Hierarchical Road Topology Learning for Urban Map-less Driving

Li Zhang **Faezeh Tafazzoli** **Gunther Krehl** **Runsheng Xu** **Timo Rehfeld**

Manuel Schier **Arunava Seal**

Mercedes-Benz R&D NA
{li.lz.zhang, faezeh.tafazzoli, gunther.krehl, runsheng.xu, timo.rehfeld, manuel.schier, arunava.seal}@daimler.com

Abstract

The majority of current approaches in autonomous vehicles rely on High-Definition (HD) maps which detail the road geometry and surrounding area. Yet, this reliance is one of the obstacles to mass deployment of autonomous vehicles due to poor scalability of such prior maps. In this paper, we tackle the problem of online road map extraction via leveraging the sensory system aboard the vehicle itself. To this end, we design a structured model where a graph representation of the road network is generated in a hierarchical fashion within a fully convolutional network. The method is able to handle complex road topology and does not require a user in the loop. We demonstrate the effectiveness of our approach in a variety of topological scenarios.

1 Introduction

Autonomous cars tend to rely on data-hungry perception algorithms to comprehend their surroundings. They are loaded with a constellation of sensors collecting data from the ambient environment, such as position and dimensions of the surrounding objects, weather condition, and traffic, but also large, detailed, and accurate global maps. Maps are an indispensable component of self-driving technology. The unique needs of autonomous vehicles necessitate a new class of HD maps for prior map-based localization, modeling the road surface at centimeter-level accuracy, enabling an autonomous vehicle to confidently deduce its position with respect to the ambient environment. With such strong prior knowledge, an autonomous vehicle is able to enhance its perception and react better to the events on the road beyond the reach of on-board sensors, facilitating its interactions with other traffic participants. As such, HD maps, typically provide 3D geometric and semantic information on static and physical parts of the world, including lane boundaries, intersections, crosswalks, parking spots, stop signs, and traffic lights. These static maps are computed offline, typically using the sensors of the self-driving car itself, although manual annotations or modifications are typically required.

While this paradigm has served to facilitate Autonomous Driving, such dependence on detailed prior maps is undesirable for global scale, as it requires a wealth of information about the geometry and traffic rules of every single road, and consequently, large volumes of data and storage space. Such maps are not only expensive to store on-board autonomous vehicles, but also very laborious to create and maintain (11). Furthermore, it is of extreme importance for the maps to reflect the latest state and components of roads, e.g. repainted road markings, blocked roads, construction sites, at all times. This further compounds the problem and renders this paradigm impractical.

Preprint. Under review.

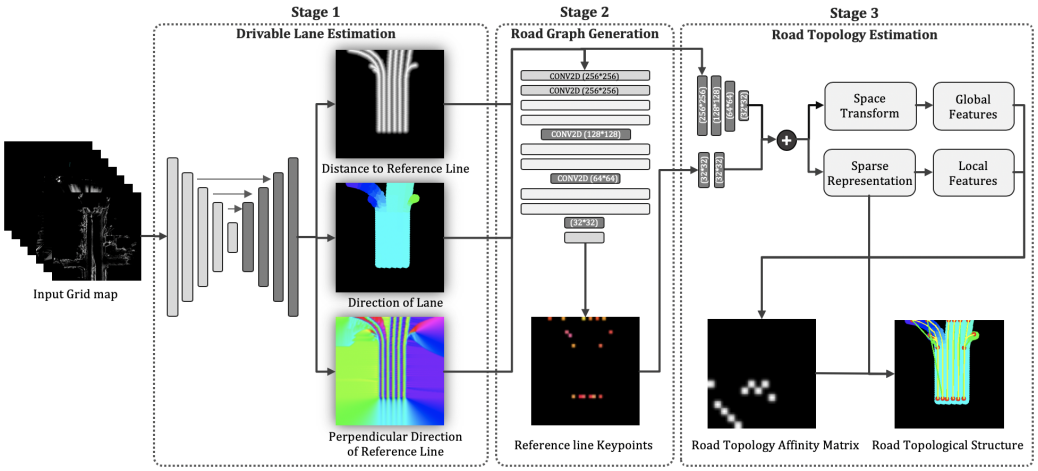


Figure 1: Overview of our road topology learning pipeline. Our model is a multi-stage, multi-task network trained to sequentially create the map components starting from the segmented lanes up to the complete road topology.

