Fine-Grained Representation for Lane Topology Reasoning

Guoqing Xu[1,2,*], **Yiheng Li**[1,2,*], **Yang Yang** [1,2,†]

[1] MAIS, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
xuguoqing2024@ia.ac.cn, liyiheng2024@ia.ac.cn, yang.yang@nlpr.ia.ac.cn

Abstract

Precise modeling of lane topology is essential for autonomous driving, as it directly impacts navigation and control decisions. Existing methods typically represent each lane with a single query and infer topological connectivity based on the similarity between lane queries. However, this kind of design struggles to accurately model complex lane structures, leading to unreliable topology prediction. In this view, we propose a Fine-Grained lane topology reasoning framework (TopoFG). It divides the procedure from bird's-eye-view (BEV) features to topology prediction via fine-grained queries into three phases, i.e., Hierarchical Prior Extractor (HPE), Region-Focused Decoder (RFD), and Robust Boundary-Point Topology Reasoning (RBTR). Specifically, HPE extracts global spatial priors from the BEV mask and local sequential priors from in-lane keypoint sequences to guide subsequent fine-grained query modeling. RFD constructs fine-grained queries by integrating the spatial and sequential priors. It then samples reference points in RoI regions of the mask and applies cross-attention with BEV features to refine the query representations of each lane. RBTR models lane connectivity based on boundary-point query features and further employs a topological denoising strategy to reduce matching ambiguity. By integrating spatial and sequential priors into fine-grained queries and applying a denoising strategy to boundary-point topology reasoning, our method precisely models complex lane structures and delivers trustworthy topology predictions. Extensive experiments on the OpenLane-V2 benchmark demonstrate that TopoFG achieves new state-of-the-art performance, with an OLS of 48.0% on *subset_A* and 45.4% on *subset_B*.

Code — https://github.com/GXmmm18/TopoFG

Introduction

In autonomous driving, environmental perception is fundamental, covering tasks such as obstacle (Li, Yang, and Lei 2025a,b) and lane detection (Wang et al. 2023b). Building on lane perception, lane topology reasoning has recently gained attention for its ability to capture structural relationships between lanes and enable more reliable navigation in complex environments (He et al. 2024). This not only involves the detection of lane centerlines and traffic elements, but also the reasoning of topological relationships such as lane connectivity and their associations with traffic elements. Traditional methods often rely on cumbersome hand-crafted rules and post-processing to obtain relationships. Recently, many methods tend to focus on using end-to-end approaches to unify the detection of lanes and traffic elements with the computation of their relationships into a single task, referred to as topology reasoning (Li et al. 2023).

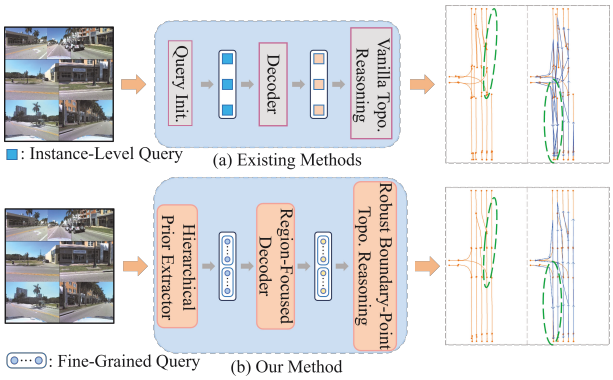

Figure 1: Comparison between existing methods and our method for lane topology reasoning. (a) **Existing Methods**: Using instance-level queries with coarse lane modeling and holistic topology reasoning, which may lead to incorrect predictions in complex scenes. (b) **Our Method**: Adopting fine-grained queries and boundary-point based topology reasoning for improved lane detection and topology prediction. The blue arrow on the right represents the lane topology connection, and the green dashed regions highlight the more reliable topology reasoning of our method.

To implement end-to-end topology reasoning, recent studies have proposed unified frameworks that jointly detect lane centerlines and traffic elements while predicting their topological relationships. A common paradigm adopted by existing methods is to first detect individual lane centerlines and traffic elements, and then infer their topological relationships based on learned representations. Approaches such as TopoNet (Li et al. 2023) and TopoLogic (Fu et al. 2024)

*These authors contributed equally.
†Corresponding author.

exemplify this framework, modeling each lane as a holistic entity and reasoning about their relationships at the instance level. LaneSegNet(Li et al. 2024b) enhances the semantic expressiveness of lane representations by modeling them as a series of lane segments. Furthermore, it employs a lane attention mechanism to model lane segments.

Despite achieving promising performance, the existing works share a common limitation. First, these approaches typically treat the entire centerline as a single unit and rely on a single query to predict all keypoints along the lane. This instance-level query modeling approach often lacks sufficient expressiveness when dealing with lanes that exhibit complex shapes or significant local variations. Furthermore, these approaches determine lane connectivity by directly computing the similarity between instance-level queries of different lanes. In practice, two lanes that are truly connected may only intersect locally at their boundary-points, making their overall similarity less obvious and thus harder to detect. For example, consider three lanes where lane a ends ahead of two parallel lanes, b and c, with similar geometries. While lane a should in fact lead to lane b, instance-level query modeling may yield similar representations for b and c, leading the model to incorrectly predict a connection from a to c.

As shown in Figure 1, we propose TopoFG. In this approach, each lane is modeled as a sequence of fine-grained queries, with each query explicitly corresponding to a specific location along the lane. This fine-grained representation allows the model to better capture local geometric variations and structural details. Based on this representation, we perform topology reasoning using boundary-point queries and introduce a denoising strategy to improve prediction reliability. We design the following three modules for the overall topology reasoning task to fully boost performance. First, we introduce a Hierarchical Prior Extractor that derives global spatial priors from BEV masks and captures local sequential priors from in-lane sequences. Next, the Region-Focused Decoder integrates spatial and sequential priors into fine-grained queries and samples reference points from RoI regions of the mask, guiding the queries to focus on lane-relevant regions. Finally, a Robust Boundary-Point Topology Reasoning module is introduced, which utilizes boundary-point queries, i.e., the start and end points of lanes, to infer topological connections and incorporates a denoising strategy to reduce ambiguity during training.

