# Using Language and Road Manuals to Inform Map Reconstruction for Autonomous Driving

Akshar Tumu*, Henrik I. Christensen*, Marcell Vazquez-Chanlatte†, Chikao Tsuchiya†, and Dhaval Bhanderi†

*Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

{atumu, hichristensen}@ucsd.edu

†Nissan North America, Santa Clara, CA, USA

{Marcell.VazquezChanlatte, Chikao.Tsuchiya, Dhaval.Bhanderi}@nissan-usa.com

*Abstract*—Lane-topology prediction is a critical component of safe and reliable autonomous navigation. An accurate understanding of the road environment aids this task. We observe that this information often follows conventions encoded in natural language, through design codes that reflect the road structure and road names that capture the road functionality. We augment this information in a lightweight manner to SMERF, a map-prior-based online lane-topology prediction model, by combining structured road metadata from OSM maps and lane-width priors from Road design manuals with the road centerline encodings. We evaluate our method on two geo-diverse complex intersection scenarios. Our method shows improvement in both lane and traffic element detection and their association. We report results using four topology-aware metrics to comprehensively assess the model performance. These results demonstrate the ability of our approach to generalize and scale to diverse topologies and conditions.

## I. INTRODUCTION

Lane-topology information [18] is crucial for Autonomous vehicles to plan and navigate complex environments safely. Lane-topology includes the geometry of the lanes and the associations of the lanes between themselves and with other traffic elements (e.g., traffic lights and signs).

High Definition (HD) Maps are a reliable source for detailed lane-topology information. However, their high maintenance cost and limited geographical coverage make them difficult to scale for real-world deployment. Recent works such as SMERF [13] show that Standard Definition (SD) Maps, such as OpenStreetMap (OSM) [1], provide valuable road-level priors which, when combined with onboard sensor data, can be used to predict the detailed lane-topology of a road.

A key limitation of these works is that they generally only use road geometry and the road-type labels (e.g, high-way=residential) from OSM while ignoring the other semantic and geometric information available for each road. Lane-level cues, such as the number of lanes and lane width, are also overlooked, even though they could aid lane-topology prediction in scenarios with camera occlusion, unmarked roads, or complex road intersections.

Notably, the missing information mirrors how the built world is constructed and named through the design guidelines implicitly or explicitly encoded in language. For example, even if the road type is unlabeled in OSM, names such as

"Main Street", "Patricia Circle", or "Oregon Expressway" act as strong cues. Similarly, many locales require infrastructure to be built according to design codes written in natural language, diagrams, and tables. We hypothesize that capturing these cues can accelerate the deployment and adaptation of topology prediction systems to diverse geographies.

*Our key insight is to use language embeddings and Retrieval Augmented Generation* to provide these cues to an existing SD-Map informed HD-map reconstruction model. We evaluate our method on two geographically distinct locations and show improvement across four metrics capturing different aspects of lane-topology, highlighting our method's generalization and real-world applicability.

## II. RELATED WORK

### A. High Definition Maps

HD Maps consist of detailed lane-level information and annotations. They are used in various downstream Autonomous Driving tasks such as Trajectory Prediction ([23], [5]), Path Planning ([8], [3]), etc., since they have detailed lane-level and traffic element annotations. Popular sources of HD Maps include nuScenes [2], Argoverse 2 [19], Waymo Open Dataset [16], etc. However, it took several months to build and annotate these HD Maps. Another issue is that these HD Maps are restricted to a few geographical areas, limiting their global scalability. For example, nuScenes only has maps for Boston, USA, and Singapore, while Argoverse 2 only covers six cities in the United States.

### B. Online Lane-Topology Prediction

Online lane-topology prediction models ([10], [12], [11]) utilize the data collected by onboard sensors, such as cameras, to predict the lane-topology of the roads on the fly.

Recent works such as SMERF [13], TopoSD [21], and ImagineMap [9] introduce SD Map as a road-level prior to help online lane-topology models do long-horizon prediction or predict lane-topology in the presence of camera occlusion. While SD Maps don't contain lane-level details, they consist of the road geometry. SMERF, in particular, encodes road centerline geometry (as polylines) and road type labels from OSM using a Transformer module and then performs cross-attention between the BEV-representation of the multi-camera features and the map features. This is followed by a decoder head

(Deformable-DeTR [24]) which outputs the lane centerlines and the topology.

