Driving Datasets Literature Review

for

BRITE: Bus RapId Transit systEm

Summer Internship 2019

Student: Charles-Éric Noël Laflamme

Supervisor: Philippe Giguère philippe.giguere@ift.ulaval.ca

Cosupervisor: François Pomerleau francois.pomerleau@ift.ulaval.ca

Laval University

Québec City, Canada

August 29th, 2019

Contents

List of Figures

List of Tables

Introduction

For the last decade, the progress made in the autonomous driving scientific community and industry has been exceptional. With the rise of deep-learning and better hardware, algorithms embodying the different aspects of driving, such as lane following, obstacle detection, semantic segmentation, tracking, and motion estimation have reached unprecedented performance. Although there are still no SAE Level-4 self-driving vehicles as of yet, recent developments in robotics and machine learning could soon make this aspiration a reality.

The availability of training data is a critical factor to the growth and success of autonomous driving. Although more powerful than traditional machine learning techniques, deep learning algorithms require a particularly massive amount of data for training and testing purposes. Moreover, in order to assimilate the entire driving process complexity and be reasonably safe, algorithms need to account for all possible real world scenarios, thus demanding highly dynamic and diverse datasets. Finally, it is often dangerous, costly and time-consuming to test driving algorithms on real vehicles.

This present document is a survey of the different autonomous driving datasets which have been published up to date. The first section introduces the many sensor types used in autonomous driving datasets. The second section investigates the calibration and synchronization procedure required to generate accurate data. The third section describes the diverse driving tasks explored by the datasets. Finally, the fourth section provides comprehensive lists of datasets, mainly in the form of tables.

1 Sensors & Hardware

In order to achieve reliability and robustness, a wide variety of sensors are usually employed in autonomous vehicles. The diversity of sensing modalities also help mitigating difficult conditions, as their failure modes will be somewhat uncorrelated. These sensors can be categorized into two main groups, namely *exteroceptive* and *proprioceptive* sensors.

1.1 Exteroceptive Sensors

Exteroceptive sensors are used to observe the environment, which in the case of autonomous vehicles means roads, buildings, cars, pedestrians, etc. The most common exteroceptive sensors for autonomous vehicles are cameras and range-sensing sensors.

1.1.1 Cameras

Cameras come in a variety of types and models. They are passive sensors meaning that they do not need to emit a signal to capture information, thereby limiting possible interference with other sensors. However, they are impacted negatively by illumination and weather conditions, due to their passive nature.

The most common type of camera is the monocular color camera. Being accessible, low-cost and straight-forward to use, monocular cameras have benefited from the majority of computer vision work of the last decades. Thus, most object detection, segmentation and tracking algorithms have been developed for these monocular cameras.

Figure 1.1: An example of monocular image from the KITTI datatset [38].

While driving, humans make use of their stereoscopic vision and focal distance information in order to judge depth and, for example, perform object avoidance. Human vision has also better resolution and a far wider field of view than most monocular cameras. In order to bridge the gap between 2D and 3D object detection and to gain more spatial information, monocular cameras are often used in stereo or multi-view systems. These setups are usually preferred because they present additional depth information, offer redundancy and provide a broader field of view. In order to be precise, these systems need to be calibrated methodically. A more detailed description of the calibration procedure can be found in Section 2. Some manufacturers also offer precalibrated stereo camera systems, which can save time.

An alternative to arrays of cameras are omnidirectional cameras. These offer panoramic 360 degrees images, and are consequently often used to gain maximum information about a surrounding area. This can be highly beneficial for tasks such as localization and mapping. However, they tend to suffer heavily from lens distortion, which can affect the accuracy of a given task. Fish eye lens are also used in a similar fashion.

Other types of specialized cameras have been used in the past. For instance, to make up for the poor camera performance at night, thermal cameras and infrared cameras have been used for tasks such as pedestrian detection [43].

Figure 1.2: An example of an omnidirectional image from [57].

Figure 1.3: An example of a thermal image from the KAIST dataset [43].

Another type of cameras gaining interest are *event cameras*, which output pixel-level brightness changes instead of standard intensity frames. They offer an excellent dynamic range and very low latency (in the order of μs), which can be quite useful in the case of highly-dynamic scenes. However, most already-developed vision algorithms cannot be readily applied to these cameras, as they output a sequence of asynchronous events rather than traditional intensity images. Nevertheless, some autonomous vehicle datasets with event cameras have now been published [11].
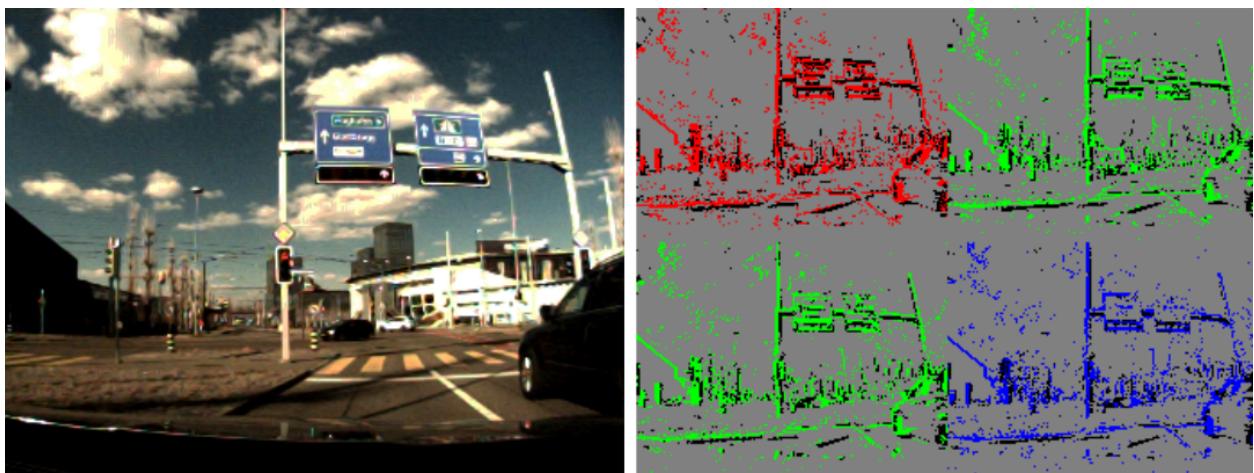
Figure 1.4: An example of an *event image* from the DAVIS dataset [11].

Finally, polarized sensors such as the *Sony Pregius 5.0 MP IMX250* sensor have also recently reached better performance, which could potentially offer a higher level of detail. Polarization channels are often less affected by illumination changes and weather. They are also quite sensitive to a surface roughness, which could help with the detection of vehicles [34]. However, no public autonomous driving datasets employing a polarization camera have been released, as of yet.
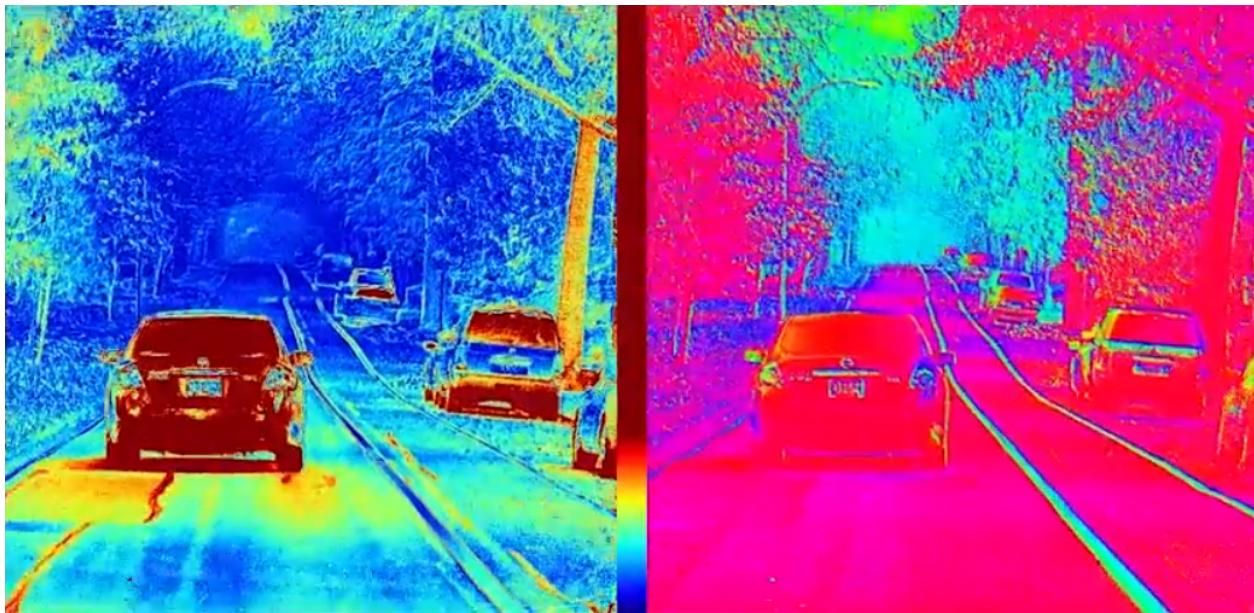
Figure 1.5: An example of a polarization image from [34].

1.1.2 Range-sensing devices

LiDARs, which stands for Light Detection and Ranging, detect objects and map their distances with great spatial coverage, in all lightning conditions. As such, they have been a sensor of choice for autonomous driving applications.

The technology works by illuminating a target with an optical pulse, and measuring the characteristics of the reflected signal return in the same way a radar would detect reflected radio waves. They are much more accurate than radars, but their performance deteriorates from weather conditions such fog, rain or snow. They can also sometimes have trouble with

detecting objects at close range.

LiDARs also come in a variety of formats, which can be split into two main families: *i)* mechanically spinning LiDARs and *ii)* solid-state LiDARs. While solid-state LiDARs are significantly cheaper, they suffer from a limited field-of-view compared to mechanical LiDARs. Solid-state LiDARs also tend to have a higher noise-to-signal ratio. They have also appeared much more recently than mechanical LiDARS. For all these reasons, most LiDARs used in autonomous datasets have been using either 2D or 3D mechanical LiDARs with often 360 degrees field-of-view. However, recent developments in solid state LiDARs are promising and they are slowly closing the performance gap.

It should be noted that LiDARs can scan a large amount of points at a very high rate, which can be a challenge for any algorithm run time.

In order to mitigate LiDAR limitations when it comes to adverse weather or close-range sensing, radars are also used as a range-sensing technology. Being a more mature sensor than LiDARs, radars are often much cheaper and lightweight, while also being able to determine the speed of its target. Nevertheless, they suffer from very poor spatial resolution, the difficulty of interpreting the received signals, and a much worse accuracy than LiDARs.

Finally, sonars are also used in the industry. While also cheap, sonars have very limited range and precision in addition to being susceptible to weather conditions. They are mainly used for nearby obstacle detection.

1.2 Proprioceptive Sensors

Proprioceptive sensors measure values internally to a given system. In the case of an autonomous vehicle, these measurements include linear and angular positions, speed and acceleration. Most modern cars are already equipped with a plethora of proprioceptive sensors. Wheel encoders are used for odometry, tachometers are used for speed, and IMUs can monitor acceleration changes. These sensors are often accessible through the vehicle's

CAN bus.