We conduct experiments on the widely used OpenLane-V2 (Wang et al. 2023a) dataset, and our method achieves superior performance by achieving 48.0% OLS on *subset_A* and 45.4% OLS on *subset_B*. Additionally, comprehensive ablation studies further demonstrate the effectiveness of each proposed module. Contributions are as follows:

- We adopt fine-grained queries to represent a single lane, thereby enhancing the ability to model complex structures and improving the accuracy of topology prediction.
- We propose a Hierarchical Prior Extractor to help the model acquire prior information, a Region-Focused Decoder to reduce the interference from irrelevant information, and a Robust Boundary-Point Topology Reasoning to enhance the robustness of the reasoning results.
- We confirm the efficacy of the proposed method by achieving state-of-the-art performance on the OpenLane-V2 benchmark and superior accuracy and robustness in complex driving scenarios.

Related Work

Lane Detection

The primary objective of lane detection is to generate reliable geometric representations of lane boundaries or centerlines within the ego vehicle's surrounding environment, serving as critical input for downstream tasks like path planning and control. Lane detection methods can generally be classified into 2D-based and 3D-based approaches. 2D lane detection typically includes segmentation-based (Lee et al. 2017; Garnett et al. 2019; Abualsaud et al. 2021) and detection-based methods (Tabelini et al. 2021b; Han et al. 2022; Wu et al. 2022). Segmentation-based approaches perform pixel-wise prediction and distinguish lane instances using masks, grids, or keypoints, while detection-based approaches localize and classify lane instances in a single stage. These methods primarily operate in the image domain and are generally suited for relatively simple road scenarios (Pan et al. 2018; Zhang et al. 2020; Qu et al. 2021). 3D lane detection incorporates depth cues or geometric constraints to better reconstruct complex lane shapes (Liu et al. 2022a). Existing methods can be grouped into BEV-based and BEV-free approaches. BEV-based methods (Garnett et al. 2019; Pittner, Janai, and Condurache 2024) transform image features into BEV space to integrate height information for accurate localization, while BEV-free methods (Tabelini et al. 2021a; Luo et al. 2023) either predict depth to project 2D lanes into 3D space or directly model lanes in 3D space. Although significant progress has been made in the aforementioned methods, they still lack the ability to model road structures and connectivity, limiting their applicability in complex driving scenarios.

Vectorized HD Map

The vectorized High-definition (HD) map method builds high-precision maps in real time using sensor data, providing a more flexible alternative to traditional HD maps. HDMapNet (Li et al. 2022) first generates BEV semantic segmentations through BEV feature extraction and a semantic segmentation model, followed by a time-consuming post-processing step to produce vectorized HD maps. VectorMapNet (Liu et al. 2023) adopts the DETR (Carion et al. 2020) architecture, using transformer attention (Vaswani et al. 2017) to decode the scene and directly predict ordered point sequences of map instances. MapTR (Liao et al. 2023) represents map instances as ordered point sets and encodes them via hierarchical queries in a transformer decoder. It further introduces a permutation-equivalent modeling approach to resolve the order ambiguity in the matching stage. MapTRv2 (Liao et al. 2024b) extends this capability by incorporating deep supervision, enhanced feature encoding, and auxiliary losses to improve overall performance. Mask2Map (Choi et al. 2024) first produces a rasterized map

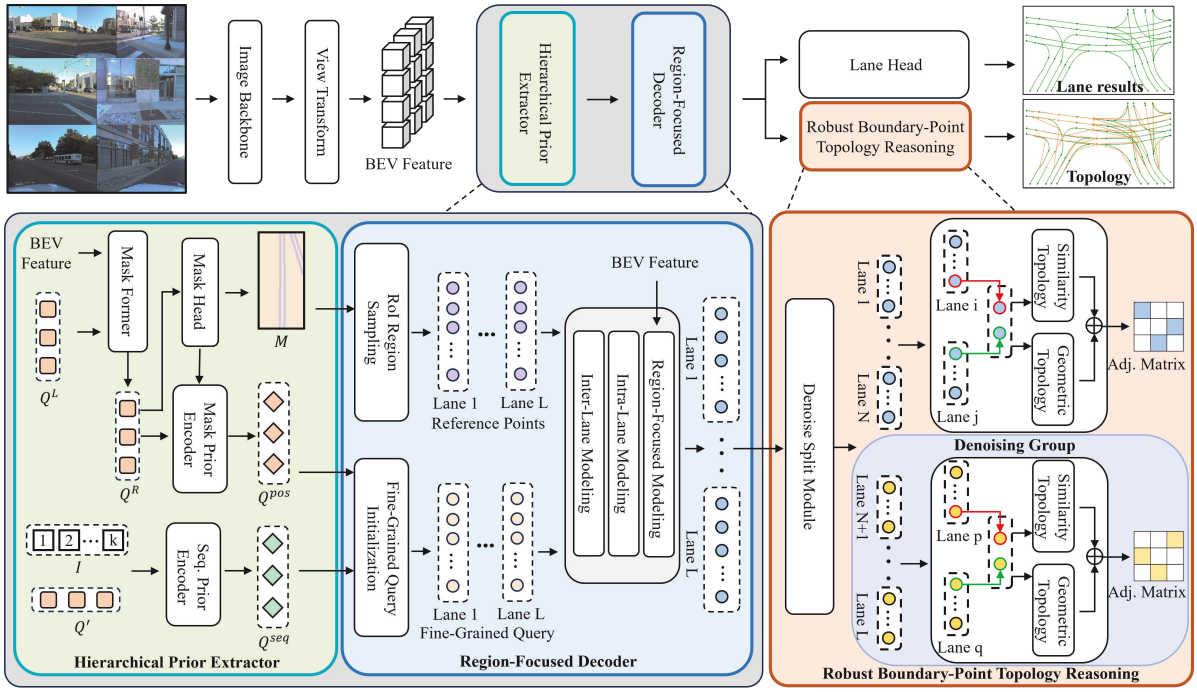

Figure 2: Overview of the TopoFG framework. The framework consists of three modules. First, the **Hierarchical Prior Extractor**, which extracts both spatial and local priors. Second, the **Region-Focused Decoder**, which enhances local geometric modeling by focusing on key lane regions. Third, the **Robust Boundary-Point Topology Reasoning** module, which constructs lane connectivity based on boundary-points and incorporates denoising training to improve structural stability.

using a segmentation network, and then refines it into a vectorized map via a mask-driven network. HD Map (Zhou et al. 2024; Li et al. 2024a) has made progress in online lane reconstruction, but it is limited to lane reconstruction and lacks a deep understanding of the lane topology in the scene, making it difficult to effectively interpret the semantic information of the lane topology in the current scene.

