

Online Inference and Detection of Curbs in Partially Occluded Scenes with Sparse LIDAR

Tarlan Suleymanov

Lars Kunze

Paul Newman

Abstract—Road boundaries, or curbs, provide autonomous vehicles with essential information when interpreting road scenes and generating behaviour plans. Although curbs convey important information, they are difficult to detect in complex urban environments (in particular in comparison to other elements of the road such as traffic signs and road markings). These difficulties arise from occlusions by other traffic participants as well as changing lighting and/or weather conditions. Moreover, road boundaries have various shapes, colours and structures while motion planning algorithms require accurate and precise metric information in real-time to generate their plans.

In this paper, we present a real-time LIDAR-based approach for accurate curb detection around the vehicle (360 degree). Our approach deals with both occlusions from traffic and changing environmental conditions. To this end, we project 3D LIDAR pointcloud data into 2D bird’s-eye view images (akin to Inverse Perspective Mapping). These images are then processed by trained deep networks to infer both visible and occluded road boundaries. Finally, a post-processing step filters detected curb segments and tracks them over time. Experimental results demonstrate the effectiveness of the proposed approach on real-world driving data. Hence, we believe that our LIDAR-based approach provides an efficient and effective way to detect visible and occluded curbs around the vehicles in challenging driving scenarios.

I. INTRODUCTION

Autonomous vehicles are required to detect road boundaries (or curbs) for understanding their surroundings [1] and for generating behaviour plans. Road boundaries separate drivable road areas from non-drivable areas in the environment. Knowing the boundaries of a road is paramount for many applications, such as autonomous parking, navigation, mapping and path planning. Road boundaries can be used for lateral guidance as part of Advanced Driver Assistant Systems (ADAS), e.g. during parking. However, road boundaries have various shapes, colours and structures, which makes road boundary detection a challenging task. Moreover, road boundaries are often occluded by other traffic participants which makes their detection even more challenging.

In this work, our goal is to address the problem of road boundary detection using methods of deep learning. Recent advances in deep learning have shown that neural networks can be used for various tasks in robotics, such as segmentation and object detection, which play important roles in solving challenging tasks for operation of self-driving cars. Here we use deep networks to detect visible and infer occluded road boundaries around the vehicle (360°)

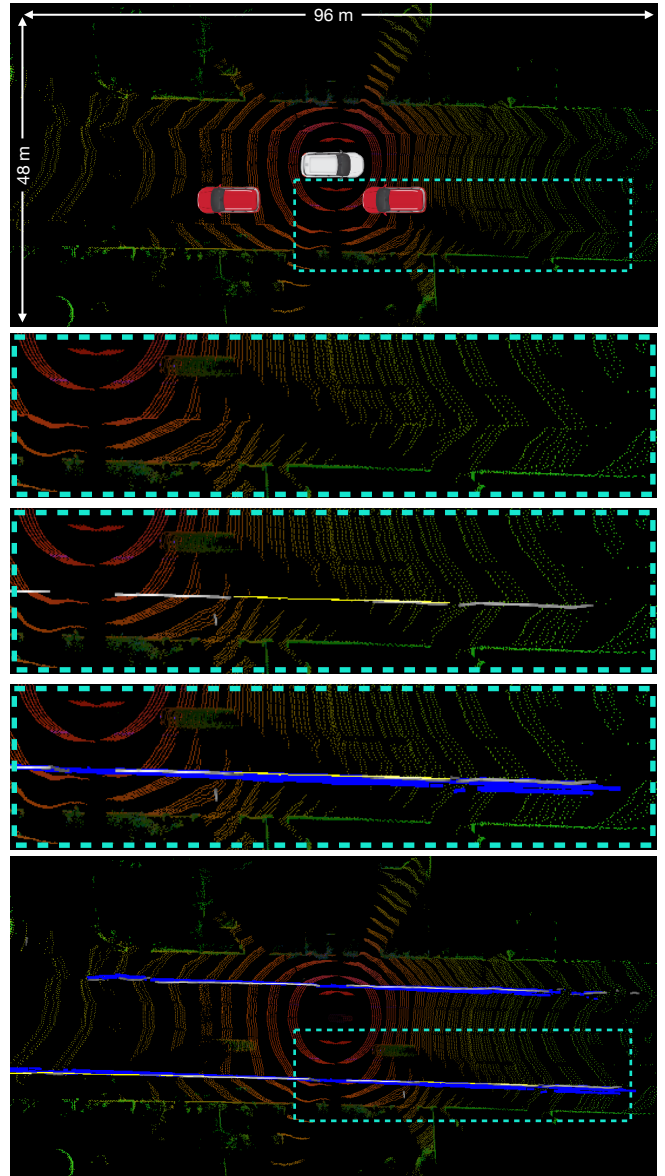


Fig. 1: Our 360° LIDAR-based curb detection approach. First, LIDAR data from the ego-vehicle (white) is transformed in bird’s-eye view images which are then processed by trained deep networks to detect visible (white) and occluded (yellow) curbs. Finally, post-processing steps filters out outliers and tracks curbs over time (blue). The result is a robust curb detection around the vehicle over a total distance of 96 metres.

from bird’s-eye view images that are generated from 3D LIDAR pointclouds.

