TopoHR: Hierarchical Centerline Representation for Cyclic Topology Reasoning in Driving Scenes with Point-to-Instance Relations

Yifeng Bai[1,2*] Zhirong Chen[3*] Bo Song[1] Erkang Cheng[3†] Haibin Ling[4]
[1]Institute of Intelligent Machines, HFIPS, Chinese Academy of Sciences
[2]University of Science and Technology of China [3]NullMax [4]Westlake University

Abstract

Topology reasoning is crucial for autonomous driving. Current methods primarily focus on instance-level learning for centerline detection, followed by a sequential module for topology reasoning that relies on simplified MLP layers. Moreover, they often neglect the importance of point-to-instance (P2I) relationships in topology reasoning. To address these limitations, we present TopoHR (Topological Hierarchical Representation), a novel end-to-end framework that establishes cyclic interaction between centerline detection and topology reasoning, allowing them to iteratively enhance each other. Specifically, we introduce a hierarchical centerline representation including point queries, instance queries, and semantic representations. These multi-level features are seamlessly integrated and fused within a hierarchical centerline decoder. Furthermore, we design a hierarchical topology reasoning module that captures both fine-grained P2I relationships and global instance-to-instance (I2I) connections within a unified architecture. With these novel components, TopoHR ensures accurate and robust topology reasoning. On the OpenLane-V2 benchmark, TopoHR refreshes state-of-the-art performance with significant improvements. Notably, compared with previous best results, TopoHR achieves +3.8 in DET_l*, +5.4 in* TOP_{ll} *on subset_A and +11.0 in* DET_l*, +7.9 in* TOP_{ll} *on subset_B, validating the effectiveness of the proposed components. The code will be shared publicly at* https://github.com/Yifeng-Bai/TopoHR.git.

1. Introduction

Topological relationships of traffic scene elements refer to the fundamental connectivity rules governing road structures, hence are crucial for high-level autonomous driving functions. Traditional lane detection methods [8] and online mapping techniques [18] focus on geometric accuracy but fail to capture scene topology relations. While HD Map can provide such topological information, it struggles with issues of freshness and scalability. Recent methods [11, 13, 15, 16, 30] adopt a sequential pipeline (Fig. 1(a)). In this pipeline, a centerline detector first extracts the instance-level centerline representation, and a topology reasoning module then analyzes the extracted centerline instances.

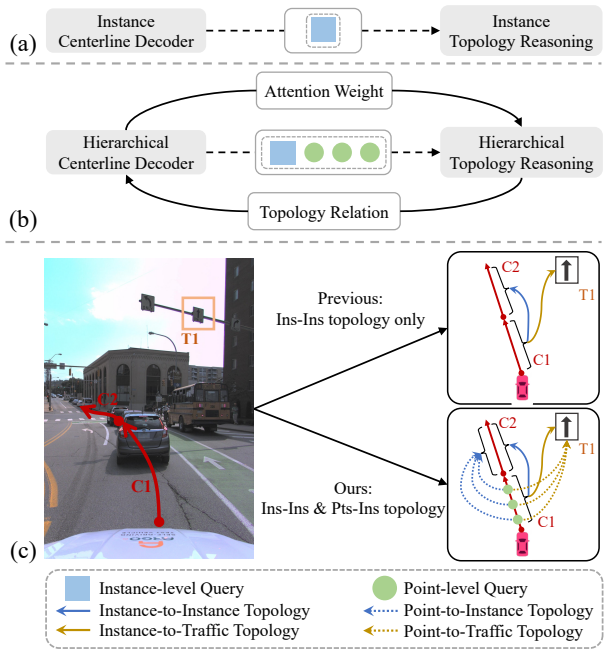

Figure 1. Different centerline detection and topology reasoning pipelines. (a) The conventional sequential pipeline. (b) Our cyclic pipeline. (c) Our motivation to integrate point-to-instance (P2I) and point-to-traffic (P2T) relations to boost topology reasoning.

Previous centerline detection methods typically use instance-level query representations in a transformer decoder, defining the task as point-set prediction [7, 15, 16, 26, 30], curve parameter estimation [12, 13], or binary segmentation [11]. For instance, TopoNet [15] represents centerlines as a point set and refines them using

*Equal contribution.
†Corresponding author. chengerkang@nullmax.ai.

a Scene Graph Neural Network (SGNN). Alternatively, segmentation-based approaches like TopoMask [12] incorporate direction prediction to derive vectorized results from segmentation outputs. However, centerlines are inherently invisible, making it challenging to accurately extract their features through direct segmentation modeling. Current implementations either use segmentation as auxiliary supervision or rely on post-processing to generate vectorized point sets, which limits the full utilization of the rich information contained in segmentation results.

Building on instance-level queries from the centerline detector, a topology reasoning module usually applies MLP functions [7, 15, 26, 30] to infer topological relationships. For instance, TopoNet [15] employs three MLP layers to reason about topological relationships using refined centerline queries. TopoMLP [30] highlights the importance of detector performance in the cascade structure and enhances topology reasoning by integrating position embeddings into the MLP layer. Similarly, TopoLogic [7] improves reasoning performance by incorporating centerline geometry priors to link two centerlines through their end points. While these methods achieve promising results in topology reasoning for driving scenes, there remains a significant gap in the accuracy of topological relationship predictions. These approaches often rely on cascaded architectures, where the detection and relation reasoning modules are optimized separately, leading to inconsistent feature representations. Additionally, the use of simplified prediction heads (e.g., MLPs) fails to adequately capture the complex spatial dependencies inherent in urban road networks.

To address the above issues, we propose TopoHR, a novel end-to-end topology reasoning framework based on a cyclic interaction structure and hierarchical centerline representation. Unlike previous cascade architectures that heavily rely on detector performance, we show that the detector and the topology reasoning module can mutually enhance each other. As shown in Fig. 1(b), we introduce a cyclic reasoning architecture where query attention weights from the detector are fed into the reasoning module as feed-forward signals, while instance relations from the topology reasoning module are fed back to the detector.

In the centerline detector, we integrate multiple centerline representations (point queries, instance queries, and semantic instances) and introduce specialized feature interaction modules. First, to simultaneously capture local and global features, we design a hierarchical integrator that facilitates information exchange between point-aware and instance-aware queries. Second, we introduce a discrete distance transform segmentation method to extract spatial centerline information, which is then used for instance-aware query interaction through masked-attention.

For topology reasoning, we design a hierarchical topology reasoning module that captures both fine-grained *point-to-instance* (P2I) and global *instance-to-instance* (I2I) topological connections. These connections are fused to predict the final topological relations, as shown in Fig. 1(c).

In the experiments, we evaluate TopoHR on the OpenLane-V2 dataset [28]. The results demonstrate significant improvements over existing methods under comparable settings. Specifically, TopoHR achieves scores of 34.6 $\mathrm{TOP}_{\mathrm{ll}}$ and 35.6 $\mathrm{TOP}_{\mathrm{lt}}$ on subset_A, and 39.7 $\mathrm{TOP}_{\mathrm{ll}}$ and 28.0 $\mathrm{TOP}_{\mathrm{lt}}$ on subset_B, thereby setting new records on OpenLane-V2. These results correspond to absolute improvements of 5.4 points in $\mathrm{TOP}_{\mathrm{ll}}$ and 3.4 points in $\mathrm{TOP}_{\mathrm{lt}}$ on subset_A, and 7.9 points in $\mathrm{TOP}_{\mathrm{ll}}$ and 2.2 points in $\mathrm{TOP}_{\mathrm{lt}}$ on subset_B over previous methods.

