Test Your Self-Driving Algorithm: An Overview of Publicly Available Driving Datasets and Virtual Testing Environments

Yue Kang, Hang Yin, and Christian Berger

***Abstract*—Many companies aim for delivering systems for autonomous driving reaching out for SAE Level 5. As these systems run much more complex software than typical premium cars of today, a thorough testing strategy is needed. Early prototyping of such systems can be supported using recorded data from onboard and surrounding sensors as long as open-loop testing is applicable; later, though, closed-loop testing is necessary—either by testing on the real vehicle or by using a virtual testing environment. This paper is a substantial extension of our work presented at the 2017 IEEE International Conference on Intelligent Transportation Systems (ITSC) that was surveying the area of publicly available driving datasets. Our previous results are extended by additional datasets and complemented with a summary of publicly available virtual testing environments to support closed-loop testing. As such, a steadily growing number of 37 datasets for open-loop testing and 22 virtual testing environments for closed-loop testing have been surveyed in detailed. Thus, conducting research toward autonomous driving is significantly supported from complementary community efforts: A growing number of publicly accessible datasets allow for experiments with perception approaches or training and testing machine-learning-based algorithms, while virtual testing environments enable end-to-end simulations.**

***Index Terms*—Driving dataset, virtual testing environment, simulation, self-driving vehicle, autonomous driving.**

I. Introduction

VEHICLES with self-driving functionality are currently entering the product portfolio of all major automotive original equipment manufacturers (OEMs). In addition, a growing number of start-ups around the world are aiming at delivering solutions towards SAE Level 5 functionality. These vehicles will substantially change the way how people will access and use mobility solutions in the future; in addition, this change in the way how mobility is consumed will also re-shape how metropolitan regions will be designed to allow for a better and more sustainable co-existence of various mobility solutions like bicycles, electric motorcycles, cars, supply vehicles, trucks, or public transportation.

The algorithms that are needed to realize autonomously acting mobility solutions are becoming increasingly complex as SAE Level 5 vehicles need to be able to act safely in any traffic situation without the need for a human driver. Therefore, careful testing and thorough evaluation of the individual software units that comprise a self-driving vehicle is mandatory including the use of open-loop stimuli from recordings to include realistic situations or for training and testing machine-learning (ML)-based algorithms. Complementary thereto, closed-loop testing using virtual testing environments is needed to enable end-to-end validation of both, individual software units as well as the complete data processing chain. Finally, new functionality is validated in prototypical vehicle platforms that are specifically instrumented to conduct measurements for systematic analysis of a functionality's behavior in real-world settings.

This article is a substantially extended version of [1] "When to Use What Data Set for Your Self-Driving Car Algorithm: An Overview of Publicly Available Driving Datasets". In contrast to our previous work, the main differences concern: (a) The presentation of publicly available datasets was updated to also include additional ten recently published datasets; (b) the description of the individual datasets was extended to include typical application scenarios for users; and (c) we complemented our previous work by additionally surveying the area of virtual testing approaches using simulations. Thereby, this article covers both, open-loop and closed-loop approaches for evaluating algorithms in the area of self-driving vehicles.

A. Background

While the work in the area of self-driving vehicle functionality dates back to 1939 at the World Fair in New York where GM was outlining a vision towards vehicles with no human intervention, only seven decades later in 2007, the first large-scale demonstration of several autonomously driving vehicles in an urban-like environment was conducted, known as the DARPA Urban Challenge (cf. [2]). Only today, the necessary perception technology, computational performance, and algorithmic approaches seem to be available to let the vision become reality.

Manuscript received December 14, 2017; revised May 18, 2018 and August 31, 2018; accepted October 17, 2018. Date of publication March 13, 2019; date of current version May 22, 2019. This work was supported by the COPPLAR Project, funded by Vinnova FFI, Diarienr: 2015-04849, and the CAIVE project at the Chalmers University of Technology and University of Gothenburg. *(Corresponding author: Hang Yin.)*

Y. Kang is with the Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden (e-mail: yuek@chalmers.se).

H. Yin is with Zenuity AB, Gothenburg 41756, Sweden (e-mail: hang.yin@zenuity.com).

C. Berger is with the Department of Computer Science and Engineering, University of Gothenburg, Gothenburg 40530, Sweden (e-mail: christian.berger@gu.se).

Digital Object Identifier 10.1109/TIV.2018.2886678

Compared to auto-pilot systems for commercial flights, automotive systems are much more complex as they have to cope with a large variety of complex traffic situations with potentially very unpredictable traffic participants. Thorough testing of such systems is then continuously necessary to cover all possible traffic scenarios. The complexity is also illustrated by the SAE classification for autonomous driving where the two topmost levels require handling of all or nearly all possible traffic scenarios even when no human driver is on-board as fall-back (Level 3 or higher). Typical vehicle testing of today as reported in [3] also includes prototypical platforms specifically instrumented for measurements; these vehicles provide the real dynamics and physical characteristics that are difficult to model accurately in pure virtual testing approaches as correct physical behavior is essential for ultimately testing or tune a specific vehicle function.

B. Problem Domain & Motivation

The traditional approach to test said algorithms mainly involves the use of recorded data from the target platform. While recorded data has the highest degree of fidelity in terms of level of realism, it can only be used for open-loop testing to stimulate perception algorithms for example. Furthermore, such approaches gained a lot of attention recently with significant achievements in AI and ML requiring large amount of data for, e.g., end-to-end learning [4].

While designing, collecting, and labeling new data to evaluate algorithms for a self-driving vehicle is resource-intensive and time-consuming, and in some case especially weather-dependent, there are many datasets publicly available to the research community. The work presented here is supporting researchers and developers to get an overview of those datasets (e.g., what datasets are available, which sensors are included, and what situations are covered) to provide guidance during the selection of existing datasets.

However, systematic limitations such as usefulness for open-loop testing only as well as of practical nature as time, weather, and vehicular recording platforms constrain the role of such data collecting approaches in practice. Therefore, complementary approaches that overcome the limitation of open-loop use on the one hand, and allow for scalability on the other hand are necessary to tackle the growing testing needs for autonomous driving systems on SAE Level 3 to 5.

C. Research Goal & Research Questions

The goal of this work is to present an extensive overview of publicly accessible datasets and instruments to support both open-loop and closed-loop testing resulting in the following two research questions:

- RQ-1 What datasets are available to support what type of testing for self-driving vehicular algorithms?[1]
- RQ-2 What virtual testing environments are available to support what type of closed-loop testing of self-driving vehicular algorithms?

[1]This research question was addressed in [1]; in this work, we have updated the previous result to also include recently published datasets.

D. Contributions

The work presented in this article is based on our previous work [1] presented at the 2017 IEEE International Conference on Intelligent Transportation Systems. We substantially updated the existing work by extending the coverage of publicly accessible datasets to 37 datasets in total and thus, include the most recently published datasets as guidance for selecting the right dataset for evaluating an algorithm.

Furthermore, we provide a survey of existing virtual testing environments to complement our existing work with approaches that enable researchers and developers to evaluate their algorithms in a closed-loop environment, i.e., receiving data from the system-under-test, adjust the simulated world accordingly, and derive the next stimulus for simulated sensors for the following time-step.

E. Scope & Limitations

The goal of our work was to carefully conduct a broad survey to provide an exhaustive overview about existing datasets and virtual testing environments supporting research and development of autonomous driving and algorithms. This work focuses particularly on datasets and virtual testing environments that could be identified using structured web searches and systematic snowballing and that are accessible to researchers and developers. For the datasets, we set our focus only on ground truth driving data collected on public roads with partial or full open access. For the virtual testing environments, we focused especially on solutions available as open source to encourage and facilitate contributions from the community. A deep analysis per data-set or virtual testing environment is very specific to particular use-cases of the development or evaluation of autonomous driving. Hence it would not contribute to the overall goal of our work but is rather suggested in specialized subsequent studies.

F. Structure of the Article

The rest of the article is structured as follows: Section II outlines relevant related work. In Section III, we outline the approach that we applied to survey the area to obtain information on publicly accessible datasets and virtual testing environments. Section IV presents and discusses the results of our findings. We conclude our work in Section V.

II. Related Work

In the work by [5], results from a research collaboration with a large European automotive OEM in Germany are presented. The authors studied how consumer tests on the example of autonomous emergency braking (AEB) can be modeled for a virtual testing environment and massively scaled in an automated way to enable a broad range of testing according to state-of-the-art test catalogs for this type of sensor-based systems. The findings demonstrated that generating hundreds of simulations with systematic variation of key parameters for the system-under-test helps to unveil unexpected anomalies to be addressed before conducting tests on a real proving ground.

In the approach of [3], the authors report about a large-scale interview study approaching scientists and industrial practitioners to explore the current state-of-the-art and future trends in the area of developing and testing active safety systems and systems aiming for self-driving functionality. The main finding for relevant future trends supported by feedback from both practitioners and researchers in the area, and relevant literature, is that the importance of virtual testing will significantly increase to improve test efficiency or even certifications. However, biggest remaining hurdles include level of fidelity of a virtual testing environment compared to data from real test runs and thus, missing clear benchmarking as well as better models for vehicle motion, sensors, and traffic situations.

In the study [6], it was demonstrated that both areas can be combined to improve the way systems are tested. It was studied how patterns from virtual testing data can be matched in data recordings from real sensors with the goal of finding interesting scenarios in reality for further analysis. However, instead of manual annotations for the datasets, matching scenarios were automatically identified.

Janai *et al.* [7] also recently surveyed a broad spectrum of datasets with a focus on computer vision in general. In contrast to our previous work [1] and to this substantially extended version, the datasets presented by Janai *et al.* are only limitedly applicable to autonomous driving.

On the other hand, virtual testing environments have been involved in a recent study [8], where various testing approaches have been classified and discussed in such virtual environments. In addition, this approach proposed a new testing framework in virtual testing environment, which combined different testing methods as a quantitative way to test the intelligence of an autonomous vehicle.

