A Multimodal Perception Driven Self Evolving Autonomous Ground Vehicle Supplementary Data

Autonomous vehicles process immense quantities of data while navigating their environment. The algorithms that drive the technology are dependent on real-world data for development, testing, and validation. Table i presents the results of the research into relevant autonomous vehicle datasets. The majority of these autonomous driving datasets are primarily addressing the challenges of (a) scene understanding, (b) simultaneous localization and mapping, and (c) object detection. Mostly they rely on sensors such as LiDAR, Radar, and camera. However, some datasets – such as the Berkley Deep Drive dataset – primarily focused on Camera GPS and Inertial Measurement Unit (IMU) data.

While these datasets are useful, their lack of modality is a major shortcoming of some of them when pursuing multimodal ML methods such as online active ML. Furthermore, some datasets rely so heavily on GPS that they were not intended indoor environments or areas where the terrain can obstruct signals. Though some of the datasets reviewed provide IMU data, such as speed and direction of travel, most do not. Usually, the gathered data is optimized for a for a single aspect of autonomous vehicle research, and therefore not the most versatile.

TABLE I: REVIEWED AV DATASETS

Name	Ref	Permission Environment	Year	Description
nuScense	[1]	Licence Structured	2019	nuScense is an AV dataset recorded using LiDAR, Radar, camera, and GPS. nuScense consists of 1000 scenes containing 1.4 million camera images, 390000 LiDAR sweeps, 1.4 million Radar sweeps. Containing 23 different classes or 1.4 million objects annotated by hand. nuScense was collected in Boston and Singapore.
Oxford Radar RobotCar Dataset	[2]	Opensource Structured	2019	The Oxford Radar RobotCar Dataset is a radar release to append the Oxford RobotCar Dataset from 2016. Sensors included a Navtech CTS350-X Millimetre radar, dual Velodyne HDL-32E LiDAR with optimized ground truth radar odometry. Data was collected around Oxford over 280km.
Brno Urban Dataset	[3]	Licence Structured	2019	Bruno urban is an AV dataset recorded in the Czech Republic over 350km using four cameras, two LiDAR's, an inertial measurement unit, an IR camera, and a differential Global Navigation Satellite System (GNSS) receiver with centimetre accuracy.
A*3D	[4]	Opensource Structured	2019	A*3D is an AV dataset recorded at different times of the day and night in sunny, cloudy, and rainy weather conditions. 230000 human ladled 3D object annotations in 39,179 LiDAR Point Cloud frames with corresponding front facing RGB images.
Waymo Open Dataset	[5]	Opensource Structured	2019	Waymo open is an AV dataset recorded in the USA using LiDAR and camera sensors. The dataset contains data from 1,000 segments collected at 10Hz in diverse scenarios and environmental conditions. The dataset shows four object classes, 12M 3D bounding box labels with tracking IDs on LiDAR data and 1.2M 2D bounding box labels with tracking IDs on camera data.
Lyft Level 5	[6]	Licence Structured	2019	Lyft level 5 is a large-scale AV dataset recorded by a multiple, high-end AV fleet containing over 55000 humans ladled 3D annotated frames. Data was captured using seven cameras and up to 3 LiDAR. A semantic map provides 4000 lane segments (2000 road segment lanes and about 2000 junction lanes), 197 pedestrian crosswalks, 60 stop signs, 54 parking zones, eight-speed bumps, and 11-speed humps.
Argoverse	[7]	Opensource Structured	2019	Argoverse is an AV dataset recorded in Pittsburgh and Miami using LiDAR and camera sensors. Split into three releases; the first contains 113 scenes with 3D tracking annotations on all objects. The second release is a dataset of 300,000-plus scenarios. The third release is a set of HD maps of several neighbourhoods.
AEV Autonomous Driving Dataset	[8]	Licence Structured	2019	AEV is a dataset consisting of 2.3 TB of the camera, LiDAR sensor data, featuring Forty thousand frames with 2D semantic segmentation, 12000 frames with 3D bounding boxes, and unlabelled 3D Point Clouds. Also, 390000 frames of unlabelled sensor data.
ApolloScape	[9]	Opensource Structured	2019	ApolloScape is an AV dataset recorded in China using camera and LiDAR sensors with pixel-by-pixel annotations, including 26 different recognizable objects – cars, bicycles, pedestrians, and buildings. The dataset offers numerous levels of complexity recorded in challenging environments, Weather, and extreme lighting conditions.
Berkeley Deep Drive	[10]	Opensource Structured	2018	Berkeley Deep Drive is an AV dataset recorded in the USA using camera and GPS sensors. The dataset contains 100000 HD video – each running 40 seconds long at 30 fps – sequences over 1100-hour of driving across many different times of day, weather conditions, and driving scenarios.
Oxford RobotCar Dataset	[11]	Opensource Structured	2016	The Oxford RobotCar Dataset was recorded in Oxford, covering a fixed path, using LiDAR, camera, and GPS. Data was captured over one year, covering a variety of Weather and traffic conditions showing road users, along with longer-term changes to the environment.
KITTI	[12]	Opensource Structured	2013	The KITTI dataset was recorded in Germany using LiDAR, camera, Radar, and GPS sensors. The KITTI dataset is regarded as a benchmark dataset upon which many of the proceeding datasets were based. The dataset contains 6 hours of diverse traffic scenarios recorded at 10-100 Hz covering autobahn, rural roads, and inner-city scenes.
CamVid	[13]	Opensource Structured	2008	CamVid dataset provides ground truth labels that associate each pixel with one of 32 semantic classes. The dataset comprises ten minutes of high-quality 30Hz footage, with corresponding semantically labelled images at 1Hz. 700 images were annotated manually and were then inspected and confirmed by a second person for accuracy.