However, the accuracy of sensors from car manufacturers are typically too low for autonomous vehicles applications, especially in the case of IMUs. For mapping purposes, the IMU measurements and odometry provide a point-matching algorithms such as Iterative Closest Point (ICP) [10] with an initial transformation guess, which is crucial to the algorithm performance (both in terms of speed and robustness). In order to reach a higher level of precision, most autonomous vehicle datasets use a navigation-grade IMU along with a GPS. Autonomous datasets dedicated to mapping and localization also often use an RTK GPS, which can provide centimeter-level accuracy, in order to compare the localization algorithms to a ground truth.

Other signals from the CAN bus protocol of a vehicle can be accessed, such as the steering angle, and the position of the accelerator and brake pedals. Such signals have been used by end-to-end learning algorithms [11, 85].

2 Calibration & Synchronization

In order to achieve coherent data alignment, every sensor needs to be calibrated and synchronized. Below we describe both spatial alignment (calibration) and temporal alignment (synchronization).

2.1 Calibration

Calibration usually refers to a spatial-referencing process by which the relative coordinate frames of all sensors are established. For cameras, the calibration is essential to accurately measure object and distances on a scene for stereo camera setups. Camera calibration or *camera resectioning* is often split into *intrinsic* and *extrinsic* parameters retrieval. Intrinsic parameters refer to the camera's inherent parameters such as focal length, principal point coordinates and distortion coefficients, to be used in image rectification. On the other hand, extrinsic parameters denote the coordinate system transformations from 3D world coordinates to 3D camera coordinates. These parameters are retrieved using referenced calibration points, called fiducial markers, usually with a checker board target with known dimensions [119].

Once camera-to-camera calibration is achieved, stereoscopic depth reconstruction can be performed for all overlapping pixels. The additional depth channel can then be exploited by various machine learning algorithms to seize additional object features which can lead to better results.

The depth channel derived from stereoscopic reconstruction can be as dense as the camera, depending on the texture in the scene. However, it can be strongly affected by the level of illumination in the environment. It is also far less precise at longer ranges, with the distance z accuracy decreasing in $1/z^2$. Range sensors, such as LiDARs, are often far more precise and reliable at estimating object distances, but their measurements significantly less dense than cameras.

In a similar fashion, camera-to-range calibration can be achieved using reference plane

surfaces [37]. However, the LiDAR-camera measurements fusion is not as straight-forward as stereoscopic reconstruction, because LiDAR data is a lot sparser than camera images. In order to create a depth mask, a LiDAR point cloud needs to be projected on the image plane and then upsampled. It is also possible to instead project the RGB channels onto the LiDAR point cloud to generate a *colored point cloud*. This approach is generally used by mapping and localization tasks, by providing extra information to a point-matching algorithm.

Finally, motion-sensing devices are also calibrated with each sensor using hand-eye calibration [41] in order to reference the measurements to the inertial navigation system (INS). This is particularly important for localization and mapping tasks such as SLAM, where the inertial and odometry measurements are used as a first estimate for the new position and orientation, which is critical to these algorithms' performance.

2.2 Synchronization

While calibration addresses the spatial alignment of sensors, synchronization temporally matches measurements together. Driving is a highly-dynamic process within a rapidly-changing environment, making the synchronization process critical for data temporal alignment. In order to synchronize these different measurements, sensors are often triggered externally. The measurements of each sensor are also timestamped with a system clock. When sensors have different acquisition rates, and thus different timestamps, the measurements can either be interpolated or the closest measurement can be selected, depending on the use case.

The exposure time of a camera is nearly instantaneous and should not yield bad data alignment. However, most rotatory LiDARs execute a full rotation in about 0.1 s. This scanning speed is considerably slower than the car speed, which can introduce distortions in the point clouds. Techniques have been developed in order to account for the vehicle's motion within a LiDAR scan [65].

3   Tasks

Most autonomous driving approaches up to date have tried to deconstruct the complex driving process into different smaller and simpler sub-tasks. With this modular approach in mind, each dataset generated for the autonomous vehicle community have more often than not revolved around one or many of these specific tasks. Below, we enumerate the most popular ones.

3.1   Stereo Vision

As mentioned before, driving algorithms can benefit from having additional depth information. One of the simplest ways to acquire such 3D information is through stereo vision. Stereo vision is the task in which the depth of a scene is triangulated by identifying common features in two images taken from cameras mounted next to each other. In the case of driving, challenges for stereo vision include reflective and shiny surfaces such as car bodies. Repetitive structures like fences and transparent surfaces (glass) are other common failure cases.

Datasets dedicated to 3D reconstruction usually offer pixel-wise depth maps as a ground truth. Often, they have been generated by interpolating 3D LiDAR point clouds and by fitting 3D CAD models onto individual objects.
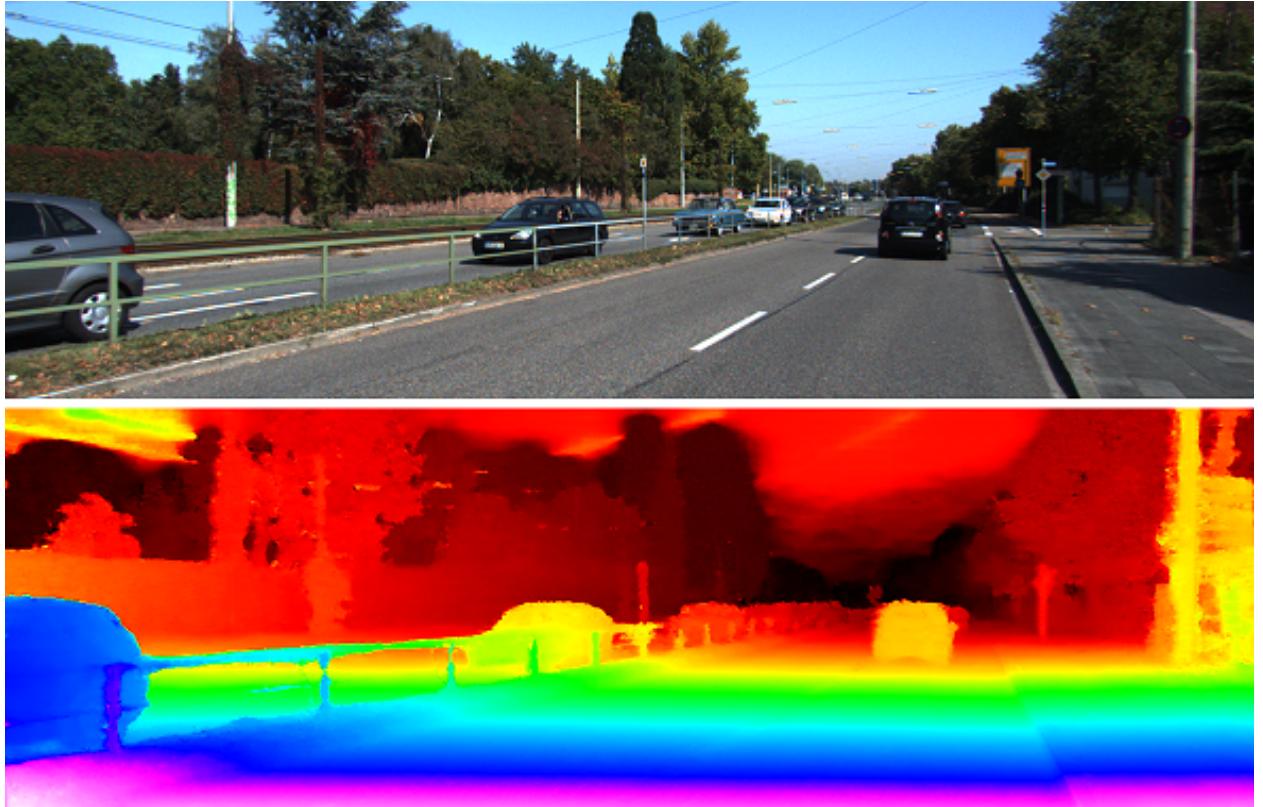
Figure 3.1: A sample depth map as ground truth for stereo vision, from the KITTI dataset [94].

It should be noted that with the rise of recent improvements in deep learning, monocular depth evaluation has also taken interest. In this case, the depth map is estimated using contextual information from a single image [55]. Readers are referred to Gabr and Elias [33] for a complete survey on stereo vision and other 3D reconstruction algorithms.

3.2   Motion Estimation

Because driving involves multiple objects moving at high speed, capturing the motion of objects in an image might yield desirable information. Optical flow, defined as finding the motion at each image location between consecutive frames, is one way of representing motion in a dense manner. Optical flow thus extracts additional motion information, which can be of particular importance for other tasks such as localization, ego-motion and tracking.

Optical flow is restricted to monocular 2D images, which makes the retrieval of 3D motion challenging. *Scene flow* is therefore defined as a generalized version of optical flow, where frames of stereo or multi-camera setups are used to establish motion.

Datasets dedicated to optical and scene flow usually offer optical flow fields, where a vector describing the motion for each pixel in the next or previous frame is provided. Retrieving ground truth for both optical flow and scene flow is a time-consuming and tedious process and is often done by matching the image objects to 3D LiDAR maps.
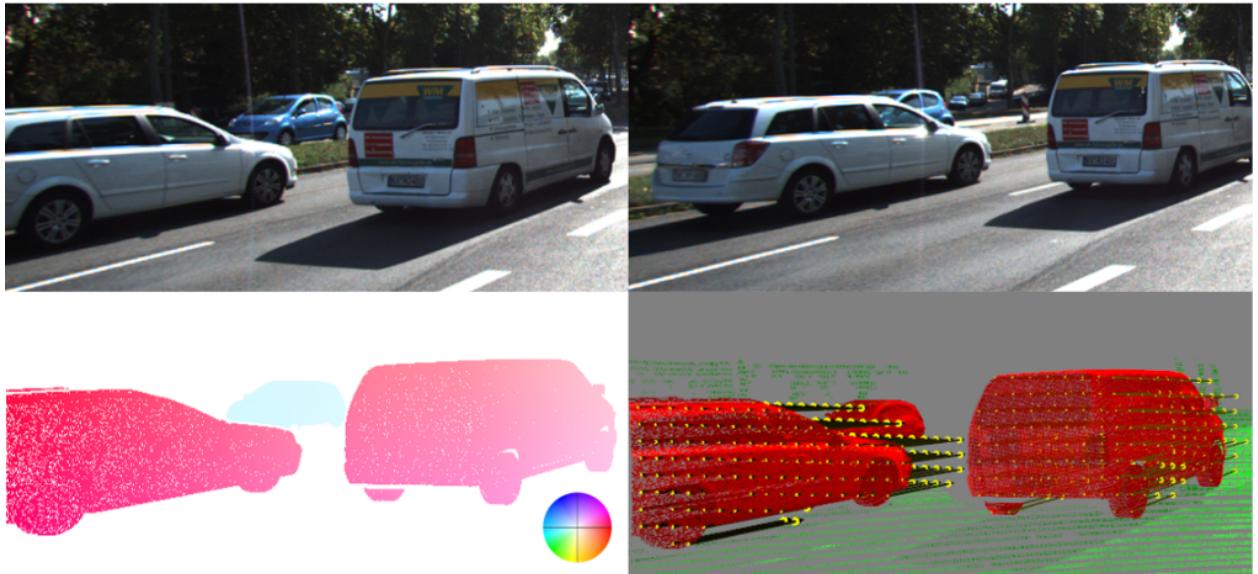
Figure 3.2: A sample of ground truth for optical flow (bottom left) and scene flow (bottom) right from the KITTI dataset [88].

