Deep Structural Feature Learning: Re-Identification of simailar vehicles In Structure-Aware Map Space

Wenqian Zhu
National Engineering Research
Center For Multimedia Software,
Wuhan University
China
zwq_pi@whu.edu.cn

Ruimin Hu
National Engineering Research
Center For Multimedia Software,
Wuhan University
China
hrm@whu.edu.cn

Zhongyuan Wang
National Engineering Research
Center For Multimedia Software,
Wuhan University
China
wzy_hope@163.com

Dengshi Li
National Engineering Research
Center For Multimedia Software,
Wuhan University
China
reallds@126.com

Xiyue Gao
National Engineering Research
Center For Multimedia Software,
Wuhan University
China
gao19920804@126.com

ABSTRACT

Vehicle re-identification (re-ID) has received more attention in recent years as a significant work, making huge contribution to the intelligent video surveillance. The complex intra-class and inter-class variation of vehicle images bring huge challenges for vehicle re-ID, especially for the similar vehicle re-ID. In this paper we focus on an interesting and challenging problem, vehicle re-ID of the same/similar model. Previous works mainly focus on extracting global features using deep models, ignoring the individual local regions in vehicle front window,such as decorations and stickers attached to the windshield, that can be more discriminative for vehicle re-ID. Instead of directly embedding these regions to learn their features, we propose a Regional Structure-Aware model (RSA) to learn structure-aware cues with the position distribution of individual local regions in vehicle front window area, constructing a FW structural map space. In this map sapce, deep models are able to learn more robust and discriminative spatial structure-aware features to improve the performance for vehicle re-ID of the same/similar model. We evaluate our method on a large-scale vehicle re-ID dataset Vehicle-1M. The experimental results show that our method can achieve promising performance and outperforms several recent state-of-the-art approaches.

KEYWORDS

Vehicle re-ID, Deep convolutional neural networks, Regional Structure Aware model

ACM Reference Format:
Wenqian Zhu, Ruimin Hu, Zhongyuan Wang, Dengshi Li, and Xiyue Gao. 2019. Deep Structural Feature Learning: Re-Identification of simailar vehicles In Structure-Aware Map Space. In *ACM Multimedia Asia (MMAsia '19), December 15–18, 2019, Beijing, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3338533.3366585

1 INTRODUCTION

Vehicle re-identification is to identify the same vehicle from massive surveillance videos or images. It is more challenging than vehicle detection or vehicle model recognition especially when identifying vehicles of the same/similar model [18] [5]. As a unique identification, license plate has been naturally and widely used for vehicle re-ID. However license plate has weak resistance to be attacked. It is easy to remove, alter or even fake a license plate, besides license plate recognition is also sensitive to view of camera and occlusion. Therefore vision-based vehicle re-ID is of great practical value in application of real-world surveillance.

There are few attempts realizing vehicle re-ID purely by vehicle's visual appearance. One of the most important reasons is that essential features of vehicles are difficult to express considering the complex variation between intra-class and inter-class and the changes happened to vehicle's appearance with time [5]. It turns to be more difficult when it comes to re-identifying vehicles of same/similar model. The similar visual features, such as the same motorcycle type, the same auto logos etc. give less identity information for vehicle re-identification making the task more challenging.

In recent years, Deep Convolutional Neural Network(DCNN) has achieved outstanding performance in various vision recognition tasks, such as large-scale objects classification [4], face recognition [10] and vehicle re-ID [5] etc. With the well-designed hierarchical structure, DCNN can automatically learn discriminative features of multiple levels and these deep features have shown obvious advantages over hand-craft features in handling visual problems [4] [14] .Therefore, more and more researchers pay attention to solve vehicle re-ID problem using CNN-based models [18] [5] [9] and some achievements have been made, while there are still

