Journal of New Media
DOI: 10.32604/jnm.2022.028775
Article

Tech Science Press

Vehicle Matching Based on Similarity Metric Learning

Yujiang Li[1,2], Chun Ding[1,2] and Zhili Zhou[1,2,*]

[1]Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[2]School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing,210044, China
*Corresponding Author: Zhili Zhou. Email: zhou_zhili@163.com

Received: 17 February 2022; Accepted: 18 March 2022

Abstract: With the development of new media technology, vehicle matching plays a further significant role in video surveillance systems. Recent methods explored the vehicle matching based on the feature extraction. Meanwhile, similarity metric learning also has achieved enormous progress in vehicle matching. But most of these methods are less effective in some realistic scenarios where vehicles usually be captured in different times. To address this cross-domain problem, we propose a cross-domain similarity metric learning method that utilizes the GAN to generate vehicle images with another domain and propose the two-channel Siamese network to learn a similarity metric from both domains (i.e., Day pattern or Night pattern) for vehicle matching. To exploit properties and relationships among vehicle datasets, we first apply the domain transformer to translate the domain of vehicle images, and then utilize the two-channel Siamese network to extract features from both domains for better feature similarity learning. Experimental results illustrate that our models achieve improvements over state-of-the-arts.

Keywords: Vehicle matching; cross-domain; similarity metric learning; two-channel siamese network

1 Introduction

With the popularization of vehicles and the rapid development of traffic, the demand for obtaining traffic information on camera equipment is also increasing. Vehicle matching entification has many practical applications like video surveillance systems that aims to identify a target vehicle in different cameras in different conditions, such as multi-viewpoint, and day or night patterns.

Previous works [1,2] mainly directly identify vehicles by the license plate. However, the license plates will not be captured clearly because of the various lightings, viewpoints, backgrounds and resolutions. the performance of the matching method based on the license plate captured by the camera drops dramatically. Therefore, some researchers [3] to solve the vehicle matching task is to utilize the similarity metric learning by the Siamese network [4]. The main idea of these methods is to learn the similarity matrices of appearance features for vehicle matching. However, these methods directly extracted features of vehicle appearance without consideration of the cross-domain problem. Unfortunately, the existing networks have limited ability to extract features of night pattern images, causing a fateful performance drop when taking corresponding images from different domains as input.

Inspired by researches on the cross-domain [5,6], this paper proposes a cross-domain similarity metric learning method to investigate the cross-domain vehicle matching task. Specifically, the network first translates the domain to the other, and then the Siamese network calculates the similarity metric by the unified domain features. Besides, the proposed two-channel Siamese network not only preserves the features with day pattern pair, but also extracts the features of night pattern pair to capture the more information details. The extensive experiments demonstrate that the proposed method improves vehicle matching accuracy significantly. In summary, our contributions can be summarized into three aspects:

1) A framework based on the cross-domain similarity metric learning is designed for vehicle matching. In this framework, we unify the domain of input image pairs and then feed it into the Siamese network to calculate their distance metrics. The proposed framework can solve the cross-domain problem of vehicle matching well.
2) To address that network has poor ability to extract features of night pattern images, the two-channel Siamese Siamese network is proposed that not only extracts the day pattern image features, but also night pattern image features. The two-channel Siamese network calculates the similarity metrics from both domains.
3) We conduct extensive experiments to show that the proposed method outperforms state-of-the-arts on the VehicleID and VERI-Wild datasets.

2 The Status of Research

With the rapid development of deep learning, the researchers use deep learning-based feature representations for vehicle matching. Li et al. [7] proposed a DJDL model that utilized a deep convolutional network to effectively extract discriminative representations for vehicle images. Wang et al. [8] proposed the Triplet Center Loss based Part-aware Model (TCPM) that leverages the discriminative features in part details of vehicles to refine the accuracy. Zhou et al. [9] learn transformations across different viewpoints of vehicles by the proposed model which is combined with CNN and LSTM. Shen et al. [10] designed the spatial graph network (SGN) to elaborately explore the spatial significance of feature maps

