

The Oxford Road Boundaries Dataset

Tarlan Suleymanov, Matthew Gadd, Daniele De Martini, and Paul Newman
 Oxford Robotics Institute, University of Oxford, United Kingdom
 {tarlan,mattgadd,daniele,pnewman}@robots.ox.ac.uk



Figure 1: Dataset examples. First row: “raw”. Second row: “classified” or “partitioned” – visible (cyan) and occluded (red). First and third columns are left lenses, second and fourth columns are right lenses. For these and other visualisations, these road boundaries are dilated, but are in fact one pixel in width in the released data.

Abstract—In this paper we present *The Oxford Road Boundaries Dataset*, designed for training and testing machine-learning-based road-boundary detection and inference approaches. We have hand-annotated two of the 10km-long forays from the *Oxford Robotcar Dataset* and generated from other forays several thousand further examples with semi-annotated road-boundary masks. To boost the number of training samples in this way, we used a vision-based localiser to project labels from the annotated datasets to other traversals at different times and weather conditions. As a result, we release 62 605 labelled samples, of which 47 639 samples are curated. Each of these samples contain both *raw* and *classified* masks for left and right lenses. Our data contains images from a diverse set of scenarios such as straight roads, parked cars, junctions, etc. Files for download and tools for manipulating the labelled data are available at: [oxford-robotics-institute.github.io/road-boundaries-dataset](https://github.com/oxford-robotics-institute/road-boundaries-dataset)

Keywords— road boundary, curb, kerb, dataset

I. INTRODUCTION

Obtaining well-generalising, high-performance deep networks typically requires large quantities of training samples. Even if easy to hand-annotate – and this is often not the case – this is an inordinate workload for the researcher. To cope with the variability of *road boundaries*, the required data should equally incorporate great variability changes in environment, scale, appearance, colour, background clutter, occlusion, perspective and illumination. Yet, fine-grained annotation of data requires time-consuming human effort where labels of different classes must be assigned to outlined distinct regions. Nevertheless, finding true road boundaries is crucial for autonomous vehicles as they legally and intentionally delimit drive-able

surfaces. Indeed, as somewhat durable parts of the road infrastructure, lane markings, traffic signs, feature points or road boundaries are often used as features for localisation [1]. Moreover, detecting only visible road boundaries is often not sufficient in urban scenarios – examples in Figure 2 show road boundaries either partially or fully occluded. Indeed, there is often no visible road boundary to detect, as in Figure 2. In these sorts of scenarios, the vehicle should still sufficiently sense the environment to move safely about it. For this reason, datasets incorporating road boundaries should include a large number of samples in order to capture the variability of road boundaries. They should also contain annotations for both visible and occluded road boundaries. To this end, we have developed a framework for efficient annotations within a reasonable amount of time. In this paper, we present *The Oxford Road Boundaries Dataset*, which we generated using this framework.

II. RELATED WORK

Road-boundary detection is useful for autonomous driving in that it can facilitate safe driving in a legally demarcated area. Work in this area includes [2]–[5].

Autonomous driving datasets abound [6]–[13] and accurate, human-derived annotations for a wide number of semantic classes are available in these releases. However, road boundaries are not included, regardless of their importance for scene understanding, lane-level localisation, mapping and planning in autonomous driving.

The work which is mostly related to ours is [14]; yet, since it is based on aerial imagery, it does not provide synchronisation to vehicle-mounted sensors – and is not released for download at the time of writing.

Works based on the proposed dataset include [15]–[17], where CNN-based deep-learning models – *Visible Road Boundary Detection*



Figure 2: Examples where road boundaries are fully occluded.

(VRBD) and *Occluded Road Boundary Inference (ORBI)* – were proposed for detection and inference of visible and occluded road boundaries and were applied for scene understanding and lateral localisation problems.

III. THE OXFORD ROAD BOUNDARIES DATASET

This section describes the vehicle, collection regime, annotation procedures and dataset partitioning.

The overall pipeline of the proposed road boundary data generation framework is shown in Figure 3. We generate camera-based annotation from the multimodal collected data (Section III-A) by projecting laser-based annotations to the stereo, forward-facing camera (Section III-B) through Visual Odometry (VO) interpolation. Visual localisation tools are used to refer ground-truth annotations to other traversals for fast extension of the labels (Section III-C). Manual curation is carried out to check the created labels (Section III-D) and a deep-learning pipeline is used to automatically classify the boundaries as either visible or occluded (Section III-E).

A. Data Collection

This release is an annotative exercise atop the original *Oxford RobotCar Dataset*, of which we used seven forays from which approximately 62 000 pairs of samples were generated. Table I summarises the forays driven.