Our main contributions can be summarized as follows:

- We propose TopoHR, a novel end-to-end framework for centerline detection and topology reasoning, where the detection and reasoning modules iteratively interact with each other to cyclically enhance both performances.
- We introduce a hierarchical centerline representation, where multi-level representations are seamlessly integrated and fused within a hierarchical centerline decoder.
- We propose a hierarchical topology reasoning module to capture both fine-grained P2I relationships and global I2I connections, leading to accurate topology results.
- Our method demonstrates promising performance on the OpenLane-V2 dataset, surpassing previous state-of-the-art models by significant margins.

2. Related Work

Map Element and Centerline Representation. Centerline representations in topology reasoning typically employ map element formulation in online map understanding. These representations can be broadly categorized into rasterization-based and vectorization-based methods.

Rasterization-based approaches formulate the extraction of map elements as a dense segmentation task. For instance, HDMapNet [14] pioneers the segmentation-based paradigm by fusing multi-view camera images and LiDAR point clouds within a Bird's-Eye View (BEV) encoder-decoder architecture, thereby predicting instance-level map elements as semantic masks. To reduce dependency on multi-modal data, MGMap [21] proposes a camera-only framework with multi-granularity decoding, enabling hierarchical segmentation of map components. Mask2Map [4] enhances point-level feature extraction by incorporating deformable attention into instance segmentation. More recently, MapVR [34] applies differentiable rasterization to vectorized outputs and exploits distance variation on raster maps to provide accurate, geometry-aware supervision, without incurring additional computation during inference.

In contrast, vectorization-based methods represent map elements as sparse point sets. VectorMapNet [22] introduces a two-stage pipeline combining polyline detection

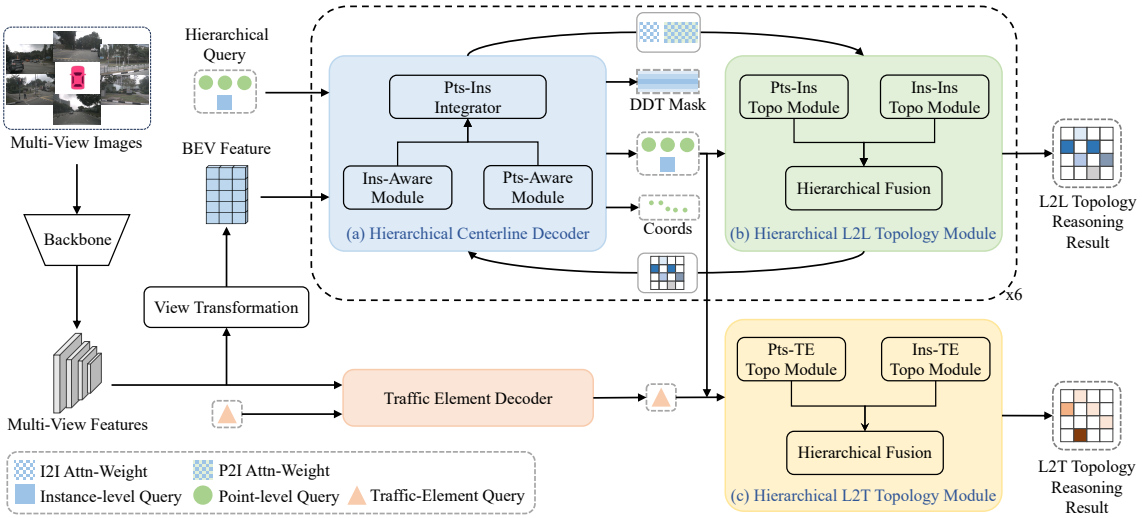

Figure 2. Overview of TopoHR. Aside from a BEV feature extractor and a traffic element decoder, TopoHR has three notable components: (a) Hierarchical Centerline Decoder, which employs a hierarchical query representation to simultaneously model point-level and instance-level features through a series of attention mechanisms and integration modules; (b) Hierarchical L2L Topology Module and (c) Hierarchical L2T Topology Module, both of which perform topology reasoning by combining both fine-grained point-to-instance relationships and global instance-to-instance topological connections.

with geometric refinement. MapTR [18] proposes a unified permutation-equivalent modeling strategy, enabling end-to-end vectorized map learning without explicit point ordering constraints. HIMap [35] unifies hierarchical queries for joint detection of lanes and traffic elements, showing robust generalization across diverse datasets. PivotNet [5] leverages sparse pivots for polyline parameterization, thereby enhancing the representation of geometric structures.

Topology Detection and Reasoning. Most topology reasoning approaches use a sequential pipeline, where a centerline detector is followed by a topology reasoning module. For centerline detection, CenterLineDet [31] represents centerlines as vertices and utilizes temporal feature fusion to achieve multi-camera perception. Beyond point-set inference, centerline detection also incorporates instance segmentation and curve parameterization. For instance, STSU [2] uses Bézier curves for centerlines. TopoMask [11] introduces instance segmentation and adds direction prediction to parse vectorized results. TopoFormer [26] utilizes the geometric distance between centerlines to guide global information aggregation and models plausible road structures under a counterfactual intervention layer. In addition, SMERF [24], TopoSD [32] and SEPT [27] propose the use of SD Map to enhance the centerline detector.

For topology reasoning, TopoNet [15] first uses MLP to reduce instance embedding dimension, then sends the features to another MLP with sigmoid activation to predict their relationship. TopoMLP [30] follows the "detection first, reasoning later" rule to design a detector and an MLP that incorporates implicit position embedding. TopoLogic [7] proposes a centerline geometry prior that explicitly preserves lane connectivity. LaneSegNet [16] embeds topological affinity fields into instance segmentation, achieving real-time lane graph extraction. Topo2Seq [33] converts the graph topology relation into a serialized representation and uses a hierarchical Transformer to achieve multiscale topological reasoning. RelTopo [25] constructs geometry-enhanced centerline relation by integrating positional embeddings and geometric distance embeddings.

3. Method

3.1. Architecture Overview

The proposed TopoHR framework is summarized in Fig. 2. Starting from input multi-view images, it first extracts multi-view features through a shared feature backbone. These features are used to generate BEV features through view transformation and to detect traffic elements through a decoder. Then, TopoHR proceeds with three notable components: Hierarchical Centerline (HC) Decoder (Fig. 2(a)), Hierarchical L2L Topology Module (Fig. 2(b)) and Hierarchical L2T Topology Module (Fig. 2(c)). The HC Decoder and Hierarchical L2L Topology Module form a cyclic and iterative block, which takes BEV features as input and eventually produces L2L topology reasoning results. By contrast, the Hierarchical L2T Topology Module produces the L2T topology reasoning results.