As a related work of combining dataset and virtual testing environment usage, [9] recently proposed a new routing algorithm for electric vehicles. The algorithm employed data mining techniques on a set of historical driving data, and eventually was tested and evaluated in virtual environment on dataset containing real vehicle state information as well.

III. Methodology

A. Surveying the Area of Publicly Accessible Datasets

To exhaustively survey the area of publicly accessible datasets, we decided to use the search engine from Google. The main motivation and also shown by our results is that all datasets are accompanied by explanatory websites, where the original contributors of the datasets provide basic information about content and links to obtain the data; these websites can be easily found and indexed by search engines. To conduct our survey in a systematic way as also described in [1], we applied the following four sequential steps:

1) Initial Google search for exploration: Keywords such as "driving" and "dataset" were used in the Google search engine to initialize the exploration of most popular dataset web pages.
   Hence, we ranked the search results by relevance and only considered the top 200 results. We observed that most relevant dataset websites were found among the top 100 while nothing relevant was found after 150 results implying a low risk of missing relevant datasets.
2) Systematic extension by forward snowballing among dataset web pages: Some dataset web pages explicitly provide reference links to other relevant datasets, thereby pointing us to more datasets not covered before.
3) Systematic validation by collection accompanying scientific publications: A majority of datasets are supported by at least one scientific publication. We collected such publications related to the datasets discovered in the first two phases.
4) Systematic extension by backward snowballing using the publications: We went through the collected publications to identify new datasets referenced by the publications. This process went recursively until no more datasets were found.

Throughout the four steps above, we selected datasets which satisfy the following inclusion criteria:

- Data must be collected from on-board sensors on public roads. We exclude datasets with synthetic data from virtual worlds as this is addressed in our summary of virtual testing environments, data collected indoors, or in confined areas such as parks and campuses.
- The dataset must contain camera, LiDAR, or radar data. It is insufficient to include GPS or Inertial Measurement Unit (IMU) data only.
- Full or partial open access.

Our previous survey [1] reported 27 datasets. This article includes ten more datasets (+37% more), most of which were released after our previous survey was published.

B. Surveying the Area of Virtual Testing Environments

Similarly, our methodology of finding existing virtual testing environments involves Google research, publication collection, and systematic snowballing. With the exploding development of intelligent vehicle functionality, driving safety, and AI/ML, a larger number of approaches in which virtual testbeds are involved can be observed. In addition, most of these approaches are accomplished by employing open-source virtual testing environments due to the consideration of availability and cost. Virtual testing environments were gathered as follows:

1) Direct Google and YouTube search: Keywords as "virtual environment" and "simulation" were similarly used in both search engines, as well as "vehicle", "traffic", "autonomous driving" and "download" for the purpose of filtering the results and also considering the accessibility issue. We observed that multiple video examples from different research groups or individuals could be found with the same keyword set, which would lead to excessively redundant results. This also implies low risk of missing relevant approaches during this phase.
2) Publication collection: A considerable number of approaches about autonomous driving, navigation, and other traffic-related research was initially gathered for virtual testing environments. We went through recent and relevant publications with focus on the simulations.
3) Snowballing among web pages of virtual testing environment: A majority of open-source projects of virtual testing

environment provide an individual website, from which we collected references and other information. This process was considered as a complement for the phases above due to the fact that not all approaches in this field have open access publications or video illustrations.

Several selection criteria were engaged during the survey of virtual testing environments:

- Relevance to autonomous or intelligent driving, vehicle and traffic simulation, etc. Approaches of irrelevant virtual testing environment examples were therefore excluded, like simulation environments designed for network models, vehicle fabrication, light condition testing, or driver training.
- Accessibility. The environment or facility should be accessible either through open access or as commercial solutions. This excludes several in-house approaches that are only mentioned in conference and journal papers, or in commercial advertisements.

IV. Results & Discussion

A. Overview of the Datasets

As an update of our previous overview [1], 37 datasets are listed below in alphabetic order with references, provider information, and highlights. A short alias based on the full name of each dataset is indicated in parentheses. In the rest of the article, the alias will be used whenever a specific dataset is referenced. The links have been verified on 2018-10-20.

- *Dataset 1: Automotive multi-sensor dataset (AMUSE) [10]* (https://goo.gl/1YbD5E)
  Provider: Linköping University, Sweden
  Highlight: Omnidirectional visual data for full surround sensing; include winter conditions with snow
- *Dataset 2: Apollo* (https://goo.gl/yy144b)
  Provider: Baidu, China
  Highlight: A huge variety of annotated data from multiple sensors suitable for deep learning and training tasks; supported with open source projects
- *Dataset 3: Berkeley DeepDrive Video dataset (BDDV) [11], [12]* (https://goo.gl/24XNzG)
  Provider: UC Berkeley, US
  Highlight: A large-scale driving video dataset with various types of annotations
- *Dataset 4: Caltech Pedestrian Detection Benchmark (Caltech) [13]* (https://goo.gl/HB7e8P)
  Provider: California Institute of Technology, US
  Highlight: The largest pedestrian dataset; pedestrian annotation; the first dataset with temporal correspondence between bounding boxes and occlusion labels
- *Dataset 5: Cambridge-driving Labeled Video Database (CamVid) [14]* (https://goo.gl/I2pbdP)
  Provider: University of Cambridge, UK
  Highlight: The first collection of videos with object class semantic labels; pixel-level annotation
- *Dataset 6: CCSAD dataset [15]* (https://goo.gl/pxr3Yc)
  Provider: Centro de Investigación en Matemáticas, Mexico
  Highlight: Stereo video captured in developing countries
- *Dataset 7: Cheddar Gorge Dataset [16]*
  Provider: BAE Systems (Operations) Limited, UK
  Highlight: Diversified sensor setup with stereo, monocular, and infrared cameras, Velodyne 64 LiDAR, GPS/IMU etc.
- *Dataset 8: Cityscapes dataset [17]* (https://goo.gl/qLM3V4)
  Provider: Daimler AG R&D, Germany; Max Planck Institute for Informatics (MPI-IS), Germany; TU Darmstadt Visual Inference Group, Germany
  Highlight: Stereo sequences from 50 cities; pixel-level annotation for semantic urban scene understanding; benchmark suite with an evaluation server; the foundation for a new dataset, CityPersons [18], with better person annotations
- *Dataset 9: CMU Visual Localization Dataset (CMU)* (https://goo.gl/0R8XX6)
  Provider: Carnegie Mellon University, US
  Highlight: Various weather and light conditions
- *Dataset 10: comma.ai driving dataset (comma.ai) [19]* (https://goo.gl/B3TWf2)
  Provider: comma.ai, US
  Highlight: Highway traffic scenarios
- *Dataset 11: Daimler Pedestrian Benchmarks (Daimler pedestrian)* (https://goo.gl/l3U2Wc)
  Provider: Daimler AG R&D, Germany; University of Amsterdam, the Netherlands
  Highlight: Encompasses multiple benchmark datasets for pedestrian detection, classification, segmentation, and path prediction based on monocular and stereo images; the first dataset with partially occluded pedestrians; include the only cyclist dataset [20] that we have encountered so far
- *Dataset 12: Daimler Urban segmentation (Daimler urban) [21]* (https://goo.gl/KRBCLa)
  Provider: 6D-Vision, Germany
  Highlight: Stereo vision sequences in urban traffic; pixel-level semantic class annotation
- *Dataset 13: DIPLECS Autonomous Driving Datasets (DIPLECS) [22]* (https://goo.gl/8isjeJ)
  Provider: University of Surrey, UK
  Highlight: Includes two datasets on public roads, one in UK, the other in Sweden; labeled frame by frame with speed and steering data (the first dataset) and driving environments and driver actions (the second dataset)
- *Dataset 14: Dr(eye)ve [23]* (https://goo.gl/45bwXr)
  Provider: ImageLab, Italy
  Highlight: The first dataset for researching driver attention, eye fixation, and visual saliency
- *Dataset 15: EISATS [24]* (https://goo.gl/ausKsL)
  Provider: University of Auckland, New Zealand; Daimler AG, Germany; Hella Aglaia Mobile Vision GmbH, Germany; HU Berlin, Germany
  Highlight: Include multiple datasets with stereo vision sequences for comparative performance evaluation of stereo vision, optic flow, motion analysis etc.
- *Dataset 16: Elektra* (https://goo.gl/GNNq0f)
  Provider: Autonomous University of Barcelona, Spain; Polytechnic University of Catalonia, Spain
  Highlight: Various types of images with annotated pedestrians; include far infrared images