In recent years, various methods have explored similarity metrics to handle the vehicle matching tasks. The main idea of similarity metric learning is features that belong to the same class are kept closed and differences are distant. Some existing networks such as Siamese [4] and pseudo-Siamese networks [11] utilize similarity metrics for vehicle matching tasks. Liu et al. [12] exploits a two-branch deep convolutional network to project raw vehicle images into Euclidean space and use the distance to measure the similarity of a pair of vehicles. Liu et al. [13] proposed the FACT model use the Euclidean and cosine distance metrics to measure the similarity between the pair of vehicles for matching. Deng et al. [14] calculate the composite similarity score of spatio-temporal patterns with Siamese neural-network-based classifier visual features. At present, many GAN-based methods are proposed to solve cross-domain problems. Taigman et al. [5] translate a sample from one domain to an analog sample from another domain such as face photos to emoji by the proposed Domain Transfer Network (DTN). In [15], the method can translate an image from a source domain to a target domain without pairing examples.

Although these works reach remarkable success in vehicle matching tasks, the performance always be terrible when these vehicle images belong to the different domains. Thus, the difference between the vehicles day and night patterns have caused difficulties and challenges.

3 The Proposed

In this section, we illustrate the details of our proposed method. Specifically, we first introduce the pattern discrimination, and then introduce domain transition. At last, the two-channel Siamese network learns the similarity metric on the basis of unified domain. As shown in Fig. 1, the framework of our proposed method was composed with three parts: the pattern discrimination, the domain transition and the similarity metric learning.

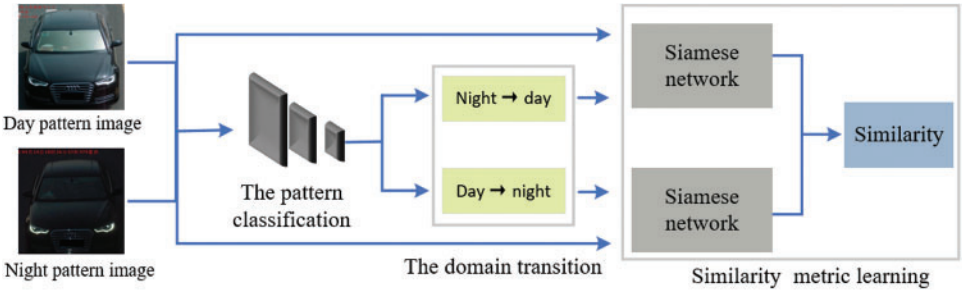

Figure 1: The framework of the proposed method

3.1 Framework

As shown in Fig. 1, the proposed framework was composed with three parts: the pattern classification, the domain transition and the similarity metric learning.

The pattern discrimination was applied at the front of the framework, which consists of a lightweight network called Resnet10. We discriminate the day-night pattern of a pair of images respectively to make different processing for different domains.

The domain transition uses a transformer which is a pre-train network called Cyclegan [15]. The generator recovers low-level features from high-level features by deconvolution layer. These low-level features will be translated to another domain. Batch normalization and ReLU are adopted for all the layers in the discriminator as well. Briefly, the discriminator extracts the feature of inputs and then discriminates whether features belong to the right class.

In the stage of similarity metric learning, we propose the two-channel Siamese structure, which could extract features from target domain without loss of source domain feature and have better generalization ability for cross-domain similarity metric learning. Inputs of the two-channel Siamese are selected by day pattern pairs and the night pattern pairs. Positive and negative samples are inputted with equal probability for the balance dataset. The two-channel Siamese network map the inputs to the new feature space respectively, and the similarity of two inputs is evaluated by calculating the loss value.

Foremost, choosing a pair of images as inputs of the gating to discriminate their pattern. The images will be fed into the domain transformer to translate their domain to the other, i.e., day pattern or night pattern. In the above manner, we ensure that each pair of the inputs are from the same domain. The feature representation from the same domain is beneficial to the similarity metric learning. The pairs with day pattern and night pattern will be fed into the two-channel Siamese network to learn the similarity metrics, respectively.

3.2 Pattern Classification

We propose the gating to discriminate the image patterns. This dataset is defined as two labels for the pattern discrimination: day and night. The pattern discrimination is critical to the next stages. On the one hand, it will make the domain transition to transform the specific domain, on the other hand, the images from two different domains will be fed into the corresponding branches correctly.