Lane Topology Reasoning

Accurate lane topology modeling is critical for autonomous vehicles to comprehend drivable paths and traffic rules. Previous methods have explored this problem from multiple perspectives. STSU (Can et al. 2021) detects lane centerlines from the BEV perspective and utilizes a Multi-Layer Perceptron (MLP) to predict connectivity between centerlines. In more complex road scenarios, TopoNet (Li et al. 2023) treats lanes and traffic elements as nodes in a graph and employs Graph Neural Networks (GNNs) (Zhou et al. 2020) to model the relationships between lanes and traffic signs. SMERF (Luo et al. 2024) incorporates Standard Definition Maps as additional information, enriching BEV feature representations and thereby improving the perception of lane centerlines. TopoMLP (Wu et al. 2024), based on PETR (Liu et al. 2022b), performs lane centerline detection and employs a lightweight MLP to predict the topological relationships between centerlines. LaneSegNet (Li et al. 2024b) models lanes as semantically enriched lane segments and introduces a Lane Attention mechanism to enhance the perception and feature representation of individual lane segments. TopoLogic (Fu et al. 2024) leverages the spatial positional relationships between lanes and proposes a concise and interpretable topology reasoning strategy, effectively modeling lane connectivity. Most of the previous works (Can et al. 2022, 2023; Liao et al. 2024a) use coarse-grained lane modeling and predict the topological relationships between lanes by calculating the overall similarity of lanes. However, this design neglects the local geometric structure of lanes, making it difficult to effectively model complex lane shapes, resulting in unreliable topological relationship predictions.

In this paper, we propose TopoFG, which integrates prior information, fine-grained modeling, and boundary-point based topology reasoning. TopoFG facilitates more reliable lane topology modeling and reasoning in dynamic driving scenes, ensuring accurate path planning and robust decision-making for vehicles in complex scenarios.

Method

Overview

As shown in Figure 2, we propose TopoFG, a novel framework designed for lane topology reasoning. First, given multi-view camera images, we extract feature maps $\boldsymbol{F}_I \in \mathbb{R}^{C\times H\times W\times D}$ using a CNN-based backbone, where C, H, W, and D denote the number of camera views, height, width, and number of channels of the image features, respectively. Following (Li et al. 2024c), we adopt deformable at-

tention (Xia et al. 2022) to transform multi-scale image features into BEV features. Inspired by previous works (Choi et al. 2024), we extend BEV features from a single scale to multiple scales. The processed BEV features are then used as input to the Hierarchical Prior Extractor module. In the Hierarchical Prior Extractor module, a mask prediction network is first employed to generate the BEV mask. We then separately extract the spatial prior from the BEV mask and the local sequential prior from lane point sequences. Next, in the Region-Focused Decoder module, fine-grained queries are used to model lane geometry. With dynamic reference point sampling from RoI regions, the model focuses on key lane areas to enhance detection accuracy and geometric representation. Finally, in the Robust Boundary-Point Topology Reasoning module, topology is modeled using start and end point queries. Boundary-point pairs are used to construct candidate connections, which are passed to a prediction head for Topology Reasoning. A denoising strategy is applied during training to guide the model toward more stable topological predictions.

Hierarchical Prior Extractor

To provide informative initialization for query embeddings, we design a Hierarchical Prior Extractor. This module extracts two complementary priors, i.e., a global spatial prior from the predicted BEV mask and a local sequential prior that captures lane order. We take the fused multi-scale BEV feature maps as input to this module. Following Mask2Former (Cheng et al. 2022), we initialize a set of learnable queries $\boldsymbol{Q}^L$, which interact with BEV features through the Mask Former module to produce refined queries $\boldsymbol{Q}^R$. These are then fed into the Mask Head to generate the final BEV lane masks $\boldsymbol{M}$. Based on the predicted BEV lane masks $\boldsymbol{M}$, we compute a weight vector $\boldsymbol{A}$. The weight for each BEV location is calculated by applying a threshold τ to the predicted probability:

$$\boldsymbol{A} = \boldsymbol{M} \cdot \mathbb{I}[\boldsymbol{M} \leq \tau] + \alpha \cdot \mathbb{I}[\boldsymbol{M} > \tau], \tag{1}$$

where $\mathbb{I}[\cdot]$ is the indicator function, and α is a scaling factor that emphasizes high-confidence regions. We adopt sine-cosine positional encoding to generate the spatial position encoding matrix $\boldsymbol{P}$ for the BEV grid. Given the weight vector $\boldsymbol{A}$ and the refined queries $\boldsymbol{Q}^R$, we derive the aggregated spatial prior queries $\boldsymbol{Q}^{\text{pos}}$ through a weighted summation:

$$\boldsymbol{Q}^{\text{pos}} = \mathcal{N}_{\text{norm}}(\boldsymbol{P}^\top \cdot \boldsymbol{A}) + \boldsymbol{Q}^R, \tag{2}$$

where $\mathcal{N}_{\text{norm}}(\cdot)$ denotes a normalization operation.

To complement the spatial prior, we further introduce a local sequential prior that captures the inherent ordering and continuity of lane lines. We first initialize a set of learnable queries $\boldsymbol{Q}'$ to represent local queries along each lane line. For each lane instance, we leverage the sequential nature of its constituent points by assigning them an ordered index. By applying positional encoding followed by linear projection, the index sequence $I = \{1, \ldots, k\}$ is transformed into an ordered embedding that preserves the local geometric structure. The sequential prior $\boldsymbol{Q}^{\text{seq}}$ is computed as:

$$\boldsymbol{Q}^{\text{seq}} = \mathcal{F}(PE(I)) + \boldsymbol{Q}', \tag{3}$$

where PE denotes a positional encoding function, e.g., sine-cosine or learnable, and $\mathcal{F}(\cdot)$ denotes an MLP.