The dataset was collected using the Oxford RobotCar platform, as in [11], [18], an autonomous-capable Nissan LEAF (Figure 4). For the purpose of data annotation (see Section III-B), the following sensors are of interest:

- 1) 1× SICK LMS-151 2D LiDAR, 270° FoV, 50 Hz, 50 m range, 0.5° resolution.
- 2) 1× Point Grey Bumblebee XB3 (BBX3-13S2C38) trinocular stereo camera, 1280×960×3, 16 Hz, 1/3'' Sony ICX445 CCD, global shutter, 3.8 mm lens, 66° HFoV, 12/24 cm baseline.

Namely, the LiDAR is used in the annotation process to project visible and occluded road boundaries into each of the camera lens’s frames.

B. Annotation

Fine-grained hand-annotation of road boundaries from images would be a very time-consuming process and it would be impossible to exactly annotate the position of occluded road boundaries from the images themselves. To tackle this problem, we annotated the 3D pointcloud data collected by the 2D laser: as shown in Figure 4, the laser is attached vertically to the front of the vehicle; as such, the light pulses not only hit the road boundaries almost perpendicularly – which made the road boundaries easily distinguishable for annotation

– but also reach road boundaries behind parked vehicles as the pulses can pass under the vehicles (see Figure 5).

3D pointclouds are constructed by accumulating the push-broom returns along a pose-chain formed by VO which is only required to be accurate in the local region around the example to be annotated – drift is not a significant problem. Specifically, to obtain a 3D pointcloud from the *Oxford RobotCar Dataset* from 2D laser, we integrated subsequent vertical laser scans in a coherent coordinate frame. We used VO to estimate the vehicle’s motion and compute the transformations between subsequent scans. Note that, as VO uses camera images to calculate the vehicle’s ego-motion, it provides transformations between camera timestamps which in general differ from the laser timestamps. For this reason, we use interpolation to obtain transformations between subsequent 2D laser scans at time frames t_l and t'_l as follows:

$$T(t'_l, t_l) = T(t'_l, t'_{vo}) \cdot T(t'_{vo}, t_{vo}) \cdot T(t_{vo}, t_l) \quad (1)$$

$$s.t. \ t'_{vo} \leq t'_l, \ t_{vo} \geq t_l$$

where t'_{vo} and t_{vo} are the closest time steps of the VO with respect to the laser frames and where $T(t'_{vo}, t_{vo})$ is defined as follows:

$$T(t'_{vo}, t_{vo}) = \prod_{i=t'_{vo}}^{t_{vo}-1} T(i, i+1) \quad (2)$$

Once we have generated 3D point clouds of the datasets, we assign the same IDs to the points lying on the same continuous boundary. This enabled us to connect consecutive points according their IDs and fully annotate road boundary segments between the points, as shown in Figure 5. A Bird’s-eye view of the annotations of the 2015-05-26-13-59-22 foray is shown in Figure 6.

Annotation error in projecting from laser to camera due to offset between these two sensors is mitigated by hand-tuning the already fairly accurate extrinsic calibration provided by [18]. Additionally, centimetre-level accuracy is available from our vision-based localiser. Finally, drift in odometry is minimal over the limited distances over which estimates are accumulated.

C. Extension to other traversals

Annotating road boundaries in laser point clouds and projecting them into images enables us to easily generate hundreds of ground-truth masks within a short period of time, approximately 750 raw ground-truth masks per hour. To boost the number of training samples, we used a vision-based localiser [19] to project labels from the annotated datasets to other traversals at different times and in different weather conditions, as shown in Figure 8.

When annotating road boundaries we did not assume that every road boundary was a curb and we annotated all road boundaries even if they were on a flat surface with no height differences between roads and pavements. Although seeing flat road boundaries in the laser pointcloud was more difficult than curbs, we were simultaneously displaying projections of annotations of road boundaries on camera images in our annotation tool and using the projections as a reference to precisely position the annotations. This makes the proposed road boundary annotation framework applicable to any type of road or road network (e.g., urban roads, dirt roads, highways and motorways). An example of annotated road boundary with no height difference is shown in Figure 7.

D. Manual Curation

We manually curated the raw masks of the dataset, removing sections where projected annotations were unreasonable, e.g. when localisation failed or where the location of road boundaries changed over time. As a result, we obtained 62 605 labelled samples, and from those, we edited 47 639 samples while curating. Note that every sample is a pair of images from the left and right lenses.