- *Dataset 17: ETH pedestrian dataset [25]* (https://goo.gl/xXDTwI)
Provider: ETH Zürich, Switzerland
Highlight: Stereo images captured in a crowed city center with many pedestrians
- *Dataset 18: EuroCity Persons Dataset (EuroCity) [26]* (Link unknown)
Provider: Daimler AG and TU Delft
Highlight: A huge number of images with person annotations in urban traffic under various weather and light conditions
- *Dataset 19: Ford Campus Vision and Lidar Data Set (Ford) [27]* (https://goo.gl/6ZkCpc)
Provider: University of Michigan, US
Highlight: Diversified sensor setup, including high precision localization devices, multiple LiDARs, omnidirectional camera etc.; full software support
- *Dataset 20: German Traffic Sign Detection Benchmark (German traffic sign)* (https://goo.gl/FqaCJQ)
Provider: Ruhr University Bochum, Germany
Highlight: Still images with traffic signs in Germany
- *Dataset 21: HCI benchmark suite (HCI)[2] [28]* (https://goo.gl/r5aRvv)
Provider: Heidelberg Collaboratory for Image Processing, Ruprecht-Karls Universität Heidelberg, and Robert Bosch GmbH, Germany
Highlight: A stereo and optical flow dataset with high accuracy for urban autonomous driving, containing a lot of manually constructed/acted scenarios on the same street
- *Dataset 22: Heidelberg benchmarks (Heidelberg) [29]* (https://goo.gl/6c2lAs)
Provider: Heidelberg University, Germany
Highlight: Associated with an event called Robust Vision Challenge; provide challenging data for stereo and optical flow, e.g., rain flares and flying snow
- *Dataset 23: Joint Attention for Autonomous Driving Dataset (JAAD) [30]* (https://goo.gl/cXoPnp)
Provider: York University, Canada
Highlight: Focus on joint attention between pedestrians and drivers for autonomous driving; provide both textual and behavioral annotations for pedestrians and vehicles
- *Dataset 24: KAIST multispectral pedestrian detection dataset (KAIST) [31]* (https://goo.gl/Tpz512)
Provider: Korea Advanced Institute of Science and Technology, South Korea
Highlight: Well-aligned color-thermal image pairs with pedestrian annotation
- *Dataset 25: Karlsruhe Dataset: Labeled Objects (Karlsruhe labeled objects) [32]* (https://goo.gl/5fk0js)
Provider: MPI-IS, Germany
Highlight: Images with object bounding boxes for cars and pedestrians; include even object orientation
- *Dataset 26: Karlsruhe Dataset: Stereo Video Sequences + rough GPS Poses (Karlsruhe stereo) [33]* (https://goo.gl/V6Q7Vx)
Provider: MPI-IS, Germany
Highlight: High-quality stereo sequences in Karlsruhe
- *Dataset 27: KITTI Vision Benchmark Suite (KITTI) [34], [35]* (https://goo.gl/cvSbGl)
Provider: Karlsruhe Institute of Technology, Germany; Toyota Technological Institute, US
Highlight: The current most prestigious dataset for self-driving; provide a number of excellent benchmarks for the evaluation of stereo vision, optical flow, scene flow, visual odometry, SLAM, object detection and tracking, road lane detection, semantic segmentation
- *Dataset 28: Málaga Stereo and Laser Urban Data Set (Malaga) [36]* (https://goo.gl/EdLHtW)
Provider: University of Málaga, Spain
Highlight: Well-documented; full tool support; message board on homepage
- *Dataset 29: Mapillary Vistas Dataset (MVD) [37]* (https://goo.gl/1L4yj1)
Provider: Mapillary research
Highlight: A huge number of images with vast geographical diversity, annotated into 66–129 object categories, covering all kinds of road/weather/light conditions
- *Dataset 30: nuScenes* (https://www.nuscenes.org/)
Provider: nuTonomy-Aptiv, US
Highlight: The first large-scale dataset to provide data from the entire sensor suite of an autonomous vehicle
- *Dataset 31: Oxford robotcar dataset (Oxford) [38]* (https://goo.gl/nJOQkq)
Provider: Oxford University, UK
Highlight: The first dataset stressing periodic long-term data collection (over a year) following predefined routes to cover long-term changes of road conditions
- *Dataset 32: Stanford track collection (Stanford) [39]* (https://goo.gl/KNOYpX)
Provider: Stanford University, US
Highlight: Velodyne 64 point cloud with object labels and GPS/IMU data
- *Dataset 33: Ground Truth Stixel Dataset (Stixel) [40]* (https://goo.gl/rf12z6)
Provider: 6D-Vision, Germany
Highlight: Heavy rain on highways; stixel annotation
- *Dataset 34: TorontoCity benchmark (TorontoCity) [41]* (Link to be released soon as the provider promised in the paper)
Provider: University of Toronto, Canada
Highlight: Data with wide range of views for mapping, reconstruction and semantic labeling
- *Dataset 35: TrafficNet [42]* (http://traffic-net.org)
Provider: University of Michigan, US
Highlight: A large-scale and extensible library of naturalistic driving scenarios
- *Dataset 36: TRoM: Tsinghua Road Marking (TRoM) [43]* (https://goo.gl/KA7DR3)
Provider: Tsinghua University, China

[2]This dataset was changed into HD1K Benchmark Suite (observed on 2018-05-14)

TABLE I
OVERVIEW OF EXISTING DRIVING DATASETS ON PUBLIC ROADS — PART 1

Dataset	Time & Venue	Traffic condition	Sensors	Data format & size	What is provided	Typical usage scenarios
AMUSE	Feb-March, 2013 Linköping (Sweden)	loop, closing, (nearly) static scene, snow, suburb, urban, low altitude of sun, water and snow on lens	omnidirectional camera; GPS+IMU; velocity sensor; 3 height sensors	7 sequences (1,169GB); png: image; liu: own format	partial raw data; API for C/C++, Python, Matlab; ROS support	visual odometry, SLAM, motion estimation, optic flow
Apollo	Mar-Nov, 2016 several major cities (China)	expressway under various weather conditions	3 monocular cameras (one front, two on sides); LiDAR with 32 layers; Velodyne HDL-64E; IMU; real time kinematic GPS	ca 270GB in total; 172GB available; jpg or png: image; txt: label; bin: Velodyne HDF5: image, curvature	raw data (training/validation/ test sets); annotation/label; benchmark; source code; demo video	3D object detection and trajectory prediction, semantic segmentation, traffic light detection, multi-sensor fusion localization
BDDV	released in 2018 (data collection time unknown) New York, Berkeley, San Francisco, Bay Area (US)	various road/weather/ light conditions	monocular color camera; sensors from a smart phone: GPS/IMU, gyroscope, magnetometer	+100,000 videos, 40s each (+1.8TB); mov: video jpg: image json: label other formats to be found by checking the dataset	raw data (training/ validation/test sets); annotations: 2D bounding box, lane marking, drivable area, pixel/instance-level segmentation	object detection, lane marking detection, drivable area detection, semantic segmentation
Caltech	before May, 2009 Los Angeles (US)	urban	monocular color camera	ca 11GB; seq: video; vbb: bounding box	videos (training/ test sets); annotation; benchmark results; Matlab code	pedestrian detection
CamVid	before 2009 Cambridge (UK)	urban	monocular color camera	4 videos (ca 8GB); png: labeled image; mxf: video; avi: video	video+label; png image extraction tool; Matlab code; paint strokes during labeling	pixel-level semantic segmentation of objects
CCSAD	May-Jul, 2014 Guanajuato (Mexico)	urban, small roads, tunnel at night, varying light conditions	stereo vision, grayscale; attitude and heading reference system; GPS-enabled smartphone	42 sequences (ca 500GB); png: image; txt: timestamp, GPS, IMU, and vehicle data; xml: calibration	raw data	stereo vision
Cheddar Gorge	2010-03-05 Cheddar Gorge (UK)	dry, sunny, clear, cold	stereo vision, color; monocular color camera; monocular infrared camera; Velodyne 64 LiDAR; professional GPS/IMU; low cost IMU; 4 wheel distance encoders; laser tracker for sensor pose measurements	329GB (57min); standard formats (no description)	raw data	SLAM, scene classification

Highlight: The first publicly available dataset related to road marking detection

- *Dataset 37: Udacity dataset* (https://goo.gl/AoxEt1)
  Provider: Udacity
  Highlight: Open source project; driving data partially including annotated objects

B. Discussion of the Datasets

After conducting a thorough study of the 37 datasets above, we herein make an intuitive comparison and provide a dataset selection guideline from the following perspectives: (1) time and venue: when and where was the data collected? (2) traffic conditions during the data collection; (3) sensor setup: what sensors were used during the data collection? (4) data format and size; (5) provided resources (e.g., raw data, annotation, benchmark, source code, and tool support); (6) license; and (7) accessibility. The comparison result is summarized in Tables I–V, where each column is a comparison factor among (1)–(5) and each row is a dataset.

Since the dataset URLs, providers, and highlights have already been provided in Section IV-A, they do not reappear in these tables. Furthermore, many datasets are similar with regard to the last two comparison factors license and accessibility. Therefore, they are skipped in the tables and will be discussed separately.

The *Time & Venue* column in the summary tables indicates that most datasets were released after 2009. Moreover, there is a growing trend in running data collections, especially since 2016. In terms of venue, most data were collected in Europe and the US. Germany is the most active country running data collections. There are about ten datasets with data collected outside Europe and the US: *CCSAD* from Mexico, several sequences in *ESATS* from New Zealand, part of *JAAD* collected from Ukraine and Canada, *TorontoCity* from Canada, three from China, including one dataset in *Daimler pedestrian*, *Apollo*, and *TRoM*, part of *MVD* collected in South America, Asia, Africa and Oceania, some data collected in South Korea in *KAIST* (this is not stated anywhere but observed by us in some sequences), and finally part of *nuScenes* collected in Singapore. Thus, we strongly urge the future release of new datasets from other regions with wider geographical distribution as each country has its unique traffic conditions. An autonomous driving algorithm that has been successfully tested using a dataset from Germany may not work similarly well in other regions of the world. More global driving data is essential towards more robust autonomous vehicles whose performance is less sensitive to geographical location.