SD Maps like OSM provide the advantage of being available open-source and for a major portion of the world. OSM also provides road metadata such as the number of lanes, service classification of the road, road direction, and speed limit, etc., which has crucial semantic information captured in natural language. This information can help disambiguate road functionality and infer lane connectivity. One key detail that is very inconsistently tagged in OSM is the lane width of a road. Lane widths can help deduce the spatial layout of multi-lane roads even in the absence of clear lane markings. Diwanji et al. [4] show that the standardized road design manuals, such as the Highway Design Manual [14], provide the lane widths for different road types, which can be used for lane-topology prediction.

Another class of methods ([20], [22], [15]) learn temporal priors from past traversals to incrementally improve the map. However, they require more training, and the priors don't generalize geographically. In contrast, we inject structural priors from globally available SD Maps and road design manuals, making it scalable.

### C. Case Study: Real-world applicability of Online Mapping Methods

We discuss the real-world applicability of Online Mapping Methods by taking the example of the 'Trajectory Prediction' downstream task. Trajectory Prediction involves predicting the trajectory of a vehicle for a few seconds into the future. Trajectory prediction models rely on HD Maps for detailed lane geometry and semantics since it improves the accuracy and reliability of the predicted vehicle paths.

Two particular works - Gu et al. [6], Gu et al. [7] evaluate the performance of DenseTNT [5], a popular Trajectory Prediction model, when trained on HD Maps and maps generated by Online Lane-Topology Prediction models. The lane-topology prediction models evaluated are variants of Map-TRv2 [11] and StreamMapNet [22]. The models train on two seconds of annotated vehicle trajectories and ego-vehicle pose from nuScenes, along with either online-predicted or ground-truth nuScenes HD maps, to predict vehicle trajectories three seconds into the future. The evaluation metrics (minADE, minFDE, Miss Rate) analyze the alignment of the predicted trajectories with the real-world motion of the vehicles.

These works report that while the StreamMapNet-based model performs very closely to the HD Map-based model, the model that uses the lane-centerline-predicting variant of MapTRv2 even outperforms the HD Map-based model. Such results demonstrate the viability of online lane-topology prediction methods as alternatives to HD Maps for downstream Autonomous Vehicles tasks.

### III. METHODOLOGY

#### A. Obtaining OSM Metadata

Each road in an OSM Map has multiple annotated metadata fields such as 'name', 'geometry', 'highway', 'lanes', etc. The map embedding generated by SMERF's map encoder consists of the 'geometry' (road polylines) and 'highway' fields. Now, we extract all other fields annotated for each road apart from 'osmid', 'name', and the fields already present in the graph embeddings. While we exclude the 'name' field, we extract the road suffix (Avenue, Street, etc.) from it, if present. Some of the fields we capture are - 'service', 'oneway', 'number of lanes', 'maxspeed', etc. We then concatenate the fields into a string and embed them using OpenAI's text-embedding-3-small model[1].

#### B. Extracting Lane Widths

To extract lane widths from the Road design manual, we follow the Retrieval Augmented Generation pipeline setup by Diwanji et al. [4] with the Llama3.3 LLM [17]. We first start by splitting the Highway Design Manual into chunks and embedding them using an embedding model. These embeddings are stored in a vector database.

```
basic_road_info + "\n Given the above basic road information
in JSON format, return the lane width in the following JSON
output format: \n\
{\n\
    'lane_width': type=int; lane width in feet for the given
road type,\n\
}\n\
Please provide the output strictly in JSON format with no
comments and no explanation"
```

Fig. 1.  RAG Prompt

Figure 1 shows the prompt that we query the LLM with. For each road, we also pass its 'basic_road_info' consisting of the fields extracted as part of the OSM metadata. When queried for each road in a map individually, the LLM outputs the corresponding lane widths by retrieving and reasoning upon relevant parts of the road design manual. We embed the lane widths using OpenAI's text-embedding-3-small model.

#### C. Passing Embeddings to SMERF

The embedded OSM metadata and Lane Widths are added since they are of the same dimension. The combined text embedding is passed to SMERF's BEV Constructor module, which also takes the Map embeddings as input. The text embeddings are first transformed into the embedding space of the map embeddings using a 2-layer MLP. Then, the Text and Map embeddings are added. Mathematically, the addition is defined as follows:

$$e_p = G(p) + MLP(t_p) \tag{1}$$

Where $p$ stands for the polyline of a road, $G(p)$ stands for the graph embedding of that polyline, and $t_p$ stands for the text embedding of that road.