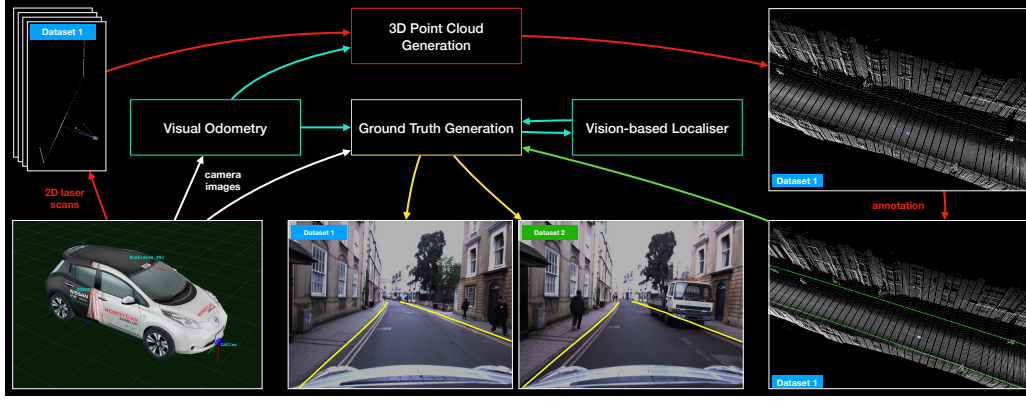


Figure 3: Road Boundary Data Annotation Framework: an efficient way of annotating road boundaries to obtain ground truth samples.

Foray	Number of frames		% of occluded labels		Download size		Annotation
	All	Curated	All	Curated	All	Curated	
2015-05-26-13-59-22	22775	14890	31%	29%	63 GB	41 GB	manual
2015-03-17-11-08-44	4869	3923	30%	30%	14 GB	10 GB	automatic
2015-05-08-10-33-09	5519	4617	31%	31%	16 GB	13 GB	automatic
2015-05-19-14-06-38	5424	4271	45%	44%	15 GB	12 GB	automatic
2018-04-30-14-43-54	15153	13353	28%	29%	44 GB	38 GB	manual
2019-01-10-11-46-21	4440	3392	39%	39%	12 GB	10 GB	automatic
2019-01-10-12-32-52	4425	3193	40%	38%	12 GB	8 GB	automatic
Total	62605	47639	33%	32%	176 GB	132 GB	N\A

Table I: Summary of data released. Manual: labels are obtained by projecting road boundaries from hand-annotated laser point cloud to images. Automatic: labels are obtained by projecting road boundaries from the hand-annotated datasets to other traversals using vision-based localiser.

E. Partitioning

Having data samples separated in two classes, visible and occluded, can be beneficial for operational safety [15]. To partition our raw road boundary masks into these two classes we trained the U-Net architecture [20] with raw masks. The resulting network can detect visible road boundaries, but fails to infer correct structure and position of occluded road boundaries for two reasons: (1) the network does not have a large enough receptive field to capture contextual information in the scene and to estimate position and structure of occluded road boundaries and (2) the network does not have any structure to force/bias outputs to be in a thin long shaped form for the occluded road boundaries. As such, the network can detect visible road boundaries when trained with raw masks, but fails to infer occluded ones and outputs blurry masks over occluding obstacles as shown in Figure 9, enabling automatic partitioning of the raw training data into two classes. We obtain masks for detected visible boundaries by applying threshold to the outputs. AND operation between the raw labels and thresholded outputs give us labels for visible road boundaries. Labels for occluded road boundaries are obtained by subtracting labels for visible from the raw labels (see Figure 1 for examples of partitioned training data masks). In total, 33% of all labels and 32% of curated labels are classified as occluded road boundaries. We refer the reader to [15] for more details on the performance of this network.

F. Recommended Usage

Our suggested usage is that all samples can be used for pretraining and that fine-tuning with curated samples is sensible. We used a subset of this dataset, approximately 23 000 samples, for training our models in [15] and achieved F_1 score of 0.9477 for visible road boundary detection and 0.8934 for occluded road boundary inference. In this application, we dilated masks before training models.

The thickness of road boundary line annotations on the ground truth masks are the same regardless of the height of road boundaries. This is because the annotated points do not contain any such information. To compensate for this, we recommend to use tolerance-based evaluation metrics – evaluation should be done by considering a sweep of thresholded pixel distances of predictions from true road boundaries.

G. Release

This section details the download format and tools related to this release.

The annotations are released alongside the associated RGB imagery as single .tar archives. The following file structure details the data tree utilised for each traversal in the dataset.

```

oxford-road-boundaries-dataset
|-- yyyy-mm-dd-HH-MM-SS
|   |-- uncurated
|   |-- raw.mp4