Traffic condition is another key factor to consider while selecting a dataset. We are particularly interested in the type of traffic (e.g., urban traffic, rural road, highway), light conditions (e.g., daylight or night), and weather conditions (e.g., sunny, overcast, rainy, other). Most datasets focus on urban traffic, daylight, and sunny weather. While perfect light and weather conditions are

TABLE II
OVERVIEW OF EXISTING DRIVING DATASETS ON PUBLIC ROADS — PART 2

Dataset	Time & Venue	Traffic condition	Sensors	Data format & size	What is provided	Typical usage scenarios
Cityscapes	before 2016-02-20 (Germany), Zürich (Switzerland), Strasbourg (France)	daytime, no adverse weather conditions	stereo vision, color	5 sequences (63.141GB, more available upon request); png: image; json: annotation	images (training/ validation/test sets); pixel-level annotation; coarse annotation; benchmark suite; evaluation server; scripts	pixel-level and instance-level semantic labeling/segmentation
CMU	from 2010-09-01 to 2011-09-02 Pittsburgh (US)	various weather/light conditions	3 monocular color cameras; 4 Sick LiDARs; GPS+IMU	16 sequences (11-22GB each); jpg: image; txt: LiDAR, GPS, and vehicle data	raw data	visual localization
comma.ai	from 2016-01-30 to 2016-06-08 San Francisco (US)	daylight, mostly highway	monocular color camera; GPS+IMU; gyroscope	11 sequences (80GB); HDF5 format	raw data; open source code	driving simulation, video prediction
Daimler pedestrian	2006-2016 Beijing (China), others unknown	urban	monocular grayscale camera; stereo vision, color or grayscale	8 datasets included; 53MB, 10-15GB, 10GB, 12GB, 6GB, 300MB, 2.5MB, 45GB images: png, pgm, mat...	raw/processed data (training/ validation/test sets); annotation	pedestrian segmentation, pedestrian path prediction, pedestrian classification, cyclist detection
Daimler urban	2014	urban	stereo vision, grayscale	5,000 stereo image pairs (7.55GB, 1024*440); pgm: image, ground truth label, disparity map; xml: camera calibration, vehicle data	video; pixel-level annotation; ego-motion data; disparity map; development kit	pixel-level semantic segmentation
DIPLECS	2015 Surrey (UK), Stockholm (Sweden)	country road, highway, urban	Surrey: - monocular color camera Stockholm: - 3 monocular grayscale cameras	4.29GB+1.06GB; Surrey: - mp4: video; - txt: vehicle data Stockholm: - avi: video; - dat: label	Surrey: - raw data Stockholm: - video; - frame label	driving action analysis, driver attention analysis
Dr(eye)ve	2016 Modena (Italy)	urban, countryside, highway; sunny, cloudy, rainy; morning, evening, night	monocular color camera; driver's eye tracking device	74 sequences (5min each); avi: video; txt: driver fixation, vehicle data, annotation; png: image	raw data; annotation	driver attention analysis

often favored for testing purposes, sometimes adverse conditions are more desired to increase the robustness of algorithms under test. Data with adverse conditions can be accessed from a number of datasets: *AMUSE*, *CCSAD*, *CMU*, *Dr(eye)ve*, *ESATS*, *Elektra*, *Heidelberg*, *JAAD*, *Oxford*, *Stixel*, *HCI*, and *TRoM*. We also observe a trend that new datasets such as *BDDV* and *MVD* focus on environmental and weather diversity, covering various traffic/weather/light conditions.

The datasets exhibit a variety of sensor setups. The core sensors are camera, LiDAR, and GPS, which is often combined with IMU. 35 out 37 datasets include at least one type of camera except for *Stanford* and *TrafficNet*. Monocular cameras are more popular than stereo cameras, while the color option is slightly preferred to grayscale. Not too much attention has been given to omnidirectional cameras, which are only used in *AMUSE* and *Ford*. To our surprise, radar is used only in *Apollo*, *TrafficNet*, and *nuScenes*, even though radar has been widely used in modern vehicles for detecting objects. Our conjecture is that most automotive radars are commercial products with proprietary data formats that cannot be easily released publicly. Other types of sensors can also be observed in specific datasets, such as the monocular infrared camera in *Cheddar Gorge*, the eye tracker device to capture driver fixation in *Dr(eye)ve*, the far infrared sensor in *Elektra*, the airborne LiDAR in *TorontoCity*, and the thermal camera and beam splitter in *KAIST*. More advanced sensing devices including multiple types of sensors are found in *HCI* and *TrafficNet*, e.g., the mobile mapping system in *HCI* and Mobileye's vision-based ADAS in *TrafficNet*. An interesting phenomenon is that new datasets like *BDDV* and *MVD* start to apply a crowd-sourcing strategy, i.e., raw data is collected by external individuals instead of dedicated teams within the organization. This is an efficient way of enlarging the scale of a dataset. The challenge is how to make the collected data consistent in terms of format, size, and other aspects. We believe that this crowd-sourcing strategy is suitable for image/video collection, optionally with low precision GPS data, which can be captured by private mobile phones and low cost personal devices. It is impractical to use this strategy to collect more professional data such as LiDAR point cloud.

Data format is an important factor for dataset selection. In general, standard data formats are more favored than proprietary data formats because standard data formats are not restricted to use a specific software, thus allowing for more flexibility. Most datasets share data in standard formats. *AMUSE*, part of *Elektra*, *Malaga*, and *Stanford* contain own data formats and tools for parsing the data. The data format in *Cheddar Gorge* is still unclear because the only currently available resource for *Cheddar Gorge* is a scientific paper where its data format is not described; the data format of *TorontoCity* and *EuroCity* is also unclear.

We also investigated the data size of each collected dataset as shown in the Data format & size column in Table I–V. The data sizes of the 27 datasets reported in our previous survey [1] were obtained around April 2017, while the data sizes of the ten new datasets were obtained in 2018. Instead of updating the data sizes of these datasets over and over again, we do not aim to show the latest data size, which will probably change over time.

TABLE III
OVERVIEW OF EXISTING DRIVING DATASETS ON PUBLIC ROADS — PART 3

Dataset	Time & Venue	Traffic condition	Sensors	Data format & size	What is provided	Typical usage scenarios
ESATS	2007-2010 Stuttgart+Lippstadt (Germany), Auckland (New Zealand)	highway, rural, urban; various weather/light conditions, adverse conditions	2-3 monocular grayscale/color cameras	8 datasets included: Set 1, 3, 4, 5, 6, 7, 9, 10: 525MB, 12GB, 705MB, 200MB, 1GB, 3MB, 5.7GB, 1.54GB; pgm: grayscale image; ppm: color image; jpg/bmp: image; binary data	raw data	semantic segmentation, optic flow, driver's eye-status detection
Elektra	Apr-Jun, 2016 Barcelona (Spain)	urban, mostly daylight; night sequences in one dataset	monocular grayscale or color camera; stereo vision, color; far infrared sensor	6 relevant subsets (CVC 01, 02, 05, 08, 09, 14): 86.9MB, 2.44GB, 280MB, 2.11GB, 1.92GB, 3.48GB; png: image; pts own format: 3D points	raw data (training/test sets); annotation	semantic segmentation, pedestrian detection, driver face monitoring
ETH pedestrian	2009 Zürich (Switzerland)	downtown	2 monocular color cameras	660MB; png: image; cal: calibration; idl: annotation	raw images; calibration; annotation; demo videos	multi-person tracking
EuroCity	released in 2018 (data collection time unknown) 31 European cities	urban traffic, day/night time, all four seasons	monocular color camera	+47,300 images (1920*1024, 20Hz); format unknown	raw data (training/validation/test sets); annotations for pedestrians and riders: bounding box, tag for occlusion/truncation, body orientation; evaluation server	pedestrian and cyclist detection
Ford	Nov-Dec, 2009 Michigan (US)	downtown, loop closure, campus	Velodyne 64 LiDAR; omnidirectional camera; 2 Riegl LMS-Q120 LiDARs; Applanix+Trimble GPS; Xsens consumer IMU	ca 100GB; mat: Velodyne scan; ppm: image; log: sensor data and timestamp; pcap: Velodyne stream; mat: calibration;	raw data; C/Matlab code	SLAM, object detection etc.
German traffic sign	available from 2012-12-01 (Germany)	unknown	monocular color camera	1.6GB; ppm: image; csv: annotation	images(training/evaluation sets); annotation; C++/Matlab code	traffic sign detection
HCI	6 days (3 seasons) Heidelberg? (Germany)	one street with a T-section, urban	Mobile Mapping System RIGEL VMX-250-CS6: - 2 calibrated laser scanners RIGEL VQ-250; - 4 cameras; Applanix POS LV 510; stereo vision: 2 pro-edge cameras	+1,000 frames (2560*1080) png: image (grayscale), masks for dynamic objects (annotation)	raw data; an interactive web application for video display; SDK with visualization; benchmark tools; pixel-level manual annotation of objects	optic flow
Heidelberg	2011-2012 Hildesheim (Germany)	city/avenue/bend/parking/village, various lighting/weather conditions	2 monocular grayscale cameras (stereo vision)	10TB in total, ca 12.6GB available; pgm: image; png and h5: image; avi: demo video	raw data	stereo disparity and optic flow estimation

The size of most datasets falls in the range of 1–100 GB. The *Oxford robotcar* dataset is currently the largest dataset with 23TB among the 37 datasets. The sizes of *TorontoCity* and *EuroCity* are unknown because their websites are still not released yet and sizes are not mentioned in the corresponding publications.

The most fundamental resource provided by a driving dataset is raw sensor data. If the data is collected from multiple sensors, all sensor data must be properly calibrated, synchronized and accompanied by timestamps. Among the 37 datasets, most of them provide much more than raw sensor data. The typical complementary resources are annotations and labels (e.g., object bounding boxes), benchmark suites, source code, toolkit, scientific publications and demo videos. The raw sensor data, in particular visual data, is often classified into training, validation, and test sets for different purposes. In addition to raw sensor data, benchmark is deemed an extremely rewarding and appreciated feature that serves as an open evaluation platform for performance comparison. Various benchmarks are available in *Caltech*, *Cityscapes*, *German traffic sign*, *KITTI*, *Apollo*, *TorontoCity*, and *HCI*, where the performance of different algorithms submitted by the dataset users is ranked.

We have summarized the typical usage scenarios of the included 37 datasets, which are shown in the last column of Tables I–V. There are a variety of usage scenarios supported by these datasets such as optic flow and SLAM. The top two usage scenarios supported by most datasets are pedestrian/vehicle detection and semantic segmentation. The most popular datasets for pedestrian detection include *Caltech*, *Daimler pedestrian*, *ETH pedestrian*, and *KITTI*. The most popular datasets for semantic segmentation are probably *Cityscapes* and *KITTI*. The emerging *MVD* also starts to gain attention for semantic segmentation. *MVD* even surpasses *Cityscapes* in terms of density of object instances per image [37]. Some datasets have a dedicated purpose, e.g., *Dr(eye)ve* for driver attention analysis, *German traffic sign* for traffic sign detection, *TRoM* for road marking detection and classification, and *TrafficNet* for traffic scenario categorization. By contrast, some datasets like *AMUSE*, *CC-SAD*, *Cheddar Gorge*, *Ford*, and *Malaga* do not clearly reflect what usage scenarios they support. *KITTI* is undoubtedly the most outstanding dataset and benchmark with the most comprehensive coverage of usage scenarios. The most recently released *Apollo* and *BDDV* look also promising by virtue of their support for various usage scenarios.