Specifically, the HC Decoder takes the hierarchical query representation $\mathbf{Q}_{\text{hcl}} \in \mathbb{R}^{N(P+1)\times C}$, where N, P, and C de-

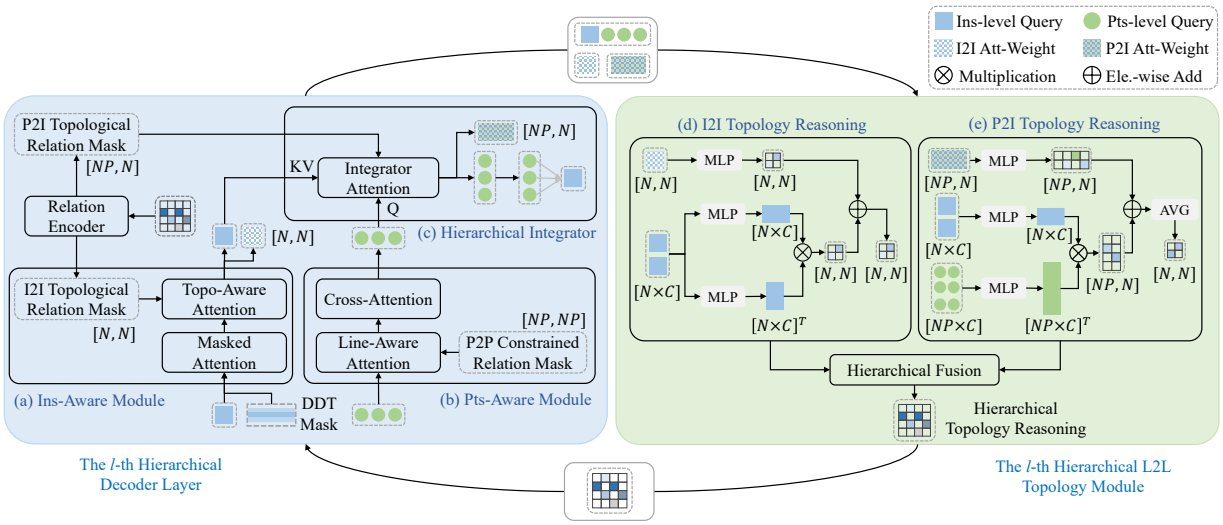

Figure 3. Illustration of the detailed architecture of hierarchical decoder Layer and hierarchical topology module: (a) an Instance-Aware Module, which combines a masked-attention module with a topo-aware attention to extract global semantic features; (b) a Point-Aware Module, which incorporates a line-aware attention module and a cross-attention module for capturing local geometric features; (c) a hierarchical Integrator, which integrates point-level and instance-level information through an integrator attention module and an aggregation module; (d) an Instance-to-Instance Topology Reasoning module; and (e) a Point-to-Instance Topology Reasoning module.

note the maximum number of centerlines, the number of points per centerline, and the number of channels, respectively. In contrast to previous approaches that rely solely on N instance-level queries, TopoHR augments the representation with an additional set of NP point-level queries, enabling joint modeling of global semantics local geometric details. The hierarchical query consists of two interrelated components: the point-level queries $\mathbf{Q}_{\text{pts}} \in \mathbb{R}^{P\times C}$ and the instance-level queries $\mathbf{Q}_{\text{ins}} \in \mathbb{R}^{C}$, which are refined through a series of attention mechanisms and a hierarchical integrator to enhance feature representations at both levels.

The hierarchical topology module performs topology reasoning by integrating updated hierarchical queries with P2I attention weights $\mathbf{W}_{\text{p2i}} \in \mathbb{R}^{NP\times N}$ from Integrator Attention and I2I attention weights $\mathbf{W}_{\text{i2i}} \in \mathbb{R}^{N\times N}$ from Topo-Aware Attention. This facilitates precise and semantically enriched topological structure understanding through both fine-grained point-level interactions and global instance-level relationships. The hierarchical L2T topology module adopts a similar design to extend the modeling to the relationships between centerlines and traffic elements.

3.2. Iterative Cyclic Enhancement

Conventional topology reasoning approaches optimize centerline detection and topology reasoning independently. Breaking this limitation, TopoLogic [7] incorporates centerline geometric priors with topology reasoning.

Different from TopoLogic, TopoHR adopts a cyclic interaction mechanism that enables iterative refinement, shown in Fig. 2. Specifically, P2I and I2I attention weights, $\mathbf{W}_{\text{p2i}}$ and $\mathbf{W}_{\text{i2i}}$, are passed from the hierarchical centerline decoder (Sec. 3.3) to the hierarchical topology reasoning module (Sec. 3.4). Conversely, the topological relations $\mathbf{T}_{\text{p2i}}$ (P2I) and $\mathbf{T}_{\text{i2i}}$ (I2I) inferred by the topology reasoning module are fed back to the centerline decoder. This bidirectional information flow establishes a mutually reinforcing loop, consequently improving the overall performance.

3.3. Hierarchical Centerline Decoder

The HC Decoder is designed to capture both local geometric details and global semantic information of centerline structures through a unified, multi-level representation, as illustrated in Fig. 3. Each centerline instance is modeled by a set of point-level queries encoding fine-grained spatial positions, as well as an instance-level query representing the overall semantic embedding of the entire centerline. By integrating information across these two levels, the decoder produces rich feature interactions via specialized attention mechanisms and facilitates effective aggregation and propagation of contextual cues. The architecture comprises three key modules, each tailored to address specific aspects of centerline detection and topology reasoning.

- The **Instance-Aware Module** processes instance-level queries combining masked-attention and topo-aware attention. The masked-attention leverages a discrete distance transform mask (Section 3.3.1) to focus on relevant spatial regions, while the topology-aware attention utilizes the I2I topological relation mask to enhance global semantic feature extraction among centerline instances.
- The **Point-Aware Module** refines point-level queries by

combining line-aware attention and cross-attention. The line-aware attention is constrained by a point-to-point (P2P) relation mask, enabling intra-instance interactions while preventing information leakage across different centerlines. The cross-attention integrates features from multiple points to enhance local geometric representation.

- The **Hierarchical Integrator Module** encodes both point- and instance-level information. It employs an integrator attention mechanism, where point-level queries interact with instance-level ones guided by the P2I topological relation mask. Then, an aggregation operation updates the instance-level queries based on the refined point-level features, ensuring coherent propagation of local and global information.

3.3.1. Discrete Distance Transform Mask

Direct binary centerline segmentation presents inherent drawbacks: a centerline is artificially defined virtual object without distinct pixel-level features. Unlike segmentation labels, which are uniform for all positive samples, the distance transform encodes spatial proximity. It effectively addresses the challenge of centerline segmentation without relying on specific visual features. Instead of directly using distance transform, we convert distance values to discrete ones to enhance computational efficiency and feature representation. Formally, given a centerline $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_P\}$ of P points, we first calculate the Euclidean distance from each pixel point $\mathbf{b}$ in the BEV map to the nearest centerline point $\mathbf{c}_i \in \mathcal{C}$. The distances are clipped within the half of the lane width L_{width} and then normalized to the range $[0, 1]$, and then discretized uniformly into 6 intervals. This produces the final discrete distance transform centerline mask $\texttt{DDT}(\mathbf{b})$.

We apply the DDT mask to the masked-attention [3] in the instance-aware module and in instance matching and loss calculation. The DDT mask can better extract the position information of the centerline and converge more easily than the standard regression method.