To enhance the practicality of a self-driving vehicle, improve its scalability both spatially, in terms of generalization to other Operational Design Domains, and temporally, in terms of always having up-to-date information, and obviate the need for human intervention, we herein propose a solution in order to transition from heavily relying on HD maps to adopting a map-less solution, thereby eliminating the need for the creation, manipulation, and maintenance of highly accurate maps. This solution, furthermore, renders the system more agile by adapting to road conditions and does not require precise localization. Fig. 1 depicts an overview of our proposed method.
We devise a learning methodology where the road condition and its components are learned in its current state, independent of the road complexity and number of the lanes, through a plethora of snapshots reflecting instantaneous environmental condition as well as the road structure and obstructions. These snapshots are formed in a hierarchical fashion; detecting the ego vehicle's drivable lanes at the low level and, subsequently, connecting this information to a global topological map for robust navigation. In the context of maps, the method produces a road network map, i.e., graph where edges are polylines corresponding to road segments, and vertices represent spatial coordinates of start, end, and fork points of each segment. This map dynamically varies with respect to ego vehicle's position and direction to contain only the relevant information for vehicle's planning.

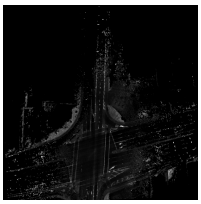
2 Related work

To reduce the dependency on maps, several techniques have been developed focusing mainly on predicting drivable routes, which can then be used for generating path proposals. Traditional methods establish connectivity by incorporating contextual priors such as color and texture information (12; 6) or road geometry (7; 13). Some approaches leverage the environment structure and rely on distinct features, such as lane markings and curbs (16; 25; 22; 4; 15). To extend these systems to complex urban environments and rural or undeveloped areas in the absence of clear or consistent lane markings, a class of approaches cast the problem as semantic segmentation to capture large spatial context (24; 17; 14; 10; 2), or estimate the lane geometry as well as the semantics of each lane (20; 11).

To enable navigation and drive in the absence of detailed maps based on a comprehensive understanding of the immediate environment while following simple higher level directions, some approaches include rough map priors as a baseline. In (23), a map-less driving framework was proposed that combined topometric maps with a LiDAR-based perception system for local navigation. The global topological localization and the corresponding graph search were performed based on the open street map (OSM) data, employing GPS data. A decision-making architecture was proposed in (1) that obtained a global route from OSM and generated driving corridors, which were then adapted and bounded using a vision-based lane detection algorithm and a probabilistic grid-based corridor


(a) Occupancy


(b) Ground markings


(c) LiDAR intensity

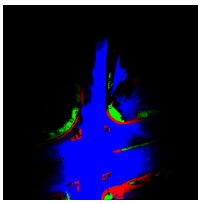

(d) Ground semantics

Figure 2: Input grid map layers

reduction. The drawback of such solutions, however, is that they cannot reason about roads absent in the initial coarse map.

A class of methods focus on road mapping from aerial images, in which pixel-level segmentation is typically combined with graph-based optimization (21; 19; 18). Such approaches, however, are usually adopted to merely provide local information about the presence of roads. The detailed information including the inter-connectivity of road segments is provided later through an error-prone post-processing stage. To eliminate such intermediate representation, some methods expand the road tree based on certain footprints or produce the road network directly from a CNN (3). Although such road topologies are very useful for routing purposes, in the context of autonomous driving, they do not provide the level of detail and accuracy required for safe localization and planning.

3 Hierarchical Road Topology Learning

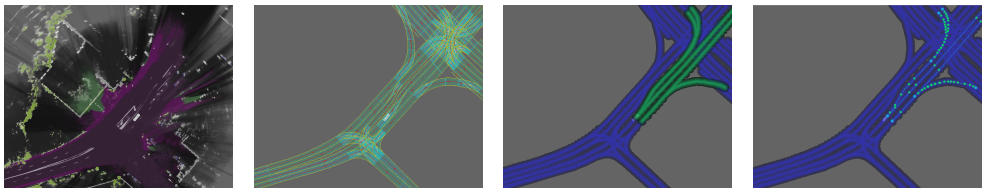
To facilitate map-less autonomous driving, we herein propose a hierarchical map-learning methodology, which does not suffer from the dependency on HD maps and enables the representation of road topology purely based on the sensory system aboard the vehicle. In this methodology, a road topology is defined as a set of keypoints and their relative connectivity, each of which representing a lane segment. As demonstrated in Fig. 1, our model is a multi-stage, multi-task network trained to sequentially create the map components starting from the drivable lanes up to the complete road topology.