Region-Focused Decoder

Accurately modeling lane topology demands capturing fine-grained local geometry. To this end, we introduce a Region-Focused Decoder, leveraging region-aware decoding for enhanced lane modeling. During fine-grained query initialization, we incorporate both spatial and sequential priors into the fine-grained query features. Given the spatial prior $\boldsymbol{Q}^{\text{pos}}$ of the i-th lane and the sequential prior $\boldsymbol{Q}^{\text{seq}}$ of the t-th keypoint, these two priors are then fused to produce the final query embedding as follows:

$$\boldsymbol{Q}^F_{i,t} = \boldsymbol{Q}^{pos}_i + \mathcal{F}(\boldsymbol{Q}^{\text{seq}}_t) \tag{4}$$

where $\boldsymbol{Q}^F_{i,t}$ denotes the fine-grained query for the t-th keypoint of lane i.

At each decoding layer m, we adopt a two-stage self-attention mechanism for fine-grained queries. We first perform inter-instance self-attention to capture interactions across different lane instances, followed by intra-instance self-attention to refine the point-level structure within each lane. Subsequently, we adopt deformable attention (Zhu et al. 2021) as cross-attention to enable effective interaction between BEV features and queries. To improve the spatial precision of query decoding, we adopt reference point sampling guided by the predicted lane mask, instead of using randomly initialized or fully learnable reference points (Fu et al. 2024). These reference points $\boldsymbol{R}$ constrain the attention to lane-relevant areas, thereby enhancing the model's ability to represent lane structures.

$$Q^{F,m+1}_{i,t} = \mathcal{A}_{\text{def}}\left(Q^{F,m}_{i,t},\, \boldsymbol{F}_B,\, \boldsymbol{R}_{i,t}\right), \tag{5}$$

where $\mathcal{A}_{\text{def}}(\cdot)$ denotes the Deformable Attention operation, and $\boldsymbol{Q}^{F,m}_{i,t}$ denotes the t-th fine-grained query of the i-th lane at the m-th layer, $\boldsymbol{F}_B$ represents the BEV feature, $\boldsymbol{R}_{i,t}$ denotes the reference point coordinates associated with the t-th fine-grained query of the i-th lane.

Robust Boundary-Point Topology Reasoning

As the lane topology is determined by the relationships between endpoints, we retain only the first and last queries of each lane to serve as structural features for topology reasoning. Specifically, each lane is represented by a sequence of fine-grained queries generated by the Region-Focused Decoder. For lane i, the final representation consists of k fine-grained queries $Lane_i = \{Q^F_{i,1}, Q^F_{i,2}, \ldots, Q^F_{i,k}\}$. The first and last queries of each lane, $Q^F_{i,1}$ and $Q^F_{i,k}$, are selected as the boundary-point features of lane i, and are denoted by f_i^{start} and f_i^{end}, respectively, for subsequent topology reasoning. Given any pair of lanes (i, j), we concatenate the endpoint feature of lane i and the start-point feature of lane j, and feed this representation into a shared multilayer perceptron to predict connectivity:

$$r_{i\to j} = \text{Concat}(f_i^{end}, f_j^{start}) \tag{6}$$

$$\boldsymbol{S}_{i\to j} = \mathcal{F}(r_{i\to j}), \tag{7}$$

where $\boldsymbol{S}_{i\to j}$ indicates whether lane i leads to lane j, corresponding to the existence of a directed topological connection from lane i to lane j. We normalize the values of $\boldsymbol{S}$ using a Sigmoid function, which produces the Similarity Topology. In parallel, we compute the Euclidean distance between the boundary-points of all lane pairs to obtain a geometric distance matrix, which is then mapped to a Geometric Topology using a mapping function following the (Fu et al. 2024) design. Finally, the Similarity Topology and Geometric Topology are summed to obtain the final adjacency matrix for lane topology reasoning.

The number of predicted lanes often exceeds that of ground-truth lanes, necessitating Hungarian matching (Kuhn 1955) to compute one-to-one correspondences for loss calculation. Based on these matches, the ground-truth adjacency matrix is mapped into a larger zero-initialized matrix to serve as supervision. However, this process depends heavily on the matching results, causing the supervision matrix for the same sample to vary across epochs. Such instability degrades the model's ability to learn topological relationships and hinders effective topology reasoning.

To improve the robustness and reliability of supervision in topology modeling, we propose a denoising training mechanism guided by boundary points. The key idea is to generate noisy queries from each ground-truth (GT) instance, allowing the supervision adjacency matrix to be fixed in advance and providing stable learning signals. Given N_{gt} GT instances and G denoising groups, we construct $N_{gt} \times G$ noisy queries, which are refined along with vanilla queries by the model. The refined queries are then split into vanilla and denoising queries, each used for separate topology reasoning. For denoising queries, the original adjacency matrix is expanded G times into a block-diagonal form, which serves as supervision during training to enhance the model's robustness and structural consistency. During inference, only vanilla queries are used for topology reasoning, and denoising queries are discarded. This strategy provides stable topology labels, thereby improving the accuracy and robustness of topology reasoning.

Lane Head

After refinement by the decoder, the queries are decoded by the lane head into lane centerline coordinates. Similar to topology reasoning, the denoising queries are discarded during inference. We follow the (Fu et al. 2024), with the only difference being that each query is responsible for predicting a single key point on the lane.

Experiments

Datasets and metric

Dataset. We conduct experiments on the OpenLane-V2 benchmark (Wang et al. 2023a). Designed specifically for autonomous driving scenarios, OpenLane-V2 is a large-scale multi-task benchmark that covers lane centerline detection, traffic element detection, and topology reasoning tasks. Built upon Argoverse2 (Wilson et al. 2021) and nuScenes (Caesar et al. 2020), the dataset consists of two subsets: *subset_A* and *subset_B*, each containing 1,000 scenes. All data are collected at 2Hz with multi-view images. These annotations include lane centerlines, traffic elements in the front-view, as well as topological relations between lanes and between lanes and traffic elements. *subset_A* and *subset_B* contain 7 and 6 camera views, respectively.