An extensive survey on the state-of-the-art algorithms for optical and scene flow can be consulted in [113].

3.3   Object Detection

One of the first and foremost aspects of driving is the awareness of its surrounding. Whether it be pedestrians, other vehicles, traffic signs or obstacles, the detection and recognition of different objects in a scene is crucial to the safety and smooth functioning of an autonomous

vehicle. Object detection addresses this task by determining the presence and localization of different predefined classes of objects in a scene.

Being an important and well-defined task, object detection has benefited from a considerable amount of attention in the computer vision and autonomous vehicle community. However, object detection still faces challenges. This is notably because of the wide variety of objects, weather conditions and illumination in a driving scene, along with heavy occlusion and truncation of objects [72].

Object detection itself can be split into subcategories depending which modality is used to detect object, or what object itself is to be detected.

Most object detection is done strictly on 2D images, hence the name 2D object detection. Each object is localized within the image, in pixel coordinates, as illustrated in Figure 3.3. However, it should be noted that recent approaches have tried to include 3D features from either point cloud data [20] or stereo reconstruction [19] in order to generate a more robust detection. Moreover, it is also possible to localize objects relative to the vehicle position in 3D space. This is commonly referred as 3D object detection, and is depicted in Figure 3.4.
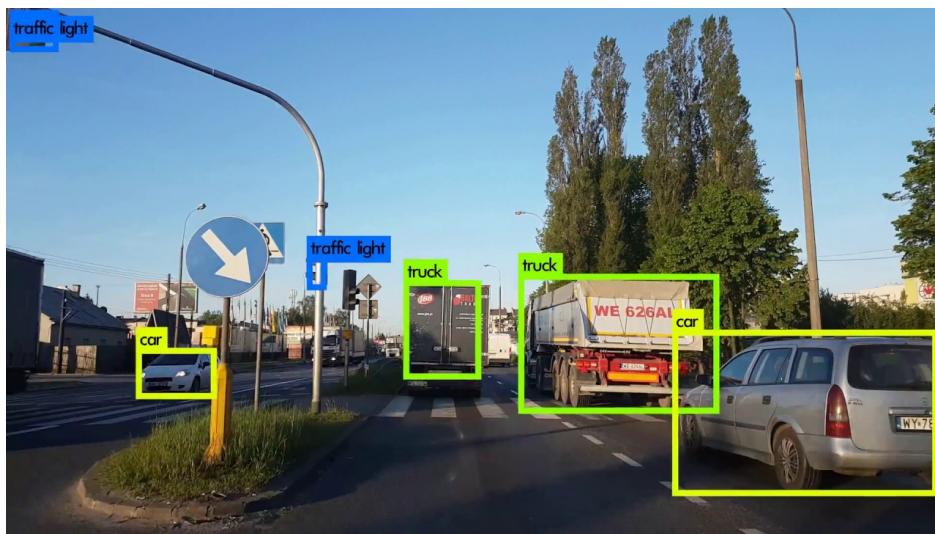
Figure 3.3: 2D detection results from the YOLO algorithm [82].
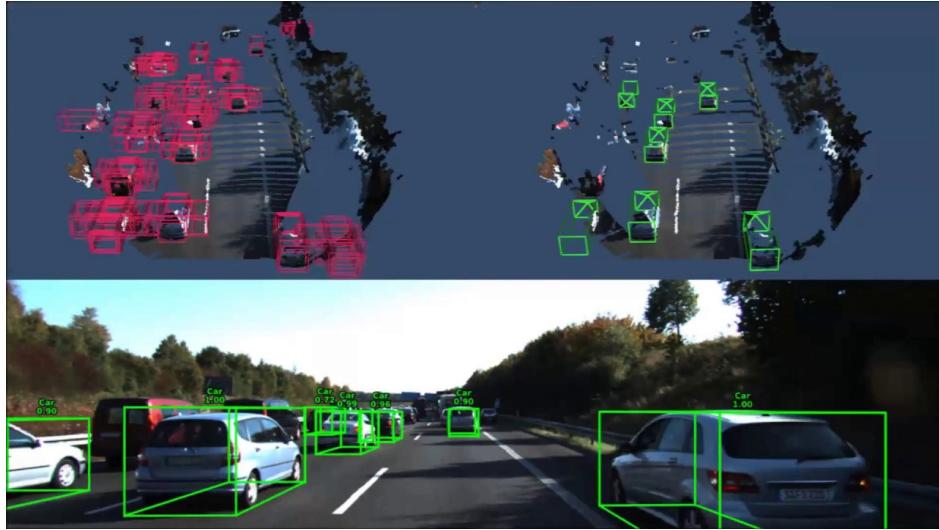
Figure 3.4: A sample ground truth frame for 3D object detection from the KITTI dataset [38].

Datasets dedicated to object detection usually contain annotated data frames with 2D or 3D bounding boxes, which encloses the different objects, as ground truth.

Extensive reviews of deep learning detection techniques can be found in [31], [69], [58] and [121].

3.4 Tracking

Driving is a dynamic process with high-speed moving objects. Therefore, object detection is often insufficient in order to avoid collisions during path planning. Driving algorithms should not only predict the location of objects in a scene, but also their velocity and acceleration. In order to do so, tracking algorithms are used, which try to predict future positions of multiple moving objects based on the history of the individual positions.

A popular and intuitive approach to tracking is *tracking-by-detection*. An object-detection algorithm is first used to detect targets in each frame, which then need to be associated with each other over multiple frames. While efficient, this approach however suffers from detection errors and from the inherent difficulties of performing data-association. Tracking can also

suffer if objects are momentarily occluded, as illustrated in Figure 3.5. It should be noted that pedestrian tracking is of particular interest, as they are the most vulnerable users of the road [28].
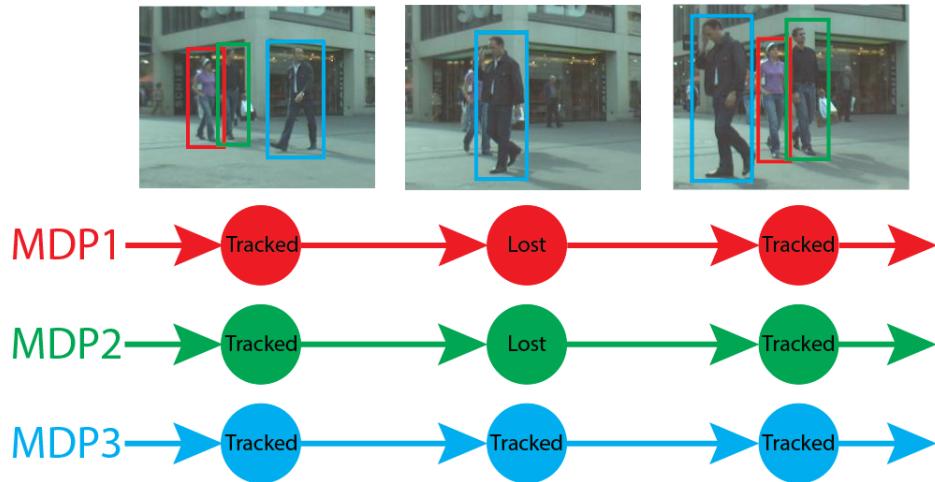
Figure 3.5: An example of tracking and occlusion challenges from [109].

In order to predict a target trajectory in 3D, range information is most certainly needed for tracking. As mentioned before, such information can be obtained either by 3D reconstruction from cameras or from LiDAR point clouds, alone or through some sensor fusion process.

Just like object detection, tracking datasets usually contain annotations of data frames in the form of bounding boxes and labels, which are coherent over multiple frames, as a ground truth.

A thorough review on state-of-the-art tracking techniques is available in Dixit et al. [27].

3.5 Semantic Segmentation

Some objects such as roads, sidewalks and traffic lines are not well-defined by bounding boxes. Consequently, they need a more flexible representation, often down at the pixel-level. This problem is referred to as *semantic segmentation*.

Semantic segmentation is indeed similar to object detection in the way that it tries to locate different predefined classes of objects in a scene. However, instead of using bounding boxes to localize objects, each pixel of an image is assigned to a class, as seen in Figure 3.6. The segmentation mask therefore offers a more dense and complex classification and localization, which can provide a better understanding of the scene. Semantic segmentation faces the same challenges as object detection such as occlusion, truncation and shadows, but also requires more complex computation. However, with model compression, pruning and hardware acceleration, it can reach real-time execution [91].

Figure 3.6: A sample ground truth frame for semantic segmentation from the Cityscapes dataset [23].

A more refined version of semantic segmentation is *instance segmentation*, which not only classifies each pixel in a class, but also separates instances of the same class. Unlike semantic segmentation, instance segmentation thus provides information about each instance such as shape and position. Instance segmentation is particularly important to assess the trajectory of individual objects, for example vehicles, cyclists or pedestrians.

It should also be noted that in the case of driving, algorithms usually have access to multiple

time frames of data. Methods which impose algorithms to be temporally coherent can improve segmentation accuracy and robustness.

Just like object detection, most of the previous work dedicated to segmentation has been done strictly on 2D images. However, shape and size are important features which cannot be exploited in the 2D space.To capture such information, LIDARs can of course be used.

