CAV Data online resources which may be of use for CAV Hackathon

Please note, participants are responsible for ensuring use of external data is in compliance with the usage conditions of the organisation that owns the data.

Organisation	Title	Link	Summary
California Institute of Technology	Caltech Pedestrian Data Benchmark	http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/index.html	The Caltech Pedestrian Dataset consists of approximately 10 hours of 640x480 30Hz video taken from a vehicle driving through regular traffic in an urban environment. About 250,000 frames (in 137 approximately minute long segments) with a total of 350,000 bounding boxes and 2300 unique pedestrians were annotated. The annotation includes temporal correspondence between bounding boxes and detailed occlusion labels. More information can be found in the PAMI 2012 and CVPR 2009 benchmarking papers.
Comma.ai	openpilot	https://github.com/commaai/openpilot	openpilot is an open source driving agent. Currently it performs the functions of Adaptive Cruise Control (ACC) and Lane Keeping Assist System (LKAS) for Hondas and Acuras. It's about on par with Tesla Autopilot at launch, and better than all other manufacturers. The openpilot codebase has been written to be concise and enable rapid prototyping. We look forward to your contributions - improving real vehicle automation has never been easier.
ELCANO	ELCANO	https://github.com/elcano/elcano	Autonomous mobile robot based on a recumbent trike

Organisation	Title	Link	Summary
Elektra	Semantic segmentation, pedestrian detection, virtual-world data, far infrared, stereo, driver monitoring	http://adas.cvc.uab.es/elektra/	**Elektra** is an autonomous vehicle project developed by research groups from the CVC research center and the UAB and UPC universities.
Institute Fur Neuroinformatik	German Road Signs	http://benchmark.ini.rub.de/index.php?section=gtsrb&subsection=dataset	50000 unique images of ~300 different road signs in a wide variety of conditions, angle, distance, clarity, resolution and light.
Karlsruhe Institute for Technology / Toyota Technological Instutute at Chicago	KITTI	http://www.cvlibs.net/datasets/kitti/raw_data.php	The dataset comprises the following information, captured and synchronized at 10 Hz: ▪ Raw (unsynced+unrectified) and processed (synced+rectified) grayscale stereo sequences (0.5 Megapixels, stored in png format) ▪ Raw (unsynced+unrectified) and processed (synced+rectified) color stereo sequences (0.5 Megapixels, stored in png format) ▪ 3D Velodyne point clouds (100k points per frame, stored as binary float matrix) ▪ 3D GPS/IMU data (location, speed, acceleration, meta information, stored as text file) ▪ Calibration (Camera, Camera-to-GPS/IMU, Camera-to-Velodyne, stored as text file) ▪ 3D object tracklet labels (cars, trucks, trams, pedestrians, cyclists, stored as xml file There is a conversion tool for ROS: https://github.com/tomas789/kitti2bag

Organisation	Title	Link	Summary
Kotseruba, Iuliia, Amir Rasouli, and John K. Tsotsos.	Joint Attention for Autonomous Driving (JAAD)	http://data.nvision2.eecs.yorku.ca/JAAD_dataset/	JAAD is a new dataset for studying joint attention in the context of autonomous driving. It contains an annotated collection of short video clips representing scenes typical for everyday urban driving in various weather conditions. JAAD dataset contains 346 high-resolution video clips (most are 5-10 sec) extracted from approx. 240 hours of driving videos filmed in several locations in North America and Eastern Europe.
Linkoping University	AMUSE	http://www.cvl.isy.liu.se/en/research/datasets/amuse/	The automotive multi-sensor (AMUSE) dataset consists of inertial and other complementary sensor data combined with monocular, omnidirectional, high frame rate visual data taken in real traffic scenes during multiple test drives.
LISA: Laboratory for Intelligent and Safe Automobiles	Vehicle Detection Dataset	http://cvrr.ucsd.edu/LISA/vehicledetection.html	Three colour video sequences captured at different times of the day and illumination settings: morning, evening, sunny, cloudy, etc. Different driving environments: highway and urban. Varying traffic conditions: light to dense traffic.
Oxford Robotics Institute (ORI)	Oxford Robocar Dataset	http://ori.ox.ac.uk/the-oxford-robotcar-dataset/	Over the period of November 2014 to December 2015 we traversed a 10km route through central Oxford twice a week on average in the Oxford RobotCar platform, an autonomous Nissan LEAF. This resulted in approximately 1000km of recorded driving with over 20 million images collected from 6 cameras mounted to the vehicle, along with LIDAR, GPS and INS ground truth. Data was collected in all weather conditions, including heavy rain, nighttime, direct sunlight and snow, and road and building works over the period of a year significantly changed sections of the route from the beginning to the end of data collection.
Peter Pinggera, Sebastian Ramos, Stefan Gehrig, Uwe Franke, Carsten Rother, Rudolf Mester	Lost and Found Dataset	http://www.6d-vision.com/lostandfounddataset	The LostAndFound Dataset addresses the problem of detecting unexpected small obstacles on the road often caused by lost cargo.