The rest of the pipeline is the same as SMERF, which involves cross-attention between the camera BEV features and the combined embeddings, followed by a decoder to generate the vectorized HD Map output.

---

[1]https://platform.openai.com/docs/models/text-embedding-3-small

## IV. Experimental Setting and Evaluation

### A. Dataset

We select two complex road scenarios from the OpenLane-V2 dataset, which are from different cities in the United States, for training and evaluating our model. The two scenarios have 64 frames in total. Figure 2 illustrates the Front-view and BEV Lane-topologies of the scenarios with annotated lanes and traffic elements. The first scenario is from Detroit, which involves an upcoming intersection that is occluded in the camera. The second scenario is from Pittsburgh, which involves a complex multi-lane merging scenario.

### B. Evaluation Metrics

We select the following evaluation metrics from the OpenLane-V2 dataset:

- $DET_l$: Average precision of 3D lane centerline detection using Fréchet distance.
- $DET_t$: Average precision of traffic element detection across different detected attributes of the traffic elements.
- $TOP_{ll}$: Accuracy of the predicted connectivity among the detected lane centerlines.
- $TOP_{lt}$: Accuracy of the predicted association between lane centerlines and traffic elements.

We also include a fifth metric, OpenLane-V2 Score (OLS), which is the weighted average of all four metrics.

These metrics evaluate the spatial accuracy of the models and their ability to comprehend traffic elements and associate them with the lanes, thus covering all lane-topology aspects for any complex road scenario in the world.

### C. Training and Fusion Configurations

*1) Model Training Configurations:* We first start by training the original SMERF model for 100 epochs. Then we further train the model for a second round of 100 epochs by including the OSM Metadata embeddings and Lane-width embeddings. We test the following three training configurations:

- F0: We freeze the entire model, except the 2-layer MLP for transforming the text embeddings, for the second round of training.
- F1: We only freeze the multi-camera BEV encoder and the map encoder for the second round of training.
- NF: No part of the model is frozen for the second round of training.

*2) Embedding Fusion Configurations:* Equation 1 defines the fusion of the graph and text embeddings for a lane. Our first fusion strategy consists of only the OSM Metadata embedding as the text embedding ($t_p = o_p$ where $o_p$ stands for the OSM metadata embedding).

Our second fusion strategy consists of a weighted addition scheme defined below:

$$e_p = G(p) + \lambda \cdot MLP(t_p) \quad \text{where} \quad t_p = o_p \quad (2)$$

We set $\lambda$ to be a learnable parameter by the network.

Our third fusion strategy consists of fusing the OSM Metadata embeddings with the Lane-width embeddings obtained from the RAG-LLM system. Mathematically, we define the fusion as follows:

$$t_p = o_p + l_p \quad (3)$$

where $l_p$ stands for the Lane-width embedding.

## V. Results

Table I shows the results over the two selected scenarios for the baseline (SMERF) and our different model configurations. For the baseline, we use the original SMERF model trained for 100 epochs as our loading checkpoint and train the model for 100 more epochs.

We can observe that our best configurations outperform the baseline for all metrics. Among F0, F1, and NF training configurations, NF has the best results for four of the five metrics, with 'NF + $\lambda$' variant performing the best for three of the five metrics. However, our 'F1 + RAG' variant gives the best lane centerline detection performance. F0 variants underperform since all other variants have a lot more trainable parameters.

Among all four metrics, we see the maximum improvement in the $TOP_{ll}$ metric by 16%, followed by $TOP_{lt}$ by 7%. This is because the OSM Metadata embeddings consist of information about the functional type and semantic properties of the lanes, which aid lane connectivity and association between lanes and traffic elements.

Figure 3 shows the lane-topology predictions in BEV view by the 'NF + $\lambda$' variant of our approach. We can observe that for the Pittsburgh map, there is a slight deviation between the ground truth and predicted lanes, especially at the joining points of the lanes. In contrast, the predictions closely match the ground truth lanes in Detroit's structured map. This highlights the general challenges in mapping curves and intersections compared to structured lanes.

## VI. Real-World Deployability and Generalization Discussion

The real-world deployability of the online lane-topology prediction methods has been discussed in Section II-C, where the models relying on these techniques match or even outperform the models that use manually annotated HD Maps for a downstream Autonomous Vehicles task.