It it also possible to train semantic segmentation models strictly on point cloud data. However, generating accurate point cloud labels, such as the one displayed in Figure 3.7, is a tedious and time consuming task.
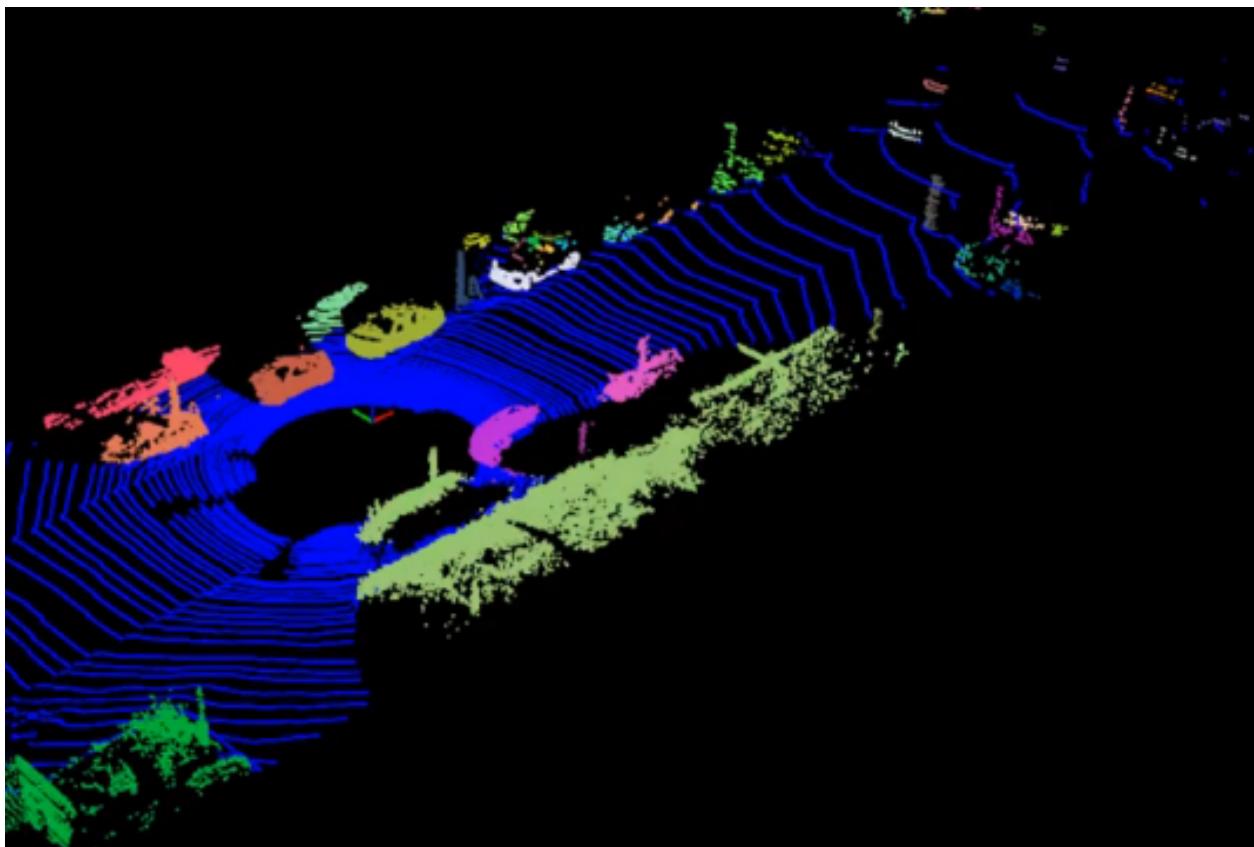
Figure 3.7: A sample ground truth frame for 3D instance segmentation from Zermas, Izzat, and Papanikolopoulos [117]

Given the importance of identifying drivable spaces, road and lane segmentation is of particular interest for autonomous vehicles. Along with spatially segmenting the road and lane itself,

some algorithms also try to establish the host and neighbor lanes along with their direction. This information is particularly useful for tasks such as lane keeping, merging and turning.

Datasets dedicated to semantic segmentation usually annotate data frames with pixel-wise segmentation masks as ground truth, or in the case of 3D segmentation, voxel-wise masks.

In order to alleviate computational burden of semantic segmentation, the *stixel* representation has been suggested [24]. Stixels create a medium-level model of the environment, compressing pixel-wise information into vertical strips. An example of a stixel segmentation can be seen in Figure 3.8.
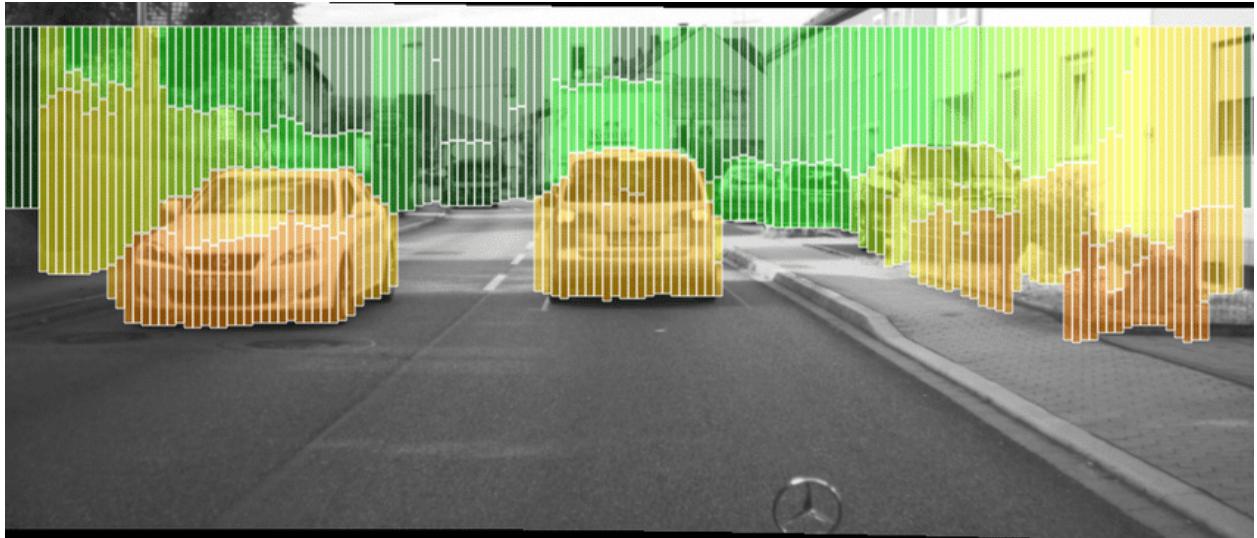
Figure 3.8: A sample ground truth frame for stixel segmentation from the Stixel dataset [24].

For more details, a review on deep-learning techniques for semantic segmentation can be found in Garcia-Garcia et al. [35]. An in-depth survey on road and lane detection can be found in Bar Hillel et al. [4].

3.6 Localization

Localization is a task critical to any mobile robot. In order to lay the appropriate path planning, a vehicle needs to know where it is exactly regarding its environment. Different

approaches have been used in the past for localization.

The most straightforward one is the use of GPS and IMU sensors. While this combination of sensors is the most accessible and low-cost approach, it lacks the requirements needed for autonomous driving. Even with dead reckoning estimation from the IMU, commercial-grade GPS are simply too inaccurate. While RTK technology offers the precision needed for autonomous driving, the accuracy of the signal is highly dependant on the environment, urban setting with high buildings being particularly prone to errors from interference.

Simultaneous localization and mapping (SLAM) is another popular approach. It tries to generate a map on the fly using a vehicle's sensors, while estimating at the same time the position of the vehicle in the constructed map. It has the advantage of not needing any prior information about the environment, meaning this approach can work in any setting. However, SLAM still faces challenges as it is computationally heavy and needs to handle large-scale environments in real-time. Moreover, SLAM is prone to diverge in difficult environments. RTKs positioning is often used as ground truth for SLAM, given an appropriate signal reception.

Using pre-constructed maps is an alternative to SLAM that alleviates the problem on generating a map on the fly. Using a point-matching algorithm or visual landmark searches approach, a priori map-based localization algorithms can be highly accurate. However, a major weakness of these approaches is the fact that roads themselves are not completely static, and therefore the maps used for localization have to be updated for construction work or weather changes.

For further details, a complete survey on state-of-the-art techniques for localization can be found in Kuutti et al. [53].

3.7 Behaviour Analysis

If driving vehicles are one day a reality, they will most likely have to interact with humans. Whether it be infering a pedestrian's intention to cross a street, identifying a driver's intent to perform a certain action or spotting potentially reckless maneuvers, autonomous vehicles need to have a high-level understanding of surrounding human behavior. The assessment of human behavior is therefore paramount for any autonomous driving applications.

While the task of behavior assessment is not as bounded as previously described tasks, recent datasets have tried to capture such human behavior. For instance, some datasets [45, 78, 79] have tried annotating each of the driver's actions. Such data can be used to develop action-predicting algorithms, which can then be used to assess if a driver's maneuver is completely safe or not.

Driver face monitoring is also a modality often used in order to predict maneuvers or visual focus, as shown in Figure 3.9. Some datasets even project the driver's gaze onto the road image, in order to know exactly on what the driver is focusing [73]. Such information can be used to established the driver's attention level for a safer driving experience.

Figure 3.9: An example of driver face monitoring and predicted maneuvers from the Brain4Cars dataset [45].

Another important behavior assessment task is driving style recognition. Driving style can be defined in various ways including fuel consumption, brake-use, distance-keeping and aggressiveness. Establishing a driver's style can be used to adjust driving strategy, such as lane merging or alert the driver if he is being reckless [9].

Finally, some datasets [52] have also collected data regarding pedestrian intention, as displayed in Figure 3.10. Algorithms can then be trained to recognize whether a pedestrian wants to cross a street or not, and help prevent collisions.
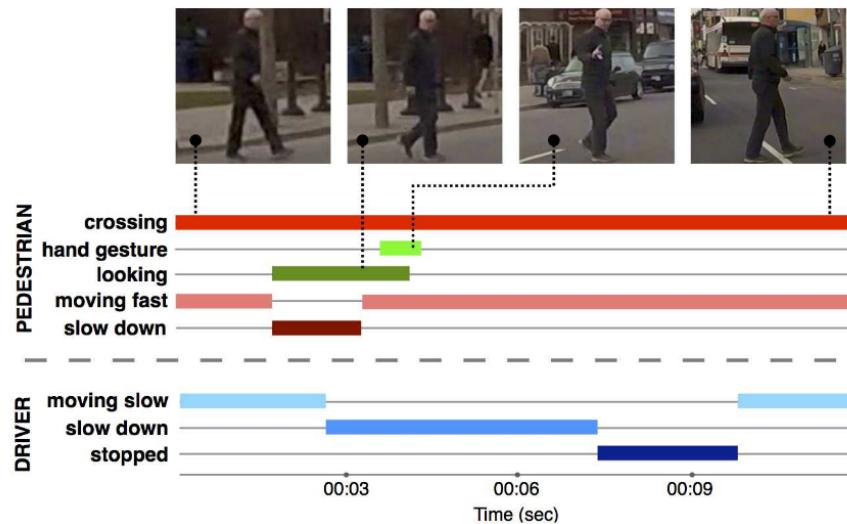
Figure 3.10: An example of labelled pedestrian and driver intention sequence from the JAAD dataset [80].

For a more complete literature review on driving style recognition, pedestrian autonomous vehicle interactions or intersection behavior, readers are invited to consult Marina Martinez et al. [62], Rasouli and Tsotsos [81] and Shirazi and Morris [90] respectively.

4   Datasets

The following section presents the available open driving datasets, sorted by their respective task. It should be noted that only datasets which are in a real-driving context were selected. Therefore, synthetic datasets are not presented. Moreover, datasets acquired from terrestrial vehicles which are not cars, such as Segways or another robotic platform are also ignored. Finally, stationary-acquired datasets are also ignored.

4.1   Object detection

The Table 4.1 table presents the autonomous driving datasets available for object detection and tracking, ordered chronologically. First of all, the shear size of the table is a great testimony to the level of attention the field has received in the autonomous driving community. Indeed, most datasets in this table have been published during the past three years, demonstrating that this is an active field. The table also attests the recognition autonomous driving has gained in the industry, with most of the major and recent datasets being published by companies such as Waymo, Aptiv, Lyft, Bosch and Hesai.