MMAsia '19, December 15–18, 2019, Beijing, China

ACM ISBN 978-1-4503-6841-4/19/12…$15.00
https://doi.org/10.1145/3338533.3366585

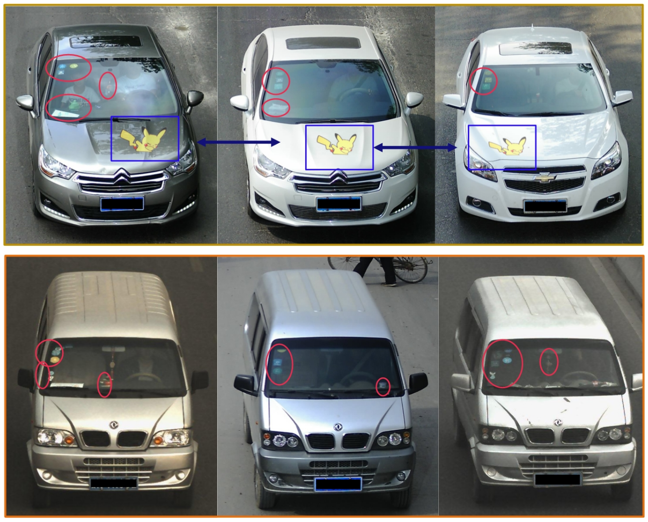

Figure 1: Each row shows three different vehicles of similar model.In the first row, the individual marks highlighted with blue rectangle can give misleading cues when performing vehicle re-ID with deep models which are quite sensitive to local texture features. The individual local regions in vehicle front window highlighted with red circles contain more discriminative structure-aware features to tell the similar vehicles apart considering their position distribution

some problems need to be solved. We list the most important two points as below:

- CNNs can make misclassification when identifying different vehicles having similar special marks (shown in Fig 1) for the reason that CNNs are more sensitive to local texture features than human do [2] and lack of ability to be spatially invariant to the input data [6].
- CNNs typically need a quite large amount of training data [13], while in the real world situations such as traffic video monitoring, there are not always enough discriminative vehicle images of multiple view for training.

How to overcome these problems to achieve better performance in vehicle re-ID has not been extensively studied. Moreover solving these problems can also facilitate us to expand the scope of application of vehicle re-ID and improve the accuracy of vehicle re-ID of the same or similar model.

As shown in Figure 1, for vehicles of the same/similar model, the front window region may contain more discriminative cues than the global appearance in many situations. We rely on this observation to extract structure-aware visual cues from front window to overcome the shortages of the CNN-based models we mentioned above. First we utilize the relative location distribution of each individual local regions in front window, such as decorations and stickers attached to the windshield, to construct the structure feature map, mapping raw vehicle image space to a structure feature map space in which CNNs can learn more robust and discriminative features. In that way, we can avoid misclassification mentioned in the first problem. And then the compactness and regularity of the structure map can facilitate us to simulate more training structure map in various visual angles by perspective transformation in a relative low cost to get over the limitation of the second problem.

In this paper, we propose a novel Regional Structure-Aware deep Model(RSA) for vehicle re-ID of the same/simailar model. As shown in Fig 2, the proposed model is composed of two parts both of which are based on CNN: Structural Map construction part and Structural Map recognition part. In the first part we learn the position distribution of each individual local regions in FW(front window) area to construct structure map for every vehicle.This process essentially map the raw vehicle image space containing massive local texture features with noise to Structural Map space that can give more robust and discriminative structure-aware cues for vehicle re-ID. In the recognition part, we make full use of the geometric property of Structural Map to simulate more Structural Map from various view angles by perspective transformation method and identify Structural Map with CNN-based model. In this way, CNN-based models could learn enough training data achieving state-of-the-art performance in vehicle re-ID of the same/similar model.

We evaluate our method on a large-scale vehicle re-ID datasets and the experimental results show our method can achieve promising results in comparison with recent works based on deep models. The contributions of our work can be summarized into two aspects:

- We first analysis the CNN-based methods which are most widely used in vehicle re-ID recently and pose the most important two "short boards" of this kind of method: strong sensibility to local texture features while weak spatial awareness and limitation of the scale of training data in real-world situations
- We propose a Structure-Aware deep Model that can learn structure-aware features which are more discriminative and robust to detailed individual cues on vehicles than global ones, succeeding to slove the two "short boards" mentioned towards a large extent.

We organize rest of the paper as follows: Related works reviewed in section 2.In section 3, we discuss Structure-Aware deep Model and give more details on structural map construction and identification.The experimental results and related analysis are presented in section 4

2 RELATED WORKS

Re-ID problem was first explored and applied in person or human face. Although vehicle re-ID has been proposed and discussed for many years, most previous works rely on a various of different non-vision based sensors [17] [8]. In recent years, several vision-based vehicle re-ID methods have been proposed and most of them rely on deep models, which achieved promising performance for vehicle re-ID.