3.3.2. Hierarchical Relation Modeling

P2P Constrained Relation. In the point-aware module (Figure 3(b)), we employ a line-aware attention mechanism to facilitate intra-instance interaction among point-level queries. Specifically, each instance-level query incorporates P associated point-level queries (*e.g.*, $P = 11$), resulting in NP point queries following sequential ordering. To construct the fixed P2P constrained relation $\mathbf{M}_{\text{p2p}} \in \mathbb{R}^{NP \times NP}$, we assign 0 to elements representing points within the same centerline instance and 1 to others. This matrix serves as an attention mask within the line-aware attention computation. Consequently, this mechanism constrains attention operations within individual centerline instances, blocks cross-instance interactions, and thus maintains instance-specific feature learning by preventing information exchange between different centerlines.

P2I & I2I Topological Relation. The inherent structural dependency and topological constraints in centerlines are particularly suitable for relation-aware feature learning. Thus motivated, we use relational learning to capture the relation between centerline detection and topology reasoning, as inspired by Relation DETR [10].

To use topology reasoning for improving the centerline detection, we introduce a relation encoder to process topological predictions from the previous layer. It employs 3-layer MLPs to generate two topological relations: (1) I2I topological relation $\mathbf{M}_{\text{i2i}} \in \mathbb{R}^{N \times N}$ and (2) P2I topological relation $\mathbf{M}_{\text{p2i}} \in \mathbb{R}^{NP \times N}$. These relations are subsequently integrated as attention masks into the topo-aware attention and integrator-attention modules. Our work represents the first attempt to establish a mutually reinforcing loop between centerline detection and topology reasoning, where each component iteratively enhances the performance of the other through information exchange and joint optimization.

3.3.3. Hierarchical Integrator

The hierarchical decoder facilitates multi-scale interactions between point-wise geometric features and instance-aware semantics in centerline modeling. Through a local-global attention architecture, it iteratively refines feature representations by fusing positional details with semantic contexts, enabling coherent feature propagation across scales.

As shown in the Figure 3(c), the hierarchical integrator module operates in two cross-level steps. Given point-level queries $\mathbf{Q}_{\text{pts}} \in \mathbb{R}^{NP \times C}$ updated from the point-aware module, and instance-level queries $\mathbf{Q}_{\text{ins}} \in \mathbb{R}^{N \times C}$ generated from the instance-aware module, the integrator runs in a cross-attention way. Specifically, point-level queries $\mathbf{Q}_{\text{pts}}$ act as Q, while instance-level queries $\mathbf{Q}_{\text{ins}}$ serve as K and V. The integrator attention is formulated as:

$$\begin{aligned} \mathbf{Q} = \mathbf{Q}_{\text{pts}}\mathbf{W}^Q, \;\; \mathbf{K} = \mathbf{Q}_{\text{ins}}\mathbf{W}^K, \;\; \mathbf{V} = \mathbf{Q}_{\text{ins}}\mathbf{W}^V \\ \hat{\mathbf{Q}}_{\text{pts}} = \text{softmax}(\mathbf{Q}\mathbf{K}^T + \mathbf{M}_{\text{p2i}})\mathbf{V}, \end{aligned} \tag{1}$$

where $\mathbf{M}_{\text{p2i}}$ denotes the P2I topological relation. This allows each point-level query to selectively focus on relevant instance-aware features, thereby enriching its representation with both local and global contexts. The refined point-level queries are then aggregated into the updated instance-level queries through learnable coefficients $\mathbf{W}_{\text{agg}} \in \mathbb{R}^P$:

$$\hat{\mathbf{Q}}_{\text{ins}} = \sum_{p=1}^{P} \hat{\mathbf{Q}}_{\text{pts}}[:, p, :] \; \text{softmax}(\mathbf{W}_{\text{agg}})_p. \tag{2}$$

This design ensures that point-level queries are enriched with both local positional details and global semantic context, while instance-level queries are dynamically updated through aggregating refined point-aware features.

3.4. Hierarchical Topology Reasoning

Existing topology reasoning approaches mostly depend on instance level query interactions, often modeled via simple MLPs. In contrast, we propose a hierarchical representation strategy that incorporates both point-to-instance and instance-to-instance topological relationships. Our key insight is grounded in the hierarchical nature of topological relationship: when considering the topology between two centerlines $\mathcal{C}_i$ and $\mathcal{C}_j$, the relationship must exist not only in the instance-level representation of the two centerlines, but also between the point-level representation of $\mathcal{C}_i$ and the instance-level representation of $\mathcal{C}_j$. This hierarchical perspective is equally applicable to the topological relationships between centerlines and traffic elements.

The hierarchical module comprises two prediction branches: (1) the I2I topology branch (Fig. 3(d)) that leverages instance-level queries and their self-attention weights through dual MLP encoding for similarity computation (inner product) and associative relationship extraction via attention-based MLP, and (2) the P2I topology branch (Fig. 3(e)) that jointly incorporates explicit feature correlations and latent dependencies. This design enables comprehensive topology reasoning by effectively integrating instance-aware features and their hidden dependencies.

Specifically, the I2I prediction process is as follows:

$$\begin{aligned} \mathbf{Q}_{\text{ins}}^{\text{sim1}}, \mathbf{Q}_{\text{ins}}^{\text{sim2}} &= \text{MLP}(\mathbf{Q}_{\text{ins}}), \text{MLP}(\mathbf{Q}_{\text{ins}}) \\ \mathbf{T}_{\text{i2i}} &= (\mathbf{Q}_{\text{ins}}^{\text{sim1}}(\mathbf{Q}_{\text{ins}}^{\text{sim2}})^{\top}) + \text{MLP}(\mathbf{W}_{\text{i2i}}). \end{aligned} \tag{3}$$

The P2I prediction follows a similar workflow, except for an averaging operation along the point dimension:

$$\begin{aligned} \mathbf{Q}_{\text{pts}}^{\text{sim}}, \mathbf{Q}_{\text{ins}}^{\text{sim3}} &= \text{MLP}(\mathbf{Q}_{\text{pts}}), \text{MLP}(\mathbf{Q}_{\text{ins}}) \\ \mathbf{T}_{\text{p2i}} &= (\mathbf{Q}_{\text{pts}}^{\text{sim}}(\mathbf{Q}_{\text{ins}}^{\text{sim3}})^{\top}) + \text{MLP}(\mathbf{W}_{\text{p2i}}). \end{aligned} \tag{4}$$

The final result is derived from the results of two hierarchical components $\mathbf{T}_{\text{i2i}}$ and $\mathbf{T}_{\text{p2i}}$.

3.5. Training Loss

Adaptive Topological Loss. We propose an adaptive topological loss $\mathcal{L}_{\text{topo}}$ for topology reasoning. We employ a dynamic weighting strategy based on reparameterized cross-entropy, where negative sample weights follow exponential scaling $e^{\lambda_{\text{neg}} x_i}$ with x_i denoting predicted positive probability, while positive samples maintain fixed weighting λ_{pos}. This mechanism creates self-adaptive gradient modulation to put more penalties for negative samples with high confidence scores, effectively mitigating false positive predictions while preserving topological coherence.