To this end, the paper makes the following contributions:

- a framework to annotate 3D pointclouds and transform

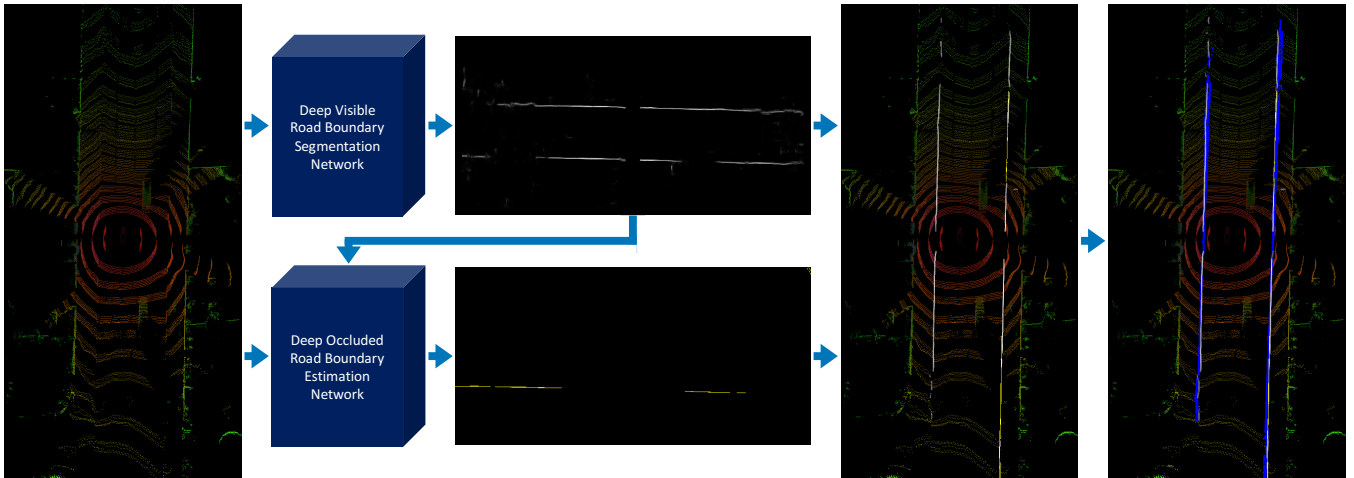


Fig. 2: Our 360° LIDAR-based curb detection approach. A pre-processing step integrates several subsequent laser scans into a coherent coordinate frame and projects them into a bird’s-eye view image with height information (left). This image is then processed by two deep segmentation networks to detect both visible and occluded road boundaries (middle). Note that the network responsible for occluded boundaries additionally considers the output of the network for visible boundaries as its input. Finally, the output of both networks is combined and tracked in order to improve the overall performance over a series of processed images (right).

them into a set of bird’s-eye view images and *raw* road boundary masks;

- an automatic method for partitioning/labelling *raw* training data into two classes (visible and occluded);
- a 360°, real-time deep learning method to detect visible and occluded road boundaries from bird’s-eye images without making any assumptions about structure and shape of the road boundaries; and
- an experimental evaluation which provides qualitative and quantitative results of our approach on real-world LIDAR data acquired while driving through Oxford.

This remainder of the paper is structured as follows. In Section II, we discuss related work on road boundary detection. An overview of our approach is given in Section III. In Section IV, we first explain how we obtained a data set for training our deep segmentation/detection networks before we describe our proposed approach in more detail in Section V. Finally, we provide a thorough qualitative and quantitative evaluation in Section VI before we conclude in Section VII.

II. RELATED WORK

Work on road boundary detection can be divided into two categories: *camera-based* and *LIDAR-based methods*.

Most camera-based methods address the problem using stereo cameras and 3D geometry to identify road boundaries in the scene [2], [3], [4], [5], [6], [7], [8]. In contrast, our previous work [9] used only a single monocular camera and deep convolutional neural networks for image processing to detect visible and occluded road boundaries. In this paper, we follow a similar machine-learning approach to infer both visible and occluded road boundaries, but we use LIDAR data as input. We do this without making any assumptions on the structure or shape of the curbs.

LIDAR-based methods often rely on more traditional information engineering techniques. In [10], a ring compression analysis on dense 3D LIDAR data followed by false-positive filters was used to detect curb points. Curb

models are estimated using Least Trimmed Squares (LTS) and describe the road shape on occluded curbs. However, the approach mostly considers simple examples where curbs are on both sides of the road. Hence, this method would likely fail in more complex scenarios, such as intersections and/or roads with fully occluded curbs. Work by [11] uses range and intensity information from 3D LIDAR to detect visible curbs on elevation data, which fails in the presence of occluding obstacles. Similarly, [12] presents a LIDAR-based method to detect visible curbs using sliding-beam segmentation followed by segment-specific curb detection, but fails to detect curbs behind obstacles.

Our proposed approach combines advantages from both research directions camera-based and LIDAR-based methods. In particular, we obtain highly accurate, 3D data about the world from LIDAR sensors which we process in real-time using deep convolutional neural networks (CNNs). 3D LIDAR data allows us to have a larger view angle than a single camera (here 360°). At the same time, LIDAR data is not restricted by lighting or weather conditions, which allows us to operate the system under various environmental conditions (e.g. rain and fog). However, LIDAR-based methods are very data intensive. To circumvent long processing times, we project 3D LIDAR data (i.e. pointclouds) into 2D bird’s-eye view images which we then process using CNNs, similar to [13], [14]. Thereby the network robustly detects road boundaries around the vehicle (360°) in urban traffic scenarios under different weather conditions.

III. LIDAR-BASED CURB DETECTION: AN OVERVIEW

In this work, we detect and infer visible and occluded road boundaries based on sparse LIDAR data using deep learning methods.

LIDAR data is used for both generating training examples and inferring road boundaries. To this end, we propose a novel framework to semi-automatically generate and boost the training data (Section IV). Moreover, we use 3D LIDAR

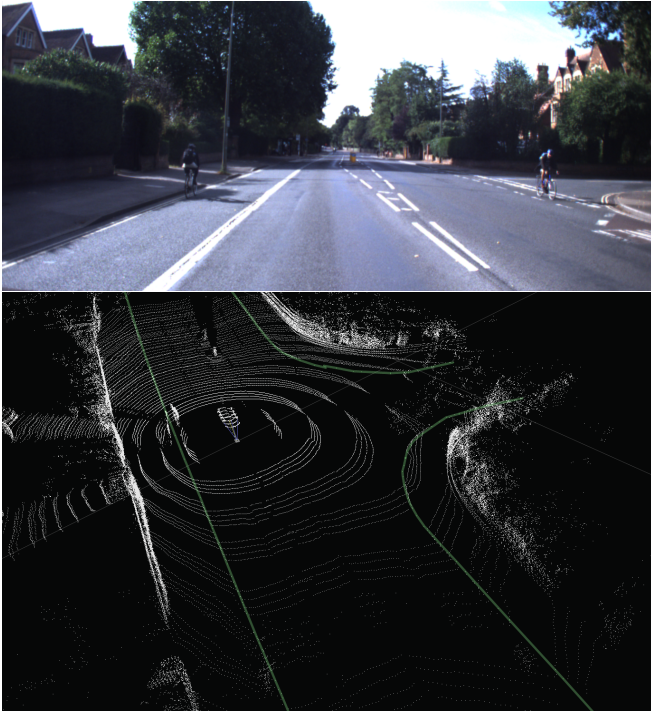


Fig. 3: Integrated 3D pointcloud data acquired by driving through Oxford and manually annotated with road boundaries.

data to detect and infer road boundaries around the vehicle (Sections V).