Moreover, by ordering the datasets by year, several interesting trends can be observed. The number of annotated frames per dataset seems to have increased considerably throughout the year. This phenomenon can probably be explained by the surge of deep learning in computer vision, which demands a vast amount of data. Furthermore, deep learning also gain from data diversity, which can explain why recent datasets seem to encompass more driving situations such as weather, time and traffic variety.

The interest in multi-modal learning and the democratization of LiDAR sensors can also be observed in this table, with most of the recent datasets incorporating LiDARs and other sensing devices. The annotations have also been refined in the recent years, with the arrival of more annotated classes and 3D bounding boxes.

Finally, the table also presents object detection datasets which are aimed at more specialized

tasks such as pedestrian detection or traffic sign detection. However, these kinds of niche datasets seem to be less frequent nowadays, which suggests that such specialized task have become trivially easy for object detection algorithms. It also seems that the state-of-the-art emphasizes rather on object detection generalization.

4.2   Object segmentation

Likewise, Table 4.2 presents the driving datasets aimed at semantic and instance segmentation. While object detection and tracking are analogous to semantic and instance segmentation, there seems to be considerably less datasets dedicated to the latter tasks. This can presumably be attributed to the fact that generating segmentation annotation is a time-consuming and costly process.

Nonetheless, comparable trends can be observed from the segmentation datasets. Recent datasets are bigger and more diverse. The use of LiDARs and point cloud annotations are also very recent. Finally, most recent datasets are, again, published by companies.

4.3   Lane detection

The Table 4.3 presents lane detection datasets, another important part of vehicle perception. It can be thought as a special case of object detection or semantic segmentation. In fact, the consensus on the annotation type among the autonomous driving community does not seem to be established. While most datasets use spline lines to describe lanes, other datasets use pixel-wise annotations, bounding boxes or point cloud annotations.

It should be noted that while recent datasets have become larger and more exhaustive, there seems to be no datasets explicitly exploring damaged lane markings and lane-detection fail cases.

4.4 Optical flow

The list of optical flow datasets presented in Table 4.4 is noticeably shorter than for the previous tasks. The lack of driving optical flow datasets can most probably be explained by two causes. First, measuring ground truth for optical flow is a complicated and precise task which can be very hard to do in a highly-chaotic environment such as driving. This is why most benchmarked optical flow datasets are usually done in a controlled environment [3] or using synthetic data [15]. Secondly, most modern computer vision algorithms do not make use of optical flow data and thus the field has been losing interest over the years. Nevertheless, the KITTI FLOW 2015 [64] dataset offers a precise optical flow benchmark, with 3D fitted CAD models as ground truth. On the other hand, the Heidelberg datasets [50, 63] provide challenging and diverse cases for optical flow to assess algorithms robustness.

4.5 Stereo

A similiar trend can be observed for stereo datasets in Table 4.5. Once again, retrieving accurate ground truth for highly-dynamic scenes can turn out to be a challenge even with the use of LIDARs, as motion distortion can have considerable effect. Consequently, commonly used benchmarks are either static [86] or synthetic [15].

Stereo datasets are nonetheless useful in order to evaluate sensor fusion, especially since multi-modal learning is becoming more important. One interesting trend which can be observed in this table is the increase of resolution over time, which is also valid for every other tasks. This improvement in resolution embodies the hardware breakthroughs made over the years, but also begs the precision-performance question which future algorithms will have to face.

4.6   Localization and mapping

It also seems that localization/mapping datasets has not benefited from the same level of attention as object detection, which can be observed by the size of Table 4.6. Most mapping and localization algorithms do not make use of deep-learning algorithms, meaning the need for large and diverse amount of data is not as critical. It can also be observed that localization and mapping datasets have been using LiDARs a lot earlier than those for object detection.

Centimeter precise measurements from DGPS or RTK are available for visual or point cloud odometry for most datasets. However, there is no datasets that provide ground truth for SLAM or mapping. Only qualitative evaluation and loop-closing can be used to evaluate the quality of a generated map.

4.7   Behavior

Finally, Table 4.7 presents datasets which focus on behavioral aspects of driving. It can be noted that these kind of datasets have only recently started to gain attention. Also, since they are so recent, there is no clear annotation or methodology defined to quantify and capture driving behaviors. These datatsets are thus highly different from one another. It should also be noted that a recent trend tries to predict steering angle, brake or gas pedal, for use in *end-to-end learning.*

TABLE 4.1: Object Detection Datasets

Name	Provider	Cit.	Year	Location	Diversity				Traffic					Sensors						Annot.		Track.	Frames (k)	Classes
					R	S	N	D	U	H	Re	Ru	C	Vi	Li	GPS	Th	If	Ra	2D	3D			
PandaSet [74]	Hesai & Scale	N/A	2019	Silicon Valley	?	?	?	?	?	?	?	?	?	X	X**	X			X	X	X	X	60	28
nuScenes [16]	Aptiv	16	2019	Boston, Singapore	X		X	X	X	X	?	?	?	X	X	X			X	X	X	X	40	23
Waymo Open [106]	Waymo Inc.	N/A	2019	United States	X		X	X	X	X	?	?	?	X	X	X				X	X	X	200	4/5
Lyft Level5 [48]	Lyft	N/A	2019	Palo Alto, London, Munich	?	?	?	?	?	X	X	X	?	X	X					X	X	X	55	23
H3D [76]	Honda Res. Inst.	4	2019	San Francisco										X	X	X					X		27	8
Boxy [7]	Bosch N.A. Res.	0	2019	Unknown	X				X					X						X			200	1
BLVD [111]	VCCIV	2	2019	Changshu			X		X					X*	X						X	X	120	3
Road Damage [61]	University of Tokyo	31	2018	Japan	X				X					X						X			9	8
KAIST Multispectral [22]	KAIST	13	2018	Seoul	X	X	X	X	X	X	X			X	X	X	X			X		X	8.9	6
BDD [115]	U. of California	103	2018	San Francisco, New York	X	X	X	X	X	X	X	X		X		X				X		X	100	10
Apollo Open Platform [2]	Baidu Inc.	N/A	2018	China	X	X	X	X	X	X				X	X				X		X	X	20	4/5
EuroCity [13]	Delft U. of Tech.	6	2018	European Cities	X	X	X	X	X	X				X						X		X	47	7
FLIR Thermal Sensing [32]	FLIR Sytems Inc.	N/A	2018	Santa Barbara		X	X	X		X				X			X			X		X	14	5
NightOwls [71]	Oxford U.	3	2018	Germany, Netherlands, UK			X	X			X			X						X		X	279	4
TuSimple [103]	TuSimple	N/A	2017	Unknown	X	X				X				X									12	1†
NEXET [100]	Nexar	N/A	2017	Around the globe	X	X	X	X	X	X	X	X		X						X			55	5
Multi-spectral Object Detection [97]	U. of Tokyo	5	2017	Tokyo	X								X	X			X			X		X	7.5	5
Bosch Small Traffic Lights [8]	Bosch N. A. Res.	45	2017	San Francisco	X						X			X						X		X	13	15
CityPersons [118]	Max Planck Inst. (Info.)	105	2017	Germany, France, Switz.					X		X			X		X				X			5	4
Udacity [104]	Udacity	N/A	2016	Mountain View					X					X						X		X	34	3
JAAD [80]	York University	16	2016	Ukraine, Canada	X	X			X					X						X		X	88	1††
Elektra (CVC-14) [95]	Auton. U. of Barcelona	N/A	2016	Barcelona	X		X		X					X				X		X		X	10	1
Tsinghua-Daimler Cyclist [110]	Daimler AG	41	2016	Beijing										X	X	X				X			15	6
KAIST Multispectral Pedestrian [44]	KAIST	195	2015	Seoul			X		X		X			X		X	X			X			95	3
Highway Workzones [89]	CMU	21	2015	United States						X				X						X		X	17	9
KITTI [94]	Karlsruhe Inst. of Tech.	5415	2013	Karlsruhe	X				X	X	X	X	X	X	X	X				X	X	X	15	2
German Traffic Sign [96]	Ruhr U.	298	2013	Germany										X*						X		X	5	43
LISA Traffic Sign [68]	U. of California	382	2012	San Diego	X	X			X					X*						X			6.6	47
TME Motorway [18]	Czech Tech. U.	127	2011	Northern Italy						X				X						X		X	30	2
Belgian Traffic Sign [101]	ETH Zürich	261	2011	Belgium										X*						X		X	9	62
LISA Vehicle Detection [93]	U. of California	333	2010	California					X					X						X		X	2.2	1
TUD-Brussels Pedestrian [107]	Max Planck Inst. (Info)	182	2009	Belgium					X					X						X		X	1.6	1
ETH Pedestrian [30]	ETH Zürich	547	2009	Zürich					X					X						X	X	X	4.8	1
Caltech Pedestrian [28]	Caltech	1009	2009	Los Angeles					X					X						X		X	250	1
Daimler Pedestrian [29]	Daimler AG	1177	2008	Beijing					X					X						X		X	21	1

R: Rain, S: Snow, N: Night, D: Dawn/Dusk, U: Urban, H: Highway, Re: Residential, Ru: Rural, C: Campus, Vi: Video, Li: LiDAR, Th: Thermal Camera, If: Infrared Camera, Ra: Radar
*Non-continuous frames   **Solid-state LiDAR   †false velocity   ††false age, gender and direction

TABLE 4.2: Object Segmentation Datasets

Name	Provider	Cit.	Year	Location	R	S	N	D	U	H	Re	Ru	C	Vi	Li	GPS	Th	Px	In	Pc	Frames (k)	Classes
PandaSet [74]	Hesai & Scale	N/A	2019	Silicon Valley	?	?	?	?			?	?	?	X	X**	X		X	?	X	60 k	37
SemanticKitti [6]	University of Bonn	1	2019	Karlsruhe					X		X			X	X	X		X		X	43 k	28
Highway Driving [49]	Korea Advanced Institute of Science and Technology	1	2019	Unknown						X				X				X			1.2 k	10
ApolloScape [42]	Baidu Inc.	7	2018	China	X		X		X	X	X	X		X	X	X		X	X	X	147 k	35
Apollo Open Platform [2]	Baidu Inc.	N/A	2018	China	X				X	X	X	X		X	X			X	X	X	17 k	26
BDD [115]	U. of California	103	2018	San Francisco, New York	X	X	X	X	X	X	X	X		X				X	X		10 k	40
Wilddash [116]	Austrian Inst. of Tech.	8	2018	Around the globe	X	X	X	X	X	X	X	X		X		X		X	X		226	30
Raincouver [102]	U. of British Columbia	3	2018	Vancouver	X		X	X	X		X			X				X			326	3
MightyAi Sample [66]	MightyAI	N/A	2018	Seattle					X		X			X				X	X		200	41
IDD [105]	IIIT Hyderabad	6	2018	Hyderabad, Bangalore				X	X		X	X		X				X	X		10 k	34
Mapillary Vistas [70]	Mapillary AB	139	2017	Around the globe	X	X	X	X	X	X	X	X		X				X	X		20 k	66
Multi-Spectral Semantic Segmentation [40]	U. of Tokyo	11	2017	Tokyo			X	X	X					X*			X	X			1.6 k	8
Cityscapes [23]	Daimler AG	1729	2016	Germany, France, Switzerland					X					X			X	X	X		5 k	30
LostAndFound [77]	Daimler AG	14	2016	Germany					X					X*		X		X	X		2 k	37
Daimler Urban Segmentation [87]	Daimler AG	67	2014	Heidelberg					X		X			X	X	X		X			500	5
KITTI Semantics [94]	Karlsruhe Institute of Technology	5415	2013	Karlsruhe					X		X			X	X	X		X	X		200	34
Standford Track [99]	Stanford University	146	2011	California					X				X	X	X					X	14 k	3
CamVid [14]	University of Cambridge	475	2009	Cambridge					X					X				X			700	32
Daimler Pedestrian [29]	Daimler AG	1177	2008	Beijing					X					X				X			785	1