In total, the training loss can be written as:

$$\mathcal{L} = \mathcal{L}_{\text{det}} + \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{topo}}, \tag{5}$$

where $\mathcal{L}_{\text{det}}$ combines a focal loss [19] for centerline detection and an ℓ_1 loss for vectorized centerline regression; and

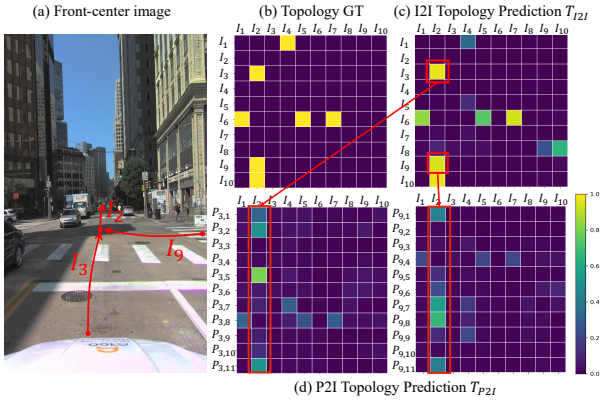

Figure 4. Instance-to-instance and point-to-instance topology reasoning results. (a) An input front-center image. (b) Groundtruth of centerline topology reasoning, where $(I_i, I_j) = 1$ denotes that the endpoint of the I_i-th centerline is connected to the start point of the I_j-th centerline. (c) Global instance-to-instance topology prediction. (d) Fine-grained point-to-instance topology prediction.

$\mathcal{L}_{\text{seg}}$ combines a dice loss and a cross-entropy loss to guide the learning of instance-aware feature based on discrete distance transform centerline mask.

4. Experiments

4.1. Experimental Settings

Datasets and Metrics. TopoHR is evaluated on the OpenLane-V2 benchmark [28], which comprehensively integrates Argoverse2 [29] and nuScenes [1]. It comprises 2,000 scenes, partitioned into subset_A and subset_B, and provides multi-view imagery at 2 Hz along with detailed annotations for 3D centerlines, traffic elements, and their topological relationships. Specifically, subset_A consists of seven camera views, while subset_B includes six.

We adopt the official evaluation metrics as in [7], including DET_{l} (mean Fréchet distance across matching thresholds), DET_{t} (IoU-based similarity for traffic elements), TOP_{ll} (similarity of the centerline topology matrix), and TOP_{lt} (centerline–traffic element topology similarity). The overall score (OLS) is computed as their average.

Implementation. We employ ResNet50 [9] and Feature Pyramid Network (FPN) [20] for multi-scale features. All input images are resized to 1024×775, as in most previous methods. Following BEVFormer [17], we project image features into the pre-defined BEV space with grid resolution 200×100. The DDT mask follows Mask2Former, initialized to zeros before the first decoder layer to avoid restricting interactions between instance queries and image features. Initial topology relations follow the TopoLogic geometric distance strategy using endpoint–startpoint distance mappings. We initialize 200 hierarchical centerline queries

Table 1. Performance comparison with other state-of-the-art methods on OpenLane-V2 subset_A benchmark. All compared methods utilize ResNet-50 as the CNN backbone. The best results are highlighted in **bold**, the second best is underline, and the third best is in *italics*. (Repr: representation of centerline; #Query: number of centerline queries.)

Method	Repr / #Query	Epoch	SD Map	DET_l	DET_t	TOP_ll	TOP_lt	OLS
TopoNet [15]	Instance / 200	24	-	28.6	48.6	10.9	23.8	39.8
SMERF [24]	Instance / 200	24	✓	33.4	48.6	15.4	25.4	42.9
TopoLogic [7]	Instance / 200	24	-	29.9	47.2	23.9	25.4	44.1
TopoLogic [7]	Instance / 200	24	✓	34.4	48.3	28.9	28.7	47.5
TopoFormer [26]	Instance / 200	24	-	34.7	48.2	24.1	29.5	46.3
SEPT [27]	Instance / 200	24	✓	34.3	*48.9*	31.2	29.7	48.4
TopoPoint [6]	Instance / 300	24	-	31.4	**55.3**	28.7	30.0	48.8
RelTopo [25]	Instance / 300	24	-	33.8	50.9	29.2	32.2	48.9
TopoHR (Ours)	Hierarchical / 200x(11+1)	24	-	36.1	48.3	*31.8*	34.6	49.9
TopoHR-L (Ours)	Hierarchical / 300x(11+1)	24	-	*35.6*	48.8	32.3	*34.3*	*49.8*
TopoHR-L (Ours)	Hierarchical / 300x(11+1)	48	-	**37.6**	47.0	**34.6**	**35.6**	**50.8**

Table 2. Performance comparison with other state-of-the-art methods on OpenLane-V2 subset_B benchmark. All compared methods utilize ResNet-50 as the CNN backbone. The best results are highlighted in **bold**, the second best is underline, and the third best is in *italics*. (Repr: representation of centerline; #Query: number of centerline queries.)

Method	Repr / #Query	Epoch	DET_l	DET_t	TOP_ll	TOP_lt	OLS
TopoNet [15]	Instance / 200	24	24.3	55.0	6.7	16.7	36.8
TopoLogic [7]	Instance / 200	24	25.9	54.7	21.6	17.9	42.3
TopoFormer [26]	Instance / 200	24	34.8	58.9	23.2	23.3	47.5
TopoPoint [6]	Instance / 300	24	31.2	**60.2**	28.3	27.1	49.2
RelTopo [25]	Instance / 300	24	32.6	*58.8*	31.8	25.8	*49.7*
TopoHR (Ours)	Hierarchical / 200x(11+1)	24	*35.3*	55.0	*32.1*	25.2	49.3
TopoHR-L (Ours)	Hierarchical / 300x(11+1)	24	38.4	53.5	35.8	*26.6*	50.8
TopoHR-L (Ours)	Hierarchical / 300x(11+1)	48	**43.6**	54.2	**39.7**	**28.0**	**53.4**

for both centerline detection and topology reasoning. In the centerline detector, the regression head consists of 3-layer MLPs with LayerNorm and ReLU, outputs 11×3 3D offsets per centerline. The proposed topology reasoning method focuses primarily on modeling the topological relationships of the centerlines. We follow the settings in [7] to detect traffic elements with no further changes. For a fair comparison, TopoHR is trained using 8 NVIDIA 4090 GPUs with a total batch size of 8 over 24 epochs. For TopoHR-L, which requires higher GPU memory, experiments are conducted on 8 NVIDIA A100 GPUs. For optimization, we adapt the AdamW optimizer [23] with an initial learning rate of 3×10^{-4} and a weight decay of 0.01. It is worth noting that TopoHR achieves 12.6 FPS on a single RTX 4090.

4.2. Comparisons with State-of-the-art Methods

We evaluate TopoHR against state-of-arts on OpenLane-V2 (Table 1). Both TopoHR and TopoHR-L substantially outperform existing methods on subset_A in DET_l, TOP_ll and TOP_lt. Notably, the TopoHR with 200 centerline queries achieves a TOP_ll score of 31.8, surpassing RelTopo [25] by 2.6 points. Given the increased complexity of subset_A, TopoHR-L exhibits limited performance gain during early training stages. To ensure full convergence of TopoHR-L, we extend its training to 48 epochs. As a result, TopoHR-L achieves improvements of 3.8 in DET_l and 5.4 in TOP_ll. Since TopoHR employs the same traffic decoder as TopoLogic [7], their performance in DET_t is nearly identical. These results clearly demonstrate that our cyclic framework, combined with P2I relation modeling, significantly enhances both centerline detection and topology reasoning.