3.1 Input Parameterization

We take advantage of different sensor measurements as input to our system in order to create a top-view representation of the vehicle's surroundings, demonstrated in an Occupancy Grid Map (OGM)-based (9) format. An OGM discretizes the space around the ego vehicle into equal cells, typically of the order of centimeters. Each cell contains a probabilistic representation of the occupancy state of its associated regions. The system builds the grid map through orthographic projection of the observations, encoding both geometric and semantic information in multiple layers. These information layers are estimated in real-time and in an online fashion according to various sensor modalities, including camera, stereo camera, LiDAR, and radar. The input, inherently, is a bird's eye view (BEV) of the vehicle's environment, referred to as $I \in \mathbb{R}^{H \times W \times C}$, where C is the number of grid map layers or channels. In our experiments, we use an encoding where the vehicle is always positioned at the bottom ¼ of the grid map.
Fig. 2 displays an example of the layers employed in our proposed approach. These sources of data are complementary. In the channels of this figure, shades of grey represent the probability of cells, with white being the maximum probability and black the minimum. Fig. 2c showcases the accumulated LiDAR intensity. Ground semantics provide a color-coded representation of the drivable road, sidewalk, and terrain.

3.2 Scope Definition

The underlying definition of the output of each stage is based on a space containing all the lanes autonomous vehicle can potentially drive in, following the intuitive definition of human's driving horizon. For the sake of simplicity, this definition applies to all lane segments topologically connected to the the lane segment ego vehicle is currently in, and are reachable by moving forward, right, or left.



(a) Input grid map (b) HD map (c) Ego drivable lanes (d) Road topology

Figure 3: Example of autonomous vehicle potential drivable's lanes and topology

To avoid cyclic graphs, particularly in the case of more complex structures, an orientation constraint is applied where each lane segment should also have an acute angle with the ego vehicle's yaw. Fig. 3 depicts the above potential drivable space in green overlayed on all the possible routes extracted from HD map depicted in blue. The corresponding road topology defined for the ego vehicle's potential drivable lanes is displayed in Fig. 3d.

3.3 Drivable Lane Estimation

The most basic information required for driving is the potential drivable lanes for the autonomous vehicle. The first stage of our architecture, hence, outlines the rough sketch of such areas in I, following the scope defined in 3.2. We adopt an encoder-decoder architecture (5), to aggregate multi-scale features and also preserve spatial information at each resolution.

The network outputs three representations of the drivable lanes with the same spatial resolution as I. The location of the lanes are encoded as a truncated inverse distance transform image $R \in \mathbb{R}^{H \times W \times 1}$ that labels each pixel in I with its relative distance to the closest reference line. To simplify the lane representation and customize the output for future planning purposes, reference line is chosen as the center line of a lane segment in the HD map. In contrast to predicting binary outputs at the lane level, the inverse distance transform of reference line encodes more information about the ideal location of the ego vehicle with respect to the road boundaries. The direction of each lane is represented as $D \in \mathbb{R}^{H \times W \times 3}$, an HSV color-encoded image in which the value of each pixel in the potential drivable lanes is defined based on the orientation of the closest reference line. Lastly, the network predicts the perpendicular direction map $P \in \mathbb{R}^{H \times W \times 3}$, encoding the normal directions to the closest reference line. These features are exploited in the later stages of the network to contribute to the hierarchical definition of the map towards road topology prediction (Fig. 1).

The parameters of the first-stage model are optimized by minimizing a weighted combination of the reference line detection loss l_R and the direction estimation losses l_D and l_P:

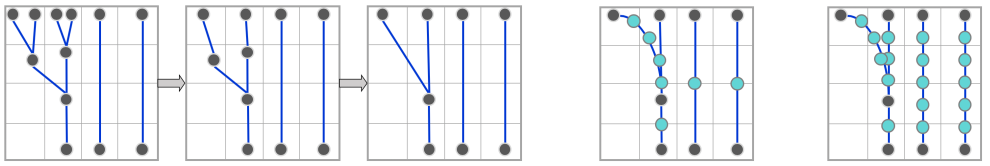
$$l_{stage1}(I) = l_R(I) + \lambda_1 l_D(I) + \lambda_2 l_P(I) \tag{1}$$

Both the inverse distance transform and direction map estimation tasks are treated as regression, and to reduce the effect of very large residual values the losses are defined based on cosine similarity and l1.

3.4 Road Graph Node Generation

Having an estimate of ego vehicle's drivable lanes and their corresponding features, the second stage serves as an approximation to the baseline of a graph representation of the road, $p(K_i|R_i, D_i, P_i)$, where K represents the corresponding keypoint grid. Towards this goal, a topology graph is characterized by a set of nodes and their connections.
The graph generated based on the driving horizon defined in 3.2 might still be very complex, for example, based on the grid map resolution some nodes might be located very close to each other. In the snapshot understanding of road topology, such level of complexity is not required and can be slightly simplified. Hence, to facilitate the learning process, the graph generated for ego vehicle's drivable lanes is pruned to keep only one keypoint per each cluster of nodes lying within 2 meter proximity of each other. Also, all the keypoints with a single child that are not a start point will be eliminated. An example of the above pruning process is depicted in Fig. 4a.