R: Rain, S: Snow, N: Night, D: Dawn/Dusk, U: Urban, H: Highway, Re: Residential, Ru: Rural, C: Campus, Vi: Video, Li: LiDAR, Th: Thermal Camera, Px: Pixel, In: Instance, Pc: Point Cloud
*Non-continuous frames **Solid-state LiDAR

TABLE 4.3: Lane Detection Datasets

Name	Provider	Cit.	Year	Location	R	S	N	D	U	H	Re	Ru	C	Vi	Li	GPS	Px	Sp	Bx	Pc	Frames	Classes
Unsupervised Llamas [47]	Bosch N. A. Research	0	2019	California	X	X	X	X	X	X	X	X	X	X		X	X			X	100 k	5
BDD [115]	U. of California	103	2018	San Francisco, New York	X	X	X	X	X	X	X	X	X	X	X	X	X	X			80 k	11
ApolloScape [42]	Baidu Inc.	7	2018	China					X	X	X	X	X	X	X	X	X			X	170 k	27
CULane [98]	University of Hong Kong	42	2018	Beijing	X				X	X	X			X			X				130 k	9
VPGNet [56]	KAIST	44	2017	Seoul	X		X		X	X	X			X		X	X				21 k	17
TuSimple [103]	TuSimple	N/A	2017	Unknown	X				X	X	X			X			X				6.4 k	4
TRoM [59]	Tsinghua University	1	2017	Beijing				X	X	X	X			X			X				712	19
KITTI Road [94]	Karlsruhe Institute of Technology	5415	2013	Karlsruhe					X	X	X			X	X	X			X		579	2
Road Marking [108]	Honda Research Institute	68	2012	California			X	X	X	X	X			X					X		1.4 k	23
CalTech Lanes [1]	California Institute of Technology	577	2008	Washington, Cordova				X	X	X	X			X			X				1.2 k	2

R: Rain, S: Snow, N: Night, D: Dawn/Dusk, U: Urban, H: Highway, Re: Residential, Ru: Rural, C: Campus, Vi: Video, Li: LiDAR, Th: Thermal Camera, Px: Pixel, Sp: Spline, Bx: Bounding Box, Pc: Point Cloud

TABLE 4.4: Optical Flow Datasets

Name	Provider	Cit.	Year	Location	R	S	N	D	U	H	Ru	Scene Flow	Optical Flow	Ground Truth Retrieval	Frames
HD1K [50]	Heidelberg Collaboratory, Bosch	20	2016	Heidelberg	X	X			X				X	Stereo/Lidar Fusion + ICP	1000
KITTI Flow 2015 [64]	Karlsruhe Institute of Technology	629	2015	Karlsruhe					X			X	X	Stereo/Lidar Fusion + 3D CAD Models	400
KITTI Flow 2012 [36]	Karlsruhe Institute of Technology	3508	2012	Karlsruhe					X				X	Stereo/Lidar Fusion + ICP	391
HCI Challenging Stereo [63]	Heidelberg Collaboratory, Bosch	57	2012	Hildesheim	X		X	X	X	X			X	Stereo + Monoscopic Tracking	10 k

R: Rain, S: Snow, N: Night, D: Dawn/Dusk, U: Urban, H: Highway, Ru: Rural

UNIVERSITÉ LAVAL

N❄R Lab

TABLE 4.5: Stereo Datasets

Name	Provider	Cit.	Year	Location	Resolution	Baseline	Frames	Env	Diversity We	Il	Ti	Se	Ground Truth
Complex Urban [46]	KAIST	14	2018	South Korea	1600x1200	47	36 k	X	X				LiDAR
ApolloScape [42]	Baidu Inc.	7	2018	China	3384x2710	30	5 k	X					Stereo/Lidar Fusion + 3D CAD Models
HD1K [50]	Heidelberg Collaboratory, Bosch	20	2016	Heidelberg	1080x2560	30	10 k		X	X			Stereo/Lidar Fusion + 3D CAD Models
Elektra (CVC-13) [5]	Autonomous University of Barcelona	73	2016	Barcelona	640x480	12	110		X	X			Manual + 3D CAD Models
CCSAD [39]	Centro de Investigación En Matemáticas	7	2015	Mexico	1096x822	50	96 k	X					-
KITTI Stereo 2015 [64]	Karlsruhe Institute of Technology	629	2015	Karlsruhe	1240x376	54	396		X	X			Stereo/Lidar Fusion + 3D CAD Models
Oxford RobotCar [60]	Oxford University	246	2015	Central Oxford	1280x960	24	1 M +		X	X	X	X	LiDAR
Málaga Stereo and Urban [12]	University of Málaga	107	2013	Málaga	1024x768	12	111 k		X	X	X		LiDAR
Stixel [24]	Daimler AG	91	2013	Germany	1024x333	21	2500		X				Manual
KITTI Stereo 2012 [36]	Karlsruhe Institute of Technology	3508	2012	Karlsruhe	1242x374	54	400			X			Stereo/Lidar Fusion + ICP
Ladicky [54]	Oxford Brookes University	91	2012	Leuven	316x256	150	70						Manual

Env: Environment, We: Weather, Il: Illumination, Ti: Daytime, Se: Seasons

TABLE 4.6: Localization and Mapping Datasets

Name	Provider	Cit.	Year	Location	Length (km)	Frames (k)	Vi	Sensor 2DLi	3DLi	Ve	If	Ev	Ground Truth Pose	Map	Distinctiveness
StreetLearn [67]	DeepMind	3	2019	New York, Pittsburgh	1100.0	114	X						GPS	X	Collection of Google Street View panoramas
UTBM RoboCar [112]	U. of Tech. of Belfort-Montbéliard	N/A	2019	Montbéliard	63.4	220	X	X	X				RTK		Long-term, day/night diversity, seasons
Multi Vehicle Stereo Event Camera [120]	U. of Pennsylvania	29	2018	Pennsylvania	9.6	36	X	X	X			X	GPS/IMU	X	Event camera, map as ground truth
Complex Urban [46]	KAIST	14	2018	South Korea	44.8	-	X		X				RTK	X	Focus on diverse and complex urban scenarios
Apollo Open Platform [2]	Baidu Inc.	N/A	2018	China	3.0	-	X		X				RTK		Multi-Sensor Localization
comma2k19 [85]	comma.ai	1	2018	San Francisco, San Jose	2000 +	2400	X						GNSS		Raw GNSS measurements, long-term
Oxford RobotCar [60]	Oxford University	246	2015	Central Oxford	1 and 010.5	1000 +	X	X	X				DGPS		Long-Term, same segment, weather diversity
AMUSE [51]	Linköping University	12	2013	Linköping	24.4	108	X			X			GPS/IMU		Velocity and height sensors and snow
KITTI Odometry [94]	Karlsruhe Inst. of Tech.	5415	2013	Karlsruhe	39.2	41	X		X				RTK		Widely-used benchmark for SLAM
Málaga Stereo and Urban [12]	University of Málaga	107	2013	Málaga	36.8	111	X	X					RTK		Direct sun
Cheddar Gorge [92]	BAE Systems	13	2011	Cheddar Gorge	31.8	86	X		X		X		DGPS		Infrared camera and Countryside mapping
The annotated laser data set [114]	CMU	11	2011	Pittsburgh	5.0	5	X	X					-		2D Lidar obstacle detection and SLAM
Ford [75]	U. of Michigan	177	2010	Michigan	5.1	7	X		X				DGPS		Omni cameras, precise camera-laser registration

Vi: Video, 2DLi: 2D LiDAR, 3DLi: 3D LiDAR, Ve: Velocity Sensor, If: Infrared Camera, Ev: Event Camera

TABLE 4.7: Behavior Datasets

Name	Provider	Cit.	Year	Location	Volume	Diversity R	S	N	D	Traffic U	H	Re	Sensor Vi	Fc	Li	GPS	IMU	CAN	Ev	Annotation	Objective
DBNet [21]	Shangai Jiao Tong University	4	2018	China	17	X	X	X	X	X	X	X	X		X	X	X	X		-	Steering/Velocity Prediction
HDD [70]	Honda Research Institute	23	2018	San Francisco	104	X	X	X		X	X	X	X		X	X	X	X		Maneuvers	Maneuver Prediction, Causal Reasoning
comma2k19 [85]	comma.ai	1	2018	San Francisco, San Jose	35					X			X			X	X	X		-	Steering Prediction
DDD17 [11]	ETH Zürich	31	2017	Switzerland, Germany	12	X		X	X	X	X		X			X		X	X	-	Steering Prediction
Brain4Cars [45]	Cornell University	35	2016	United States	22					X	X		X	X		X				Maneuvers, Street Maps	Maneuver Prediction
Dr(eye)ve [73]	University of Modena	24	2016	Modena	6	X		X	X	X	X	X	X			X				Gaze Maps, Pupil Dilation	Attention Monitoring
JAAD [80]	York University	16	2016	Ukraine, Canada	240	X	X	X	X	X		X	X							Pedestrian/Driver Behavior, Context	Pedestrian/Driver Behavior
UAH [83]	University of Alcalá	29	2016	Madrid	8					X	X		X			X	X			Driver Behavior, Maneuvers	Driver Behavior, Maneuver Prediction
Elektra (DrivFace) [26]	Autonomous University of Barcelona	10	2016	Barcelona	-							X	X	X						Gaze Direction, Facial Keypoints	Gaze Monitoring
Udacity [104]	Udacity	N/A	2016	Mountain View	8					X			X		X	X	X			-	Steering Prediction
comma.ai [84]	comma.ai	90	2016	San Francisco	7					X			X			X	X	X		-	Steering Prediction
DIPLECS Surrey [78]	University of Surrey	28	2015	Surrey	0.5					X	X	X	X							-	Steering Prediction
DIPLECS Sweden [78]	Autoliv	36	2015	Stockholm	3					X	X	X	X					X		Maneuvers, Context	Maneuver Prediction
EISATS (Set 11) [25]	University of Auckland	5	2010	Germany, New Zealand	<1					X			X							-	Gaze Monitoring

R: Rain, S: Snow, N: Night, D: Dawn/Dusk, U: Urban, H: Highway, Re: Residential, Vi: Video, Fc: Face Camera, Li: LiDAR, CAN: Controller Area Network Data, Ev: Event-camera

Acknowledgments

References

[1]   Mohamed Aly. "Real time Detection of Lane Markers in Urban Streets". In: *CoRR* abs/1411.7113 (2008). arXiv: 1411.7113.