Liu et al.[11] utilize both hand-crafted features like SIFT and deep features extracted from CNN model to realize vehicle re-ID. Experimental results showed that deep features have obvious advantages over hand-crafted features. Yet they use CNN model trained

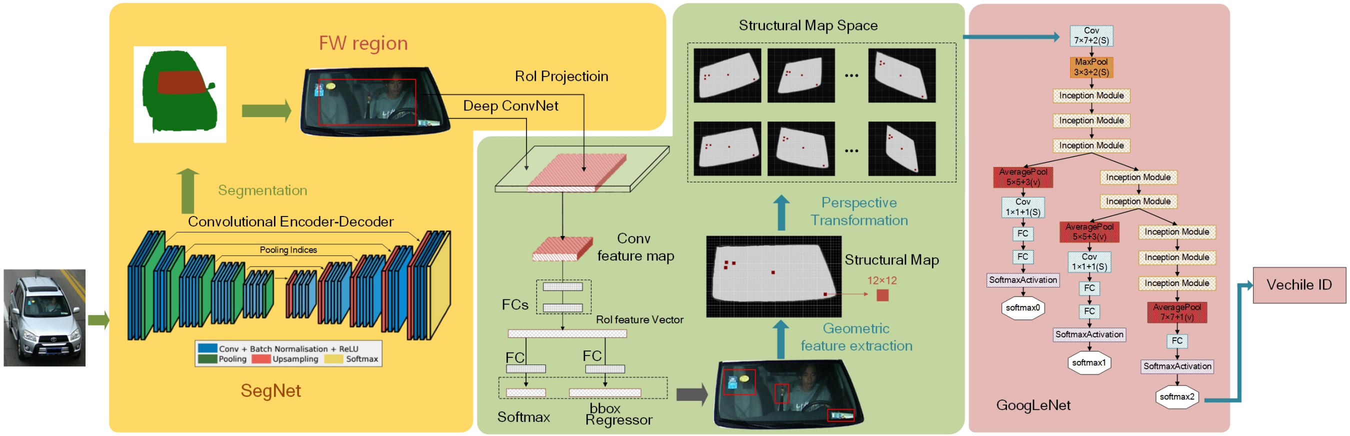

Figure 2: An overview of the architecture of RSA which consists of three CNN-based Deep model. SegNet is utilized to separate vehicle front region(FW region) precisely apart from raw vehicle images and individual local regions detection is done in FW region with Fast R-CNN. Each individual local region is uniformly represented by a 12×12 pixel area which has the same geometric center with the raw local region. The pixel areas and the front window region boundary constitute the structure map . With the compactness and regularity of the structure map, sufficient amount of structure map from multiple view can be simulated as training data by perspective transformation in a relative low cost. Finally, we exploit an efficient and classic deep model,GoogLeNet, to learn more rubost and discriminative structure-aware features in structure map space, realizing vehicle re-ID of the same/similar model

for vehicle model classification to extract features for vehicle re-ID. The learnt deep features could identify vehicles from different vehicle models, but have difficulties in telling difference between similar vehicles. Benefiting from new dataset for vehicle re-ID, Liu et al.[9] explored a two-branch Deep Convolutional Neural Network supervised by a coupled clusters loss and succeed to design a deep relative distance learning method for learning the difference between similar vehicles. Inspired by Liu's work, Guo et al.[5] proposed a coarse-to-fine feature embedding method to build a more efficient distance learning space for vehicle re-ID. They posed stronger constrains on the compactness between images of the same vehicle and enforced discrimination between different vehicle models by utilizing a series of well designed loss terms, mapping vehicle images to a deep feature embedding space where the distance between feature points directly reflects the semantic similarity of vehicles. As most CNN-based works take the global vehicle images as input, the descriptions they learned tend to depict the global features and may fail to acquire the discriminative power to local details. Realizing this significant fact, Liu et al. [12] propose a four branches CNN model sharing several convolutional layers to learn local features from vehicle regions, and fuse global and regional features to improve the performance for vehicle re-ID. Our approach shares some similarities with it. However, we not only utilize the local features that can give crucial cues to vehicle re-ID, but also extract the structure-aware features which are more compact and robust to express vehicle's identity from vehicle local regions. Moreover,the Structural-Aware deep Model we proposed have better performance than most other CNN-based models with a relative small amount training data for vehicle re-ID of the same/similar model.