Metric. We follow the official evaluation protocol of OpenLane-V2 to assess both perception and topology reasoning capabilities. The evaluation metrics include DET_l, DET_t, TOP_{ll}, and TOP_{lt}. Specifically, DET_l measures alignment between predicted and ground-truth lane centerlines using the Fréchet distance. DET_t measures traffic-element detection quality using Intersection over Union (IoU). TOP_{ll} and TOP_{lt} evaluate the structural correctness of lane-lane and lane–traffic element connections, respectively. The OpenLane-V2 Score (OLS) provides a unified evaluation by aggregating metrics from various perspectives of the primary task:

$$\text{OLS} = \frac{1}{4}\left[\text{DET}_l + \text{DET}_t + \sqrt{\text{TOP}_{ll}} + \sqrt{\text{TOP}_{lt}}\right]. \quad (8)$$

All TopoFG evaluations are conducted using the metrics defined in the latest OpenLane-V2 v2.1.0 release.

Implementation Details

For image inputs, the resolution is set to 800×1024 for *subset_A* and 480×800 for *subset_B*. A ResNet-50 (He et al. 2016) backbone is used to extract features from surround-view images, followed by a Feature Pyramid Network (Lin et al. 2017) to generate multi-scale representations. In the lane centerline detection task, the model predicts 200 lanes, with each lane represented by $k = 11$ fine-grained queries. A decoder with $m = 6$ layers is used to refine the queries. We use $G = 5$ denoising groups, each containing N_{gt} queries. For the Hierarchical Prior Extractor, we set the threshold τ to 0.3 and the scaling factor α to 1.0. The AdamW optimizer is adopted with a weight decay of 0.01 and an initial learning rate of 2×10^{-4}. A cosine annealing strategy is used for learning rate scheduling, with 500 warmup steps. All experiments are conducted on 8 NVIDIA A100 GPUs for 24 epochs, with a batch size of 1 per GPU.

Performance Comparison

As shown in Table 1, the main results on *subset_A* and *subset_B* are presented. We compare the proposed TopoFG with existing state-of-the-art methods on the OpenLane-V2 dataset. The results show that TopoFG consistently achieves superior performance across both subsets. On *subset_A*, TopoFG achieves an OLS score of 48.0%, significantly outperforming TopoLogic (44.1%) and TopoMLP (44.1%). For lane centerline detection, it attains 33.8% on DET_l. In terms of topology reasoning, it achieves 30.8% on TOP_{ll} and 30.9% on TOP_{lt}, indicating stronger structural modeling capability compared to existing methods. On *subset_B*, TopoFG also achieves the best overall performance with an OLS of 45.4%. It obtains 27.2% on TOP_{ll} and 21.7% on TOP_{lt}, demonstrating stable and generalizable topology reasoning in complex scenarios. As shown in Table 2, the performance of TopoFG is compared with existing methods on the OpenLane-V2 lane segment detection benchmark. TopoFG outperforms all other methods in all metrics,

Dataset	Method	Venue	OLS↑	DET_l ↑	DET_t ↑	TOP_{ll} ↑	TOP_{lt} ↑
subset_A	STSU (Can et al. 2021)	ICCV2021	29.3	12.7	43.0	2.9	19.8
	VectorMapNet (Liu et al. 2023)	ICML2023	24.9	11.1	41.7	2.7	9.2
	MapTR (Liao et al. 2023)	ICLR2023	31.0	17.7	43.5	5.9	15.1
	TopoNet (Li et al. 2023)	Arxiv2023	39.8	28.6	<u>48.6</u>	10.9	23.8
	TopoMLP (Wu et al. 2024)†	ICLR2024	<u>44.1</u>	28.5	**49.5**	21.7	<u>26.9</u>
	TopoLogic (Fu et al. 2024)	NeurIPS2024	<u>44.1</u>	<u>29.9</u>	47.2	<u>23.9</u>	25.4
	TopoFG(Ours)	-	**48.0(+3.9)**	**33.8(+3.9)**	47.2	**30.8(+6.9)**	**30.9(+4.0)**
subset_B	STSU (Can et al. 2021)	ICCV2021	-	8.2	43.9	-	-
	VectorMapNet (Liu et al. 2023)	ICML2023	-	3.5	49.1	-	-
	MapTR (Liao et al. 2023)	ICLR2023	-	15.2	54.0	-	-
	TopoNet (Li et al. 2023)	Arxiv2023	36.8	24.3	**55.0**	6.7	16.7
	TopoLogic (Fu et al. 2024)	NeurIPS2024	<u>42.3</u>	<u>25.9</u>	<u>54.7</u>	<u>21.6</u>	<u>17.9</u>
	TopoFG(Ours)	-	**45.4(+3.1)**	**30.0(+4.1)**	53.0	**27.2(+5.6)**	**21.7(+3.8)**

Table 1: Comparisons of our model and existing state-of-the-art methods on *subset_A* and *subset_B*. "-" denotes the absence of relevant data. The best performances are highlighted in **bold**, while the second one is <u>underlined</u>. "+" indicates the absolute improvements compared with the second one. "†" indicates the result evaluated using the officially released model.

Method	mAP↑	AP_{ls} ↑	AP_{ped} ↑	TOP_{ls} ↑
TopoNet	23.0	23.9	22.0	-
MapTR	27.0	25.9	28.1	-
MapTRv2	28.5	26.6	30.4	-
LaneSegNet	32.6	32.3	32.9	25.4
TopoLogic	33.2	33.0	33.4	<u>30.8</u>
Topo2Seq	<u>33.6</u>	<u>33.7</u>	<u>33.5</u>	26.9
TopoFG (Ours)	**34.4(+0.8)**	**33.8(+0.1)**	**35.1(+1.6)**	**31.3(+0.5)**

Table 2: Performance comparison with state-of-the-art methods on OpenLane-V2 benchmark on lane segment. "-" denotes the absence of relevant data. The best performances are highlighted in **bold**, while the second one is <u>underlined</u>. "+" indicates the absolute improvements compared with the second one.

achieving an mAP score of 34.4%, surpassing TopoLogic and Topo2Seq (Yang et al. 2025). Additionally, TopoFG excels in topology reasoning compared to other methods.