[2]   *Apollo Data Open Platform.* 2018. URL: http://data.apollo.auto/.

[3]   Simon Baker et al. "A Database and Evaluation Methodology for Optical Flow". In: *International Journal of Computer Vision* 92.1 (Mar. 2011), pp. 1–31.

[4]   Aharon Bar Hillel et al. "Recent progress in road and lane detection: a survey". In: *Machine Vision and Applications* 25.3 (Apr. 2014), pp. 727–745.

[5]   F. Barrera Campo, F. Lumbreras Ruiz, and A. D. Sappa. "Multimodal Stereo Vision System: 3D Data Extraction and Algorithm Evaluation". In: *IEEE Journal of Selected Topics in Signal Processing* 6.5 (Sept. 2012), pp. 437–446.

[6]   Jens Behley et al. "A Dataset for Semantic Segmentation of Point Cloud Sequences". In: *CoRR* abs/1904.01416 (2019). arXiv: 1904.01416.

[7]   Karsten Behrendt. *Boxy Vehicle Detection in Large Images.* 2019. URL: https://boxy-dataset.com/boxy/.

[8]   Karsten Behrendt and Libor Novak. "A Deep Learning Approach to Traffic Lights: Detection, Tracking, and Classification". In: *IEEE International Conference on Robotics and Automation (ICRA).* 2017, pp. 1370–1377.

[9]   Luis Bergasa et al. "DriveSafe: an App for Alerting Inattentive Drivers and Scoring Driving Behaviors". In: *IEEE Intelligent Vehicles Symposium.* June 2014.

[10]   P. J. Besl and N. D. McKay. "A method for registration of 3-D shapes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (Feb. 1992), pp. 239–256.

[11]   Jonathan Binas et al. "DDD17: End-To-End DAVIS Driving Dataset". In: *CoRR* abs/1711.01458 (2017). arXiv: 1711.01458.

[12]   José-Luis Blanco-Claraco, Francisco Angel Moreno, and J. M. Hurtado Gonzalez. "The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario". In: *International Journal of Robotics Research (IJRR)* 33 (2014), pp. 207–214.

[13]   Markus Braun et al. "EuroCity Persons: A Novel Benchmark for Person Detection in Traffic Scenes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.8 (2019), pp. 1844–1861.

[14]   Gabriel J. Brostow, Julien Fauqueur, and Roberto Cipolla. "Semantic Object Classes in Video: A High-Definition Ground Truth Database". In: *Pattern Recognition Letters* 30 (2009), pp. 88–97.

[15]   D. J. Butler et al. "A naturalistic open source movie for optical flow evaluation". In: *European Conference on Computer Vision (ECCV).* Oct. 2012, pp. 611–625.

[16]   Holger Caesar et al. "nuScenes: A multimodal dataset for autonomous driving". In: *CoRR* abs/1903.11027 (2019). arXiv: 1903.11027.

[17]   Luca Caltagirone et al. "LIDAR–camera fusion for road detection using fully convolutional neural networks". In: *Robotics and Autonomous Systems* 111 (2019), pp. 125–131.

[18] Claudio Caraffi et al. "A System for Real-time Detection and Tracking of Vehicles from a Single Car-mounted Camera". In: *IEEE International Conference on Intelligent Transportation Systems*. Sept. 2012, pp. 975–982.

[19] Xiaozhi Chen et al. "3D Object Proposals using Stereo Imagery for Accurate Object Class Detection". In: *CoRR* abs/1608.07711 (2016). arXiv: 1608.07711.

[20] Xiaozhi Chen et al. "Multi-View 3D Object Detection Network for Autonomous Driving". In: *CoRR* abs/1611.07759 (2016). arXiv: 1611.07759.

[21] Yiping Chen et al. "LiDAR-Video Driving Dataset: Learning Driving Policies Effectively". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2018.

[22] Y. Choi et al. "KAIST Multi-Spectral Day/Night Data Set for Autonomous and Assisted Driving". In: *IEEE Transactions on Intelligent Transportation Systems* 19.3 (Mar. 2018), pp. 934–948.

[23] Marius Cordts et al. "The Cityscapes Dataset for Semantic Urban Scene Understanding". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[24] Marius Cordts et al. "The Stixel world: A medium-level representation of traffic scenes". In: *CoRR* abs/1704.00280 (2017). arXiv: 1704.00280.

[25] Michał Daniluk et al. "Eye Status Based on Eyelid Detection: A Driver Assistance System". In: *Lecture Notes in Computer Science*. Vol. 8671. Sept. 2014, pp. 171–178.

[26] Katerine Diaz-Chito, Aura Hernández-Sabaté, and Antonio M. López. "A Reduced Feature Set for Driver Head Pose Estimation". In: *Appl. Soft Comput.* 45.C (Aug. 2016), pp. 98–107.

[27] Shilp Dixit et al. "Trajectory planning and tracking for autonomous overtaking: State-of-the-art and future prospects". In: *Annual Reviews in Control* (Mar. 2018).

[28] P. Dollár et al. "Pedestrian Detection: A Benchmark". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2009.

[29] M. Enzweiler and D. M. Gavrila. "Monocular Pedestrian Detection: Survey and Experiments". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.12 (Dec. 2009), pp. 2179–2195.

[30] A. Ess et al. "A mobile vision system for robust multi-person tracking". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2008, pp. 1–8.

[31] Di Feng et al. "Deep Multi-modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges". In: *CoRR* abs/1902.07830 (2019). arXiv: 1902.07830.

[32] *FREE FLIR Thermal Dataset for Algorithm Training*. 2018. URL: https://www.flir.com/oem/adas/adas-dataset-form/.

[33] Mohamed Karam Gabr and Rimon Elias. "3D Reconstruction Algorithms Survey". In: *Intelligent Multidimensional Data and Image Processing*. IGI Global, 2018, pp. 1–17.

[34] Missael Garcia et al. "Bioinspired polarization imager with high dynamic range". In: *Optica* 5.10 (Oct. 2018), pp. 1240–1246.

[35] Alberto Garcia-Garcia et al. "A Review on Deep Learning Techniques Applied to Semantic Segmentation". In: *CoRR* abs/1704.06857 (2017). arXiv: 1704.06857.

[36]   A. Geiger, P. Lenz, and R. Urtasun. "Are we ready for autonomous driving? The KITTI vision benchmark suite". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2012, pp. 3354–3361.

[37]   Andreas Geiger et al. "Automatic Camera and Range Sensor Calibration using a single Shot". In: *IEEE International Conference on Robotics and Automation (ICRA)*. May 2012.

[38]   Andreas Geiger et al. "Vision meets Robotics: The KITTI Dataset". In: *International Journal of Robotics Research (IJRR)* 32.11 (Sept. 2013), pp. 1231–1237.

[39]   Roberto Guzman, Jean-Bernard Hayet, and Reinhard Klette. "Towards Ubiquitous Autonomous Driving: The CCSAD Dataset". In: *Lecture Notes in Computer Science* (Sept. 2015), pp. 582–593.

[40]   Q. Ha et al. "MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Sept. 2017, pp. 5108–5115.

[41]   Radu Horaud and Fadi Dornaika. "Hand-Eye Calibration". In: *The International Journal of Robotics Research (IJRR)* 14.3 (1995), pp. 195–210.

[42]   Xinyu Huang et al. "The ApolloScape Dataset for Autonomous Driving". In: *CoRR* abs/1803.06184 (2018). arXiv: 1803.06184.

[43]   Soonmin Hwang et al. "Multispectral Pedestrian Detection: Benchmark Dataset and Baselines". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.

[44]   S. Hwang et al. "Multispectral pedestrian detection: Benchmark dataset and baseline". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 1037–1045.

[45]   Ashesh Jain et al. "Brain4Cars: Car That Knows Before You Do via Sensory-Fusion Deep Learning Architecture". In: *CoRR* abs/1601.00740 (2016). arXiv: 1601.00740.

[46]   Jinyong Jeong et al. "Complex urban dataset with multi-level sensors from highly diverse urban environments". In: *The International Journal of Robotics Research (IJRR)* (2019), pp. 642–657.

[47]   Ryan Soussan Karsten Behrendt. *Unsupervised Labeled Lane Marker Dataset Generation Using Maps*. Tech. rep. Bosch Automated Driving, 2019.

[48]   R. Kesten et al. *Lyft Level 5 AV Dataset 2019*. https://level5.lyft.com/dataset/. 2019.

[49]   Byungju Kim, Junho Yim, and Junmo Kim. "Highway Driving Dataset for Semantic Video Segmentation". In: *British Machine Vision Conference (BMVC)*. 2018.

[50]   D. Kondermann et al. "The HCI Benchmark Suite: Stereo and Flow Ground Truth with Uncertainties for Urban Autonomous Driving". In: *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2016, pp. 19–28.

[51]   P. Koschorrek et al. "A Multi-sensor Traffic Scene Dataset with Omnidirectional Video". In: *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2013, pp. 727–734.

[52]   Iuliia Kotseruba, Amir Rasouli, and John K. Tsotsos. "Joint Attention in Autonomous Driving (JAAD)". In: *CoRR* abs/1609.04741 (2016). arXiv: 1609.04741.

[53] S. Kuutti et al. "A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications". In: *IEEE Internet of Things Journal* 5.2 (Apr. 2018), pp. 829–846.

[54] Lubor Ladický et al. "Joint Optimization for Object Class Segmentation and Dense Stereo Reconstruction". In: *International Journal of Computer Vision* 100.2 (Nov. 2012), pp. 122–133.

[55] I. Laina et al. "Deeper Depth Prediction with Fully Convolutional Residual Networks". In: *International Conference on 3D Vision (3DV)*. Oct. 2016, pp. 239–248.

[56] Seokju Lee et al. "VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition". In: *CoRR* abs/1710.06288 (2017). arXiv: 1710.06288.

[57] C. Li et al. "Multi-lane detection in urban driving environments employing omni-directional camera". In: *IEEE International Conference on Information and Automation (ICIA)*. July 2014, pp. 284–289.

[58] Li Liu et al. "Deep Learning for Generic Object Detection: A Survey". In: *CoRR* abs/1809.02165 (2018). arXiv: 1809.02165.

[59] X. Liu et al. "Benchmark for road marking detection: Dataset specification and performance baseline". In: *IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. Oct. 2017, pp. 1–6.

[60] Will Maddern et al. "1 Year, 1000km: The Oxford RobotCar Dataset". In: *The International Journal of Robotics Research (IJRR)* 36.1 (2017), pp. 3–15.