The pruned topology graph might be very sparse depending on the input grid map resolution. For instance, a 256×256 grid map would cover the surrounding of ego vehicle with $0.26\ m/pixel$



(a) Input graph topology pruning process (b) HD map points (c) Interpolated points

Figure 4: Reference line points definition

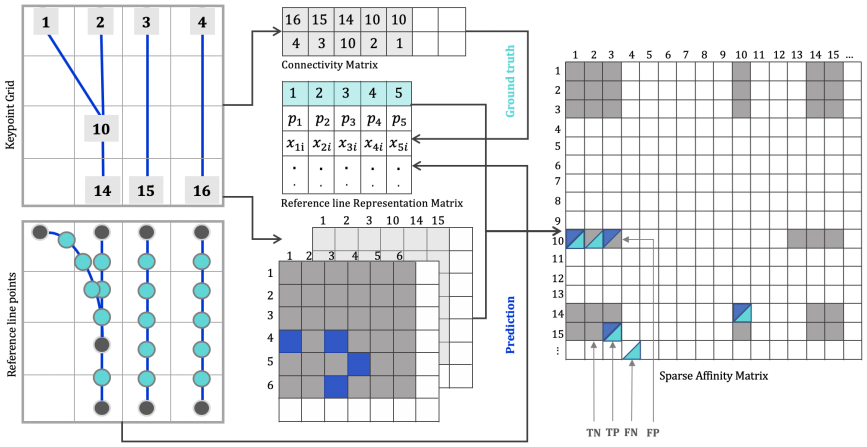


Figure 5: Graph affinity matrix prediction

(edge length). Thus, we treat the learning problem as combination of segmentation and regression in downsampled grid space. A lightweight CNN with a combination of 2D convolution layers and residual blocks is designed to predict the keypoint grid $K \in \mathbb{R}^{H' \times W' \times 3}$ given the outputs of previous stage. The output keypoint grid essentially encodes the probability of existence of a keypoint in each grid cell and its relative position within the cell. It is noteworthy that there is no distinction between different types of nodes in this process. Given the fact that the standard lane width in the United States is $\sim 3.65\, m$, to optimize the network and reduce the number of parameters and assure that the keypoints in neighboring lanes do not fall into the same grid cell, we downsize the keypoint grid to 32×32, reducing the resolution down to $2.08\, m/pixel$. As such, each point in the keypoint grid will cover a larger area making the existence prediction more robust to the potential noise accumulated from the previous stage. Consequently, relative positions are used to refine actual position of key points within each grid.

To train the network, the sum of losses over the pixelwise sigmoid cross entropy to estimate the likelihood of a cell containing a keypoint and mean square error (MSE) of the keypoints coordinates are minimized:

$$\begin{aligned} l_{stage2}(R, D, P) =& \lambda_{conf}(-\frac{1}{H'W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'}[p_{ij}\log\hat{p}_{ij} + (1-p_{ij})\log(1-\hat{p}_{ij})]) \\ &+ \lambda_{coord}(\frac{1}{N_k}\sum_{i=1}^{N_k}(c_i - \hat{c}_i)^2) \end{aligned} \tag{2}$$

where N_k represents the expected number of keypoints, p_{ij}^n denotes the ground truth map of the n^{th} keypoint at pixel location (i, j) and $p_{ij}^{\hat{n}}$ is the corresponding sigmoid output at the same location.

3.5 Road Topology and Reference Line Prediction

To complete the graph with the entailing connecting edges to create the road topological structure, the keypoint grid predicted in the previous stage is passed to the third stage of the network to estimate the graph affinity matrix, $p(C_i, L_i|K_i, R_i, D_i, P_i)$, where C and L denote the corresponding connection

and lane information of the grid. Hence, in the last stage, in addition to the connections–which is inherently a probability estimation of the existence of a reference line between two keypoints–an accurate localization of the reference line is predicted which is essential to the final task of map-less driving.

3.5.1 Reference Line Definition

For the raw topology generated in Fig. 4a, the definition of a reference line in the HD map might contain a varying number of points depending on the curvature of the underlying line (Fig. 4b). To simplify the regression task, taking into account the drivable lane orientation constraint explained in 3.2, the reference line is represented as a set of anchor points evenly distributed vertically between two given keypoints. This way, as the coordinates of keypoints have been estimated in the previous stage, and the direction of the reference lines and grid map are known, the y coordinates of the reference line points can be calculated and the prediction task is reduced to the regression of x coordinates only. Fig. 4c exhibits the above ground truth definition of reference line.