An overview of our 360° curb detection approach is given in Figure 2. In a pre-processing step, a sequence of LIDAR scans are integrated and projected into a bird’s-eye view image. Using this image, a first network detects visible road boundaries with a fully convolutional neural network and passes the output to a second network. The purpose of the second network is to infer road boundaries given both the original bird’s-eye view image and a mask of detected visible road boundaries. The generated output represents segments of road boundaries in a hybrid, discrete-continuous form. To improve the overall performance, a post-processing step filters out noise by consolidating detections in subsequent images and by tracking detected curbs over time.

IV. OBTAINING TRAINING DATA AND GROUND TRUTH

Training deep networks to obtain high-performance and well generalised models requires large datasets with ground truth labels. Moreover, the variability of road boundaries in shape and structure requires training data to include samples from different environments. Although obtaining *raw* data is relatively simple, fine-grained annotation of data requires human interaction and can be extremely time consuming. For example, outlined distinct regions in an image must be associated with corresponding class labels. To obtain ground truth labels we annotated road boundaries in 3D integrated pointcloud data, which was collected by a test car fitted with a 3D LIDAR (Velodyne VLP-32C) on the roof (Figure 3). To register and integrate a set of subsequent LIDAR scans in a coherent coordinate frame, we estimate the vehicle’s motion using Visual Odometry (VO) [15]. In this work, we used images acquired by a Point Grey Bumblebee XB3

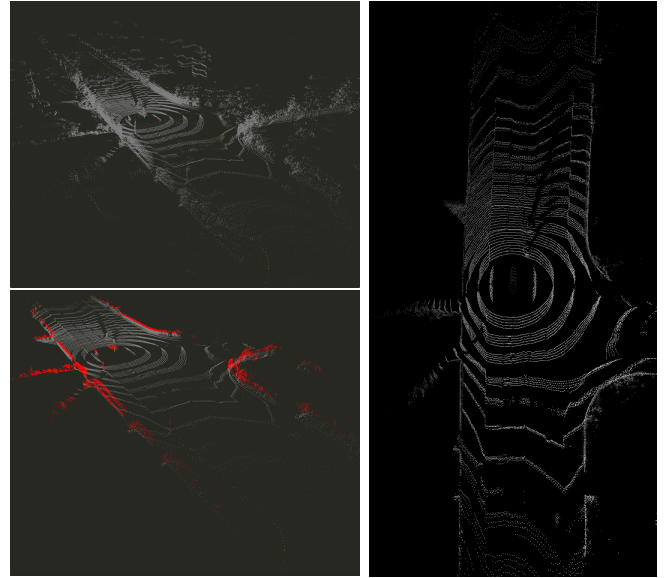


Fig. 4: Top left: integrated LIDAR pointcloud. Bottom left: Filtered and trimmed pointcloud. Right: Projected bird’s-eye view image.

camera, mounted on the front of the platform facing towards the direction of motion. In particular, our implementation of VO uses FAST corners [16] combined with BRIEF descriptors [17], RANSAC [18] for outlier rejection, and nonlinear least-squares refinement.

Based on integrated 3D pointclouds we then generate 2D projections which are transformed into bird’s-eye view (IPM) images. To do this, we take a laser scan and remove all points above the LIDAR device. As the LIDAR is mounted on top of the car, points above the LIDAR cannot be part of road boundaries. Similarly, we remove points that are below the LIDAR by more than 3.55 metres because some points may appear below the road surface as reflections. Also, we only keep points if they are located within 24 metres away from the car in x direction and within 48 metres in z direction. Thus we obtain trimmed pointclouds that we transform into bird’s-eye view images (Figure 4). In similar way, we obtain image masks by projecting annotations of road boundaries. Note that input bird’s-eye view images consist of three channels, range, intensity and height. In this work, we used LIDAR data from the OxfordRobotcar dataset [19] to obtain a new data set of bird’s-eye view images and semi-annotated (labelled) road boundary masks.

A. Partitioning Training Data: Visible and Occluded Curbs

To detect all road boundaries in a given scene we split the problem into two tasks: first we detect road boundaries that are visible from the laser and then infer road boundaries that are occluded by other road users. To achieve this, we partition the training data into visible and occluded classes.

However, note that the obtained masks contain both visible and occluded road boundaries as a single class as they were generated by projecting 3D annotations into bird’s-eye view images. To determine which points are visible and which are occluded we use the hidden point removal operator as described in [20]. The operator determines all visible points in a pointcloud when observed from a given viewpoint. This