3 PROPOSED METHOD

Our proposed Structure-Aware model is illustrated in Fig. 2. It consists two CNN-based work parts: Structure Map construction and Structure Map identification. Structure Map space is constructed to reduce the CNNs' sensitive to local texture features and enhance their spatial awareness, facilitating CNNs to learn more discriminative and robust structure feature of vehicles. We extract the Structure Map from vehicle front window area(FW area), considering the fact that vehicle front window naturally contains plenty of individual identity information such as decorations and stickers on it [12]. In other words, the Structure Map of a vehicle can be viewed as the relative position distribution of individual local regions in FW area. SegNet [1] and Fast R-CNN [3] are introduced to detect FW boundary and individual local regions respectively to realize the construction of Structure Map. In the identification part, we first generate more Structure Maps of multiple views from the original Structure Map using perspective transformation, aiming to ensure the CNN-based models are able to extract multi-level features from sufficient amount training data from different viewpoint. We finally train the GoogLeNet using both original and generated Structural Map to learn robust and discriminative structure-aware deep features to achieve a state-of-the-art performance on vehicle re-ID of the same/similar model.

3.1 Structural Map

Vehicles of the same/similar model shares quite lots of similar parts in appearance, making it more challenging in distinguish them using current visual-based methods. We map raw vehicle image space to the Structural Map space which contains more robust and efficient feature expression to facilitate solving problems mentioned in Introduction.

Let$\{(x_i, d_i, m_i)\}_{i=1}^{N}$ be the training set in raw vehicle image space, where x_i denotes the i-th vehicle images, N is the total number of training image samples, D is the total number of training vehicles, d_i and m_i are vehicle ID and vehicle model label respectively. We focus on solving the problem that given an image x_i of model m_j as probe, how the other images with $d = d_i$ could be identified from the dataset $\{(x_k, d_k, m_k)|m_k = m_j\}$.Instead of directly learning deep features from raw image sapce using deep model, we introduce a structure-aware feature space,named Structural Map space, constructed by the position distribution of individual local regions in vehicle FW area and the boundary of this area, as shown in Fig.2.

We separate the FW region from the vehicle image x_j using SegNet. SegNet [1] is a VGG-based deep model with symmetrical Encoder-Decoder architecture that achieves end-to-end image segmentation. The SegNet is trained with labeled images of different vehicle regions. Images of FW region are label as 1 and other region images are label as 0. Take SegNet as a function denoted as f_{seg} and the FW region x_j^{area} can be calculated:

$$x_j^{area} = f_{seg}(x_j) \tag{1}$$

$boun(x_j^{area})$,the boundary of front window region, can be easily get from x_j^{area} using the segmentation. In order to make precise detection of individual local regions in FW area, we build a large set of individual local region images for deep model learning which contains 12,742 images of kinds of regions located in FW area such as car stickers, ornaments and other special objects. All the images are manually extracted from the real-world vehicles. With the dataset, we utilize Fast R-CNN [3] that is an another classic and efficient CNN-based model in object detection and classification with VGG-based feature extractors and Multitask loss functions, to detect the individual local regions in x_j^{area}. We take FW area images x_j^{area} as input and the output is the location area of each individual regions. Every individual local region is then uniformly replaced by a 12×12 pixel area which shares the same geometric center with individual region area.Let f_{RNN} denote the trained Fast R-CNN and $\{pix_{ji}\}$ denote the i-th pixel area of x_j^{area}, then the structure map denoted as Smp_j can be constructed by the combination of $boun(x_j^{area})$ and $\{pix_{ji}\}$:

$$\begin{aligned} Smp_K &= \bigcup_{i \in K, T \in \{T_{xyz}\}} T(Smp_i) \\ &= \bigcup_{i \in K, T \in \{T_{xyz}\}} T(boun(f_{seg}(x_i)) \cup pix(f_{RNN}(x_i))) \end{aligned} \tag{2}$$

where K is vehicle identity number and Smp_K is the structure map set of vehicle images whose identity is K. The Structure Map space is subsequently constructed for vehicle re-id of the same/similar model

