A Large-Scale Car Dataset for Fine-Grained Categorization and Verification

Linjie Yang, Ping Luo, Chen Change Loy, Xiaoou Tang
Department o Information Engineering, The Chinese University of Hong Kong.

This paper aims to highlight vision related tasks centered around "car", which has been largely neglected by vision community in comparison to other objects. Cars present several unique properties that other objects cannot offer, which provides more challenges and facilitates a range of novel research topics in object categorization. Specifically, cars own large quantity of models that most other categories do not have, enabling a more challenging fine-grained task. In addition, cars yield large appearance differences in their unconstrained poses, which demands viewpoint-aware analyses and algorithms. Importantly, a unique hierarchy is presented for the car category, which is three levels from top to bottom: make, model, and released year. This structure indicates a direction to address the fine-grained task in a hierarchical way, which is only discussed by limited literature. Apart from the categorization task, cars reveal a number of interesting computer vision problems. Firstly, different designing styles are applied by different car manufacturers and in different years, which opens the door to fine-grained style analysis and fine-grained part recognition (see Fig. 2). Secondly, the car is an attractive topic for attribute prediction. In particular, cars have distinctive attributes such as car class , seating capacity, number of axles, maximum speed and displacement, which can be inferred from the appearance of the cars. Lastly, in comparison to human face verification [3], car verification, which targets at verifying whether two cars belong to the same model, is an interesting and under-researched problem. The unconstrained viewpoints make car verification arguably more challenging than traditional face verification.

We believe the lack of high quality datasets greatly limits the exploration of the community in this domain. To this end, we collect and organize a large-scale and comprehensive image database "CompCars". The latest dataset can be downloaded at http://mmlab.ie.cuhk.edu.hk/datasets/comp_cars/index.html. The "CompCars" dataset is much larger in scale and diversity compared with the current car image datasets [1]. In particular, the CompCars dataset contains data from two scenarios, including images from *web-nature* and *surveillance-nature*. The images of the web-nature are collected from car forums, public websites, and search engines (see Fig. 1). The images of the surveillance-nature are collected by surveillance cameras. The data of these two scenarios are widely used in the real-world applications. They open the door for cross-modality analysis of cars. Specifically, the web-nature data contains 161 car makes with 1,687 car models, covering most of the commercial car models in the recent ten years. There are a total of 136,727 images capturing the entire cars and 27,618 images capturing the car parts, where most of them are labeled with attributes and viewpoints. The surveillance-nature data contains 50,000 car images captured in the front view.

We highlight the key features of the dataset: **1) Car Hierarchy:** The car models can be organized into a large tree structure, consisting of three layers , namely car make, car model, and year of manufacture. **2) Car Attributes:** Each car model is labeled with five attributes, including maximum speed, displacement, number of doors, number of seats, and type of car. These attributes provide rich information while learning the relations or similarities between different car models. **3) Viewpoints:** We also label five viewpoints for each car model, including front (F), rear (R), side (S), front-side (FS), and rear-side (RS). **4) Car Parts:** We collect images capturing the eight car parts for each car model, including four exterior parts (i.e. headlight, taillight, fog light, and air intake) and four interior parts (i.e. console, steering wheel, dashboard, and gear lever). These images are roughly aligned for the convenience of further analysis.

We further demonstrate a few important applications exploiting the dataset, namely car model classification, car model verification, and attribute prediction. Here we highlight the results on fine-grained car model classification. Please refer to our full paper for the full experiments. Specifically, we fine-tune Overfeat [2], a CNN model which is pretrained on ImageNet, for the task of classifying 431 car models. Different models are respectively fine-tuned on images of specific viewpoints and all the viewpoints, denoted as "front (F)", "rear (R)", "side (S)", "front-side (FS)", and "rear-side (RS)", and "All-View". The performances of these six models are summarized in Table 1, where "FS" and "RS" achieve better performances than the performances of the other viewpoint models. CNN trained on all views performs surprisingly well. Detailed analyses are provided in the full paper.

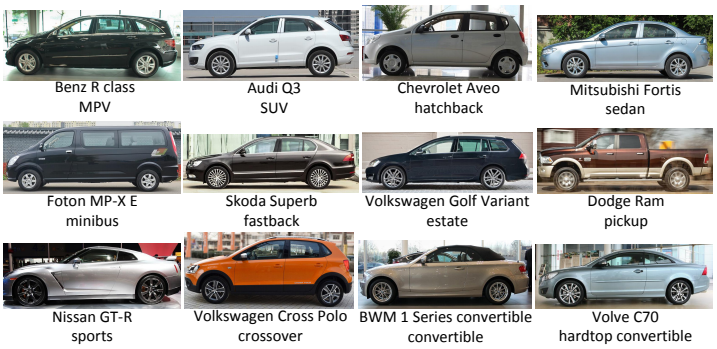

Figure 1: Each image displays a car from the 12 car types. The corresponding model names and car types are shown below the images.

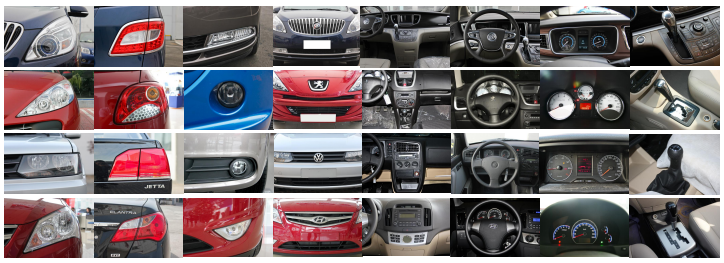

Figure 2: Each row displays 8 car parts from a car model. The corresponding car models are Buick GL8, Peugeot 207 hatchback, Volkswagen Jetta, and Hyundai Elantra from top to bottom, respectively.

We believe that the new dataset can foster more sophisticated and robust computer vision models and algorithms. There are many other potential tasks that can exploit CompCars, such as part detection, image ranking, and 3D reconstruction. It is also helpful to consider the rich attributes and different depths of the semantic hierarchy for learning semantic relations between different fine-grained car categories.

Table 1: Fine-grained classification results.

Viewpoint	F	R	S	FS	RS	All-View
Top-1	0.524	0.431	0.428	0.563	0.598	0.767
Top-5	0.748	0.647	0.602	0.769	0.777	0.917

[1] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3D object representations for fine-grained categorization. In *ICCV Workshops*, 2013.

[2] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[3] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *CVPR*, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.