3.5.2 Graph Affinity Matrix Prediction

Theoretically, the keypoint grid can have up to $H' \times W'$ keypoints. In most of the existing road structures, however, this number has a lower limit $N_{k'}$. Hence, rather than having a sparse affinity matrix of size $(H' \times W')by(H' \times W')$ to represent all the graph connections, a dense representation of affinity matrix is introduced which keeps the indices of $N_{k'}$ keypoints and their connectivity information. For a given set of keypoints which are represented as a matrix of size $H' \times W'$ labeled with indices $1 \dots N_{k'}$, the dense representation would be in the form of two matrices entailing the original connected indices $C \in \mathbb{Z}^{2 \times N_{k'}}$ and corresponding reference line information $L \in \mathbb{Z}^{(N_{r_{max}}+2) \times N_{k'}}$, where $N_{r_{max}}$ denotes the maximum number of points to model a reference line segment. For that, the affinity matrix of predictions is initialized with the fully connected set of predicted keypoints. During the loss calculation process, both ground truth and predicted affinity matrix are mapped back to the sparse matrix. This transformation, also, would prevent accumulation of error from previous stage; i.e. in case of having false predictions, this step ensures that all the correctly predicted keypoints are indexed to the right position in the sparse matrix before calculating the loss. Fig. 5 depicts this process for both stage 3 ground truth and predictions.

To construct the road topology and refine the road structure estimated from earlier stages, the following loss function is defined for reference line classification and localization:

$$\begin{aligned} l_{stage3}(K, R, D, P) =& \lambda_{conf}(-\frac{1}{N_k^2}\sum_{i=1}^{N_k}\sum_{j=1}^{N_k}[p_{ij}\log\hat{p}_{ij} + (1-p_{ij})\log(1-\hat{p}_{ij})]) \\ &+ \lambda_{coord}(\frac{1}{N_c N_r}\sum_{i=1}^{N_c}\sum_{j=1}^{N_r}(l_{ij}-\hat{l}_{ij})^2) \end{aligned} \tag{3}$$

where p_{ij} denotes the likelihood of existence of connection between keypoints i and j, l represents the x coordinates of reference line anchor points, and N_c is the number of connections.

4 Experimental Evaluations

Data The experiments are done on datasets recorded from Santa Clara county, United States. The data has been collected from multiple passes of several autonomous vehicles equipped with the same set of sensors. To increase diversity of road structures, we also generated another set of data using the CARLA driving simulator (8), covering 4 different towns. The datasets consist of 25,000 frames, with 97 forks/intersections. We leave out one simulation town and one real-world route, with 800 and 1,000 frames respectively, for test of generalization. The recordings are categorized for train and validation with 92:8 ratio with same distribution of simulation and real-world data. The ground truth is generated from HD map of the corresponding area. Instead of doing manual labeling, we implemented a fully automatic map annotation tool that, given a pre-defined mission for the vehicle, extracts potential ego drivable lanes, directional features, and sparse road topology.

Table 1: Quantitative evaluation of different stages wrt. grid map resolution tested on real-world data.

Res.	Distance to ref. line		Direction of lane		Perp. direction of ref. line		Graph keypoints				Graph connectivity			
	MAE	SSIM	MAE	SSIM	MAE	SSIM	Prec.	Recall	F1	IOU	Prec.	Recall	F1	IOU
128	0.026	0.922	0.016	0.931	0.030	0.883	0.87	0.85	0.85	0.78	0.71	0.97	0.79	0.71
256	0.015	0.956	0.010	0.966	0.026	0.878	0.80	0.71	0.75	0.61	0.57	0.83	0.66	0.51

Table 2: Quantitative evaluation of different stages wrt. input grid map layers tested on real-world data. The experiments contain the following channels: (a) Ground semantics+Ground markings, (b) Ground semantics+LiDAR intensity and (c) All.