Ablation Study

Contributions of Main Components. To thoroughly assess the contribution of each component in TopoFG, we carry out ablation experiments on the OpenLane-V2 *subset_A*. As shown in Table 3, each module in TopoFG consistently improves performance. Compared to the baseline (Fu et al. 2024), our proposed HPE achieves significant improvements in OLS and DET_l, while RFD further enhances both detection and topology-related metrics. RBTR yields the best overall performance, boosting OLS to 48.0% and significantly improving TOP_{ll} and TOP_{lt}, demonstrating the effectiveness of TopoFG in topology reasoning.

Contributions of Submodules. As shown in Table 4, we progressively introduce the local sequential prior and global spatial prior to evaluate their effects. The introduction of local sequential prior leads to notable improvements in OLS and DET_l, and further adding global spatial prior

	OLS↑	DET_l ↑	DET_t ↑	TOP_{ll} ↑	TOP_{lt} ↑
baseline	44.1	29.9	47.2	23.9	25.4
+HPE	45.4	31.3	47.5	26.1	26.6
+RFD	45.8	31.8	47.2	26.8	27.7
+RBTR	48.0	33.8	47.2	30.8	30.9

Table 3: Contributions of Main Components. The main components include the Hierarchical Prior Extractor (HPE), Region-Focused Decoder (RFD), and Robust Boundary-Point Topology Reasoning (RBTR). Each module brings consistent performance gains.

LP	GP	OLS↑	DET_l ↑	DET_t ↑	TOP_{ll} ↑	TOP_{lt} ↑
		44.1	29.9	47.2	23.9	25.4
✓		45.1	31.2	45.5	26.7	27.3
	✓	45.2	31.0	45.4	26.7	28.0
✓	✓	45.4	31.3	47.5	26.1	26.6

Table 4: Ablation study for evaluating the contributions of the Hierarchical Prior Extractor, including the local sequential prior (LP) and global spatial prior (GP).

continuously enhances all metrics, with OLS increasing to 45.4%. These results indicate that the two priors are complementary and jointly contribute to performance gains. As shown in Table 5, fine-grained query initialization effectively enhances lane representation capability. When combined with sampled reference points, the model achieves the best performance across all metrics. These results indicate that the design improves both the accuracy of lane modeling and the perception of key regions. As shown in Table 6, the boundary-point topology reasoning effectively enhances the recognition of topological connections. When combined with denoised topology reasoning, the overall performance is further improved, with OLS reaching 48.0%, while TOP_{ll}

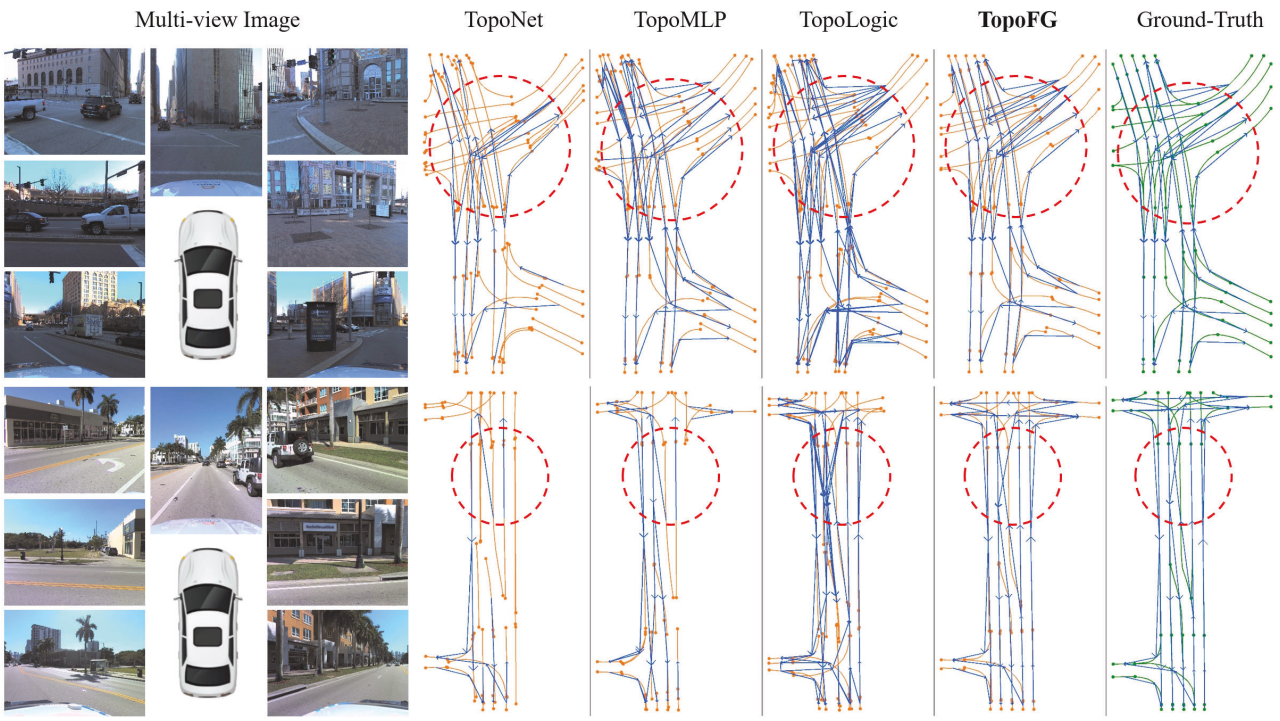

Figure 3: Qualitative comparison of different methods on the lane topology reasoning task. From left to right: multi-view input images, TopoNet, TopoMLP, TopoLogic, our proposed TopoFG, and the ground truth. The figure shows the predicted lane centerlines in the bird's-eye view, where orange lines represent the predicted lane centerlines and blue arrows indicate the topological connections between lanes.