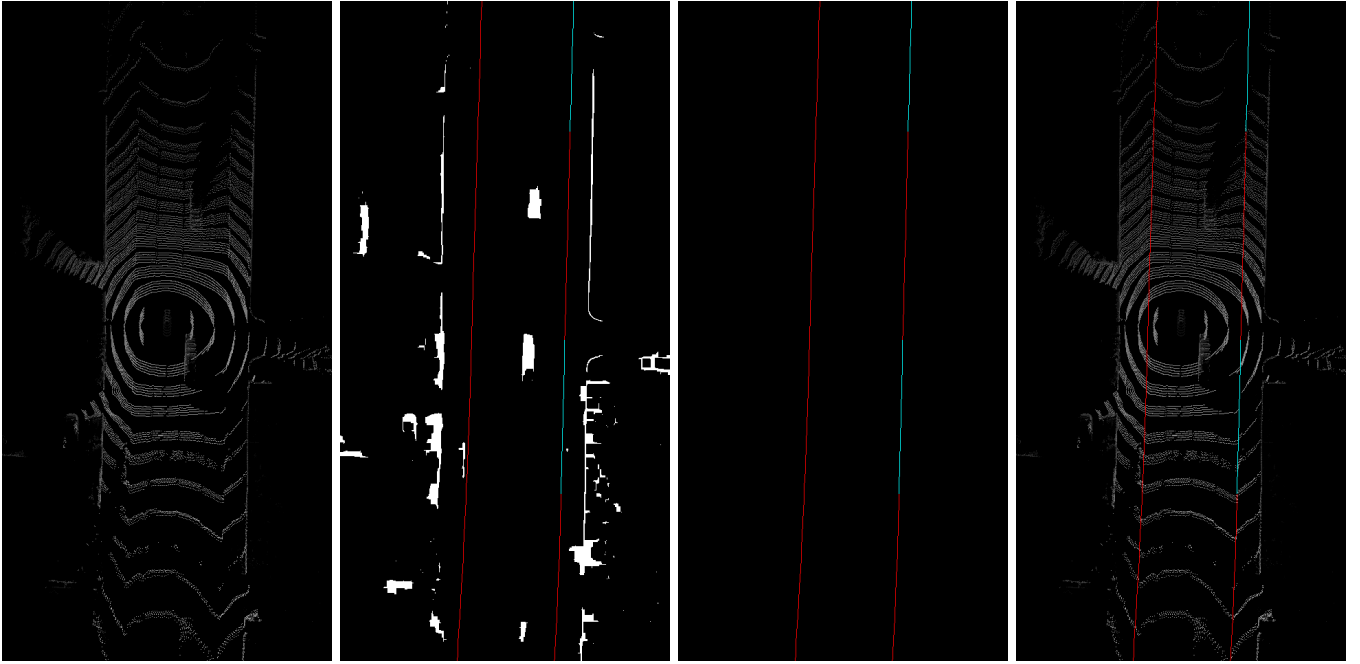


Fig. 5: Partitioning of training data. From left to right: bird’s-eye view image, detected obstacles as well as visible and occluded curbs, curbs labels only, bird’s-eye view image with labels.

is achieved by extracting all points residing on the convex hull of a transformed pointcloud. These points resemble the visible points, all other (labeled) points are considered as *hidden* (or occluded). We take the previously trimmed pointclouds and create binary bird’s-eye view images by taking the height of points from the ground into account. The points that are within a predefined height difference from the LIDAR roughly correspond to the points (obstacles) that are blocking the view. By putting together raw labels and binary masks of obstacles, obtained by running the hidden point removal algorithm, we obtain separate masks for visible and occluded road boundaries (Figure 5).

V. OUR APPROACH

Separating the training data into two classes (visible and occluded curbs) allows us to address the problem in two steps. Note, however, that these steps are linked (see Figure 2). In a first step we detect only visible curbs. These detected curbs then provide additional information for the second step, the detection of occluded curbs. In the following we briefly describe the architectures used for both steps.

A. Detecting visible curbs

To detect visible curbs we use the U-net architecture [21]. U-net is a fully convolutional network which concatenates higher resolution “input-side” features from convolution layers with up-sampled outputs from deconvolution layers. This enables the network to detect and segment objects such as curbs more precisely. Although this approach has been successful for visible curbs, it did not generate the desired outcome for occluded curbs. The reasons for this are twofold: first, the network’s limited receptive field, which is not big enough to capture context around large obstacles to estimate the position of curbs behind them, and second, the lack of

structure (model-free) which prevents the network to infer very thin curves of occluded road boundaries within an image. Hence, we have employed a different approach for the detection of occluded curbs which we explain in the next section.

B. Inferring occluded curbs

In this section, we explain our approach on inferring curbs in partially occluded scenes.

Our model consists of several convolutional layers that produce an output of detected curbs as discrete lines at multiple scales. Instead of pixels, a trained network estimates parameters of lines that correspond to cells in a grid (at different scales). It discretises the output space of lines into a set of default (anchor) lines over different orientation angles. At inference time, the network then generates probabilities for the presence of occluded curbs for each anchor line and its orientation. Making predictions of occluded curbs at multiple scales is important due to the different sizes and shapes of occluding obstacles.

In this work, we have selected three scales of parameterised outputs due to run-time constraints and a given accuracy target. Pixel-wise curb labels are converted to parameterised labels by dividing curb masks (at each scale) into a grid. Lines in each grid cell are parameterised in a discrete-continuous form: first, fitted lines are assigned to one of four types of anchor lines, and secondly, offsets between fitted and anchor lines are calculated. Anchor lines pass through the centre of a grid cell at different angles (22.5° , 67.5° , 112.5° and 157.5°). During fitting, lines are assigned to the closest anchor line. Once a fitted line is discretised, two continuous parameters are calculated: (1) an angle offset between a fitted and the respective anchor line ($\omega_{i,j,gt}^k$), and