As of 30th January 2020, the LboroLdnAV dataset consists of 45.6 hours of Video, LiDAR, and Ultrasound data collected over 1.2 km of unstructured indoor and outdoor environments under various scenarios. In total, 2.5 million frames captured by four cameras, 672k frames captured by the 360Fly Wide-angled Camera; 1.2 million frames captured by the Ricoh Theta V 360° Camera; and 624k frames captured by the two Wansview IP cameras. In addition, both the LiDAR and Ultrasonic sensor array captured a total of 252k and 220k scans, respectively.

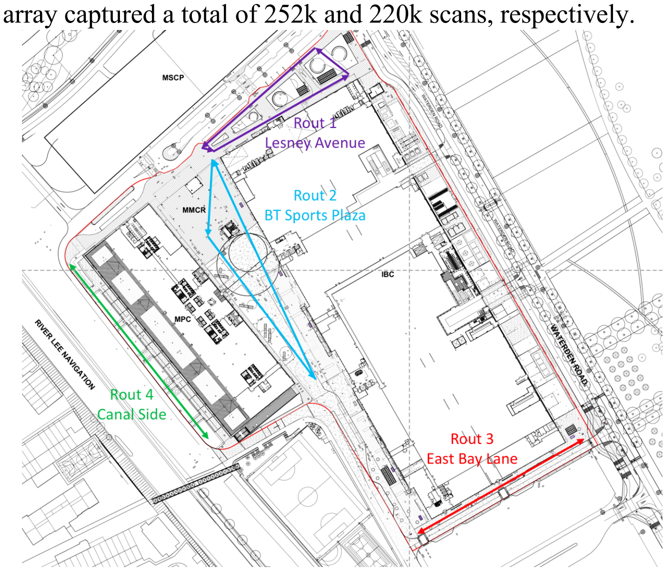

Figure I: The primary outdoor data collection routes. The traversable distance was 1.2 km over four different locations, Route 1 Lesney Avenue, Route 2 BT Sports Plaza, Route 3 East Bay Lane, and Route 4 Canal Side.

Seven object classes were annotated - People, Bus; Van; Car; Motorbike; Cyclist, and Traversable Surfaces. While this research's focus was FSD, it was felt that the additional object classes were beneficial in improving the value of the dataset and were therefore annotated. Each class was labelled with bounding boxes or polygons at 5Hz intervals. Although four cameras were used during the data collection period, only two of the data streams were used for the FSD process – the Ricoh Theta V 360° Camera and the ultrasonic sensor array. Data were annotated by hand, and a ground truth label for the seven classes is appended to the dataset.

TABLE II: DESCRIPTION OF THE LBOROLDNAV DATASET

Title	LBORO Dataset
Collection Periods	28/05/2018 to 18/08/2018 & 01/11/2019 to 30/01/2020
Location	East Bay Lane, Lesney Avenue, BT Sports Plaza, Canal-Side
Total Size	45.6 hours over 1.2 km
Class	People, Bus, Van, Car, Motorbike, Cyclist, Traversable Surfaces

The primary prerequisites of the LboroLdnAV dataset were to facilitate the development of multimodal machine learning algorithms for intelligent mobility. Therefore, the data collection route, shown in Figure , was chosen to cover various unstructured environments and traversable surfaces. While extreme weather conditions are desired, they were not the primary prerequisite when collecting the data. It should be noted that Figure I does not include the traversals for the indoor environment as these routes were not planed due to the changing environment.