We further assess TopoHR on subset_B (Table 2). TopoHR achieves outstanding performance in topology reasoning, attaining a notable 35.3 DET_l and outperforming RelTopo that utilizes a larger number of centerline queries. Under the same number of centerline queries, TopoHR-L maintains a clear advantage over RelTopo, achieving 43.6 DET_l and 39.7 TOP_ll. Additional experimental and qualitative results are provided in Suppl. Materials.

4.3. Ablation Study

Four ablation experiments are conducted on OpenLane-V2 subset_A to evaluate contributions of individual components. To construct a memory-efficient baseline, we adapt TopoLogic [7] by removing its GNN module. As TopoHR adopts the same traffic decoder as TopoLogic, traffic detec-

Table 3. Ablation of hierarchical centerline representation. (P2P Constrain: P2P constrained relation mask; DT: distance transform mask; DDT: discrete distance transform mask.)

Repr	P2P Constrain	Hierarchical Integrator	Seg GT	DET_{l}	TOP_{ll}
Ins	-	-	-	26.8	23.1
Ins+Pts	-	-	-	19.6	24.9
Ins+Pts	✓	-	-	20.1	25.1
Ins+Pts	✓	✓	-	32.2	26.3
Ins+Pts+Seg	✓	✓	0/1	32.6	29.8
Ins+Pts+Seg	✓	✓	DT	32.0	29.8
Ins+Pts+Seg	✓	✓	DDT	34.6	30.6
Inprovement	-	-	-	7.8↑	7.5↑

Table 4. Ablation of cyclic pipeline and P2I relation using hierarchical centerline representation and adaptive topological loss.

Forward Query	Forward Weight	Backward Topo	DET_{l}	TOP_{ll}	TOP_{lt}
$\mathbf{Q}_{\text{ins}}$	-	-	34.8	31.0	31.1
$\mathbf{Q}_{\text{ins}}$	$\mathbf{W}_{\text{i2i}}$	$\mathbf{T}_{\text{i2i}}$	35.6	31.5	32.9
$\mathbf{Q}_{\text{ins}}$+$\mathbf{Q}_{\text{pts}}$	$\mathbf{W}_{\text{i2i}}$+$\mathbf{W}_{\text{p2i}}$	$\mathbf{T}_{\text{i2i}}$+$\mathbf{T}_{\text{p2i}}$	36.1	31.8	34.6
Inprovement	-	-	1.3↑	0.8↑	3.5↑

tion metrics are omitted from the ablation study. All experiments use 200 centerline queries with focal loss for topology reasoning and are trained for 24 epochs by default.

Hierarchical Centerline Representation. We evaluate various centerline representations using the topology reasoning module of TopoLogic [7]. As reported in Table 3, incorporating point-level queries reduces DET_{l} to 19.6, while TOP_{ll} slightly increases. Introducing the point-to-point (P2P) constrained relation, which enables intra-instance interaction, leads to a modest performance improvement. Including the hierarchical integrator yields a significant boost, achieving 32.2 DET_{l} and 26.3 TOP_{ll}, underscoring its critical role in ensuring coherent propagation of both local and global information. Further, we enhance the representation by incorporating masked-attention and supervision with a segmentation loss. Using binary segmentation GT achieves 32.6 DET_{l} and 29.8 TOP_{ll}, while DT yields similar results. Our DDT achieves the best performance, bringing 7.8 DET_{l} and 7.5 TOP_{ll} gain over the baseline. These results collectively validate that each hierarchical component contributes meaningfully to the overall performance. It is also noteworthy that the incorporation of point-level queries and DDT segmentation within the hierarchical centerline representation only result in a 13.8% increase in parameters.

Cyclic Pipeline and P2I Relation. Table 4 demonstrates the effectiveness of cyclic information flow and hierarchical relation modeling, using our proposed hierarchical centerline representation and adaptive topological loss. Using only instance-level queries without cyclic feedback yields 34.8 DET_{l} and 31.0 TOP_{ll}. Introducing I2I attention weights in the forward path and backward feedback of I2I topological result improves the DET_{l} and TOP_{ll} by 0.8 and 0.5, demonstrating the benefit of iterative cyclic refinement. Finally, incorporating both instance- and point-level queries, along with both I2I and P2I relations in the cyclic pipeline, achieves the best results of 36.1 DET_{l}, 31.8 TOP_{ll} and 34.6 TOP_{lt}. These results confirm that the proposed cyclic architecture, especially with P2I relation modeling, significantly enhances both detection and topology reasoning performance. We present detailed qualitative results in Figure 4, which illustrate that point-to-instance topology prediction enables more fine-grained connectivity. These findings validate the hierarchical topology concept proposed in this work: topological relationships are manifested not only at the instance-level between centerlines, but also through their point-level representations.

Table 5. Ablation of adaptive topological loss in TopoHR. (FC: focal loss; Dice: dice loss; ATL: adaptive topological loss.)

Topo Loss	λ_{neg}	λ_{pos}	DET_{l}	TOP_{ll}	TOP_{lt}
FC	-	-	34.9	31.2	32.3
Dice	-	-	30.4	26.0	29.3
ATL	5	200	33.8	31.0	33.6
ATL	5	400	36.1	31.8	34.6
ATL	5	800	35.7	31.3	33.9
ATL	10	400	34.0	31.1	33.2
Inprovement	-	-	1.2↑	0.6↑	2.3↑

Adaptive Topological Loss (ATL). The efficacy ATL of our TopoHR is validated in Table 5. Replacing the standard Focal Loss with our proposed ATL results in improvements of 1.2 DET_{l} 0.6 TOP_{ll} and 2.3 TOP_{lt}. We limit the exploration of λ_{neg} and λ_{pos} combinations, as our primary objective is to validate the importance of emphasizing false negative predictions in topological loss formulation rather than pursuing hyperparameter optimization.

5. Conclusion

We present TopoHR, an end-to-end topology reasoning approach integrating cyclic detector-topology interactions and hierarchical centerline representations. TopoHR overcomes the limitations of sequential pipelines by enabling detector-topology co-evolution, while the hierarchical representation fuses multi-level centerline features through discrete distance transform. A unified hierarchical topology module simultaneously captures fine-grained point-to-instance relationships and global topological connections. Extensive experiments on OpenLane-V2 clearly validates the effectiveness of design and components of TopoHR.

References

[1] 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*, pages 11621–11631, 2020. 6

[2] Yigit Baran Can, Alexander Liniger, Danda Pani Paudel, and Luc Van Gool. Structured bird's-eye-view traffic scene understanding from onboard images. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15641–15650. IEEE Computer Society, 2021. 3

[3] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022. 5

[4] Sehwan Choi, Jungho Kim, Hongjae Shin, and Jun Won Choi. Mask2map: Vectorized hd map construction using bird's eye view segmentation masks. In *European Conference on Computer Vision*, pages 19–36. Springer, 2024. 2

[5] Wenjie Ding, Limeng Qiao, Xi Qiu, and Chi Zhang. Pivotnet: Vectorized pivot learning for end-to-end hd map construction. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3649–3659. IEEE, 2023. 3

[6] Yanping Fu, Xinyuan Liu, Tianyu Li, Yike Ma, Yucheng Zhang, and Feng Dai. Topopoint: Enhance topology reasoning via endpoint detection in autonomous driving. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*. 7

[7] Yanping Fu, Wenbin Liao, Xinyuan Liu, Hang Xu, Yike Ma, Yucheng Zhang, and Feng Dai. Topologic: An interpretable pipeline for lane topology reasoning on driving scenes. *Advances in Neural Information Processing Systems*, 37:61658–61676, 2024. 1, 2, 3, 4, 6, 7, 8

