing. If we just concatenate two features together, the top-1 matching rate of concatenation method is less than that of using license plate features by about 5%. The result shows that feature fusion can lead to a worse result without a good algorithm. Our method produces the best performance in each level. It beats the second-rank method by 1.48 percent in top-1 matching rate. We can get from the experiment results that our method is effective in the real environment.

Table 3. CMC of Vehicle Re-identification task

Method	Top1	Top3	Top5	Top10
Vehicle	58.43%	68.09%	72.67%	79.13%
License plate	92.55%	95.99%	97.22%	98.45%
Vehicle + plate	86.17%	91.82%	94.03%	95.99%
Ours	94.03%	97.22%	98.61%	98.94%

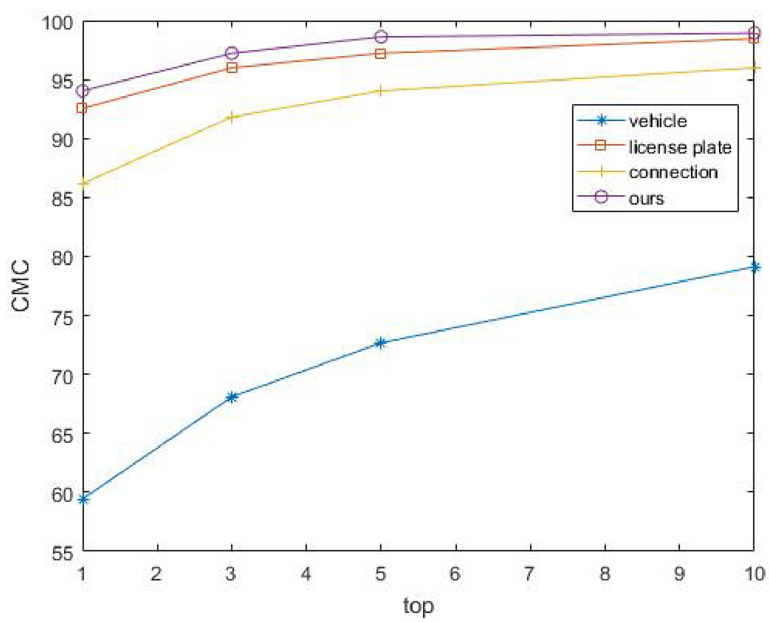


Fig. 5. CMC of vehicle re-identification

Vehicle features are the most basic features for vehicle re-identification. However, in the real world, vehicle features between similar vehicles are hard to be distinguished. Besides, the background, viewpoint and illuminations will also influence the results. License plate features are suitable for this work and it gets a high accuracy rate. But for these reasons mentioned in Section 1, license plate features are useless in certain scenarios. Feature fusion is a good idea, but a proper fusion algorithm is needed. With our method, we can get a weight with each input pair of images.

When the license plate information is missing, the vehicle information will be more important. On the contrary, if the license plate information can re-identify vehicles with high confidence the weight of vehicle information will be smaller.

5. CONCLUSION

In this paper, a Multi-View Feature Fusion Network is proposed for vehicle re-identification. This model aims to solve two problems in the real environment: vehicles of the same model is difficult to re-identify and in some cases the license plate information is missing. This model combines two kinds of features (vehicle features and license plate features) with a weight w. The weight can be trained according to different situations. Compared with other features, the fused features achieve higher predict accuracy. Experimental results show that the fusion strategy is effective and useful.

REFERENCES

[1] Nijhuis, J. A. G., Ter Brugge, M. H., Helmholt, K. A., Pluim, J. P. W., Spaanenburg, L., & Venema, R. S., et al. (1995). Car license plate recognition with neural networks and fuzzy logic. IEEE International Conference on Neural Networks, 1995. Proceedings (Vol.21, pp.2232-2236 vol.5). IEEE Xplore.

[2] Naito, T., Tsukada, T., Yamada, K., & Kozuka, K. (2000). Robust license-plate recognition method for passing vehicles under outside environment. Vehicular Technology IEEE Transactions on, 49(6), 2309-2319.

[3] Anagnostopoulos, C. N. E., Anagnostopoulos, I. E., Loumos, V., & Kayafas, E. (2006). A license plate-recognition algorithm for intelligent transportation system applications. IEEE Transactions on Intelligent Transportation Systems, 7(3), 377-392.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems (Vol.25, pp.1097-1105). Curran Associates Inc.

[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D., et al. (2014). Going deeper with convolutions. 1-9.

[6] Arora, S., Bhaskara, A., Ge, R., & Ma, T. (2013). Provable bounds for learning some deep representations. Eprint Arxiv, 584-592.

[7] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(6), 1137-1149.

[8] Wu, S., Chen, Y. C., Li, X., Wu, A. C., You, J. J., & Zheng, W. S. (2016). An enhanced deep feature representation for person re-identification. 1-8.

[9] Lai, J., Lai, J., & Lai, J. (2015). Mirror representation for modeling view-specific transform in person re-identification. International Conference on Artificial Intelligence (pp.3402-3408). AAAI Press.

[10] Farenzena, M., Bazzani, L., Perina, A., Murino, V., & Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. Computer Vision and Pattern Recognition (Vol.23, pp.2360-2367). IEEE.

[11] Li, Z., Chang, S., Liang, F., Huang, T. S., Cao, L., & Smith, J. R. (2013). Learning Locally-Adaptive Decision Functions for Person Verification. IEEE Conference on Computer Vision and Pattern Recognition (Vol.9, pp.3610-3617). IEEE Computer Society.

[12] Yi, D., Lei, Z., & Li, S. Z. (2014). Deep metric learning for practical person re-identification. Computer Science, 34-39.

[13] Liu, H., Tian, Y., Wang, Y., Pang, L., & Huang, T. (2016). Deep Relative Distance Learning: Tell the Difference between Similar Vehicles. Computer Vision and Pattern Recognition (pp.2167-2175). IEEE.

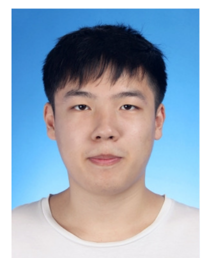
[14] Ahmed, E., Jones, M., & Marks, T. K. (2015). An improved deep learning architecture for person re-identification. Computer Vision and Pattern Recognition (pp.3908-3916). IEEE.

[15] Bottou, L. (2012). Stochastic Gradient Descent Tricks. Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg.

[16] Gray, D., Brennan, S., & Tao, H. (2007). Evaluating appearance models for recognition, reacquisition, and tracking.

[17] Liu, X., Liu, W., Mei, T., & Ma, H. (2016). A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance. European Conference on Computer Vision (pp.869-884). Springer International Publishing.

AUTHORS



Haoran Wu is a student of Tianjin University, and this work was done when Haoran Wu was a postgraduate student.