The qualitative and quantitative results for our approach suggest that integrating multi-camera features with an SD-Map prior and the linguistic priors provided by the SD Maps for each road helps generalize online lane-topology prediction (or, online HD Map building) to even those geometric locations in the world for which manual HD Maps haven't been designed. The evaluation setup, consisting of scenarios from different geometric regions, confirms the generalization ability of our method.

## VII. Conclusion

We introduce a lightweight, deployable augmentation to SMERF, which integrates semantic and geometric metadata presented in natural language in SD Maps and lane-width priors extracted from road design manuals using RAG to

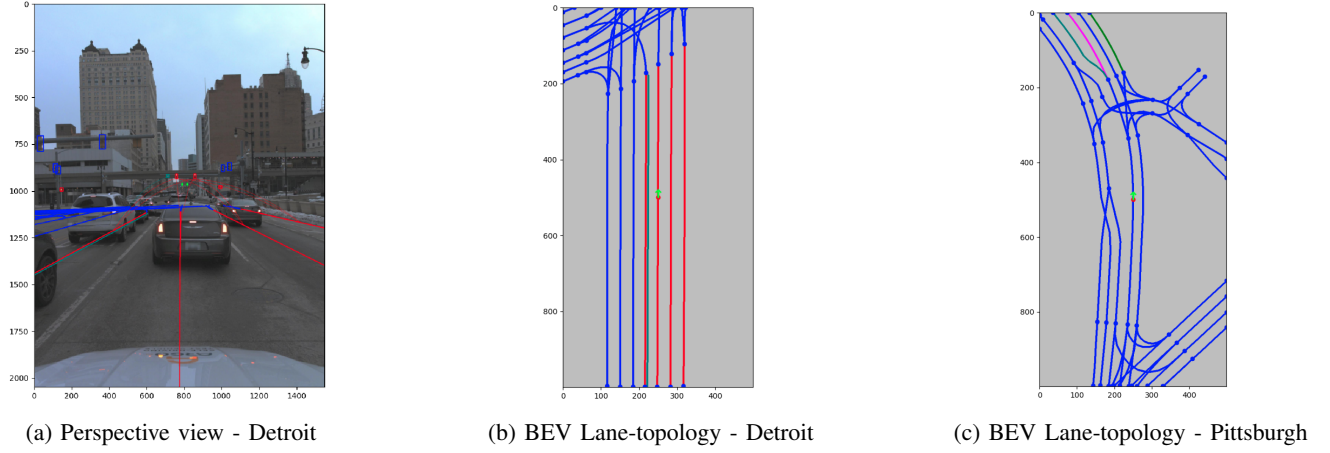(a) Perspective view - Detroit      (b) BEV Lane-topology - Detroit      (c) BEV Lane-topology - Pittsburgh

Fig. 2. Comparison of the front-view image and corresponding Ground-Truth BEV lane-topologies. Lane colors in the BEV Map indicate intended actions at intersections, such as left turn (blue), right turn (green), etc. The red dot and the green arrow jointly indicate the ego-pose of the vehicle.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MODEL AND FUSION CONFIGURATIONS. '$\lambda$' CONFIGURATION STANDS FOR THE EMBEDDING FUSION
DEFINED IN EQUATION 2. 'RAG' CONFIGURATION STANDS FOR THE COMBINED TEXT EMBEDDING DEFINED IN EQUATION 3.

| Model | $\lambda$ | #Trainable Params | $DET_l$ | $DET_t$ | $TOP_{ll}$ | $TOP_{lt}$ | OLS |
|-------|-----------|-------------------|---------|---------|------------|------------|-----|
| SMERF [13] | - | 39.62M | 0.5364 | <u>0.9734</u> | 0.0595 | 0.3645 | 0.5894 |
| F0 | - | 132K | 0.4134 | 0.9392 | 0.0141 | 0.2125 | 0.4831 |
| F0 + $\lambda$ | 0.869 | 132K | 0.4138 | 0.9392 | 0.0137 | 0.2122 | 0.4827 |
| F0 + RAG | - | 132K | 0.414 | 0.9392 | 0.0141 | 0.2123 | 0.4831 |
| F1 | - | 26.28M | 0.5267 | 0.9703 | 0.0538 | 0.3519 | 0.5806 |
| F1 + $\lambda$ | 0.869 | 26.28M | <u>0.5391</u> | 0.9726 | 0.0583 | 0.3462 | 0.5854 |
| F1 + RAG | - | 26.28M | **0.5468** | 0.9725 | 0.0628 | 0.3692 | <u>0.5944</u> |
| NF | - | 39.75M | 0.5327 | 0.9652 | **0.0691** | 0.3452 | 0.5871 |
| NF + $\lambda$ | 0.869 | 39.75M | <u>0.5391</u> | **0.9745** | <u>0.0674</u> | **0.3905** | **0.5995** |
| NF + RAG | - | 39.75M | 0.5376 | 0.9671 | 0.0669 | <u>0.3748</u> | 0.5939 |