[8] Noa Garnett, Rafi Cohen, Tomer Pe'er, Roee Lahav, and Dan Levi. 3d-lanenet: End-to-end 3d multiple lane detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2921–2930. IEEE, 2019. 1

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[10] Xiuquan Hou, Meiqin Liu, Senlin Zhang, Ping Wei, Badong Chen, and Xuguang Lan. Relation detr: Exploring explicit position relation prior for object detection. In *European Conference on Computer Vision*, pages 89–105. Springer, 2024. 5

[11] M Kalfaoglu, Halil Ibrahim Ozturk, Ozsel Kilinc, and Alptekin Temizel. Topomask: Instance-mask-based formulation for the road topology problem via transformer-based architecture. *arXiv preprint arXiv:2306.05419*, 2023. 1, 3

[12] M Kalfaoglu, Halil Ibrahim Ozturk, Ozsel Kilinc, and Alptekin Temizel. Topomaskv2: Enhanced instance-mask-based formulation for the road topology problem. *arXiv preprint arXiv:2409.11325*, 2024. 1, 2

[13] Muhammet Esat Kalfaoglu, Halil Ibrahim Ozturk, Ozsel Kilinc, and Alptekin Temizel. Topobda: Towards bezier deformable attention for road topology understanding. *Neurocomputing*, page 132360, 2025. 1

[14] 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. IEEE, 2022. 2

[15] Tianyu Li, Li Chen, Huijie Wang, Yang Li, Jiazhi Yang, Xiangwei Geng, Shengyin Jiang, Yuting Wang, Hang Xu, Chunjing Xu, et al. Graph-based topology reasoning for driving scenes. *Transactions on Machine Learning Research*. 1, 2, 3, 7

[16] Tianyu Li, Peijin Jia, Bangjun Wang, Li Chen, Kun Jiang, Junchi Yan, and Hongyang Li. Lanesegnet: Map learning with lane segment perception for autonomous driving. In *12th International Conference on Learning Representations, ICLR 2024*, 2024. 1, 3

[17] Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Qiao Yu, and Jifeng Dai. Bevformer: learning bird's-eye-view representation from lidar-camera via spatiotemporal transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 6

[18] Bencheng Liao, Shaoyu Chen, Xinggang Wang, Tianheng Cheng, Qian Zhang, Wenyu Liu, and Chang Huang. Maptr: Structured modeling and learning for online vectorized hd map construction. In *The Eleventh International Conference on Learning Representations*. 1, 3

[19] TY Lin, P Goyal, R Girshick, K He, and P Dollar. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 2018. 6

[20] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 6

[21] Xiaolu Liu, Song Wang, Wentong Li, Ruizi Yang, Junbo Chen, and Jianke Zhu. Mgmap: Mask-guided learning for online vectorized hd map construction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14812–14821. IEEE Computer Society, 2024. 2

[22] Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. Vectormapnet: End-to-end vectorized hd map learning. In *International Conference on Machine Learning*, pages 22352–22369. PMLR, 2023. 2

[23] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 7

[24] 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. IEEE, 2024. 3, 7

[25] Yueru Luo, Changqing Zhou, Yiming Yang, Erlong Li, Chao Zheng, Shuqi Mei, Shuguang Cui, and Zhen Li. Reltopo: Enhancing relational modeling for driving scene topology reasoning. *arXiv preprint arXiv:2506.13553*, 2025. 3, 7

[26] Changsheng Lv, Mengshi Qi, Liang Liu, and Huadong Ma. T2sg: Traffic topology scene graph for topology reasoning in autonomous driving. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17197–17206. IEEE Computer Society, 2025. 1, 2, 3, 7

[27] Muleilan Pei, Jiayao Shan, Peiliang Li, Jieqi Shi, Jing Huo, Yang Gao, and Shaojie Shen. Sept: Standard-definition map enhanced scene perception and topology reasoning for autonomous driving. *IEEE Robotics and Automation Letters*, 2025. 3, 7

[28] Huijie Wang, Tianyu Li, Yang Li, Li Chen, Chonghao Sima, Zhenbo Liu, Bangjun Wang, Peijin Jia, Yuting Wang, Shengyin Jiang, et al. Openlane-v2: A topology reasoning benchmark for unified 3d hd mapping. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 6

[29] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, et al. Argoverse 2: Next generation datasets for self-driving perception and forecasting. *arXiv preprint arXiv:2301.00493*, 2023. 6

[30] Dongming Wu, Jiahao Chang, Fan Jia, Yingfei Liu, Tiancai Wang, and Jianbing Shen. Topomlp: An simple yet strong pipeline for driving topology reasoning. *arXiv preprint arXiv:2310.06753*, 2023. 1, 2, 3

[31] Zhenhua Xu, Yuxuan Liu, Yuxiang Sun, Ming Liu, and Lujia Wang. Centerlinedet: Centerline graph detection for road lanes with vehicle-mounted sensors by transformer for hd map generation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3553–3559. IEEE, 2023. 3

[32] 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. *arXiv preprint arXiv:2411.14751*, 2024. 3

[33] Yiming Yang, Yueru Luo, Bingkun He, Erlong Li, Zhipeng Cao, Chao Zheng, Shuqi Mei, and Zhen Li. Topo2seq: Enhanced topology reasoning via topology sequence learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9318–9326, 2025. 3

[34] Gongjie Zhang, Jiahao Lin, Shuang Wu, Zhipeng Luo, Yang Xue, Shijian Lu, Zuoguan Wang, et al. Online map vectorization for autonomous driving: A rasterization perspective. *Advances in Neural Information Processing Systems*, 36:31865–31877, 2023. 2

[35] Yi Zhou, Hui Zhang, Jiaqian Yu, Yifan Yang, Sangil Jung, Seung-In Park, and ByungIn Yoo. Himap: Hybrid representation learning for end-to-end vectorized hd map construction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15396–15406. IEEE, 2024. 3

TopoHR: Hierarchical Centerline Representation for Cyclic Topology Reasoning in Driving Scenes with Point-to-Instance Relations

Supplementary Material

In this supplementary material, we provide additional analysis of the proposed TopoHR, including:

- Analysis of computational complexity.
- More ablation studies on the OpenLane-V2 subset_B split.
- More point-to-instance topology prediction results.
- Qualitative analysis on OpenLane-V2 dataset.
- Future works.

A. Analysis of Computational Complexity

Table 6 reports model size, inference speed, and accuracy on the OpenLane-V2 subset_A validation set. TopoLogic [7] serves as the original baseline. To build a memory-efficient reference, we re-implement it without the GNN module while keeping the same topology reasoning design ("TopoHR, Ins"). This reduces parameters from 62.1MB to 54.9MB and slightly improves inference speed (24.4 $\rightarrow$ 25.2 FPS), forming a lighter baseline. Building upon this baseline, introducing point-level representation ("Ins+Pts") increases parameters only marginally (54.9MB $\rightarrow$ 59.7MB) while yielding clear accuracy gains. Further adding segmentation supervision ("Ins+Pts+Seg") increases parameters by only 13.8% (54.9MB $\rightarrow$ 62.5MB) but substantially improves both $\mathrm{DET}_{\mathrm{l}}$ and $\mathrm{TOP}_{\mathrm{ll}}$. Notably, even without cyclic enhancement, this variant already outperforms TopoLogic with comparable parameters (62.1MB vs. 62.5MB), highlighting the effectiveness of the hierarchical representation. Enabling the cyclic pipeline and P2I relation further improves topology reasoning, achieving the best overall accuracy while maintaining competitive speed. For comparison, the second row increases the decoder depth of TopoLogic to match the parameter scale but yields only marginal gains, confirming that TopoHR's improvements stem from hierarchical representation and cyclic reasoning rather than model size. Overall, each component demonstrate that each architectural enhancement offers favorable accuracy–efficiency trade-offs, with the segmentation-augmented representation offering particularly strong gains for its modest parameter increase.

