Densely Annotated Video Driving (DAVID) Data Set

Christopher Kuhn, Markus Hofbauer, Murong Xu, and Eckehard Steinbach

christopher.kuhn@tum.de, markus.hofbauer@tum.de, eckehard.steinbach@tum.de

Technical University of Munich

Chair of Media Technology

Arcisstr. 21, 80333 Munich

1. Introduction

This document summarizes the contents and structure of the Densely Annotated Driving (DAVID) data set. The DAVID data set consists of 28 video sequences of urban driving recorded in the CARLA simulator[1]. For each frame, the ground truth pixel-wise semantic class labels were recorded as well. An example image and the corresponding semantic labels are shown in Figure 1. The data set is intended to further facilitate research of semantic video segmentation by providing a diverse and large scale video corpus. When using this work, please cite our article "Pixel-Wise Failure Prediction for Semantic Video Segmentation" submitted to the 2021 *IEEE International Conference on Image Processing* where the data set was first introduced.
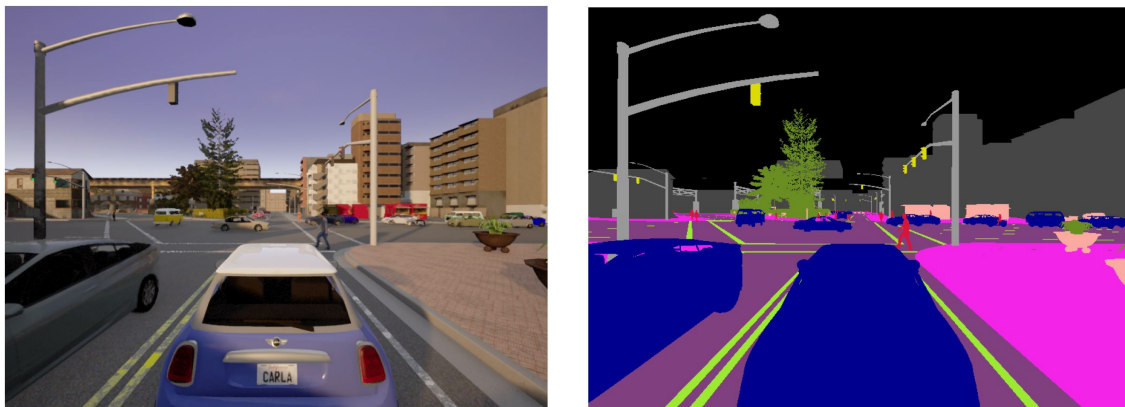

Figure 1: Example image (left) and the corresponding semantic labels (right) from the DAVID data set.

---

1   Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. "Carla: An open urban driving simulator." In *Conference on robot learning*, pages 1–16. PMLR, 2017

2. Video Contents

The DAVID data set consists of 28 video sequences that consist of a total of $10\,767$ frames. An equal number of $10\,767$ pixel-wise semantic labels is provided. The videos were recorded at a frame rate of $10\,\text{Hz}$. The average sequence duration is $38.4\,\text{s}$. Half the sequences were recorded in sunny weather. Nine sequences were recorded in rain and the remaining five were recorded in cloudy conditions. The recorded driving scenarios include regular driving, traffic jams as well as stopping and starting at traffic lights. The contents of the individual video sequences are summarized in Table 1.

Index	Duration	Weather	Driving Scenario	Number of Images
0	30.4 s	Sunny	Driving	305
1	34.1 s	Sunny	Driving + Traffic Jam	342
2	36.2 s	Sunny	Driving	363
3	35.0 s	Sunny	Driving + Traffic Jam	350
4	33.4 s	Sunny	Driving	335
5	38.9 s	Sunny	Driving + Traffic Jam	390
6	39.6 s	Sunny	Traffic Light	397
7	44.2 s	Sunny	Traffic Light	443
8	39.3 s	Sunny	Driving	393
9	36.3 s	Sunny	Driving	364
10	43.7 s	Sunny	Driving	438
11	34.5 s	Sunny	Driving	346
12	37.9 s	Sunny	Driving + Traffic Light	380
13	32.5 s	Sunny	Driving	325
14	35.7 s	Rainy	Driving + Traffic Light	357
15	38.3 s	Cloudy	Driving	384
16	34.9 s	Rainy	Driving + Traffic Light	350
17	35.9 s	Rainy	Driving	360
18	37.4 s	Rainy	Traffic Light	374
19	43.0 s	Cloudy	Driving	431
20	36.5 s	Rainy	Driving	366
21	44.7 s	Rainy	Traffic Light	448
22	35.1 s	Rainy	Traffic Light	351
23	45.0 s	Rainy	Driving	451
24	44.3 s	Cloudy	Driving + Traffic Light	443
25	43.7 s	Cloudy	Driving	438
26	38.5 s	Rainy	Driving + Traffic Light	386
27	45.6 s	Cloudy	Driving	457

Table 1: Summary of the video sequences of the DAVID data set.

3. Semantic Labels

For each RGB image frame, a png file of the same resolution is provided that contains the semantic class of each pixel of the corresponding RGB image. The semantic labels were obtained using the ground truth available from the CARLA simulator itself. A total of 12 semantic classes are distinguished by the simulator. An additional unlabeled void class consists of the sky as well as objects not belonging to the other categories. The RGB color values of each class and the relative frequency among the pixels of all images are given in Table 2.

Index	Semantic Class	RGB Color Values	Relative Frequency
0	Traffic Sign	[220, 220, 0]	0.05 %
1	Building	[70, 70, 70]	14.08 %
2	Fence	[190, 153, 153]	0.28 %
3	Other	[250, 170, 160]	0.29 %
4	Pedestrian	[220, 20, 60]	1.02 %
5	Pole	[153, 153, 153]	0.74 %
6	Road Line	[157, 234, 50]	1.86 %
7	Road	[128, 64, 128]	33.80 %
8	Sidewalk	[244, 35, 232]	5.85 %
9	Vegetation	[107, 142, 35]	4.78 %
10	Car	[0, 0, 142]	3.63 %
11	Wall	[102, 102, 156]	0.82 %
12	Unlabeled	[0, 0, 0]	32.81 %

Table 2: The RGB color values corresponding to the 12 semantic classes plus an unlabeled void class.