The input sample is defined as x_c, $c \in \{day, night\}$ means the class of sample. As shown in the following function, the affinity score $F_k(x)$ is usually calculated by linearly transforming the feature vector as:

$$F_k(x) = w_k^T x + b_k, \quad k \in \{day, \ night\} \tag{1}$$

The linear functions of all the k classes are combined to form a linear transformation layer, where the w_k and b_k are trainable parameters. The posterior probability of belonging to a certain class is computed as:

$$P\ (c|x) = \frac{e^{F_c(x)}}{e^{F_{day}(x)} + e^{F_{night}(x)}} \tag{2}$$

The larger value of the affinity score $F_k(x)$ indicates the higher posterior probability of x belonging to the class c.

3.3 Domain Transition

We utilize the pre-train network to translate the domain, which based on the pix2pix framework of Isola et al. This framework uses conditional generative adversarial networks to learn the mapping from input to output images. This network learns the mapping functions between two domains A and B.

The input of GAN is defined by x^d, where $d \in \{A, B\}$ means day and night pattern. We denote the data distribution as $x \sim pdata(x^A)$ and $x \sim pdata(x^B)$. The network is composed of two mapping $G_A{:}B \rightarrow A$, $G_B{:}A \rightarrow B$, and two adversarial discriminators D_A, D_B, where G aims to translate the image domain and D aims to distinguish between original images and translated images. We aim to map each input image of domain A x_i^A to domain B representations Fx_i^B by the following mapping function:

$$\mathcal{L}_{GAN}(G_A,\ D_B,\ x^A,\ x^B) = E_{x\sim pdata(x^B)}[\log D_B(x^B)] + E_{x\sim pdata(x^A)}]\log(1 - D_B(Fx^B)] \tag{3}$$

The network combines the two discriminators and two generators, it is beneficial to the network to translate different input images to different output images. The same set of input images may be inputted to any random permutation of images in another domain, where any of the learned mappings can induce an output distribution that matches the target distribution. The full loss function is:

$$\mathcal{L}(G_A,\ G_B,\ D_A,\ D_B) = \mathcal{L}_{GAN}(G_A,\ D_{B,}x^A,\ x^B) + \mathcal{L}_{GAN}(G_B,\ D_A,\ x^B,\ x^A) + \lambda\mathcal{L}_{cyc}(G_A,\ G_B) \tag{4}$$

3.4 Similarity Metric Learning

The pairs of day pattern images and night pattern images were inputted into the two-channel Siamese network, and then measure the distances of input pairs which belong to the same domain. To make it clear, these samples are defined by x_1, x_2, x_3, x_4, x_1, $x_2 \in \{x^A, x^B\}$ and x_3, $x_4 \in \{x_+, x_-\}$. x^A means the day pattern image, x^B means the night pattern image, x_+ and x_- are positive and negative sample.

One of the distance D is defined as follows:

$$D(x_1,\ x_3) = \|x_1 - x_3\|_2 = (\Sigma_{i=1}^{P}(x_1 - x_3)^2)^{\frac{1}{2}} \tag{5}$$

We aim to pull features from the same class closer to each other and push features from different classes away by adopting the contrastive loss as follows:

$$L(W,\ (Y,\ x_1,\ x_2)) = \frac{1}{2N}\sum_{n=1}^{N} YD_W^2 + (1 - \mathrm{Y})\max(\mathrm{m} - D_w,\ 0)^2 \tag{6}$$

This loss function can expression of matching degree of paired samples. If the two inputs are the same vehicle, the output features will be spatially close. If not, the output features will be spatially distant. Y means whether the input samples belong to the same class and the loss function will be:

$$L_s = \frac{1}{2N}\sum_{n=1}^{N} YD_W^2 \tag{7}$$

$$L_d = (1 - \mathrm{Y})\max(\mathrm{m} - D_w,\ 0)^2 \tag{8}$$

When the samples belong to the different vehicles, if the Euclidean distance of its feature space is decline, the loss value will be increase. Thus, we train the network to learn the similarity metric of vehicle images by reducing the contrastive loss value.

We propose the two-channel Siamese network to learn a similarity metric for both day pattern and night pattern. The feature of the night pattern image also contains some meaningful information even it is not sufficient to support the similarity learning for the network. The domain transition will cause the loss of some feature details. Thus, we calculate the similarity metric of images and fuse with the similarity metric of another domain, and it is beneficial to the generalization ability of network by calculating the loss.