Input	Distance to ref. line		Direction of lane		Perp. direction of ref. line		Graph keypoints				Graph connectivity			
	MAE	SSIM	MAE	SSIM	MAE	SSIM	Prec.	Recall	F1	IOU	Prec.	Recall	F1	IOU
(a)	0.027	0.911	0.016	0.925	0.031	0.878	0.86	0.83	0.84	0.74	0.68	0.96	0.77	0.66
(b)	0.029	0.910	0.017	0.924	0.028	0.900	0.86	0.81	0.84	0.73	0.65	0.95	0.77	0.64
(c)	0.026	0.922	0.016	0.931	0.030	0.883	0.87	0.85	0.85	0.78	0.71	0.97	0.79	0.71

Experimental Setup For stage 1, the model was trained using Adam with a learning rate of 1e-4, with decay rate and step of 0.96 and 1e+5. In stage 2, the original input to the model was downsized by 8. Finally, in stage 3, for both maximum keypoint and maximum connections value of 16 was chosen based on the distribution of training data. As for the reference point definition, we chose a 4 pixel step. Since all the stages are differentiable, the network is optimized end-to-end to predict the parameters and trained for 200 epochs over the entire dataset.
We use Mean Absolute Error (MAE) and Structural Similarity Index (SSIM) as evaluation metrics for stage 1 tasks. For stage 2 and 3, we use precision, recall, IOU, and F1-score for evaluation. The average offset from predicted reference line points to ground truth is used to evaluate the performance of points position prediction.

Quantitative Analysis Due to the lack (to the best of our knowledge) of a public road topology benchmark, comparison with other existing approaches could not be undertaken. Nevertheless, we present our results based on evaluations with our ground truth test data. In our ablation study, we evaluate the importance of grid map resolution and input channels. Table 1 presents the detailed results of the effect of the resolution of grid map on each stage for two values of 128×128 and 256×256. As expected, the results of stage 2 and 3 drop by increasing the resolution, since the sensor data gets sparser in the further range and the keypoint prediction has more mis-detections and hence reduced graph connectivity performance. Therefore, for the rest of experiments we only report the values in 128 resolution. Table 2 outlines the fundamental importance of utilizing different sensor modalities represented as different input channels explained in 3.1.
We also evaluate the performance of keypoint and reference line point prediction at different scene difficulty levels depending on the complexity of topology in straight roads vs. intersections/forks. As can be seen in Table 3, even though the increasing number of keypoints affects the performance of system, it is still able to recover the underlying topology.
The ultimate goal of our method is to be able to navigate in the areas with no HD map to facilitate self driving at scale. To evaluate the generalization of our method, we evaluate the model on an unseen simulation town (Table 4). Although the performance has dropped as expected in comparison to Table 1 where we evaluated on the areas with similar topology, considering the fact that the input is not as noisy as real-world data and is coming from a different distribution, we still obtain very promising results.

Qualitative Results Fig. 6 depicts results obtained from inference of real-world data coming partially from similar areas in training set but driven in the opposite direction with increasing topology complexity. In row 1 and 2, we showcase how our model correctly infers the change of topology by spawning a new lane boundary at a fork. In rows 3 and 4, we demonstrate the behavior of our model at left turn with the option of u-turn in row 4. Fig. 7 displays some failure cases of the system which root in accumulated error of earlier stages or limited sensor information.

Running Time For the end-to-end inference of the system the pipeline is optimized to run at 20 Hz. The running time breakdown for stage 1, 2, and 3 are 7.4, 9.6, and 13.0 (ms), respectively, on a

Table 3: Quantitative evaluation of stage 2 and 3 wrt. scene complexity defined based on the total number of keypoints, evaluated on real-world data.

Complexity (# keypoints)	Graph keypoints				Graph connectivity				
	Prec.	Recall	F1	IOU	Prec.	Recall	F1	IOU	Avg. offset (cm)
Easy (1-5)	0.88	0.90	0.89	0.84	0.80	0.97	0.86	0.79	15
Medium (6-10)	0.85	0.77	0.80	0.70	0.61	0.95	0.69	0.60	26
Difficult (11-15)	0.72	0.62	0.68	0.51	0.40	0.78	0.53	0.40	42