Conditions of the license and permit granted by the management company restricted the speed at which the automatous platform could operate – less than 4kph. While the autonomous platform can operate at speeds of up 22kph, there would be little point since changing the sensors' frequency of operation would return inaccurate measurements. For example, decreasing the frequency of operation of either the LiDAR or Ultrasonic sensors array can result in ghosting. Ghosting is a replica of a recorded image, offset in position. Although possible to increase the frequency of operation to prevent ghosting, the resolution of the data captured by the LiDAR would significantly reduce. In the case of the ultrasound, it would result in crosstalk of transmitted and detected signals. While it would be possible to overcome these issues with sensors with a higher resolution, there will always be a degree of give and take in terms of frequency of operation.

TABLE III: SUMMARY STATISTICS FOR THE LBOROLDNAV DATASET

Sensor	Type	Size
360Fly Wide-angled Camera	Image	2.34 GB
Ricoh Theta V 360° Camera	Image	4.05 GB
Wansview IP Camera (x 2)	Image	0.24 GB
HC-SR04 Ultrasonic Array	2D Scan	5.6 MB
VLP-16 LiDAR	3D Scan	12.8 GB
Delphi ESR	3D Scan	N/A

The primary data collection routes were in unstructured indoor and outdoor environments. If we were to change the data collection environment to a structured environment, where the autonomous platform operated at a higher speed, the instruments and frequency of operation would need to be re-evaluated. Table ii details the date, location, classes captured, and data streams of the LboroLdnAV dataset, while Table iii lists summary statistics for the dataset so far. Data collection periods were chosen to encompass many classes in as many different environmental conditions as possible. However, since the dataset appended to this research is a partial release, most data was gathered during fine conditions. It should be noted that the data release reported in this article is the first part of a more extensive project that is currently in the process of expanding into Sri Lanka, in addition to further experiments around the Olympic Park.

The self-evolving autonomous ground vehicle used during the data collections period required a specific type of dataset. One that can fulfil the requirements of multimodality while providing optical and range data from at least two sensor streams of a known location. To determine the effectiveness of the proposed FSD framework, we visually compare the self-evolving component against the baseline SVM classifier and present the results in Figure II. Figure II a depicts the results of the baseline SVM classifier. Figure II b depicts the results of the self-evolving component. Figure II c depicts the results of the self-evolving component and the fusion of the image and ultrasonic range data.