3.2 Re-identification In Structure Map Space

In this part, training phase and identification phase are both performed in Structure Map space. Specifically, images of training set and probe set are first mapped into Structure Map space, then deep model(GoogLeNet [16]) trained with Structure Map is exploited to identify the ID of the probe vehicles in Structure Map expression. Before training GoogLeNet, we have to solve the problem that there are usually not sufficient amount of vehicle images for training in many real-world tasks.With the compact representation and plane geometry-like structure of the Structure Map, we exploit perspective transformation F_{trans} to simulate more structure maps from raw ones. Perspective transformation is a projective transformation of the central projection, it essentially project images to a new visual plane using the transformation formula:

$$[x, y, w'] = [u, v, w] \cdot T \tag{3}$$

$[u, v, w]$ is the homogeneous coordinate of pixel in original image, $[u, v, w']$ is the homogeneous coordinates of the corresponding pixel in projected image, and T is a 3×3 matrix for transformation. In order to simulate Structural Maps of various visual angles as many as possible, we rotate the structural map around axis: $\{X, Y, Z\}$ of 3D Cartesian coordinate system respectively from 30 degree to -30 degree, and capture a transformation every 10 degree. $6 \times 6 \times 6$ transformation matrixes T are subsequently constructed from three dimensions, denoted by $\{T_x yz\}$.

In the raw vehicle image space, $vech(id = k) = \{(x_i, d_i, m_i)|m_i = k\}$ represents the vehicle set consists of vehicles whose id is k, the set of subscript of x in $vech(id = k)$ is denoted by K. Then the structural map set of $vech(id = k)$ can be constructed using method proposed in section 3.1. The probe set for identified is mapped into Structure Map space in the same way, denoted as $\{Smp\}_{pr}$.The union set of original structure map of $vech(id = k)$ and its simulated set, denoted by $\{Smp_j^u\}$,are used as training data for identification procedure.In the identification part, GoogLeNet is exploited here to do classification task learning from structural map space $\{(Smp_i^u, d_i)\}_{i=1}^{D}$. Specifically we take the structural map $\{Smp_j^u\}$ labeled with the corresponding ID number d_i as input and outputs its corresponding ID number d_i. With sufficient amount and structure-aware representation of training data, GoogLeNet we trained have a promising performance for vehicle re-ID of the same/similar model.

4 EXPERIMENT

4.1 Dataset

We evaluate our method on a large-scale dataset for vehicle re-ID named Vehicle-1M [5]. "Vehicle-1M" was build from real-world surveillance videos containing 936,051images of 55,527 vehicles, and each image is annotated with one of 400 refined vehicle models. The difference between vehicles of the model can be quite small which is just similar with the real-world vehicle re-ID situations. This dataset is very suitable for vehicle re-ID of the same/similar model.

We select vehicle images of the same/similar model, like "BMW 3-Series-2008" has little different from "BMW 3-Series-2012" in appearance, as training sets. Each set of vehicles of the same/similar model has a corresponding training dataset labeled by vehicles' ID for vehicle re-ID. 274,208 images covering all the vehicle modes from 45,000 vehicles are selected for training and 152,422 images are selected for testing. All the images we choose to pick are vehicle images captured from head. Inspired by [9], we extract a small test set denoted as "Small" containing 600 vehicles of 11,334 images and 36 vehicle models, a medium test set denoted as "Medium" containing 1200 vehicles of 24,727 images and 69 vehicle models, and a large test set denoted as "Large" containing 2400 vehicles of 39,211 images and 122 vehicle models.

4.2 Structural Map Space Construction

Boundary extraction of vehicle FW region and individual local region detection in FW area are the most import parts in construction of Structural Map. Deep models are introduced to handle these two tasks for their state-of-the-art performance in object recognition and classification. Specifically, we utilize SegNet [1] to extract front window region contour and Fast R-CNN [3] to detect individual local regions in FW area respectively.