(a) Pittsburgh      (b) Detroit
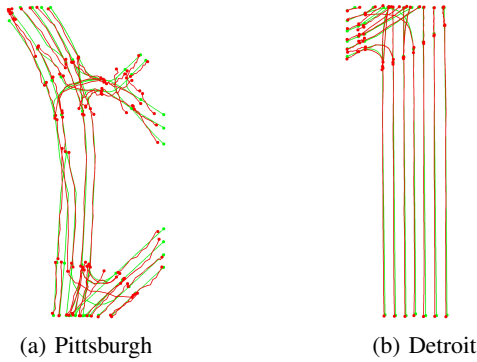
Fig. 3. Lane-topology predictions in the BEV view for the two scenarios. The green lines correspond to the Ground Truth lanes, and the red lines correspond to the Predicted lanes. The dots correspond to the joining points between lanes.

improve online lane-topology prediction. By encoding the said information with an LLM and combining it with the road polyline encodings from SD Maps, our model enhances the SD Maps without modifying the model architecture or using HD Maps.

Our evaluation spans two geographically diverse real-world intersection scenarios from the OpenLane-V2 dataset. Through carefully-chosen topology-aware metrics, we show that our method achieves consistent improvement in all aspects of lane-topology prediction for the chosen complex scenarios. These results demonstrate the practical viability of integrating lightweight, globally available priors into robot perception systems. By operating without the expensive and geometrically constrained HD Maps and handling scenarios with camera occlusion, our method shows scalability and generalizability across geographical regions and conditions.

## REFERENCES

[1] Jonathan Bennett. *OpenStreetMap*. Packt Publishing Ltd, 2010. URL https://www.researchgate.net/publication/262206447_OpenStreetMap.
[2] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A Multimodal Dataset for Autonomous Driving. In *Proceedings of the*

*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Caesar_nuScenes_A_Multimodal_Dataset_for_Autonomous_Driving_CVPR_2020_paper.html.

[3] Alejandro Diaz-Diaz, Manuel Ocaña, Ángel Llamazares, Carlos Gómez-Huélamo, Pedro Revenga, and Luis M. Bergasa. HD maps: Exploiting OpenDRIVE potential for Path Planning and Map Monitoring. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, page 1211–1217. IEEE Press, 2022. doi: 10.1109/IV51971.2022.9827297. URL https://doi.org/10.1109/IV51971.2022.9827297.

[4] Hitvarth Diwanji, Jing-Yan Liao, Akshar Tumu, Henrik I. Christensen, Marcell Vazquez-Chanlatte, and Chikao Tsuchiya. SD++: Enhancing Standard Definition Maps by Incorporating Road Knowledge using LLMs. In *Intelligent Vehicles - 2025*, June 2025. URL https://arxiv.org/abs/2502.02773.

[5] Junru Gu, Chen Sun, and Hang Zhao. DenseTNT: End-to-End Trajectory Prediction From Dense Goal Sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15303–15312, October 2021.

[6] Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, and Boris Ivanovic. Producing and leveraging online map uncertainty in trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14521–14530, June 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Gu_Producing_and_Leveraging_Online_Map_Uncertainty_in_Trajectory_Prediction_CVPR_2024_paper.html.

[7] Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, and Boris Ivanovic. Accelerating online mapping and behavior prediction via direct bev feature attention. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part LXXXI*, page 412–428, Berlin, Heidelberg, 2024. Springer-Verlag. ISBN 978-3-031-73003-0. doi: 10.1007/978-3-031-73004-7_24. URL https://doi.org/10.1007/978-3-031-73004-7_24.