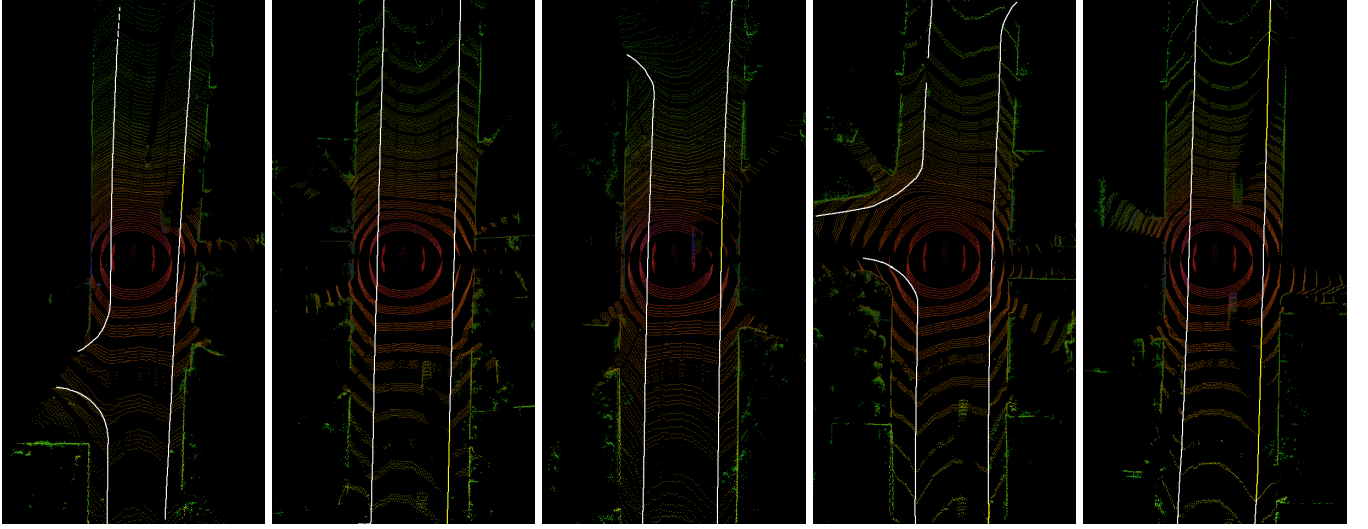


Fig. 6: Examples of labelled training data. Visible curbs are marked in white, while occluded curbs are marked in yellow.

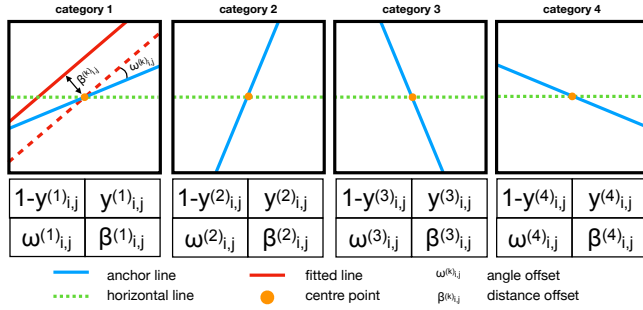


Fig. 7: Parameterisation of curb lines in discrete-continuous form. The four categories correspond to the four anchor lines.

(2) a distance from the centre of the cell to the fitted line ($\beta^{(k)}_{i,j,gt}$). As a result, we obtain 16 numbers for each grid cell, 4 numbers for each line category.

Estimating the presence of a curb line is a classification problem, but estimating adjustments to that line is a regression problem. To teach our network to perform the classification and regression at the same time, a discrete-continuous loss is applied during the training process. The total loss of the model L_t is defined as:

$$L_t = L_d + \alpha L_c = \sum_{i=1}^S L_{d_i} + \alpha \sum_{i=1}^S L_{c_i} \quad (1)$$

where L_d is a discrete loss of the curb line category classification, L_c is a continuous loss of the curb line parameters' regression, α is a weight term, and S denotes the number of scales (here 3).

Let $\hat{p}^k_{i,j}$ be a softmax output of the network for the k -th anchor line category in the j -th cell of the i -th scale, then the discrete loss for the i -th scale is:

$$L_{d_i} = - \sum_{j=1}^{C_i} \sum_{k=1}^A (y^k_{i,j} \log(\hat{p}^k_{i,j}) + (1 - y^k_{i,j}) \log(1 - \hat{p}^k_{i,j})) \quad (2)$$

where A is the number of anchor line categories (there are 4 categories), C_i is the number of cells in the i -th scale and

$y^k_{i,j}$ is the ground truth for the k -th anchor line category in j -th cell of the i -th scale.

The continuous loss is a smooth L1 loss between the predicted line ($\omega^k_{i,j,pr}$, $\beta^k_{i,j,pr}$) and the ground truth line ($\omega^k_{i,j,gt}$, $\beta^k_{i,j,gt}$) parameters. The continuous loss for the i -th scale is defined as:

$$L_{c_i} = \sum_{j=1}^{C_i} \sum_{k=1}^A (y^k_{i,j} (\text{smooth}_{L1}(\omega^k_{i,j,pr} - \omega^k_{i,j,gt}) + \text{smooth}_{L1}(\beta^k_{i,j,pr} - \beta^k_{i,j,gt}))) \quad (3)$$

where smooth_{L1} is define as in [22].

To increase the receptive field of the model we added intra-layer convolutions [23] before the multi-scale parameter estimation layers. Traditional layer-by-layer convolutions are applied between feature maps, but intra-layer convolutions are slice-by-slice convolutions within feature maps. Hence, intra-layer convolutions capture aspects across the whole image and can thereby capture spatial relationships over longer distances. For example, there is a strong correlation between the length of the occluded curbs and the size of objects which are obstructing the view (ranging from 10-15 pixels through occlusions by traffic cones to 200-300 pixels through occlusions by several parked cars).

C. Post-processing

Although the generated bird's-eye view images integrate five subsequent laser scans, our models do not take any temporal information into account explicitly. That is, each inference step is independent from the previous one. However, using temporal information can be useful in two ways: filtering out false positives and tracking true positives. To achieve this, we transform inputs at different time steps, t_l and t'_l , into a common reference frame using the following transformation:

$$T(t'_l, t_l) = \begin{bmatrix} R(t'_l, t_l) & t(t'_l, t_l) \\ 0 & 1 \end{bmatrix} \quad (4)$$

