

DVM-CAR: A Large-Scale Automotive Dataset for Visual Marketing Research and Applications

User Manual

Jingmin Huang
University of Glasgow
2421107h@student.gla.ac.uk

Bowei Chen*
University of Glasgow
bowei.chen@glasgow.ac.uk

Lan Luo
University of Southern California
lluo@marshall.usc.edu

Shigang Yue
University of Lincoln
syue@lincoln.ac.uk

Iadh Ounis
University of Glasgow
iadh.ounis@glasgow.ac.uk

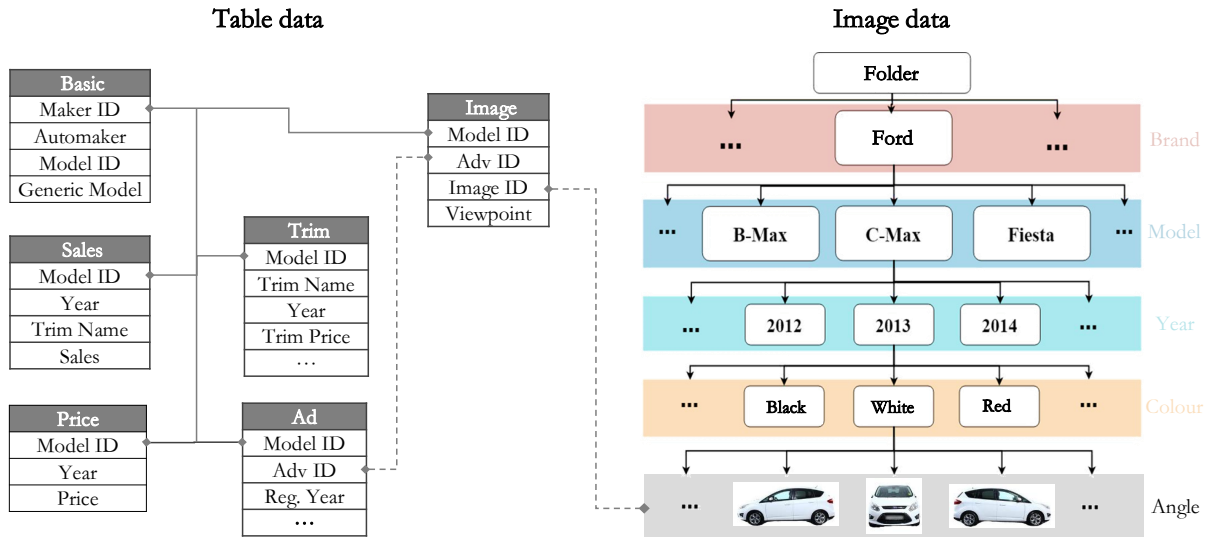


Figure 1: The structure of DVM-Car dataset.

Welcome to use the DVM-Car dataset! This dataset is a UK automotive market based research dataset created for business analytics and applications purposes. Specifically, it can be used for topics such as car appearance design, consumer analytics and used for used car sales modelling.

Structure of the dataset

As the above figure shows, the DVM-Car is comprised of two main parts: the image part (13.6 GB in a single zip file) and the table part (six CSV tables).

Images: Our dataset contains 1,451,784 car images that deliberately collected from automotive classified advertising platforms. It covers 899 car models which have been sold in the UK market over the last 20 years. All provided images are stored under the categorisation “brand-model-year-colour”. This structure allows researchers to easily locate images. A table for image indexing (i.e., Image Table) is provided. For shared version 2.0, all the images are provided without background. In case you need the images in original sizes or other requirements, you can contact us about these needs.

Tables: The model/car specification information is provided in the data tables. Users can ‘join’ different tables via the ‘Genmodel ID’ or the ‘Adv ID’ attribute. Following are the details of these tables.

- **Basic Table.** This table is mainly for indexing other tables. It includes 1,011 generic models from 101 automakers.
- **Sales Table.** It contains over 20 years car sales data of the UK market (based on the government released statics of new car registrations). In sum, it covers the sales of 773 car models from 2001 to 2020.
- **Price Table.** It is designed for users who only need the basic price data of historical models. This table contains the entry-level (i.e., the cheapest trim price) new car prices of 647 models since 1998. For users who are interested in the specific prices of trims, they can use the price values in the trim table.
- **Trim Table.** It includes 0.33 million trim-level information such as selling price, fuel type and engine size.
- **Ad Table.** It shows more than 0.27 million used car advertisements information collected from online sources. It consists of variables like the advertisement’s creation time and the car’s registration year, cumulate mileage and selling price.
- **Image Table.** It contains images information like predicted viewpoint and quality check result. The ‘P’ in the quality-check column means ‘positive’, which means the angle prediction of the image is right and the image quality is satisfying. While ‘N’ means incorrect angle prediction. Currently, it includes information of 1,451,784 car images, while expansions of this table are applicable in the future.

You are welcome to contact us for any this dataset related issues!

*Corresponding author.