[8] Xiaomin Guo, Yongxing Cao, Jian Zhou, Yuanxian Huang, and Bijun Li. HDM-RRT: A Fast HD-Map-Guided Motion Planning Algorithm for Autonomous Driving in the Campus Environment. *Remote Sensing*, 15(2), 2023. ISSN 2072-4292. doi: 10.3390/rs15020487. URL https://www.mdpi.com/2072-4292/15/2/487.

[9] Yishen Ji, Zhiqi Li, and Tong Lu. ImagineMap: Enhanced HD Map Construction with SD Maps, 2024. URL https://arxiv.org/abs/2412.16938.

[10] Qi Li, Yue Wang, Yilun Wang, and Hang Zhao. HDMapNet: An Online HD Map Construction and Evaluation Framework. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 4628–4634, 2022. doi: 10.1109/ICRA46639.2022.9812383. URL https://doi.org/10.1109/ICRA46639.2022.9812383.

[11] Bencheng Liao, Shaoyu Chen, Yunchi Zhang, Bo Jiang, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Maptrv2: An end-to-end framework for online vectorized hd map construction. *International Journal of Computer Vision*, pages 1–23, 2024. URL https://doi.org/10.1007/s11263-024-02235-z.

[12] Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. VectorMapNet: end-to-end vectorized HD map learning. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023. URL https://proceedings.mlr.press/v202/liu23ax.html.

[13] Katie Z Luo, Xinshuo Weng, Yan Wang, Shuang Wu, Jie Li, Kilian Q Weinberger, Yue Wang, and Marco Pavone. Augmenting Lane Perception and Topology Understanding with Standard Definition Navigation Maps. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4029–4035, 2024. doi: 10.1109/ICRA57147.2024.10610276. URL https://doi.org/10.1109/ICRA57147.2024.10610276.

[14] California. Department of Transportation. California Department of Transportation (Caltrans): Highway Design Manual: U.S. Customary Units: Seventh Edition. https://rosap.ntl.bts.gov/view/dot/15907, 2020.

[15] Nan Peng, Xun Zhou, Mingming Wang, Guisong Chen, and Wenqi Xu. Uni-PrevPredMap: Extending PrevPredMap to a Unified Framework of Prior-Informed Modeling for Online Vectorized HD Map Construction, 2025. URL https://arxiv.org/abs/2504.06647.

[16] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Sun_Scalability_in_Perception_for_Autonomous_Driving_Waymo_Open_Dataset_CVPR_2020_paper.html.

[17] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. LLaMA: Open and Efficient Foundation Language Models, 2023. URL https://arxiv.org/abs/2302.13971.

[18] Huijie Wang, Tianyu Li, Yang Li, Li Chen, Chonghao Sima, Zhenbo Liu, Bangjun Wang, Peijin Jia, Yuting Wang, Shengyin Jiang, Feng Wen, Hang Xu, Ping Luo, Junchi Yan, Wei Zhang, and Hongyang Li. OpenLane-V2: A Topology Reasoning Benchmark for Unified 3D HD Mapping. In A. Oh,

T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 18873–18884. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/3c0a4c8c236144f1b99b7e1531debe9c-Paper-Datasets_and_Benchmarks.pdf.

[19] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting, 2023. URL https://arxiv.org/abs/2301.00493.

[20] Xuan Xiong, Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. Neural Map Prior for Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17535–17544, June 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Xiong_Neural_Map_Prior_for_Autonomous_Driving_CVPR_2023_paper.html.

[21] Sen Yang, Minyue Jiang, Ziwei Fan, Xiaolu Xie, Xiao Tan, Yingying Li, Errui Ding, Liang Wang, and Jingdong Wang. TopoSD: Topology-Enhanced Lane Segment Perception with SDMap Prior, 2024. URL https://arxiv.org/abs/2411.14751.

[22] Tianyuan Yuan, Yicheng Liu, Yue Wang, Yilun Wang, and Hang Zhao. StreamMapNet: Streaming Mapping Network for Vectorized Online HD Map Construction. In *WACV*, pages 7341–7350, 2024. URL https://doi.org/10.1109/WACV57701.2024.00719.

[23] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu. HiVT: Hierarchical Vector Transformer for Multi-Agent Motion Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8823–8833, June 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Zhou_HiVT_Hierarchical_Vector_Transformer_for_Multi-Agent_Motion_Prediction_CVPR_2022_paper.html.

[24] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection, 2021. URL https://arxiv.org/abs/2010.04159.