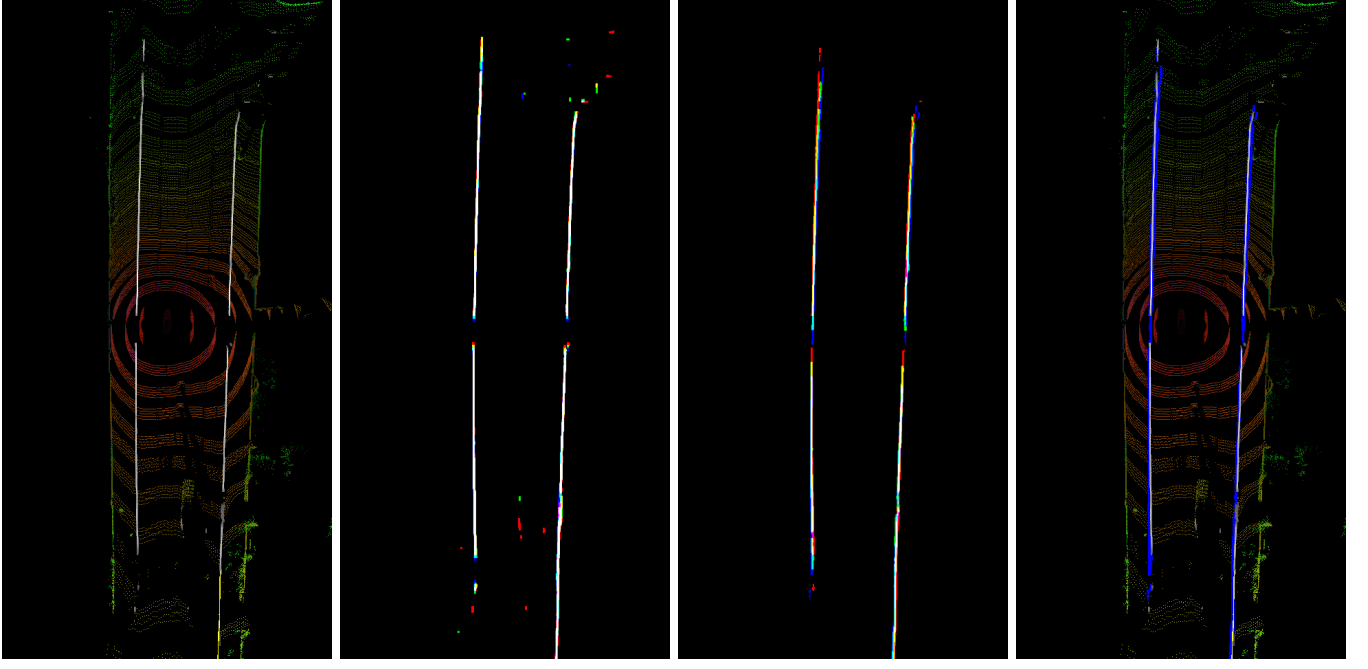


Fig. 8: Post-processing. From left to right: Output of curb detection networks (visible and occluded). Filtering step in which noise (highlighted in colour) is removed. Tracking step in which features tracked over time (in colour) are added. Finally, curb detection result after post-processing.

where $R(t'_l, t_l)$ is a rotation matrix and $t(t'_l, t_l)$ is a translation vector. As we explained in Section IV, VO is used to estimate the vehicle's ego-motion that allows us to obtain transformations between the vehicle's poses at different time steps. However, VO provides transformations corresponding to a camera frame rate. To obtain transformations between successive laser scans we use interpolation and calculate the transformation between time frames t_l and t'_l as follows [24]:

$$T(t'_l, t_l) = T(t'_l, t'_{vo}) \cdot T(t'_{vo}, t_{vo}) \cdot T(t_{vo}, t_l) \quad (5)$$

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

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

$$T(t'_{vo}, t_{vo}) = \prod_{i=t'_{vo}}^{t_{vo}-1} T(i, i+1) \quad (6)$$

To obtain $T(t'_l, t'_{vo})$ and $T(t_{vo}, t_l)$ we interpolate in $[t'_{vo}, t'_{vo} +$

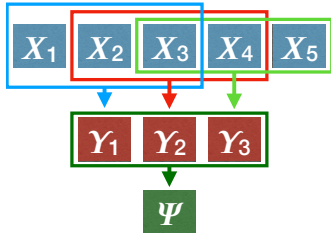


Fig. 9: Post-processing steps: A first step consolidates detection results of several subsequent scans (X_i) and generates a filtered output (Y_i). A second step, tracks detected segments over a series of filtered images (Y_i) and generates a final output (Ψ).

1] and $[t_{vo} - 1, t_{vo}]$ respectively. Using these transformations we apply two post-processing steps as follows:

Filtering. In the first step, we transform the last three output masks of detected road boundaries into a common reference attached to the current frame. Then we construct a histogram of output mask size (480x960) by counting the number of overlapping pixels with a value greater than threshold of 0.7 (which was determined experimentally). We keep the detected road boundaries that appear in all three frames above the threshold and disregard the rest. As a result noise is filtered out from the output mask as shown on histogram in Figure 8.

Tracking. In the second step, we perform a similar procedure as outlined above. However, this time we consider road boundary masks from the last three frames that were generated by the first step (as shown in Figure 9). By taking the union of these masks we track the detected road boundaries over the time. Integrating temporal information helps to close gaps between boundary segments (Figure 8).

VI. EXPERIMENTAL RESULTS

In this section, we show some qualitative results and provide an extensive quantitative evaluation of our approach. Qualitative results are presented in Figure 10 where outputs of curb detection networks and post-processing steps are included. The examples show that our approach is able to generate accurate masks for detected visible and inferred occluded road boundaries. To the best of our knowledge there is no public road boundary detection benchmark. Hence, we could not undertake a quantitative comparison of our approach with other existing approaches. However, we present an extensive quantitative evaluation based on our ground truth data.