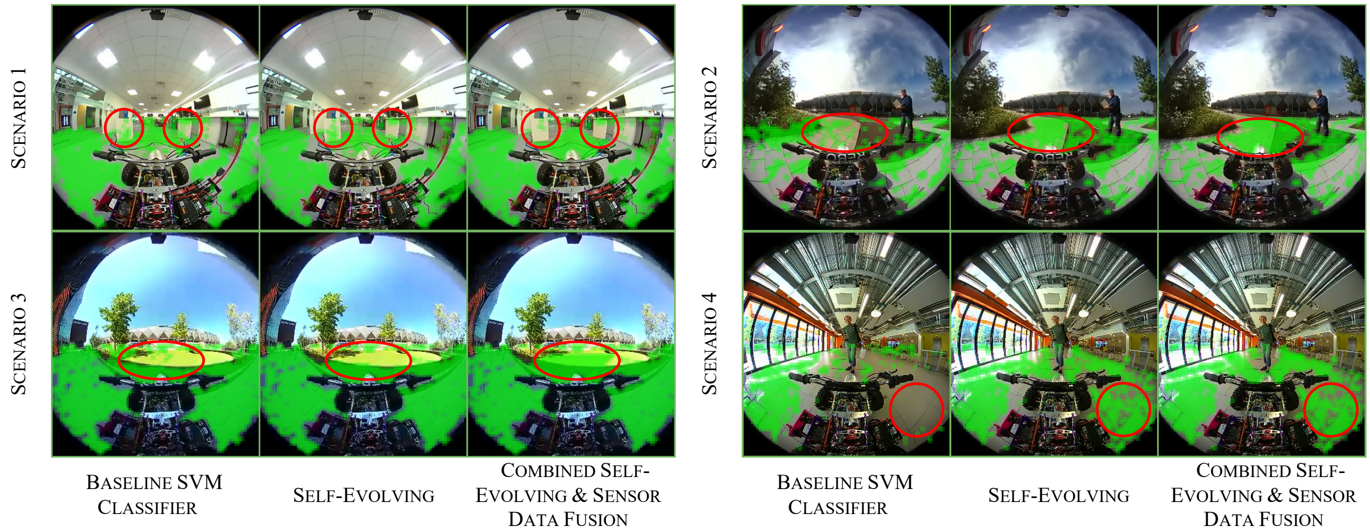

Figure II: From top left to right. Scenario 1: Indoor environment with stationary obstacle traversing lino. Scenario 2: Outdoor environment traversing changing surface (Concrete to Tarmac). Scenario 3: Outdoor environment traversing changing surface (Concrete/AstroTurf). Scenario 4: Indoor environment with stationary obstacle traversing tiled surface.

In consonance with the results presented in Figure II Scenario 1 (c), 2 (c), 3 (c) and 4 (c), the combination of the self-evolving component and sensor data fusion is better at detecting free space than the baseline SVM classifier. The combined approach correctly classifies the obstacles. For example, in Figure II Scenario 1 (a) & (b), the white boxes - positioned in front of the autonomous platform - are largely classified as free space. This corresponds to a situation where the baseline classifier performs poorly due to high saturation.

Similarly, for Figure II Scenario 2 (a), the baseline classifier fails to detect different surfaces as free space and miss-classifies a flower bed. Although there is an improvement in Figure II Scenario 2 (b), it is not until Figure II Scenario 2 (c) that the flowerbed and curbstone are correctly classified as free space. Figure II Scenario 3 (a) & (b) correspond to a situation where the baseline classifier and the self-evolving component is not performing very well. Due to high saturation, the AstroTurf area to the front of the testbed gets classified as an obstacle. Contrary to this, in Figure II Scenario 3 (c), the AstroTurf area to the front of the testbed gets correctly classified as traversable space. It should be noted there are areas of the AstroTurf that remain classified as occupied space. This can be attributed to these areas falling outside the FoV of the ultrasonic sensor array.

Figure II Scenario 4 (a) illustrates a different situation where the area to the left of the image is classified correctly, and the area to the right of the image is miss-classified. Conversely, in Figure II Scenario 4 (b) & (c), there is a marked improvement when using the self-evolving and the combined approach to free space segmentation.

References

[1] H. Caesar *et al.*, "nuScenes: A multimodal dataset for autonomous driving," *arXiv:1903.11027*, 2019.

[2] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The Oxford Radar RobotCar Dataset: A Radar Extension to the Oxford RobotCar Dataset," *arXiv:1909.01300*, p. 5, 2019.

[3] A. Ligocki, A. Jelinek, and L. Zalud, "Brno Urban Dataset -- The New Data for Self-Driving Agents and Mapping Tasks," *arXiv:1909.06897*, p. 7, 2019.

[4] Q.-H. Pham *et al.*, "A*3D Dataset: Towards Autonomous Driving in Challenging Environments," *arXiv:1909.07541*, p. 7, 2019.

[5] Waymo, "The Waymo Open Dataset," *Waymo*, 2019. [Online]. Available: https://waymo.com/open/. [Accessed: 16-Nov-2019].

[6] Lyft, "Lyft Level 5," *Lyft*, 2019. [Online]. Available: https://level5.lyft.com/dataset/. [Accessed: 16-Nov-2019].

[7] M.-F. Chang *et al.*, "Argoverse: 3D Tracking and Forecasting With Rich Maps," in *CVPR*, 2019.

[8] Audi Electronics Ventures, "AEV Autonomous Driving Dataset," *Audi*, 2019. [Online]. Available: https://www.audi-electronics-venture.de/aev/web/en/driving-dataset.html. [Accessed: 16-Nov-2019].

[9] P. Wang, X. Huang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, "The ApolloScape Open Dataset for Autonomous Driving and its Application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 10, pp. 2702–2719, 2019.

[10] F. Yu *et al.*, "BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling," *arXiv:1805.04687*, p. 16, 2018.

[11] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," *Int. J. Rob. Res.*, vol. 36, no. 1, pp. 3–15, 2016.

[12] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[13] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 88–97, 2009.