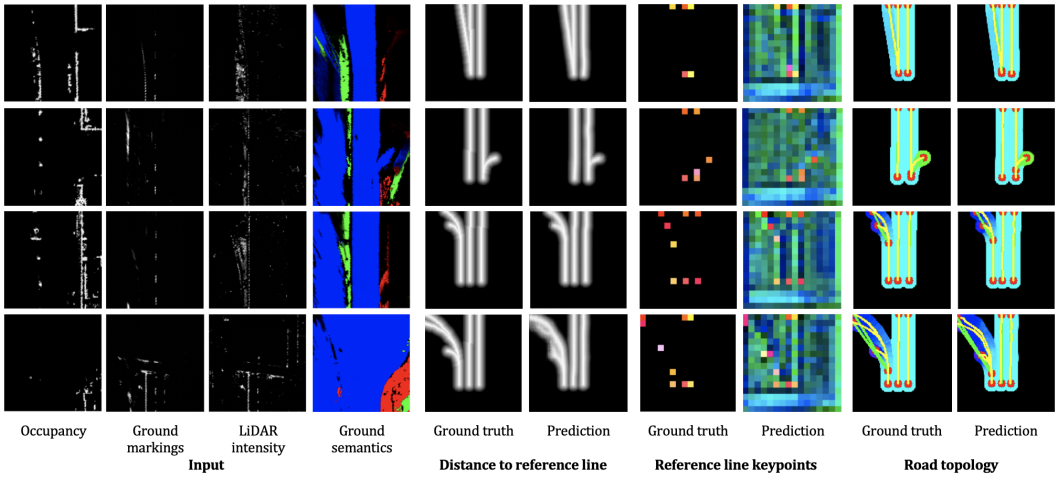


Figure 6: Qualitative results of road topology predictions on the real-world data

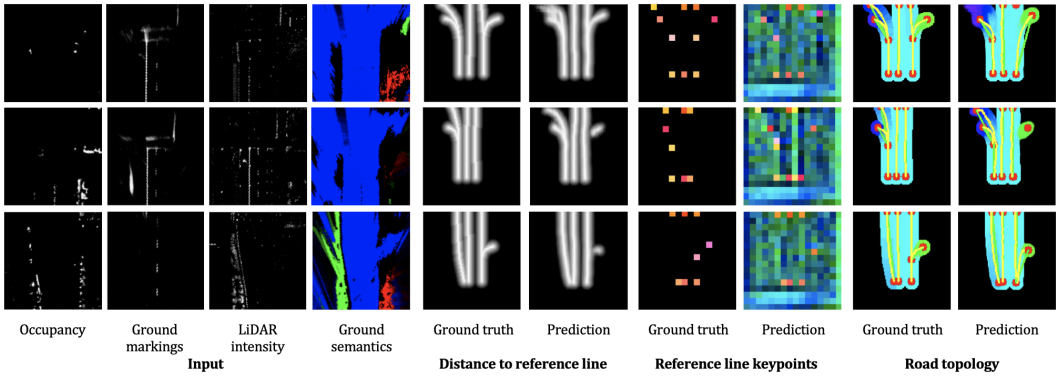


Figure 7: Failure cases

Table 4: Performance analysis of model generalization on simulation data.

Dataset	Distance to ref. line		Direction of lane		Perp. direction of ref. line		Graph keypoints				Graph connectivity			
	MAE	SSIM	MAE	SSIM	MAE	SSIM	Prec.	Recall	F1	IOU	Prec.	Recall	F1	IOU
Town 10	0.028	0.920	0.022	0.911	0.040	0.842	0.62	0.59	0.60	0.45	0.37	0.81	0.56	0.35

single Tesla P100 GPU. These values are comparable to the experiments with 256×256 grid map resolution taking 9.0, 8.1, and 15.4 (ms).

5 Conclusion

In this work, we have presented an approach that directly estimates structured road topology from an autonomous vehicle's on-board sensors. Our approach copes with difficult road structures, including one-way streets and large intersections with multiple lanes. Compared to existing approaches, which

often require some level of post-processing to improve graph connectivity, or even have a human in the loop, our approach estimates the road topology in real time and yields a structure that is directly usable by a behavior planning system, providing an affordable and scalable map solution with fast adaptability and low maintenance. In the future, to boost the generalization, we plan to invest more into procedural generation of vast amounts of topological scenarios in simulation to address corner cases and create more balanced and diverse datasets. Additionally, we will extend the method to include other semantic map elements, such as stop lines and traffic lights.

Broader Impact

To advance the practicality of self-driving technology, we need to seek approaches that enable scalability over time and towards various geographical locations, with minimum reliance on human intervention. For that, we need to transition from heavily relying on HD maps to adopting a map-less solution. The contribution is manifold; it supersedes the huge effort of creating, manipulating and maintaining highly accurate maps, and it pushes the system to adapt to road changes much more quickly and conveniently, and, hence, also can act as a map validator. Furthermore, in situations where the map fails or is not available, it can be used as an alternative solution to guide the system to a safe stopping point or contribute to the planner and continue the drive safely.