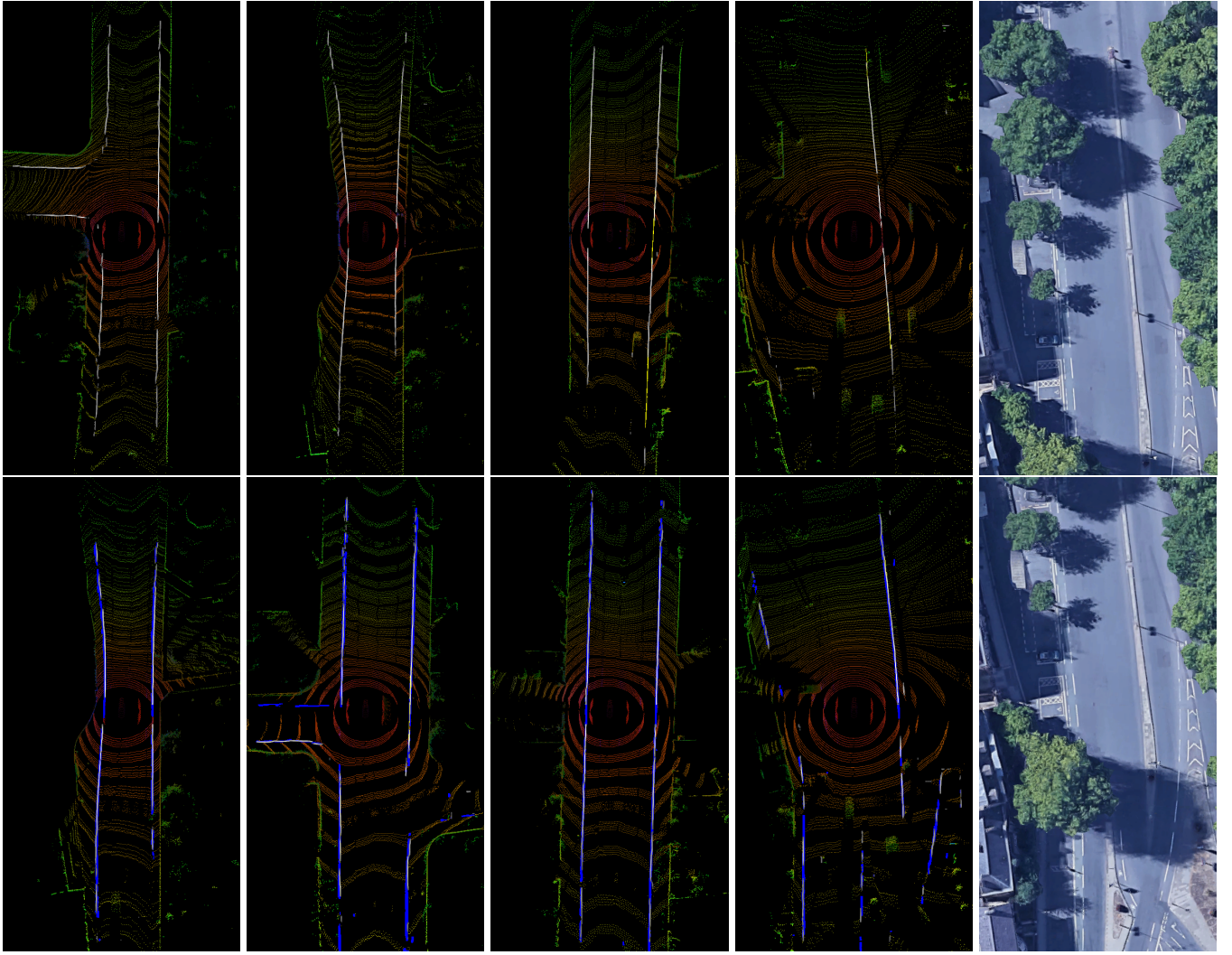


Fig. 10: First row: Sample outputs from the networks of detected and inferred road boundaries. Second row: Sample outputs after post-processing steps. The satellite images on the right depict the complexity of the scenes next to them.

1) *Visible only*: as we described in Section V, our approach tackles the detection of visible and inference of occluded road boundaries in two steps with a continuity constraint. The visible detector model is trained with 21K samples, which are generated by augmenting 1300 bird's-eye view images. Table III summarises the accuracy of the model with respect to 500 test images. We present precision, recall and F1 score for different areas around the vehicle (images size) and different tolerances. With minimum tolerance and maximum distance of 96 metres the model achieves 0.8168 F1 score. The model performs better in areas close to the car and achieves up to 0.9437 F1 score.

2) *Visible and occluded*: the occluded road boundary inference model is also trained with 21K samples that are obtained from 600 bird's-eye images through augmentation. The model infers road boundaries that are occluded by other road users and achieves 0.7867 F1 score. Note that for occluded road boundaries having higher tolerance is acceptable as their exact location is unknown. Combining outputs of both networks gives 0.893 F1 score with respect

TABLE I: Precision, recall and F1 score of the visible road boundary detection model

Image size / area	Tolerance	Precision	Recall	F1 Score
480x960 / 48x96 metres	1px	0.8371	0.7974	0.8168
	2px	0.9193	0.8387	0.8772
	3px	0.9455	0.8572	0.8992
	4px	0.9579	0.8698	0.9117
480x720 / 48x72 metres	1px	0.8631	0.8721	0.8676
	2px	0.9310	0.9050	0.9178
	3px	0.9520	0.9183	0.9349
	4px	0.9618	0.9272	0.9442
480x480 / 48x48 metres	1px	0.8819	0.8921	0.8870
	2px	0.9343	0.9021	0.9179
	3px	0.9541	0.9157	0.9345
	4px	0.9632	0.9250	0.9437

to all road boundaries.

3) *Post-processing*: The importance of the post-processing step is to track detected and inferred road boundaries and fill in the gaps that were not detected initially. This happens on both sides of the vehicle as can be seen in Figure 8 (left). Those gaps are closed after the

TABLE II: Precision, recall and F1 score of the visible and occluded road boundary detection and inference models (size: 480x960)

Model	Labels	Tolerance	Precision	Recall	F1 Score
Occluded	Occluded	1px	0.5099	0.6420	0.5684
		2px	0.6414	0.7441	0.6889
		3px	0.7090	0.7938	0.7490
		4px	0.7523	0.8245	0.7867
Vis + Occ	Vis + Occ	1px	0.8318	0.7531	0.7905
		2px	0.9184	0.7980	0.8540
		3px	0.9477	0.8191	0.8787
		4px	0.9619	0.8333	0.8930

TABLE III: Precision, recall and F1 score with and without post-processing step (size: 480x480)

Pipeline	Tolerance	Precision	Recall	F1 Score
Vis + Occ	1px	0.8922	0.8387	0.8646
	2px	0.9470	0.8649	0.9041
	3px	0.9627	0.8784	0.9186
	4px	0.9706	0.8887	0.9279
Vis + Occ + Post	1px	0.5961	0.9246	0.7249
	2px	0.8773	0.9524	0.9133
	3px	0.9542	0.9594	0.9568
	4px	0.9741	0.9630	0.9685

post-processing step (right). Note that when generating filtered outputs (Y_i) dilation is applied to the output masks to increase the overlap between output masks. As a result, the thickness of detected and inferred road boundaries increases, which leads to a decrease in precision and F1 score but an increase in recall for the minimum tolerance. However, whenever the tolerance is above 1px, the F1 score increases as the post-processing step fills in the gaps.