4 Experiments

4.1 Datasets and Evaluation Criteria

VehicleID dataset [13]. VehicleID is captured by multiple non-overlapping cameras and there are 221,763 images of 26,267 vehicles in total. Each image is either captured from the front view or back view. In VehicleID, only 250 vehicle models are included, which means many different identities share same vehicle model.

VERI-Wild dataset [16]. VERI-Wild is a large-scale vehicle ReID dataset containing 416,314 vehicle images of 40,671 identities. It is captured by a traffic surveillance camera system consisting of 174 cameras across one month under unconstrained scenarios. The images not only captured by complex backgrounds and various viewpoints, but also under various weather and illumination conditions. This dataset is randomly divided into a training set with 277,797 of 30,671 vehicle IDs and a testing set with 138,517 images of 10,000 vehicle IDs.

The mean average precision (MAP) which is computed from its precision-recall curve (P-R) is adopted in our experiments.

4.2 Performance Evaluation and Comparison

Here we test the influence of different parameter combinations on the experimental results. There are five combinations: {0, 1}, {0.1, 0.9}, {0.2, 0.8}, {0.3, 0.7}, {0.4, 0.6}, {0.5, 0.5}and we set $w_1 = 0$ is equivalent to remove distance metric of night pattern images. The effects of W of the proposed method are shown in Tab. 1. Compared to other combines, the accuracy of our method dramatic decline when the values of W are set as {0, 1} and {0.5, 0.5}. The features of the day pattern vehicle images are not informative enough. While the fused distance contains too much similarity metric of night pattern images also take negative impacts on similarity metric learning. We set W as {0.3, 0.7}to achieve the optimal performance.

Table 1: The effects of parameter combinations

W_1, W_2)	VehicleID	VERI-wild
(0, 1)	94.55	90.73
(0.1, 0.9)	94.21	90.25
(0.2, 0.8)	94.71	91.72
(0.3, 0.7)	95.74	92.55
(0.4, 0.6)	95.12	91.85
(0.5, 0.5)	92.68	89.12

The quantity of night pattern images is much less than the day pattern images. Thus, we translate the images from day pattern to night pattern to change the percentage of the images. It is clearly observed in the Fig. 2, the performance of proposed the method without domain transition drops when the percentage of night pattern image increase. Because the features of night pattern images contain less information, and the network will have less ability to learn the similarity metric.

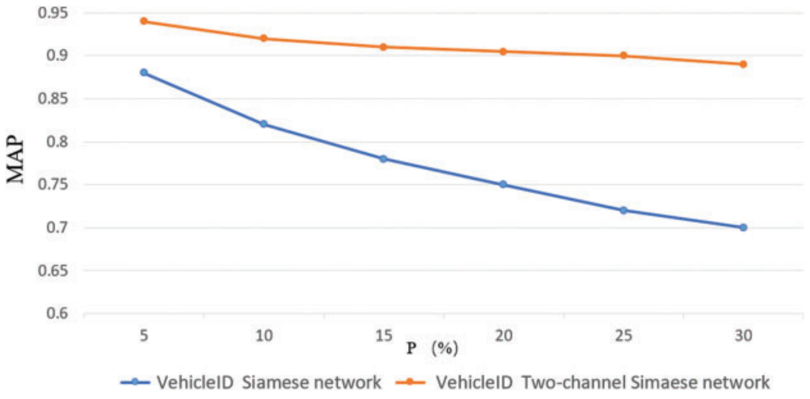

Figure 2: Accuracy comparison of two different methods

We compare our method on VehicleID and VERI-Wild datasets with several state-of-the-art methods, including LABNet [17], LABNet-50 [17], PVEN [18] and DMML [19]. As shown in Tabs. 2 and 3, our method obtains the superior results. There are lots of night pattern images in these datasets, while other methods have not taken consideration into the cross domain problem.

Table 2: The detection accuracies (MAP) of different methods on the VERI-wild dataset

Methods	VERI-wild MAP
LABNet [17]	82.61
LABNet-50 [17]	81.05
PVEN [18]	82.53
Ours	92.55

Table 3: The detection accuracies (MAP) of different methods on the VehicleID dataset

Methods	VehicleID MAP
LABNet [17]	89.63
LABNet-50 [17]	87.54
DMML [19]	87.37
Ours	95.72

5 Conclusions

This paper proposes the cross-domain similarity metric learning method for vehicle matching through a two-channel Siamese network. In the proposed method, we first discriminate the day-night pattern of a pair of images respectively and translate their domain to the other. Then, the network can learn the similarity metric between pairs of vehicle images whether they belong to the same domain or not because we calculate the distance matrix of both domains. Experimental results confirm that that the proposed method brings substantial improvements to vehicle matching accuracy. However, the proposed method relies on an extra network to distinguish the domain. In the future, we will make the domain transition generate images based on demand, rather than requiring an additional discriminant.