References

[1] Artuñedo, A., Godoy, J., Villagra, J.: A decision-making architecture for automated driving without detailed prior maps. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp. 1645–1652. IEEE (2019)

[2] Barnes, D., Maddern, W., Posner, I.: Find your own way: Weakly-supervised segmentation of path proposals for urban autonomy. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp. 203–210. IEEE (2017)

[3] Bastani, F., He, S., Abbar, S., Alizadeh, M., Balakrishnan, H., Chawla, S., Madden, S., DeWitt, D.: Roadtracer: Automatic extraction of road networks from aerial images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4720–4728 (2018)

[4] Beck, J., Stiller, C.: Non-parametric lane estimation in urban environments. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings. pp. 43–48. IEEE (2014)

[5] Chaurasia, A., Culurciello, E.: Linknet: Exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing (VCIP). pp. 1–4. IEEE (2017)

[6] Chiu, K.Y., Lin, S.F.: Lane detection using color-based segmentation. In: IEEE Proceedings. Intelligent Vehicles Symposium, 2005. pp. 706–711. IEEE (2005)

[7] Dickmanns, E.D., Mysliwetz, B.D.: Recursive 3-d road and relative ego-state recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence (2), 199–213 (1992)

[8] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: Carla: An open urban driving simulator. arXiv preprint arXiv:1711.03938 (2017)

[9] Elfes, A., et al.: Occupancy grids: A stochastic spatial representation for active robot perception. In: Proceedings of the Sixth Conference on Uncertainty in AI. vol. 2929, p. 6 (1990)

[10] He, B., Ai, R., Yan, Y., Lang, X.: Accurate and robust lane detection based on dual-view convolutional neutral network. In: 2016 IEEE Intelligent Vehicles Symposium (IV). pp. 1041–1046. IEEE (2016)

[11] Homayounfar, N., Ma, W.C., Kowshika Lakshmikanth, S., Urtasun, R.: Hierarchical recurrent attention networks for structured online maps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3417–3426 (2018)

[12] Kong, H., Audibert, J.Y., Ponce, J.: General road detection from a single image. IEEE Transactions on Image Processing **19**(8), 2211–2220 (2010)

[13] Kühnl, T., Kummert, F., Fritsch, J.: Spatial ray features for real-time ego-lane extraction. In: 2012 15th International IEEE Conference on Intelligent Transportation Systems. pp. 288–293. IEEE (2012)

[14] Lee, S., Kim, J., Shin Yoon, J., Shin, S., Bailo, O., Kim, N., Lee, T.H., Seok Hong, H., Han, S.H., So Kweon, I.: Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. In: Proceedings of the IEEE international conference on computer vision. pp. 1947–1955 (2017)

[15] Lee, U., Jung, J., Jung, S., Shim, D.H.: Development of a self-driving car that can handle the adverse weather. International journal of automotive technology **19**(1), 191–197 (2018)

[16] Li, J., Mei, X., Prokhorov, D., Tao, D.: Deep neural network for structural prediction and lane detection in traffic scene. IEEE transactions on neural networks and learning systems **28**(3), 690–703 (2016)

[17] Liang, J., Homayounfar, N., Ma, W.C., Wang, S., Urtasun, R.: Convolutional recurrent network for road boundary extraction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9512–9521 (2019)

[18] Marmanis, D., Wegner, J.D., Galliani, S., Schindler, K., Datcu, M., Stilla, U.: Semantic segmentation of aerial images with an ensemble of cnss. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016 **3**, 473–480 (2016)

[19] Máttyus, G., Luo, W., Urtasun, R.: Deeproadmapper: Extracting road topology from aerial images. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3438–3446 (2017)

[20] Meyer, A., Salscheider, N.O., Orzechowski, P.F., Stiller, C.: Deep semantic lane segmentation for mapless driving. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 869–875. IEEE (2018)

[21] Mnih, V., Hinton, G.E.: Learning to detect roads in high-resolution aerial images. In: European Conference on Computer Vision. pp. 210–223. Springer (2010)

[22] Neven, D., De Brabandere, B., Georgoulis, S., Proesmans, M., Van Gool, L.: Towards end-to-end lane detection: an instance segmentation approach. In: 2018 IEEE intelligent vehicles symposium (IV). pp. 286–291. IEEE (2018)

[23] Ort, T., Jatavallabhula, K., Banerjee, R., Gottipati, S.K., Bhatt, D., Gilitschenski, I., Paull, L., Rus, D.: Maplite: Autonomous intersection navigation without a detailed prior map. IEEE Robotics and Automation Letters (2019)

[24] Suleymanov, T., Amayo, P., Newman, P.: Inferring road boundaries through and despite traffic. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp. 409–416. IEEE (2018)

[25] Töpfer, D., Spehr, J., Effertz, J., Stiller, C.: Efficient road scene understanding for intelligent vehicles using compositional hierarchical models. IEEE Transactions on Intelligent Transportation Systems **16**(1), 441–451 (2014)