4) *Running time*: The system runs at 8.53 Frames Per Second (FPS) with input images of size 480x960 on a NVIDIA 1080 Ti GPU.

Overall, the experiments demonstrate that our approach is capable of detecting and inferring all road boundaries in complex road scenes over a total distance of 96 metres around the vehicle and in real-time.

VII. CONCLUSIONS

In this paper, we have presented a LIDAR-based approach for curb detection around the vehicle. Integrated LIDAR pointclouds are transformed into bird's-eye images which are then processed by trained convolutional networks. Instead of processing the whole pointcloud we only process a projected pointcloud in form of an image which allows our method to operate in real-time. Finally, we post-process the network outputs, i.e. the detected curbs (visible and occluded), by filtering out noise and tracking detections over time. This last step increases the overall performance of curb detection as we have shown in an extensive evaluation (Section VI).

Overall, we conclude that LIDAR data can be used to accurately detect curbs at high processing rates. Compared to camera-based approaches, LIDAR has not only a greater coverage and range around the vehicle, but it is also more robust to environmental changes due to lighting and/or weather. The experiments demonstrated that our approach achieves a high performance in detecting and inferring visible and occluded road boundaries around the vehicle and

achieves an F1 score of 0.9685 with post-processing. Hence, we strongly believe that our proposed deep learning approach based on LIDAR data can have a wide impact on a range of different applications in the autonomous driving domain.

Acknowledgment: The work has been supported by the EPSRC/UK Research and Innovation Programme Grant EP/M019918/1 (Mobile Autonomy) and the NVIDIA Corporation with the donation of Titan Xp and Titan V GPUs.

REFERENCES

- [1] L. Kunze, T. Bruls, T. Suleymanov, and P. Newman, "Reading between the lanes: Road layout reconstruction from partially segmented scenes," in *IEEE ITSC*, Maui, Hawaii, USA, November 2018.
- [2] V. Prinnet, J. Wang, J. Lee, and D. Wettergreen, "3D road curb extraction from image sequence for automobile parking assist system," *IEEE ICIP*, pp. 3847–3851, 2016.
- [3] M. Kellner, U. Hofmann, M. E. Bouzouraa, and N. Stephan, "Multi-cue, model-based detection and mapping of road curb features using stereo vision," in *IEEE ITSC*, Sept 2015, pp. 1221–1228.
- [4] L. Wang, T. Wu, Z. Xiao, L. Xiao, D. Zhao, and J. Han, "Multi-cue road boundary detection using stereo vision," in *IEEE ICVES*, July 2016.
- [5] F. Oniga, S. Nedeveschi, and M. M. Meinecke, "Curb detection based on a multi-frame persistence map for urban driving scenarios," in *IEEE ITSC*, Oct 2008, pp. 67–72.
- [6] M. Kellner, M. E. Bouzouraa, and U. Hofmann, "Road curb detection based on different elevation mapping techniques," in *IEEE IV*, June 2014, pp. 1217–1224.
- [7] J. Siegemund, U. Franke, and W. Förstner, "A temporal filter approach for detection and reconstruction of curbs and road surfaces based on conditional random fields," in *IEEE IV*, June 2011, pp. 637–642.
- [8] M. Enzweiler, P. Greiner, C. Knöppel, and U. Franke, "Towards multi-cue urban curb recognition," in *IEEE IV*, June 2013, pp. 902–907.
- [9] T. Suleymanov, P. Amayo, and P. Newman, "Inferring road boundaries through and despite traffic," in *IEEE ITSC*, November 2018.
- [10] A. Y. Hata and D. F. Wolf, "Feature detection for vehicle localization in urban environments using a multilayer lidar," *T-ITS*, vol. 17, no. 2, pp. 420–429, Feb 2016.
- [11] W. Yao, Z. Deng, and L. Zhou, "Road curb detection using 3d lidar and integral laser points for intelligent vehicles," in *SCIS/ISIS*, Nov 2012, pp. 100–105.
- [12] Y. Zhang, J. Wang, X. Wang, and J. M. Dolan, "Road-segmentation-based curb detection method for self-driving via a 3d-lidar sensor," *T-ITS*, vol. 19, no. 12, pp. 3981–3991, 2018.
- [13] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," *IEEE CVPR*, Jul 2017.
- [14] L. Caltagirone, S. Scheidegger, L. Svensson, and M. Wahde, "Fast lidar-based road detection using fully convolutional neural networks," *IEEE IV*, Jun 2017.
- [15] D. Nistér, O. Naroditsky, and J. R. Bergen, "Visual odometry," in *IEEE CVPR*. IEEE Computer Society, 2004, pp. 652–659.
- [16] E. Rosten, G. Reitmayr, and T. Drummond, "Real-time video annotations for augmented reality," in *Advances in Visual Computing*. Springer, 2005, pp. 294–302.
- [17] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE TPAMI*, vol. 34, no. 7, pp. 1281–1298, July 2012.
- [18] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [19] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," *IJRR*, vol. 36, no. 1, pp. 3–15, 2017.
- [20] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007.
- [21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.
- [22] R. B. Girshick, "Fast R-CNN," in *ICCV*, 2015.
- [23] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," *arXiv preprint arXiv:1712.06080*, 2017.
- [24] R. Aldera, D. De Martini, M. Gadd, and P. Newman, "Fast radar motion estimation with a learnt focus of attention using weak super-vision," *IEEE ICRA*, 2019.