Apart from these summary tables, we also investigated the license and accessibility of these datasets. Regarding the legal constraint of using these datasets, 11 datasets have declared the licenses under which they were published. Creative Commons Attribution-NonCommercial-ShareAlike 3.0 (CC-BY-NC-SA) is the most adopted license used by *CMU*, *comma.ai*, *Karlsruhe labeled objects*, *Karlsruhe stereo*, and *KITTI*. *nuScenes* is under the same license yet with version 4.0. *Elektra* and *Oxford* adopted for Creative Commons Attribution-NonCommercial

TABLE IV
OVERVIEW OF EXISTING DRIVING DATASETS ON PUBLIC ROADS — PART 4

Dataset	Time & Venue	Traffic condition	Sensors	Data format & size	What is provided	Typical usage scenarios
JAAD	Before Nov, 2016 Toronto (Canada) Kremenchuk+ Lviv (Ukraine), Hamburg (Germany), New York (US)	mainly urban, a few rural roads, most daytime, occasional night, sunset and sunrise, various weather conditions	monocular color camera	347 videos, 5-15s each; mp4: video; seq: video; vbb/tsv: textual annotation; xml: bounding box annotation	videos; textual and bounding box annotations; bash script for splitting videos	pedestrian and vehicle detection, pedestrian and driver behavior study
KAIST	before 2015 South Korea?	mainly urban, day/night time	monocular color camera; thermal camera; beam splitter	95,328 color-thermal image pairs (640*480, 20Hz); 35.9GB video, 48MB annotation; jpg: image; vbb: annotation	raw data: thermal and color image pairs (training/test sets); pedestrian annotation with occlusion tags; toolbox for viewing data in Matlab	pedestrian detection
Karlsruhe labeled objects	2011 Karlsruhe (Germany)	urban, daylight	monocular grayscale camera	631.2MB (ca 1,800 images with labels); png: image; mat: label	images; object labels; object orientation	vehicle detection, pedestrian detection
Karlsruhe stereo	2009-2010 Karlsruhe (Germany)	urban, rural, daylight	stereo vision, grayscale; GPS+IMU	20 sequences (0.2-1.4GB each); png: image; txt: GPS/IMU data	raw data; camera calibration	stereo vision
KITTI	Sep-Oct, 2011 Karlsruhe (Germany)	urban, rural, highway	2 monocular grayscale cameras; 2 monocular color cameras; Velodyne 64 LiDAR; GPS+IMU	180GB; png: image; txt: Velodyne and GPS/IMU data, calibration; xml: bounding box label	raw data; object annotation (3D bounding box); calibration; various benchmarks: stereo, optic flow, visual odometer, SLAM, 3D object detection/ tracking; development kit; Matlab/C++ code	stereo vision, optic flow, scene flow, depth map, visual odometry, SLAM, 3D object detection and tracking, road/lane detection, pixel/instance-level semantic segmentation
Malaga	Before Feb, 2014 Malága (Spain)	urban, highway, loop closure, direct sun etc.	stereo vision, color; 3 Hokuyo UTM-30LX laser scanners; 2 Sick LiDARs; GPS+IMU	15 sequences (+70GB); txt: raw laser scan, GPS/IMU data, camera calibration; jpg: image; rawlog: own format binary; kml: Google earth file to represent path	raw data; C++ example code for parsing rawlog files; demo videos; support for posting public messages by users	stereo vision, visual odometry, visual SLAM, object detection
MVD	released in 2018 (data collection time unknown) Europe, North/ South America, Asia, Africa, Oceania	various road/ weather/light conditions	cameras of different devices: mobile phones, tablets, action cameras, professional capturing rigs...	25,000 images (25.6GB); jpg, png: image	raw data (training/ validation/test sets); object annotations	semantic segmentation, object detection

4.0 (CC-BY-NC) License. *AMUSE* is licensed under Creative CommonsAttribution-NonCommercial-NoDerivatives 3.0 Unported (CC-BY-NC-ND) License. *Udacity* applies the MIT license for its data, while everything else is licensed under GPLv3. *MVD* applies Mapillary Vistas Dataset Research Use License. The licenses of the remaining datasets are not clearly specified.

Although our dataset inclusion criteria require either partial or full open access, accessibility still varies a lot among these datasets. While most datasets allow convenient and direct data download, some datasets make data access more complicated to various degrees. *Cityscapes*, *Elektra*, *Heidelberg*, *Dr(eye)ve*, *BDDV*, *MVD* and *nuScenes* require a valid email address to obtain the download links. *Cityscapes* and *MVD* are relatively more stringent in the sense that a data user must register an account with work email as private email is not accepted, and any new registration will be manually inspected and it takes a few days to get it approved. *Apollo* provides sample data which can be accessed with mobile phone registration. Access to the entire dataset would require an online application where more detailed information and motivation must be indicated. *TRoM* does not have an official dataset web page, but instead shares the raw data on Baidu Cloud web disk with Chinese as its only language option. The publication of *TRoM* [43] states that a toolkit for road marking annotation is also available together with the dataset. However, this toolkit was not found on the web disk as investigated at the time of writing. *TrafficNet* classifies driving data into eight different scenarios; though, only six scenarios are reported in [42]. Nevertheless, only the two scenarios "lane change" and "car following" are available to the public, whereas the other six scenarios are only available to Mcity members. *HCI* requires a user to install the SDK toolkit provided on its website to download data. Meanwhile, part of *HCI* is used for the Stereo Geometry Challenge 2016, where data can be directly downloaded. However, since May 2018 we observed dramatic change of the *HCI* dataset website. Data for Stereo Geometry Challenge 2016 was no longer available. Instead, data for Robust Vision Challenge 2018 was provided upon registration by email. The data of *TorontoCity* was still not released at the time of writing, though its authors have promised that it will come soon. *Cheddar Gorge* can only be obtained by sending hard disks to the provider. *EuroCity* allows free download for non-commercial use as stated in [26], however, the website of *EuroCity* is not given yet by the time of writing (August 2018).

Despite the comprehensive summary and comparison of datasets, completeness is still a potential threat to validity. It is

TABLE V
OVERVIEW OF EXISTING DRIVING DATASETS ON PUBLIC ROADS — PART 5

Dataset	Time & Venue	Traffic condition	Sensors	Data format & size	What is provided	Typical usage scenarios
Oxford	from 2014-05-06 to 2015-12-13 Oxford (UK)	various light/ weather conditions	stereo vision, color; 3 monocular color cameras; 2 Sick 2D LiDARs; Sick 3D LiDAR; GPS/INS	133 sequences (23.15TB); png: image; bin: LiDAR data; csv: GPS/INS data	raw data; calibration; Matlab/Python tools	long-term localization and mapping
Stanford	2009-2010 San Francisco (US)	urban, campus, intersections	Velodyne 64 LiDAR; Applanix (GPS/IMU)	33 files (5.72GB); tm: Velodyne and Applanix data (own format)	raw data; background data without objects(training and testing sets); object labels; code in ROS package	3D object detection and classification
Stixel	2013	highway, good weather, heavy rain	stereo vision, grayscale	12 sequences (3.1GB); pgm: image; xml: ground truth stixel	videos; ground truth stixel (own novel concept); vehicle data including timestamps	stereo vision
TorontoCity	2009, 2011-2013 GTA (Canada)	densely populated area with many buildings	Vehicle: - PointGray Bumblebee3 stereo camera; - Velodyne HDL-64E LiDAR; - GoPro Hero 4 RGB camera; - Applanix POS LV; Airplane: airborne LiDAR; Drone: 3D camera	Size: unknown; image: format not mentioned	raw data; benchmark; annotated map	building height estimation, road extraction, building segmentation, building recognition, semantic labeling, scene type classification
TrafficNet	from 2012-10-01 to 2013-04-30 Ann Arbor, Michigan (US)	various road conditions	Mobileye's vision-based ADAS; Wireless Safety Unit (WSU); radar unit (part of the vehicle's integrated safety device unit)	10.3GB; csv for all data	raw data indexed by 8 scenarios; source code of scenario categorization methods	traffic scenario categorization
TRoM	from 2016-05-02 to 2016-06-03 Beijing (China)	different weather, time-of-the-day, illumination, traffic road conditions	PointGray color camera; GPS receiver	+700 scenes; jpg: image (RGB color)	raw data (training/validation/ test sets); pixel-level annotation of 19 road marking categories; annotation toolkit in Matlab; preliminary evaluation results	road marking detection and classification
Udacity	Sep-Oct, 2016 Mountain View (US)+around	sunny, overcast, daylight	monocular color camera; Velodyne 32 LiDAR; GPS+IMU	223GB (>10h); png or jpg: image; log: GPS and vehicle motion; csv: label; ROSBAG	videos; labels: vehicle, pedestrian, traffic lights; open source code; tools for ROSBAG files	vehicle/pedestrian/ traffic light detection, steering angle prediction

difficult to rule out the possibility of missing other existing relevant datasets; however, we adopted and applied a thorough and rigor approach to explore and create this overview of datasets. In addition, the factors for dataset comparison are defined on the basis of our expertise and experience conducting research and development for more than ten years in this area. Other dataset users may be interested in certain aspects or use cases of the datasets, which cannot be discussed in detail in this overview but would require specific subsequent studies.

C. Overview of the Virtual Testing Environments

Complementary to our previous work [1], we have added an overview of virtual testing environments to enable closed-loop testing during the development of algorithms for self-driving vehicles. 22 virtual testing environments are listed below in alphabetic order with references, links, and highlights. Names of the providers are also indicated in parentheses where applicable.