FQI	SRP	OLS↑	DET_l ↑	DET_t ↑	TOP_{ll} ↑	TOP_{lt} ↑
		45.4	31.3	47.5	26.1	26.6
✓		45.6	31.0	45.8	28.6	27.2
✓	✓	45.8	31.8	47.2	26.8	27.7

Table 5: Ablation study for evaluating the contributions of the Region-Focused Decoder, including fine-grained query initialization (FQI) and sampled reference points (SRP).

BTR	DTR	OLS↑	DET_l ↑	DET_t ↑	TOP_{ll} ↑	TOP_{lt} ↑
		45.8	31.8	47.2	26.8	27.7
✓		46.6	33.0	45.8	28.3	29.6
	✓	47.3	33.4	46.4	29.0	30.7
✓	✓	48.0	33.8	47.2	30.8	30.9

Table 6: Ablation study to assess the effectiveness of Robust Boundary-Point Topology Reasoning, which is based on boundary-point topology reasoning (BTR) and denoised topology reasoning (DTR).

and TOP_{lt} increase to 30.8% and 30.9%, respectively. These results demonstrate the robustness and reliability of the proposed method in performing topology reasoning.

Qualitative Analysis

We compare several mainstream methods on the lane topology reasoning task, as shown in Figure 3. From left to right, the results include multi-view input images, TopoNet (Li et al. 2023), TopoMLP (Wu et al. 2024), TopoLogic (Fu et al. 2024), our TopoFG, and the ground truth. All predictions are visualized in a bird's-eye view. Orange lines represent the predicted lane centerlines, and blue arrows indicate the topological connections between lanes. It can be observed that TopoNet, TopoMLP, and TopoLogic suffer from missing lanes and incorrect connections, particularly at intersections. In contrast, TopoFG better recovers complete lane structures and accurately captures topological relationships.

Conclusion

In this work, we present an end-to-end fine-grained lane topology reasoning framework that addresses the limitations of previous methods relying on a single query to represent each lane. By representing each lane with multiple spatially-aware queries, our proposed method captures local geometric variations and enables more accurate topological inference. The proposed denoised topology reasoning strategy effectively improves the robustness and reliability of learning topological relationships between lanes. Extensive experiments on the OpenLane-V2 benchmark validate the effectiveness of our method, which achieves 48.0% OLS on *subset_A* and 45.4% OLS on *subset_B*, outperforming previous state-of-the-art methods in topology reasoning.

Acknowledgments

This work was supported in part by Beijing Natural Science Foundation under no. L221013 and Chinese National Natural Science Foundation Projects 62206276.

References

Abualsaud, H.; Liu, S.; Lu, D. B.; Situ, K.; Rangesh, A.; and Trivedi, M. M. 2021. Laneaf: Robust multi-lane detection with affinity fields. *IEEE Robotics and Automation Letters*, 6(4): 7477–7484.

Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631.

Can, Y. B.; Liniger, A.; Paudel, D. P.; and et al. 2021. Structured bird's-eye-view traffic scene understanding from onboard images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 15661–15670.

Can, Y. B.; Liniger, A.; Paudel, D. P.; and Van Gool, L. 2022. Topology preserving local road network estimation from single onboard camera image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17263–17272.

Can, Y. B.; Liniger, A.; Paudel, D. P.; and Van Gool, L. 2023. Improving online lane graph extraction by object-lane clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8591–8601.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-end object detection with transformers. In *European conference on computer vision*, 213–229. Springer.

Cheng, B.; Misra, I.; Schwing, A. G.; Kirillov, A.; and Girdhar, R. 2022. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1290–1299.

Choi, S.; Kim, J.; Shin, H.; and Choi, J. W. 2024. Mask2Map: Vectorized HD Map Construction Using Bird's Eye View Segmentation Masks. In *European Conference on Computer Vision*.

Fu, Y.; Liao, W.; Liu, X.; Xu, H.; Ma, Y.; Zhang, Y.; and Dai, F. 2024. TopoLogic: An Interpretable Pipeline for Lane Topology Reasoning on Driving Scenes. In *Advances in Neural Information Processing Systems*, volume 37, 61658–61676.

Garnett, N.; Cohen, R.; Pe'er, T.; Lahav, R.; and Levi, D. 2019. 3d-lanenet: end-to-end 3d multiple lane detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2921–2930.

Han, J.; Deng, X.; Cai, X.; Yang, Z.; Xu, H.; Xu, C.; and Liang, X. 2022. Laneformer: Object-aware row-column transformers for lane detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 799–807.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

He, X.; Guo, H.; Zhu, K.; Zhu, B.; Zhao, X.; Fang, J.; and Wang, J. 2024. Monocular lane detection based on deep learning: A survey. *arXiv preprint arXiv:2411.16316*.

Kuhn, H. W. 1955. The Hungarian method for the assignment problem. 2(1-2): 83–97.

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

Li, Q.; Wang, Y.; Wang, Y.; and Zhao, H. 2022. Hdmapnet: An online hd map construction and evaluation framework. In *2022 International Conference on Robotics and Automation (ICRA)*, 4628–4634. IEEE.

Li, S.; Lin, J.; Shi, H.; Zhang, J.; Wang, S.; Yao, Y.; Li, Z.; and Yang, K. 2024a. DTCLMapper: Dual Temporal Consistent Learning for Vectorized HD Map Construction. *IEEE Transactions on Intelligent Transportation Systems*.

Li, T.; Chen, L.; Wang, H.; Li, Y.; Yang, J.; Geng, X.; Jiang, S.; Wang, Y.; Xu, H.; Xu, C.; et al. 2023. Graph-based topology reasoning for driving scenes. *arXiv preprint arXiv:2304.05277*.

Li, T.; Jia, P.; Wang, B.; Chen, L.; Jiang, K.; Yan, J.; and Li, H. 2024b. LaneSegNet: Map Learning with Lane Segment Perception for Autonomous Driving. In *ICLR*.