We use SegNet to just separate the Fw region form other vechicle regions, making the segmentation much easier. 15,200 images of Fw region and 32,000 images of other vehicle regions are extracted from 20,000 vehicle images for SegNet training. We test the classification performance of the SegNet on 12,000 vehicle images and the global accuracy (G) which measures the percentage of pixels correctly classified in the dataset turns to be 91.3%, the mean of the predictive accuracy over all classes (C) is 88.6%. SegNet turns out to have a satisfactory performance in segmentation for the Fw region.

For individual local regions detection in FW area, we build a dataset containing 12,742 images of kinds of individual regions located in Fw region extracted from real-world vehicles. The individual local regions can be efficient detected with Fast R-CNN trained by this dataset.The detection is performed within the Fw area extracted by SetNet to further improve its accuracy and efficiency. We test the detectioin performance on 5,800 FW images extracted by SegNet and the MAP(Mean Average Precision) reaches 87.4%

The construction of structural map is essentially a process mapping the original image space to structural map space where deep model can learn robust and compact structural features. After the construction, perspective transformation introduced in section 3.2 is exploited to simulate sufficient amount of Structure Map. For a training set consisted by n_j vehicles of the same model $(m_k = m_j)$ in raw image space $\{x_k | m_k = m_j\}$, there will be $6^3 \cdot n_j$ Structure Map simulated for training.

4.3 FW Structure Map space vs FW region image space

The Structure Map space is the core conception of our proposed method. Before evaluating the re-ID results, we first demonstrate the effectiveness of the deep models trained in FW Structure Map space. We perform vehicle re-ID of the same/similar model in both FW Structure Map space and FW region image space with deep models on "Small", "Medium" and "Large" datasets respectively. Classic and efficient Deep Learning models in classification and recognition are introduced for the experiment including: AlexNet [7],VGG16 [15], and GoogLeNet [16]

Following the practice in previous works [5] [9], Mean Average Precision (MAP) , the Top-1 and Top-5 match rate are used to evaluate the performance of Deep Learning methods in the two spaces.We use "Deep model name"+"Space name" to denote the models, for example, "VGG16+RIS" and "VGG16+SMP represent re-identify vehicles of the same/simailar model in FW region image space and FW Structure Map space using VGG16 respectively. Table.1 and Table.2 illustrate the final results. All the deep models achieved relatively good performance, showing that deep features extracted from FW region are quite distinguish for vehicle re-ID of the same/similar model. And the deep models achieved better performance when trained in FW Structure Map space

Table 1: Match rate of vehicle re-ID of the same model task on "Vehicle-1M" in two space

Match Rate		Small	Medium	Large
AlexNet+RIS	Top-1	0.582	0.544	0.508
AlexNet+SMP		**0.735**	**0.714**	**0.685**
VGG16+RIS		0.637	0.601	0.544
VGG16+SMP		**0.761**	**0.738**	**0.683**
GoogLeNet+RIS		0.654	0.622	0.593
GoogLeNet+SMP		**0.787**	**0.751**	**0.714**
AlexNet+RIS	Top-5	0.68	0.639	0.572
AlexNet+SMP		**0.825**	**0.798**	**0.743**
VGG16+RIS		0.713	0.681	0.649
VGG16+SMP		**0.846**	**0.813**	**0.785**
GoogLeNet+RIS		0.739	0.698	0.642
GoogLeNet+SMP		**0.892**	**0.868**	**0.847**

Table 2: MAP of vehicle re-ID of the same model task on "Vehicle-1M" in two space

Method	Small	Medium	Large
AlexNet+RIS	0.635	0.592	0.563
AlexNet+SMP	**0.697**	**0.653**	**0.614**
VGG16+RIS	0.682	0.638	0.597
VGG16+SMP	**0.733**	**0.680**	**0.651**
GoogLeNet+RIS	0.688	0.648	0.581
GoogLeNet+SMP	**0.861**	**0.824**	**0.773**

4.4 Experimental Results on Vheicle-1M

Considering CNN-based methods have better performance than traditional models using hand-crafted features for vehicle re-ID , we mainly compare our approach to the CNN-based methods recently proposed, and these methods include: VGG+TL, Mixed Diff+CCL [9] which extract deep features from raw image space by the VGGNet pre-trained on CompCars. "C2F-Rank" method [5] in which a novel coarse-to-fine ranking loss is designed to facilitate pulling images of the same vehicle as close as possible and achieve discrimination between images from different vehicles.RAM method [12] which extracts deep features from a series of local regions as well as global features to achieve vehicle re-ID.Table 3 and Table 4 shows the match rate and MAP results on "Vehicle-1M". RSA outperforms the other current methods on both of the experiment indicators. Especially, compared with RAM method which shares the similar spirit that taking use of local region cues in vechicle for re-ID, the RSA model beats RAM by about 7% and 4%, on match rate and MAP respectively.