Before coming into the following discussion, we would like to comment that the datasets as listed in Tables I–V are typically providing video stream, single image files, or lists of character-separated values (CSV) to be used for offline data processing (e.g, for training neural networks (NN)). Simulation and virtual testing environments in the following sections, though, typically serve use-cases where an algorithm is connected with the simulation system to get stimuli data from the simulation and to relay back its output into the simulation environment. Using a dataset from the aforementioned tables in simulation systems is usually not the intended use-cases and thus, this paper mainly focuses on the overview of the testing environments on use-case specified aspect. Nevertheless, the combination of the two ends is indeed practical and might be discussed and evaluated in detail in future work. The links have been verified on 2018-10-20.

- *Environment 1: AirSim (Microsoft) [44]* (https://goo.gl/XV6e23)
  Highlight: Multirotor drone simulator using Unreal Engine; also supports vehicle simulation
- *Environment 2: ASM Traffic (dSpace) [45]* (https://goo.gl/CcsmMD)
  Highlight: Integrated real-time environment simulation for ADAS controllers with traffic and infrastructure
- *Environment 3: CARLA [46]* (https://goo.gl/vDKKEp)
  Highlight: Urban driving simulator for autonomous approaches with flexible specification of sensors and environmental conditions
- *Environment 4:CarMaker (IPG)* (https://goo.gl/HQ7QAx)
  Highlight: Simulation solution specifically for testing passenger cars and light-duty vehicles
- *Environment 5: DYNA4 (TESIS)* (https://goo.gl/GEq6rR)

TABLE VI
OVERVIEW OF EXISTING VIRTUAL TESTING ENVIRONMENTS — PART 1

Virtual testing environments	Latest release	Accessibility	Platform	Use-cases	Language
AirSim (Microsoft)	June. 2018 v1.2	Open source (MIT License)	Linux, Windows	Drone and car simulation 3D visual environment HIL controller support	C++, Python, C# and Java
ASM Traffic (dSpace)	2017	Commercial	N/A	DIL traffic environment simulation for ADAS controllers	N/A
CARLA	Jul. 2018 v0.9.0	Open source (MIT License)	Linux (Ubuntu 16.04 or later)	3D urban environment Camera and sensor simulation	Python
CarMaker (IPG Automotive)	N/A	Commercial Free trial on demand	N/A	Virtual testing driving	N/A
DYNA4 (TESIS)	2017 V2.8	Commercial	Windows	Modular simulation SIL and HIL functional testing Report and analysis generation	C/C++, Matlab/Simulink
Gazebo for ROS	Jan. 2018 v9.0.0	Open source (Apache 2.0)	Linux, Mac OS X, Windows	Robot dynamics simulation 3D visual environment Sensor data generation	C++
(Simulator of) Hank Virtual Env. Lab	N/A	Access on demand	N/A (Hardware platform)	Bicycling and pedestrian simulator	N/A
Legion for Aimsun	N/A	Commercial	(Aimsun Plug-in)	Integrated pedestrian and traffic simulation for traffic engineering and planning	N/A
OpenDaVINCI & OpenDLV	Sep. 2017 v4.16.0	Open source (GPLv2, LGPLv2)	POSIX-compatible OS, Windows	Environment visualization Sensor model Autonomous driving	C++, Python
PELOPS (fka)	2011	Commercial	Linux	Traffic simulation combining sub-microscopic vehicle model and microscopic traffic model	N/A
PreScan (Tass)	2018 v8.5	Commercial Free trial on demand	Windows	Sensor simulation for ADAS HIL driving simulation	Matlab/Simulink
PTV Vissim	v10.0	Commercial Free trial available	Windows	Road junction geometry Public transport simulation Active traffic management	N/A
Racer	Aug. 2014 v0.9.0	Free for Non-commercial use	Linux, Mac OS X, Windows	3D car racing simulation High DOF car modeling	C++
SCANeR Studio (OKTAL)	Oct. 2017 v1.7	Commercial	Windows	Traffic scenario simulation Vehicle dynamics Autonomous driving	C++, Matlab/Simulink
Sim IV (VTI)	N/A	Commercial	N/A (Hardware platform)	2-Axe driving simulator facility with 210° forward FOV	N/A
Speed Dreams	Dec. 2015 v2.2 Beta	Open source (GPL)	Linux, Mac OS X, Windows (32-bit)	3D car racing simulation (TORCS alternative) Simu V3 physics engine	C/C++
SUMO	Dec. 2017 v0.32.0	Open source (EPL v2)	Linux, Windows	Urban traffic flow simulation Vehicular communication	C++
TORCS	Mar. 2017 v1.3.7	Open source (GPL v2)	Linux, FreeBSD, OpenSolaris, Mac OS X, Windows	3D car racing simulation Programmable AI for racing	C/C++
VDrift	Oct. 2014	Open source (GPL v2)	Linux, FreeBSD, Mac OS X, Windows	3D car racing simulation Driving physics	C++
V-Rep (Coppelia)	Feb. 2018 v3.5.0	Commercial Free educational license possible	Linux, Mac OS X, Windows	Virtual robot simulator Robotic dynamics and kinematics Sensor simulation	C/C++, Python, Matlab, Octave, Java and Lua
VTD (Vires)	May. 2018	Commercial	N/A	Driving simulation tool-chain Free data standards	N/A

Highlight: Modular simulation software with integration of vehicle variants and driving tasks

- *Environment 6: Gazebo for ROS [47]* (https://goo.gl/NaD5pF)
  Highlight: Physics engine and visualization for robot simulation
- *Environment 7: Hank Virtual Environments Lab [48]* (https://goo.gl/7FyM2B)
  Highlight: Simulator facility for bicycles and pedestrian
- *Environment 8: Legion for Aimsun* (https://goo.gl/Dc99a2)
  Highlight: Integrated simulation of pedestrian and traffic
- *Environment 9: OpenDaVINCI & OpenDLV [49]* (https://goo.gl/1HK5nj)
  Highlight: Enabling deterministic, distributed, and repeatable simulations with transparent control of communication, time, and scheduling; provides models for vehicle motion and sensors; allows for combining existing independently running distributed simulations
- *Environment 10: OpenDS* (https://goo.gl/ph7rgt)
  Highlight: 3D driving simulation with integration of physics and road semantics
- *Environment 11: PELOPS [50]* (https://goo.gl/danSWm)
  Highlight: Microscopic, vehicle-orientated traffic simulation program
- *Environment 12: PreScan (Tass International)* (https://goo.gl/csDkzZ)

TABLE VII
OVERVIEW OF EXISTING VIRTUAL TESTING ENVIRONMENTS — PART 2

Virtual testing environments	Supplementary links for use-cases *Examples of recent research projects (if applicable)	Comments
OpenDaVINCI & OpenDLV	http://youtu.be/hWHFiaaHjSk	Validated results of autonomous driving functionality in reality
	http://youtu.be/sGuOqQEV97c	Simulation of lane detection
	*https://goo.gl/dQeW8K	Project COPPLAR: Campus shuttle perception and planning
	*https://goo.gl/gBqHDS	Project CAIVE: Cyberphysical training grounds for artificial intelligence-based vehicular functions
VDrift	http://youtu.be/n7370dzCZ1o	Example of Deep Q-learning agent
	*https://goo.gl/BdJ9QW	Reinforcement learning for End-to-End autonomous driving
TORCS	http://youtu.be/qOvEz3-PzRo	Autonomous control based on fuzzy logic
	https://goo.gl/hm4XT1	Tutorial of AI driver (robot) for TORCS
	*https://goo.gl/vpKvZK	Project EphemeCH: Bio-inspired algorithms in complex ephemeral environments
	*https://goo.gl/qNPAh7	Project DeepDriving: Learning affordance for direct perception in autonomous driving
MS AirSim	http://youtu.be/fv-oFPAqSZ4	Reinforcement Learning for vehicles
	https://goo.gl/u8UxR6	Microsoft Research Blog archive
	*https://goo.gl/uRvM1g	Reinforcement Learning framework
CARLA	http://youtu.be/cFtnflNe5fM	End-to-end driving via conditional imitation learning
Gazebo for ROS	http://youtu.be/tQpD9GCn7Ts	Validated results of autonomous driving on robot in reality
	https://goo.gl/rujqdn	Tutorial and API provided by Gazebo
	*https://goo.gl/9QeR9R	Methods for identification of interaction models of robot
V-Rep	http://youtu.be/xI-ZEewIzzI	Line-following vehicular robot
	*https://goo.gl/fdDgGc	Project Poppy: Open source platform for interactive 3D-printed robots
CarMaker	http://youtu.be/UOzT02EjRvw	Driving environment and scenario rebuilding
	*https://goo.gl/cFh888	Formula Student Simulator for Driver Training
SCANeR Studio	http://youtu.be/nKm1D-z1cac	Hardware-in-the-loop driving simulator
	*https://goo.gl/X6N5EW	Project VERVE: Novel vehicle dynamics control technique for enhancing active safety of intelligent electric vehicles
PreScan	https://goo.gl/Ysq5q9	Tutorial of V2V and V2P Autonomous Emergency Braking
	*https://goo.gl/Cga7HW	Assistive software benefits in the process of autonomous driving
PELOPS	https://goo.gl/rQn6mc	Field Operational Tests (FOT) wiki for PELOPS
PTV Vissim	http://youtu.be/lF_RoSlofyA	Simulation of non-lane-Based traffic
	http://youtu.be/gUwS_9KCoT0	Tutorial of scenario and traffic flow design
SUMO	http://youtu.be/OEyzdEqacko	Traffic simulation example with OpenStreetMap
	https://goo.gl/RbLqxZ	Wiki and tutorial for SUMO
	*https://goo.gl/sBEMNd	Proceedings of SUMO User Conference 2017
Legion for Aimsum	http://youtu.be/MvcjfhCzUjI	Barcelona Camp Nou Stadium: crowd and vehicle modeling
DYNA4	http://youtu.be/2fJVDl86drM	Virtual ADAS testing environment: Berlin City
	http://youtu.be/emKzNi8F73A	Tire model integration in vehicle simulation
VTD	http://youtu.be/Om5VfKpw2pQ	Introduction for the development of automated vehicles
ASM Traffic	http://youtu.be/6VgKTxmy0-0	Test scenario for validating intersection assistants
Sim IV	http://youtu.be/FXwGTgZHqzA	Presentation of VTI driving simulator
Hank Virtual Env. Lab	https://goo.gl/a3kdjg	Simulation examples of pedestrian interaction

Note: Virtual testing environments are listed according to the order of usage.