Li, Y.; Yang, Y.; and Lei, Z. 2025a. CoreNet: Conflict Resolution Network for point-pixel misalignment and sub-task suppression of 3D LiDAR-camera object detection. *Information Fusion*, 118: 102896.

Li, Y.; Yang, Y.; and Lei, Z. 2025b. Rctrans: Radar-camera transformer via radar densifier and sequential decoder for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 5048–5056.

Li, Z.; Wang, W.; Li, H.; Xie, E.; Sima, C.; Lu, T.; Yu, Q.; and Dai, J. 2024c. Bevformer: learning bird's-eye-view representation from lidar-camera via spatiotemporal transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Liao, B.; Chen, S.; Jiang, B.; Cheng, T.; Zhang, Q.; Liu, W.; Huang, C.; and Wang, X. 2024a. Lane graph as path: Continuity-preserving path-wise modeling for online lane graph construction. In *European Conference on Computer Vision*, 334–351. Springer.

Liao, B.; Chen, S.; Wang, X.; Cheng, T.; Zhang, Q.; Liu, W.; and Huang, C. 2023. MapTR: Structured Modeling and Learning for Online Vectorized HD Map Construction. In *International Conference on Learning Representations*.

Liao, B.; Chen, S.; Zhang, Y.; Jiang, B.; Zhang, Q.; Liu, W.; Huang, C.; and Wang, X. 2024b. Maptrv2: An end-to-end framework for online vectorized hd map construction. *International Journal of Computer Vision*, 1–23.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Liu, R.; Chen, D.; Liu, T.; Xiong, Z.; and Yuan, Z. 2022a. Learning to predict 3d lane shape and camera pose from a single image via geometry constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 1765–1772.

Liu, Y.; Wang, T.; Zhang, X.; and Sun, J. 2022b. Petr: Position embedding transformation for multi-view 3d object detection. In *European conference on computer vision*, 531–548. Springer.

Liu, Y.; Yuan, T.; Wang, Y.; Wang, Y.; and Zhao, H. 2023. Vectormapnet: End-to-end vectorized hd map learning. In *International Conference on Machine Learning*, 22352–22369. PMLR.

Luo, K. Z.; Weng, X.; Wang, Y.; Wu, S.; Li, J.; Weinberger, K. Q.; Wang, Y.; and Pavone, M. 2024. Augmenting lane perception and topology understanding with standard definition navigation maps. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 4029–4035. IEEE.

Luo, Y.; Zheng, C.; Yan, X.; Kun, T.; Zheng, C.; Cui, S.; and Li, Z. 2023. Latr: 3d lane detection from monocular images with transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7941–7952.

Pan, X.; Shi, J.; Luo, P.; Wang, X.; and Tang, X. 2018. Spatial as deep: Spatial cnn for traffic scene understanding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Pittner, M.; Janai, J.; and Condurache, A. P. 2024. Lanecpp: Continuous 3d lane detection using physical priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10639–10648.

Qu, Z.; Jin, H.; Zhou, Y.; Yang, Z.; and Zhang, W. 2021. Focus on local: Detecting lane marker from bottom up via key point. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14122–14130.

Tabelini, L.; Berriel, R.; Paixao, T. M.; Badue, C.; De Souza, A. F.; and Oliveira-Santos, T. 2021a. Keep your eyes on the lane: Real-time attention-guided lane detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 294–302.

Tabelini, L.; Berriel, R.; Paixao, T. M.; Badue, C.; De Souza, A. F.; and Oliveira-Santos, T. 2021b. Polylanenet: Lane estimation via deep polynomial regression. In *2020 25th international conference on pattern recognition (ICPR)*, 6150–6156. IEEE.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, H.; Li, T.; Li, Y.; Chen, L.; Sima, C.; Liu, Z.; Wang, B.; Jia, P.; Wang, Y.; Jiang, S.; et al. 2023a. Openlane-v2: A topology reasoning benchmark for unified 3d hd mapping. *Advances in Neural Information Processing Systems*, 36: 18873–18884.

Wang, R.; Qin, J.; Li, K.; Li, Y.; Cao, D.; and Xu, J. 2023b. Bev-lanedet: An efficient 3d lane detection based on virtual camera via key-points. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1002–1011.

Wilson, B.; Qi, W.; Agarwal, T.; Lambert, J.; Singh, J.; Khandelwal, S.; Pan, B.; Kumar, R.; Hartnett, A.; Pontes, J. K.; et al. 2021. Argoverse 2: Next generation datasets for self-driving perception and forecasting. In *NeurIPS*.

Wu, D.; Chang, J.; Jia, F.; Liu, Y.; Wang, T.; and Shen, J. 2024. TopoMLP: An Simple yet Strong Pipeline for Driving Topology Reasoning. *ICLR*.

Wu, D.; Liao, M.-W.; Zhang, W.-T.; Wang, X.-G.; Bai, X.; Cheng, W.-Q.; and Liu, W.-Y. 2022. Yolop: You only look once for panoptic driving perception. *Machine Intelligence Research*, 19(6): 550–562.

Xia, Z.; Pan, X.; Song, S.; Li, L. E.; and Huang, G. 2022. Vision transformer with deformable attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4794–4803.

Yang, Y.; Luo, Y.; He, B.; Li, E.; Cao, Z.; Zheng, C.; Mei, S.; and Li, Z. 2025. Topo2seq: Enhanced topology reasoning via topology sequence learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 9318–9326.

Zhang, Y.; Lu, Z.; Ma, D.; Xue, J.-H.; and Liao, Q. 2020. Ripple-GAN: Lane line detection with ripple lane line detection network and Wasserstein GAN. *IEEE Transactions on Intelligent Transportation Systems*, 22(3): 1532–1542.

Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; and Sun, M. 2020. Graph neural networks: A review of methods and applications. *AI open*, 1: 57–81.

Zhou, Y.; Zhang, H.; Yu, J.; Yang, Y.; Jung, S.; Park, S.-I.; and Yoo, B. 2024. Himap: Hybrid representation learning for end-to-end vectorized hd map construction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15396–15406.

Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; and Dai, J. 2021. Deformable DETR: Deformable Transformers for End-to-End Object Detection.