Acknowledgement: Thanks to the teacher and students of my team for their guidance in the process of completing this article. We gratefully acknowledge the support of Engineering Research Center of Digital Forensics for providing us the RTX3090 GPU which was used in this research.

Funding Statement: This work is supported in part by the National Natural Science Foundation of China under Grant 61972205, in part by the National Key R&D Program of China under Grant 2018YFB1003205.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] J. Špaňhel, J. Sochor and R. Juránek, "Holistic recognition of low quality license plates by CNN using track annotated data," in *2017 14th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Lecce, Italy, pp. 1–6, 2017.

[2] V. Jain, Z. Sasindran, A. Rajagopal, S. Biswas *et al.,* "Deep automatic license plate recognition system," in *Proc. of the Tenth Indian Conf. on Computer Vision, Graphics and Image Processing*, Guwahati, Assam, India, pp. 1–8, 2016.

[3] I. O. de Oliveira, K. V. Fonseca and R. Minetto, "A Two-stream siamese neural network for vehicle re-identification by using non-overlapping cameras," in *2019 IEEE Int. Conf. on Image Processing (ICIP)*, Taipei, China, pp. 669–673, 2019.

[4] S. Chopra, R. Hadsell and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, pp. 539–546, 2005.

[5] Y. Taigman, A. Polyak and L. Wolf, "Unsupervised cross-domain image generation," arXiv preprint arXiv:1611.02200, 2016.

[6] T. Kim, M. Cha, H. Kim, J. K. Lee and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Int. Conf. on Machine Learning*, Not Wiki, pp. 1857–1865, 2017.

[7] Y. Li, Y. Li, H. Yan and J. Liu, "Deep joint discriminative learning for vehicle re-identification and retrieval," in *2017 IEEE Int. Conf. on Image Processing (ICIP)*, Beijing, China, IEEE, pp. 395–399, 2017.

[8] H. Wang, J. Peng and G. Jiang, "Discriminative feature and dictionary learning with part-aware model for vehicle re-identification," *Neurocomputing*, vol. 438, pp. 55–62, 2021.

[9] Y. Zhou, L. Liu and L. Shao, "Vehicle re-identification by deep hidden multi-view inference," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3275–3287, 2018.

[10] F. Shen, J. Zhu and X. Zhu, "Exploring spatial significance via hybrid pyramidal graph network for vehicle re-identification," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[11] S. Zheng, Y. Song, T. Leung and I. Goodfellow, "Improving the robustness of deep neural networks via stability training," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 4480–4488, 2016.

[12] H. Liu, Y. Tian, Y. Yang, L. Pang and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 2167–2175, 2016.

[13] X. Liu, W. Liu, H. Ma and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *2016 IEEE Int. Conf. on Multimedia and Expo (ICME)*, Seattle, WA, USA, pp. 1–6, 2016.

[14] J. Deng, J. Cai and M. U. Aftab, "Visual features with spatio-temporal-based fusion model for cross-dataset vehicle Re-identification," *Electronics*, vol. 9, no. 7, pp. 1083, 2020.

[15] J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 2223–2232, 2017.

[16] Y. Lou, Y. Bai, J. Liu, S. Wang and L. Duan, "Veri-wild: A large dataset and a new method for vehicle re-identification in the wild," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 3235–3243, 2019.

[17] A. M. N. Taufique and A. Savakis, "LABNet: Local graph aggregation network with class balanced loss for vehicle re-identification," *Neurocomputing*, vol. 463, pp. 122–132, 2021.

[18] D. Meng, L. Li and X. Liu, "Parsing-based view-aware embedding network for vehicle re-identification," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 7103–7112, 2020.

[19] G. Chen, T. Zhang, J. Lu and J. Zhou, "Deep meta metric learning," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, Korea, pp. 9547–9556, 2019.