Highlight: Physics-based simulation platform for sensor- and V2X-based ADAS development

- *Environment 13: PTV Vissim* (https://goo.gl/WNruwK)
  Highlight: Traffic simulation for junction design and active traffic management
- *Environment 14: Racer (Cruden)* (https://goo.gl/C4CDTX)
  Highlight: High-quality rendering engine and motion formulas from actual engineering documents
- *Environment 15: SCANeR Studio (OKTAL)* (https://goo.gl/paQW5H)
  Highlight: Simulation of driving scenarios with fine-tuning dynamic models and data analysis
- *Environment 16: Sim IV environment (VTI)* (https://goo.gl/1wkjcm)
  Highlight: Simulator facility with simultaneous longitudinal and lateral motion and wide field of view
- *Environment 17: Speed Dreams* (https://goo.gl/1ZjbaV)
  Highlight: A fork of TORCS (Environment 19), aiming at realization of proposals that the official update failed to respond
- *Environment 18: SUMO [51]* (https://goo.gl/f6CUiz)
  Highlight: Microscopic and multi-modal traffic simulator about mobility in urban road network
- *Environment 19: TORCS [52]* (https://goo.gl/bCZxyH)
  Highlight: Vast number of programmable (customizable) AI-agents; large user group
- *Environment 20: VDrift* (https://goo.gl/kMnRFK)
  Highlight: Pre-built real-world models, realistic physics simulations, and scenarios
- *Environment 21: V-Rep (Coppelia Robotics)* (https://goo.gl/ja86if)
  Highlight: Virtual robot simulator and development based on distributed control architecture
- *Environment 22: VTD (Virtual Test Drive) environment (Vires)* (https://goo.gl/Y4C1Z8)

Highlight: Complete tool-chain for driving simulation including ADAS and automated systems

D. Discussion of the Virtual Testing Environments

Similar to Tables I–V, we provide a summary of our overview of the virtual testing environments focusing on the following perspectives: (1) Latest release: what was the most recent version available for the virtual testing environment and when was it released? (2) Accessibility: is the environment open source or available through commercial access? (3) Platform(s) on which the virtual testing environment is supported; (4) Typical use-cases / examples presenting the usage as well as levels of fidelity; and (5) Language: which programming language(s) does the virtual testing environment support for developers? The overview is summarized in Table VI, where each row represents a virtual testing environment and each column corresponds to the perspectives (1)–(5) above. The "N/A" in this table denotes cases where related information is limited due to accessibility.

We believe such survey upon virtual testing environments by considering the perspectives listed above is of critical significance, especially in comparison with experiments on real vehicles. Virtual environments with easy accessibility enable to a large extent the progress of testing autonomous algorithms in a swift and safe sandbox before engaging the invalidated thus potentially dangerous situations in real driving scenarios [53]. Taking a step further, open-source virtual testing environments also outrun in reducing the facility cost (experimental vehicles, sensors, maintenance of testing field, etc.) to the minimum, while those with multi-platform and/or programming languages support a large range of choice for the researchers by reducing the barrier and unnecessary dependencies.

A majority of the virtual testing environments in our survey were, and still are, recently supported and updated. For most cases of the open source projects, a full list of update and release history is commonly available on the website or on GitHub platform respectively. We can observe that several virtual testing environments in our survey are outdated in terms of recent updates. These projects were mainly created and supported as part-time work by individual developers, nonetheless they still serve as alternative options in case a free, light-weighted virtual testing environment is in need. Simulation environments such as OpenDaVINCI, TORCS, ROS-based Gazebo, and V-Rep have relatively larger user groups, which leads to frequent and regular updates. On the other hand, commercial software and integrated solutions do not necessarily provide release history or previous versions to customers. Instead, it is more common to receive news about latest updates from their website, commercial mailing list, or RSS feeds in such case.

Accessibility is one of the major concerns during our survey. It is widely accepted that simulation platforms that aim at non-commercial, scientific research purpose should be approachable as open source under certain public licenses. Most common licenses that open source simulators follow are GPL, MIT, EPL, and Apache. In addition, a considerable amount of commercial software provides free trial version or educational license, e.g., the V-Rep robot simulator.

It is also crucial to note that quite a few virtual testing environments are unable to be cited in our survey because of accessibility issues. For instance, DeepDrive, which aimed at self-driving AI development based on virtual testing environment of the game GTA-V, was shut down for legal reasons. Also, many virtual testing environments only appeared respectively in papers of conferences and workshops, and thus, are not publicly accessible for peer researchers.

In terms of platforms or operating systems for the virtual testing environments, open source instances are naturally cross-platforms projects, as they tend to be developed on Linux, and in many cases Mac OS, FreeBSD and others. For example, OpenDaVINCI provides compilable source code for a variety of POSIX-compatible operating systems, and CARLA as a young project runs on Ubuntu only. Both open source and commercial software provide releases on Windows, and in latter cases mostly on Windows only.

The column of use cases contains examples of typical usage for each virtual testing environment respectively. Due to the limited accessibility, it is impossible to inspect each in detail. Nonetheless, we have compared the use cases according to their usage. We summarize the survey result in categories as illustrated in Table VII and include further examples of use cases of each environment that we surveyed.

The following is a summary of our experience upon different virtual testing environments and their fidelity. OpenDAVINCI leads the survey list with the highlight of not only providing a virtual testing environment for simulations, but also being a realtime middleware that has been demonstrated on several self-driving vehicles ranging from the 2007 DARPA Urban Challenge up to the 2016 Grand Cooperative Driving Challenge. VDrift and TORCS stand out as 3D environments with virtual vehicle and AI drivers, the latter of which outperforms by larger user groups and regular update. The recent approach of Microsoft AirSim enables autonomous vehicle simulation for the formal drone-dedicated simulator, which yields promising performance by dynamics modeling and Unreal rendering engine. The same engine is also involved in the CARLA project, which provides open digital assets for establishing an urban environment. Last but not least in the list of open source environments for autonomous driving simulation, ROS-based Gazebo stands out with support for robotics and vehicle dynamics.

On the other hand, commercial testing environments also hold a significant place for researchers and industrial users as integrated solution providers as they could adapt to the demands of specific purposes. Both SCANeR Studio (OKTAL) and DYNA4 (TESIS) offer customized training courses for better utilization of their products. The widely acknowledged and mature PELOPS has always been supporting industrial and technical research projects since several years. VTD (Vires) contributes and manages OpenDRIVE standard for the description of road networks in driving simulations.

In terms of virtual testing environments for traffic flow simulation, the open source project SUMO attracts a large number of users by its microscopic, multi-modal traffic characteristics, while PTV Vissum is also competitive by integrating active traffic management and geometry of road and intersection design. From the point of view of ADAS controller applications, ASM

Traffic (dSpace) is typically designed for HIL testing of electronic control units (ECUs) or for early function validation by offline simulation.

The two hardware testing environments with virtual reality approach that are listed in our survey are Sim IV (VTI) and simulation environment of Hank Virtual Env. Lab. The simulator facility of VTI is dedicated to realistic and simultaneous simulation of lateral and longitudinal acceleration, as well as a wide synthetic forward field of vision. Bicycle/pedestrian simulators powered by Hank Lab enables direct perception and interaction between human users and virtual objects in the environment, thus creating a creditable platform for research of traffic safety.

Alongside the survey of particular use-cases that are provided by or accomplished on the virtual testing environments, we suppose that validity is also a significant aspect worth consideration in general. "How close is the modeled realism in the virtual simulation compared to the real world (level of fidelity)" has always been one of the major interests and in various cases, a threat to validity for all approaches related to virtual testing environments and simulation work on them [54]. Lighting conditions of synthetic scenarios are another widely recognized issue that limits the validity of virtual testing environments [55], especially for approaches that are highly related to computer vision and sensor simulation.

V. Conclusion & Future Work

The global race to develop, evaluate, and deploy algorithms and solutions to realize self-driving vehicles has significantly heated up – first solutions at SAE Level 3 are being made available to customers addressing automated driving on highways for example. The community around this comprises researchers, major automotive OEMs, as well as young start-ups. They all have in common that they need open-loop and closed-loop solutions to systematically develop, test, and evaluate their approaches. Especially for researchers and young start-ups, the threshold for contributing to the field is high as collecting own datasets is resource-consuming and time-intensive.

Our work presents a combined survey for publicly available datasets next to an overview of virtual testing environments to support the research, development, and evaluation of algorithms from the field of autonomous driving. We present 37 datasets from different perspectives such as included driving situations, sensor setups, data format; additionally, up to 22 virtual testing environments are presented to support the closed-loop testing. To the best of our knowledge, our work is the first and most comprehensive survey of this kind providing guidance about existing and publicly available assets to researchers and developers. Future work in this area should evaluate what combination of publicly available datasets and virtual testing environments will result in the best fidelity level in terms of resulting performance in reality compared to what has been achieved with open-loop and closed-loop testing. Recent approaches combining dataset collection and virtual environment usage [56] and [57] have been observed encouraging future work. Furthermore, commonalities between existing datasets could be studied in greater details to find overlapping or complementary parts; in that regard, a standardized representation or encoding of all datasets could be proposed to enable a simplified comparison of a system-under-test using various datasets. Such standardized representation could also serve as joint interface to various test environments to enable better modularity and reuse. Furthermore, an empirical study involving many dataset users will help to identify the most crucial factors to be considered during dataset selection.

Acknowledgment

The authors would like to thank O. Benderius and Y. Yu for their valuable comments.

References

[1] H. Yin and C. Berger, "When to use what data set for your self-driving car algorithm: An overview of publicly available driving datasets," in *Proc. 20th IEEE Int. Conf. Intell. Transp. Syst.*, Yokohama, Japan, Oct. 2017, pp. 1–8.