[61] Hiroya Maeda et al. "Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone". In: *CoRR* abs/1801.09454 (2018). arXiv: 1801.09454.

[62] C. Marina Martinez et al. "Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey". In: *IEEE Transactions on Intelligent Transportation Systems* 19.3 (Mar. 2018), pp. 666–676.

[63] S. Meister, B. Jähne, and D. Kondermann. "Outdoor stereo camera system for the generation of real-world benchmark data sets". In: *Optical Engineering* 51.2 (2012), p. 021107.

[64] M. Menze and A. Geiger. "Object scene flow for autonomous vehicles". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 3061–3070.

[65] Pierre Merriaux et al. "LiDAR point clouds correction acquired from a moving car based on CAN-bus data". In: *CoRR* abs/1706.05886 (June 2017). arXiv: 1706.05886.

[66] *Mighty Ai SampleSegmentation Dataset*. https://info.mighty.ai/semantic-segmentation-data. 2018.

[67] Piotr Mirowski et al. "The StreetLearn Environment and Dataset". In: *CoRR* abs/1903.01292 (2019). arXiv: 1903.01292.

[68] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund. "Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey". In: *IEEE Transactions on Intelligent Transportation Systems* 13.4 (Dec. 2012), pp. 1484–1497.

[69]   A. Mukhtar, L. Xia, and T. B. Tang. "Vehicle Detection Techniques for Collision Avoidance Systems: A Review". In: *IEEE Transactions on Intelligent Transportation Systems* 16.5 (Oct. 2015), pp. 2318–2338.

[70]   Gerhard Neuhold et al. "The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes". In: *International Conference on Computer Vision (ICCV)*. 2017.

[71]   L. Neumann et al. "NightOwls: A Pedestrians at Night Dataset". In: *Asian Conference on Computer Vision*. 2018.

[72]   Eshed Ohn-Bar and Mohan Manubhai Trivedi. "Are All Objects Equal? Deep Spatio-Temporal Importance Prediction in Driving Videos". In: *Pattern Recognition* 64 (Sept. 2016), pp. 425–436.

[73]   Andrea Palazzi et al. "Predicting the Driver's Focus of Attention: the DR(eye)VE Project". In: *CoRR* abs/1705.03854 (2017). arXiv: 1705.03854.

[74]   *PandaSet: Public large-scale dataset for autonomous driving provided by Hesai and Scale*. 2019. URL: https://scale.com/open-datasets/pandaset.

[75]   Gaurav Pandey, James McBride, and Ryan Eustice. "Ford Campus vision and lidar data set". In: *The International Journal of Robotics Research (IJRR)* 30 (Oct. 2011), pp. 1543–1552.

[76]   Abhishek Patil et al. "The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes". In: *CoRR* abs/1903.01568 (2019). arXiv: 1903.01568.

[77]   Peter Pinggera et al. "Lost and Found: Detecting Small Road Hazards for Self-Driving Vehicles". In: *CoRR* abs/1609.04653 (2016). arXiv: 1609.04653.

[78]   N. Pugeault and R. Bowden. "How Much of Driving Is Preattentive?" In: *IEEE Transactions on Vehicular Technology* 64.12 (Dec. 2015), pp. 5424–5438.

[79]   Vasili Ramanishka et al. "Toward Driving Scene Understanding: A Dataset for Learning Driver Behavior and Causal Reasoning". In: *CoRR* abs/1811.02307 (2018). arXiv: 1811.02307.

[80]   A. Rasouli, I. Kotseruba, and J. K. Tsotsos. "Are They Going to Cross? A Benchmark Dataset and Baseline for Pedestrian Crosswalk Behavior". In: *IEEE International Conference on Computer Vision Workshops (ICCVW)*. Oct. 2017, pp. 206–213.

[81]   A. Rasouli and J. K. Tsotsos. "Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice". In: *IEEE Transactions on Intelligent Transportation Systems* (2019), pp. 1–19.

[82]   Joseph Redmon and Ali Farhadi. "YOLOv3: An Incremental Improvement". In: *arXiv* (2018).

[83]   E. Romera, L. M. Bergasa, and R. Arroyo. "Need data for driver behaviour analysis? Presenting the public UAH-DriveSet". In: *IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Nov. 2016, pp. 387–392.

[84]   Eder Santana and George Hotz. "Learning a Driving Simulator". In: *CoRR* abs/1608.01230 (2016). arXiv: 1608.01230.

[85]   Harald Schafer et al. *A Commute in Data: The comma2k19 Dataset*. 2018. eprint: arXiv:1812.05752.

[86] D. Scharstein, R. Szeliski, and R. Zabih. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms". In: *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*. Dec. 2001, pp. 131–140.

[87] Timo Scharwächter et al. "Efficient Multi-cue Scene Segmentation". In: *Pattern Recognition*. 2013, pp. 435–445.

[88] René Schuster et al. "Combining Stereo Disparity and Optical Flow for Basic Scene Flow". In: *CoRR* abs/1801.04720 (2018). arXiv: 1801.04720.

[89] Y. Seo et al. "Recognition of Highway Workzones for Reliable Autonomous Driving". In: *IEEE Transactions on Intelligent Transportation Systems* 16.2 (Apr. 2015), pp. 708–718.

[90] M. S. Shirazi and B. T. Morris. "Looking at Intersections: A Survey of Intersection Monitoring, Behavior and Safety Analysis of Recent Studies". In: *IEEE Transactions on Intelligent Transportation Systems* 18.1 (Jan. 2017), pp. 4–24.

[91] Mennatullah Siam et al. "A Comparative Study of Real-Time Semantic Segmentation for Autonomous Driving". In: *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2018.

[92] R Simpson, J Cullip, and J Revell. *The Cheddar Gorge Data Set*. Tech. rep. BAE Systems (Operations) Limited, UK, Feb. 2011.

[93] S. Sivaraman and M. M. Trivedi. "A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking". In: *IEEE Transactions on Intelligent Transportation Systems* 11.2 (June 2010), pp. 267–276.

[94] Nikolai Smolyanskiy, Alexey Kamenev, and Stan Birchfield. "On the Importance of Stereo for Accurate Depth Estimation: An Efficient Semi-Supervised Deep Neural Network Approach". In: *CoRR* abs/1803.09719 (2018). arXiv: 1803.09719.

[95] Yainuvis Socarras et al. "Adapting Pedestrian Detection from Synthetic to Far Infrared Images". In: *ICCV – Workshop on Visual Domain Adaptation and Dataset Bias*. Sydney, Australia, 2013.

[96] J. Stallkamp et al. "The German Traffic Sign Recognition Benchmark: A multi-class classification competition". In: *International Joint Conference on Neural Networks*. July 2011, pp. 1453–1460.

[97] Karasawa Takumi et al. "Multispectral Object Detection for Autonomous Vehicles". In: *Thematic Workshops of ACM Multimedia*. Thematic Workshops '17. 2017, pp. 35–43.

[98] Xiaoou Tang. "Spatial As Deep: Spatial CNN for Traffic Scene Understanding". In: *AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2018.

[99] Alex Teichman, Jesse Levinson, and Sebastian Thrun. "Towards 3D Object Recognition via Classification of Arbitrary Object Tracks". In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2011.

[100] *The Nexar challenge and Nexet dataset*. 2017. URL: https://www.getnexar.com/challenge-2.

[101] R. Timofte, K. Zimmermann, and L. V. Gool. "Multi-view traffic sign detection, recognition, and 3D localisation". In: *2009 Workshop on Applications of Computer Vision (WACV)*. Dec. 2009, pp. 1–8.

[102] F. Tung et al. "The Raincouver Scene Parsing Benchmark for Self-Driving in Adverse Weather and at Night". In: *IEEE Robotics and Automation Letters* 2.4 (Oct. 2017), pp. 2188–2193.

[103] *TuSimple Benchmark*. 2018. URL: http://benchmark.tusimple.ai/#/.

[104] Udacity. *Udacity Self-Driving Car Driving Data 10/3/2016 (dataset-2-2.bag.tar.gz)*. 2016. URL: https://github.com/udacity/self-driving-car.

[105] Girish Varma et al. "IDD: A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments". In: *CoRR* abs/1811.10200 (2018). arXiv: 1811.10200.

[106] *Waymo Open Dataset: An autonomous driving dataset*. 2019. URL: https://www.waymo.com/open.

[107] C. Wojek, S. Walk, and B. Schiele. "Multi-cue onboard pedestrian detection". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2009, pp. 794–801.

[108] T. Wu and A. Ranganathan. "A practical system for road marking detection and recognition". In: *IEEE Intelligent Vehicles Symposium*. June 2012, pp. 25–30.

[109] Y. Xiang, A. Alahi, and S. Savarese. "Learning to Track: Online Multi-object Tracking by Decision Making". In: *IEEE International Conference on Computer Vision (ICCV)*. Dec. 2015, pp. 4705–4713.

[110] Xiaofei Li et al. "A new benchmark for vision-based cyclist detection". In: *IEEE Intelligent Vehicles Symposium (IV)*. June 2016, pp. 1028–1033.

[111] Jianru Xue et al. "BLVD: Building A Large-scale 5D Semantics Benchmark for Autonomous Driving". In: *CoRR* abs/1903.06405 (2019). arXiv: 1903.06405.

[112] Zhi Yan et al. *EU Long-term Dataset with Multiple Sensors for Autonomous Driving*. URL: https://epan-utbm.github.io/utbm_robocar_dataset/.

[113] Zike Yan and Xuezhi Xiang. "Scene Flow Estimation: A Survey". In: *CoRR* abs/1612.02590 (2016). arXiv: 1612.02590.

[114] Shao-Wen Yang, Chieh-Chih Wang, and Charles Thorpe. "The annotated laser data set for navigation in urban areas". In: *The International Journal of Robotics Research (IJRR)* 30.9 (2011), pp. 1095–1099.

[115] Fisher Yu et al. "BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling". In: *CoRR* abs/1805.04687 (2018). arXiv: 1805.04687.

[116] Oliver Zendel et al. "WildDash - Creating Hazard-Aware Benchmarks". In: *The European Conference on Computer Vision (ECCV)*. Sept. 2018.

[117] D. Zermas, I. Izzat, and N. Papanikolopoulos. "Fast segmentation of 3D point clouds: A paradigm on LiDAR data for autonomous vehicle applications". In: *IEEE International Conference on Robotics and Automation (ICRA)*. May 2017, pp. 5067–5073.

[118] Shanshan Zhang, Rodrigo Benenson, and Bernt Schiele. "CityPersons: A Diverse Dataset for Pedestrian Detection". In: *CoRR* abs/1702.05693 (2017). arXiv: 1702.05693.

[119] Zhengyou Zhang. "A Flexible New Technique for Camera Calibration". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 22.11 (Nov. 2000), pp. 1330–1334.

[120]   A. Z. Zhu et al. "The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception". In: *IEEE Robotics and Automation Letters* 3.3 (July 2018), pp. 2032–2039.

[121]   Zhengxia Zou et al. "Object Detection in 20 Years: A Survey". In: *CoRR* abs/1905.05055 (2019). arXiv: 1905.05055.