Table 3: Match rate of Vehicle re-ID of the same model Task

Match Rate		Small	Medium	Large
VGG+TL	Top-1	0.362	0.346	0.305
Mixed Diff+CCL		0.512	0.467	0.431
C2F-Rank		0.663	0.627	0.549
RAM		0.746	0.703	0.637
RSA(ours)		**0.787**	**0.751**	**0.714**
VGG+TL	Top-5	0.627	0.523	0.488
Mixed Diff+CCL		0.746	0.687	0.631
C2F-Rank		0.757	0.701	0.647
RAM		0.837	0.767	0.723
RSA(ours)		**0.892**	**0.868**	**0.847**

Table 4: MAP of Vehicle re-ID of the same model Task

Method	Small	Medium	Large
VGG+TL	0.492	0.447	0.384
Mixed Diff+CCL	0.534	0.463	0.437
C2F-Rank	0.841	0.776	0.724
RAM	0.832	0.784	0.716
RSA(ours)	**0.861**	**0.824**	**0.773**

5 CONCLUSION

In this paper, we propose a Regional Structure Aware model (RSA) for vehicle re-ID of the same/similar model. We exploit two classic and efficient deep model, SegNet and Fast R-CNN, to jointly extract the relative position distribution of individual local regions in vehicle front window area, mapping vehicle images into a structure-aware map space (FW Structural Map Space). This encourages deep models to reduce their over-reliance on large amount of training data and enhance their ability of identifying spatial relationship, learning more discriminative and robust structure-aware features for vehicle re-ID. Experiments on the large dataset "Vehicle-1M" demonstrate that RAS achieves the state-of-the-art performance on vehicle re-ID of the same /similar model.

ACKNOWLEDGEMENTS

This work was supported by National Nature Science Foundation of China No. U1736206 and No. 61701194

REFERENCES

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39, 12 (2017), 2481–2495.

[2] Nicholas Baker, Hongjing Lu, Gennady Erlikhman, and Philip J Kellman. 2018. Deep convolutional networks do not classify based on global object shape. *PLoS computational biology* 14, 12 (2018), e1006613.

[3] Ross Girshick. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

[4] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

[5] Haiyun Guo, Chaoyang Zhao, Zhiwei Liu, Jinqiao Wang, and Hanqing Lu. 2018. Learning coarse-to-fine structured feature embedding for vehicle re-identification. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[6] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. 2015. Spatial transformer networks. In *Advances in neural information processing systems*. 2017–2025.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[8] Wei-Hua Lin and Daoqin Tong. 2011. Vehicle re-identification with dynamic time windows for vehicle passage time estimation. *IEEE Transactions on Intelligent Transportation Systems* 12, 4 (2011), 1057–1063.

[9] Hongye Liu, Yonghong Tian, Yaowei Yang, Lu Pang, and Tiejun Huang. 2016. Deep relative distance learning: Tell the difference between similar vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2167–2175.

[10] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. 2017. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 212–220.

[11] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. 2016. Large-scale vehicle re-identification in urban surveillance videos. In *2016 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

[12] Xiaobin Liu, Shiliang Zhang, Qingming Huang, and Wen Gao. 2018. Ram: a region-aware deep model for vehicle re-identification. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

[13] Gary Marcus. 2018. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631* (2018).

[14] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.

[15] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

[17] Jiankai Wang, Nakorn Indra-Payoong, Agachai Sumalee, and Sakda Panwai. 2013. Vehicle reidentification with self-adaptive time windows for real-time travel time estimation. *IEEE transactions on intelligent transportation systems* 15, 2 (2013), 540–552.

[18] Y. Zhou, L. Liu, and L. Shao. 2018. Vehicle Re-Identification by Deep Hidden Multi-View Inference. *IEEE Transactions on Image Processing* PP, 99 (2018), 1–1.