```

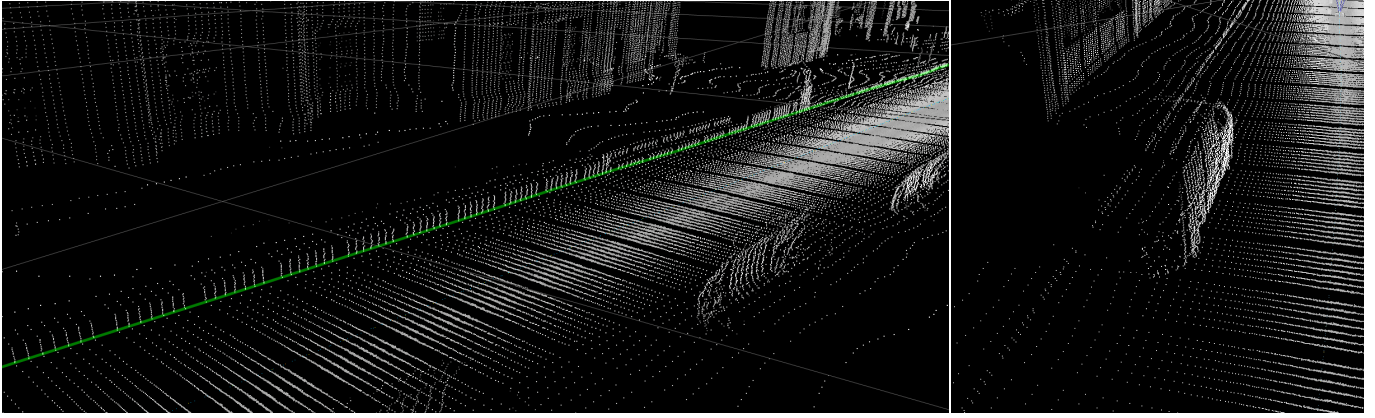



Figure 5: Left: lines are drawn between consecutive points with the same ID to annotate road boundary segments between the points. Right: laser light pulses reach the road boundary behind the parked car as they pass under the car.

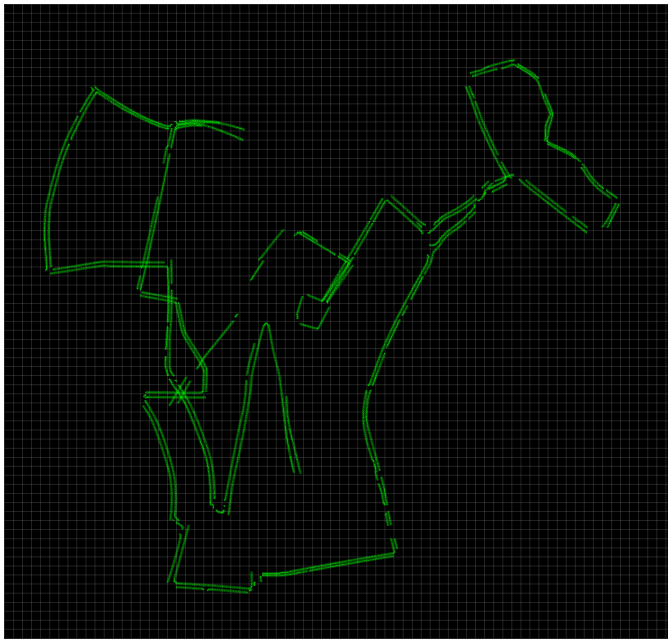


Figure 6: Bird's-eye view of the annotations of the 2015-05-26-13-59-22 dataset, 10 kilometres were annotated.



Figure 7: The road boundary on the left hand side of the road with no height difference between road and pavement was precisely annotated in the laser pointcloud by simultaneously displaying its projection on the camera image and using it as a reference while annotating.



Figure 8: Boosting data samples: annotate a dataset once, project the annotations to the images of the same dataset, then run the training data generation tool using outputs from the vision-based classifier to automatically project the annotations to other traversals.

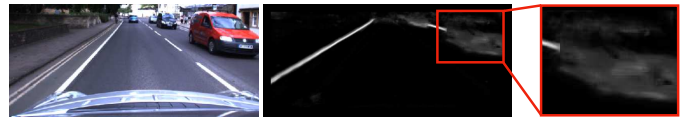


Figure 9: Blurry masks over occluding obstacles as understood by U-Net.

1000km: The Oxford RobotCar Dataset,” *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.

- [19] Winston Churchill and Paul Newman, “Experience-based navigation for long-term localisation,” *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1645–1661, 2013.
- [20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” *CoRR*, vol. abs/1505.04597, 2015.
- [21] T. Suleymanov, L. Kunze, and P. Newman, “Online inference and detection of curbs in partially occluded scenes with sparse lidar,” in *IEEE International Conference on Intelligent Transportation Systems*, Oct. 2019.



Figure 10: Generated “raw” road boundary examples: road boundary masks overlaid on top of RGB images. These masks are generated by projecting labels from annotated 3D point clouds into the corresponding images and they contain both visible and occluded road boundaries.



Figure 11: Partitioned road boundary data examples with two classes: visible (cyan) and occluded (red).