The dataset comprises 112 stereo video sequences with 2104 annotated frames (picking roughly every tenth frame from the recorded data). |

Organisation	Title	Link	Summary
Self Racing Cars	Data collected at Self Racing Cars events	http://selfracingcars.com/	Data collected at Self Racing Cars events
SHRP2 Naturalistic Driving Study	InSight Data Access Website	https://insight.shrp2nds.us/	Information describing the 3,400+ drivers and vehicles that participated in the naturalistic driving study. 5,400,000+ Trip summary records that describe individual trips recorded during the study. SHRP2 NDS status information including data collection and processing progress. 36,000+ Crash, near crash, and baseline driving events. Background information about the project and data being collected. Discussion forums for questions about the project and available data.
SYNTHIA Dataset	SYNTHIA Dataset	http://synthia-dataset.net/dataset/	**SYNTHIA**, The *SYNTHetic collection of Imagery and Annotations*, is a dataset that has been generated with the purpose of aiding semantic segmentation and related scene understanding problems in the context of driving scenarios. SYNTHIA consists of a collection of photo-realistic frames rendered from a virtual city and comes with precise pixel-level semantic annotations for 13 classes: misc, sky, building, road, sidewalk, fence, vegetation, pole, car, sign, pedestrian, cyclist, lanemarking.

Organisation	Title	Link	Summary
TensorFlow™	An open-source software library for Machine Intelligence	https://www.tensorflow.org/	TensorFlow™ is an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows you to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API. TensorFlow was originally developed by researchers and engineers working on the Google Brain Team within Google's Machine Intelligence research organization for the purposes of conducting machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.
The Rawseeds Project	Benchmarking Toolkit	http://www.rawseeds.org/	The datasets are composed of raw sensor data. The Benchmark Problems include quantitative performance metrics that can be applied to the output of the Benchmark Solutions. Therefore, once they are put in the form of Benchmark Solutions, algorithms can be quantitatively assessed and compared. Rawseeds' Benchmarking Toolkit is mainly targeted at the problems of localization, mapping and SLAM in robotics; but its use is not limited to them.
Tier iV	Tier IV software	https://github.com/CPFL/Autoware	Open-source software for urban autonomous driving.
Transport Systems Catapult	Various transport related datasets	http://imdata.co.uk/dataset	A directory of Intelligent Mobility related sources of data.
Udacity	223GB of Mountain View Driving Data	https://medium.com/udacity/open-sourcing-223gb-of-mountain-view-driving-data-f6b5593fbfa5#.qyz39werw	223GB of image frames and log data from 70 minutes of driving in Mountain View on two separate days, with one day being sunny, and the other overcast.

Organisation	Title	Link	Summary
UK Environment Agency	LIDAR data of English landscape	https://environmentagency.blog.gov.uk/2015/09/18/laser-surveys-light-up-open-data/	For the last 17 years the Environment Agency has used lasers to map and scan the English landscape from above to help us carry out work such as flood modelling and tracking changing coastal habitats. This month, for the first time, we are making our LIDAR (Light Detection and Ranging) data available for everyone to use for free. The aim is to help organisations, businesses, and individuals to do everything from manage forests, discover hidden archaeological features, and even create virtual reality worlds for gaming.
UK Government	National Public Transport Data Repository	https://data.gov.uk/dataset/nptdr	The NPTDR database contains a snapshot of every public transport journey in Great Britain for a selected week in October each year. Data is available for October 2004 to October 2011.
US Department for Transportation	Research Data Exchange	https://www.its-rde.net/	US Department of Transportation's data sharing system, which includes archived and real time transport data.