[2] F. W. Rauskolb *et al.*, "Caroline: An autonomously driving vehicle for urban environments," *J. Field Robot.*, vol. 25, no. 9, pp. 674–724, Sep. 2008.

[3] A. Knauss, J. Schroeder, C. Berger, and H. Eriksson, "Paving the roadway for safety of automated vehicles: An empirical study on testing challenges," in *Proc. 28th IEEE Intell. Veh. Symp.*, May 2017.

[4] M. Bojarski *et al.*, "End to end learning for self-driving cars," arXiv:1604.07316, 2016.

[5] C. Berger *et al.*, "Simulations on consumer tests: A systematic evaluation approach in an industrial case study," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst.*. Qingdao, China, Oct. 2014, pp. 1474–1480.

[6] C. Berger, "Scenario pattern matching in large sensor recordings with simulation models for cyber-physical systems," in *Proc. Summer Comput. Simul. Conf.*, Monterey, CA, USA, Jul. 2014, p. 8.

[7] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: Problems, datasets and state-of-the-art," arXiv:1704.05519, 2017.

[8] L. Li, W. Huang, Y. Liu, N. Zheng, and F. Wang, "Intelligence testing for autonomous vehicles: A new approach," *IEEE Trans. Intell. Veh.*, vol. 1, no. 2, pp. 158–166, Jun. 2016.

[9] A. M. Bozorgi, M. Farasat, and A. Mahmoud, "A time and energy efficient routing algorithm for electric vehicles based on historical driving data," *IEEE Trans. Intell. Veh.*, vol. 2, no. 4, pp. 308–320, Dec. 2017.

[10] P. Koschorrek, T. Piccini, P. Öberg, M. Felsberg, L. Nielsen, and R. Mester, "A multi-sensor traffic scene dataset with omnidirectional video," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, Jun. 2013, pp. 727–734.

[11] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3530–3538.

[12] F. Yu *et al.*, "BDD100K: A diverse driving video database with scalable annotation tooling," arXiv:1805.04687, 2018.

[13] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 304–311.

[14] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognit. Lett.*, vol. 30, pp. 88–97, 2009.

[15] R. Guzmán, J.-B. Hayet, and R. Klette, "Towards ubiquitous autonomous driving: The CCSAD dataset," in *Proc. 16th Int. Conf. Comput. Anal. Images Patterns*, Sep. 2015, pp. 582–593.

[16] R. Simpson, J. Cullip, and J. Revell, "The cheddar gorge data set," BAE Systems (Operations) Limited, UK, Tech. Rep., 2011.

[17] M. Cordts *et al.*, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3213–3223.

[18] S. Zhang, R. Benenson, and B. Schiele, "CityPersons: A diverse dataset for pedestrian detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4457–4465.

[19] E. Santana and G. Hotz, "Learning a driving simulator," arXiv:1608.01230, 2016.

[20] X. Li *et al.*, "A new benchmark for vision-based cyclist detection," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2016, pp. 1028–1033.

[21] T. Scharwächter, M. Enzweiler, U. Franke, and S. Roth, "Stixmantics: A medium-level model for real-time semantic scene understanding," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 533–548.
[22] N. Pugeault and R. Bowden, "How much of driving is preattentive?" *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5424–5438, Dec. 2015.
[23] S. Alletto, A. Palazzi, F. Solera, S. Calderara, and R. Cucchiara, "DR(eye)VE: A dataset for attention-based tasks with applications to autonomous and assisted driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2016, pp. 54–60.
[24] Z. Liu and R. Klette, "Performance evaluation of stereo and motion analysis on rectified image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1831–1846, Oct. 2009.
[25] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "Robust multiperson tracking from a mobile platform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1831–1846, Oct. 2009.
[26] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, "The eurocity persons dataset: A novel benchmark for object detection," arXiv:1805.07193, 2018.
[27] G. Pandey, J. R. Mcbride, and R. M. Eustice, "Ford campus vision and lidar data set," *The Int. J. Robot. Res.*, vol. 30, no. 13, pp. 1543–1552, Nov. 2011.
[28] D. Kondermann *et al.*, "The HCI benchmark suite: Stereo and flow ground truth with uncertainties for urban autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2016, pp. 19–28.
[29] S. Meister, B. Jähne, and D. Kondermann, "Outdoor stereo camera system for the generation of real-world benchmark data sets," *J. Opt. Eng.*, vol. 51, no. 2, 2012, Art. no. 021107.
[30] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, "Joint attention in autonomous driving (JAAD)," arXiv:1609.04741, 2016.
[31] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baselines," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, Art. no. 1037–1045.
[32] A. Geiger, C. Wojek, and R. Urtasun, "Joint 3D estimation of objects and scene layout," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 1467–1475.
[33] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3d reconstruction in real-time," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2011, pp. 963–968.
[34] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The KITTI Dataset," *The Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.
[35] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.
[36] J.-L. Blanco, F.-A. Moreno, and J. González-Jiménez, "The Málaga urban dataset: High-rate stereo and lidars in a realistic urban scenario," *Int. J. Robot. Res.*, vol. 33, no. 2, pp. 207–214, 2014.
[37] G. Neuhold, T. Ollmann, S. R. Bulò, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 5000–5009.
[38] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000 km: The oxford robotcar dataset," *The Int. J. Robot. Res.*, vol. 36, no. 1, pp. 3–15, 2017.
[39] A. Teichman, J. Levinson, and S. Thrun, "Towards 3D object recognition via classification of arbitrary object tracks," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 4034–4041.
[40] D. Pfeiffer, S. Gehrig, and N. Schneider, "Exploiting the power of stereo confidences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 297–304.
[41] S. Wang *et al.*, "TorontoCity: Seeing the World with a Million Eyes," arXiv:1612.00423, 2016.
[42] D. Zhao, Y. Guo, and Y. J. Jia, "TrafficNet: An open naturalistic driving scenario library," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst.*, Oct. 2017, pp. 1863–1870.
[43] X. Liu, Z. Deng, H. Lu, and L. Cao, "Benchmark for road marking detection: Dataset specification and performance baseline," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst.*, Oct. 2017, pp. 1871–1876.
[44] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "Airsim: High-fidelity visual and physical simulation for autonomous vehicles," in *Field and Service Robotics*. New York, NY, USA: Springer, 2018, pp. 621–635.
[45] M. Beine and A. Geburzi, "Test environment for safety-critical driver assistance systems," in *Elektronik Automotive, Special Edition*, Sep. 2016.
[46] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. 1st Annu. Conf. Robot Learn.*, 2017, pp. 1–16.
[47] N. P. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 4, 2004, pp. 2149–2154.
[48] Y. Jiang *et al.*, "Acting together: Joint pedestrian road crossing in an immersive virtual environment," in *Proc. IEEE Virtual Reality*, Mar. 2016, pp. 193–194.
[49] C. Berger, *Automating Acceptance Tests for Sensor- and Actuator-Based Systems on the Example of Autonomous Vehicles*. Aachen, Germany: Shaker, Aug. 2010.
[50] FKA GmbH, Aachen, Germany, *PELOPS White Paper*, 2007. [Online]. Available: fka.de/images/pelops-whitepaper-e.pdf
[51] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "SUMO (simulation of urban MObility)-an open-source traffic simulation," in *Proc. 4th Middle East Symp. Simul. Modelling*, 2002, pp. 183–187.
[52] B. Wymann, E. Espié, C. Guionneau, C. Dimitrakakis, R. Coulom, and A. Sumner, TORCS, the open racing car simulator, 2015.
[53] H.-P. Schöner, "Simulation in development and testing of autonomous vehicles," in *Proc. 18th Int. Stuttgarter Symp.*, 2018, pp. 1083–1095.
[54] Y. Kang and A. C. Victorino, "Human-vehicle cooperative driving using image-based dynamic window approach: System design and simulation," in *Proc. 19th IEEE Int. Conf. Intell. Transp. Syst.*, 2016, pp. 2487–2492.
[55] V. Veeravasarapu, R. N. Hota, C. Rothkopf, and R. Visvanathan, "Simulations for validation of vision systems," arXiv:1512.01030, 2015.
[56] J. Morton, T. A. Wheeler, and M. J. Kochenderfer, "Closed-loop policies for operational tests of safety-critical systems," *IEEE Trans. Intell. Veh/.*, vol. 3, no. 3, pp. 317–328, Sep. 2018.
[57] M. Toromanoff, E. Wirbel, F. Wilhelm, C. Vejarano, X. Perrotton, and F. Moutarde, "End to end vehicle lateral control using a single fisheye camera," arXiv:1808.06940, 2018.

Yue Kang received the Ph.D. degree from the Université de Technologie de Compiègne, Sorbonne Universités, Paris, France, in 2016. Since September 2017, he has been a Postdoc with the Chalmers University of Technology, Gothenburg, Sweden, in the area of virtual simulation for autonomous driving. His research interests include autonomous driving, system modeling, sensor-based navigation, HIL cooperative driving, software engineering, etc.

Hang Yin received the Ph.D. degree from Mälardalen University, Väster's, Sweden, in 2015. From 2015 to 2017, he was a Postdoc with the Chalmers University of technology, Gothenburg, Sweden, in the area of software/system architecture for autonomous driving. Since January 2018, he has been with Zenuity AB, Gothenburg, Sweden, as a Software Development Engineer. His research interests include autonomous driving, embedded systems, real-time systems, software engineering, etc.

Christian Berger received the Ph.D. degree from RWTH Aachen University, Aachen, Germany, in 2010. He coordinated the research project for the vehicle "Caroline," which participated in the 2007 DARPA Urban Challenge Final. He also co-led Chalmers Truck Team during the 2016 Grand Cooperative Driving Challenge (GCDC), and is one of the two leading architects behind OpenDLV. He is an Associate Professor with the Department of Computer Science and Engineering, University of Gothenburg, Gothenburg, Sweden. His research expertise is on distributed realtime software, microservices for embedded systems and cyber-physical systems, and continuous integration/deployment/experimentation for embedded systems.