Table 6. Comparison with different methods on OpenLane-V2 subset_A val set. FPSs are measured on one NVIDIA RTX 4090 GPU with batch size as 1. #L: number of decoder layers.

Method	Repr	Cyclic	#L	Params	FPS	$\mathrm{DET}_{\mathrm{l}}$	$\mathrm{TOP}_{\mathrm{ll}}$
TopoLogic	Ins	-	6	62.1MB	24.4	29.9	23.9
TopoLogic*	Ins	-	12	89.5MB	10.8	30.2	24.8
TopoHR	Ins	-	6	54.9MB	25.2	26.8	23.1
TopoHR	Ins+Pts	-	6	59.7MB	23.0	32.2	26.3
TopoHR	Ins+Pts+Seg	-	6	62.5MB	13.3	34.6	30.6
TopoHR	Ins+Pts+Seg	✓	6	83.1MB	12.6	36.1	31.8

B. More Abltion Study

We similarly validated the effectiveness of each TopoHR component in the split of subset_B. The setup for the ablation study remains consistent with the Section 4.3.

Hierarchical Centerline Representation. Table 7 shows that adding point-level constraints and the hierarchical integrator already provides a clear improvement over the instance-only baseline, reaching 28.0 $\mathrm{DET}_{\mathrm{l}}$ and 25.2 $\mathrm{TOP}_{\mathrm{ll}}$. This confirms that aligning local geometric cues with global instance structure is essential. Incorporating segmentation supervision further boosts performance. Binary (0/1) masks raise the scores to 30.6 $\mathrm{DET}_{\mathrm{l}}$ and 29.6 $\mathrm{TOP}_{\mathrm{ll}}$, and Distance Transform (DT) masks offer additional gains. Our Discrete Distance Transform (DDT) mask performs best, improving the baseline by 5.9 $\mathrm{DET}_{\mathrm{l}}$ and 9.1 $\mathrm{TOP}_{\mathrm{ll}}$. Overall, each component of the hierarchical centerline representation contributes meaningfully, and the full design yields the strongest results.

Table 7. Ablation of hierarchical centerline representation on subset_B split. (P2P Constrain: P2P constrained relation mask; DT: distance transform mask; DDT: discrete distance transform mask.)

Repr	P2P Constrain	Hierarchical Integrator	Seg GT	$\mathrm{DET}_{\mathrm{l}}$	$\mathrm{TOP}_{\mathrm{ll}}$
Ins	-	-	-	25.9	21.1
Ins+Pts	✓	✓	-	28.0	25.2
Ins+Pts+Seg	✓	✓	0/1	30.6	29.6
Ins+Pts+Seg	✓	✓	DT	31.1	29.7
Ins+Pts+Seg	✓	✓	DDT	31.8	30.2
Inprovement	-	-	-	5.9↑	9.1↑

Cyclic Pipeline and P2I Relation. Table 8 evaluates the cyclic information flow and hierarchical relation modeling under our hierarchical centerline representation and adaptive topological loss. Using only instance-level queries without cyclic feedback achieves 31.8 $\mathrm{DET}_{\mathrm{l}}$ and 31.0 $\mathrm{TOP}_{\mathrm{ll}}$. Introducing I2I attention in the forward path and feeding back I2I topological predictions improves $\mathrm{DET}_{\mathrm{l}}$ and $\mathrm{TOP}_{\mathrm{ll}}$ by 2.9 and 0.8, showing the benefit of iterative refinement. The full model which combines instance- and

point-level queries with both I2I and P2I relations, achieves the best results, 35.3 $\mathrm{DET_l}$, 32.1 $\mathrm{TOP_{ll}}$, and 25.2 $\mathrm{TOP_{lt}}$. These findings confirm that cyclic modeling, especially the P2I relation, substantially strengthens both detection and topological reasoning. They further support our claim that topology emerges from complementary instance- and point-level representations.

Table 8. Ablation of cyclic pipeline and P2I relation using hierarchical centerline representation and adaptive topological loss on subset_B split.

Forward Query	Forward Weight	Backward Topo	$\mathrm{DET_l}$	$\mathrm{TOP_{ll}}$	$\mathrm{TOP_{lt}}$
$\mathbf{Q}_{\mathrm{ins}}$	-	-	31.8	31.0	22.8
$\mathbf{Q}_{\mathrm{ins}}$	$\mathbf{W}_{\mathrm{i2i}}$	$\mathbf{T}_{\mathrm{i2i}}$	34.7	31.8	23.5
$\mathbf{Q}_{\mathrm{ins}}$+$\mathbf{Q}_{\mathrm{Pts}}$	$\mathbf{W}_{\mathrm{i2i}}$+$\mathbf{W}_{\mathrm{p2i}}$	$\mathbf{T}_{\mathrm{i2i}}$+$\mathbf{T}_{\mathrm{p2i}}$	35.3	32.1	25.2
Inprovement	-	-	3.5↑	1.1↑	2.4↑

C. More Point-to-Instance Topology Prediction Results

Figure 5 provides additional qualitative examples demonstrating the effect of point-to-instance topology prediction. Compared with global instance-to-instance reasoning, the point-level formulation captures finer structural cues and yields more granular connectivity patterns. These observations further support our hierarchical topology formulation: topological relationships emerge not only between centerline instances but also through their underlying point-level representations.

D. Qualitative Analysis

We provide further qualitative results to illustrate the effectiveness of TopoHR in both detection and topological reasoning. Figures 6 and 7 show diverse driving scenes from the OpenLane-V2 validation set. For each scene, the front-view image is shown on the left and the surround-view representation on the right. Detected centerlines are drawn as solid blue curves, while traffic elements are visualized as bounding boxes. Centerlines associated with traffic elements of a specific color are highlighted in the matching color to indicate predicted topological relations. TopoHR consistently delivers strong performance across varied scenarios, producing accurate centerline and traffic-element detections alongside reliable topology reasoning for both centerline-to-centerline and centerline-to-traffic relations. These results demonstrate that the combination of our cyclic pipeline, hierarchical centerline representation, and point-to-instance relation modeling yields robust improvements in both geometric perception and structured reasoning.

E. Limitations and Future Work

Although our method demonstrates strong performance in centerline topology reasoning, building more robust and accurate associations between traffic elements and centerlines remains a challenging problem. These associations are often influenced by the complexity of the scene, occlusion, and the diverse geometric layouts encountered in real driving scenarios. Improving the stability and precision of such cross-entity reasoning will therefore be an important direction of our future work. To this end, we are exploring differentiable neural geometric interfaces that can better integrate structural and semantic cues, with the goal of further enhancing cross-modal reasoning capabilities.

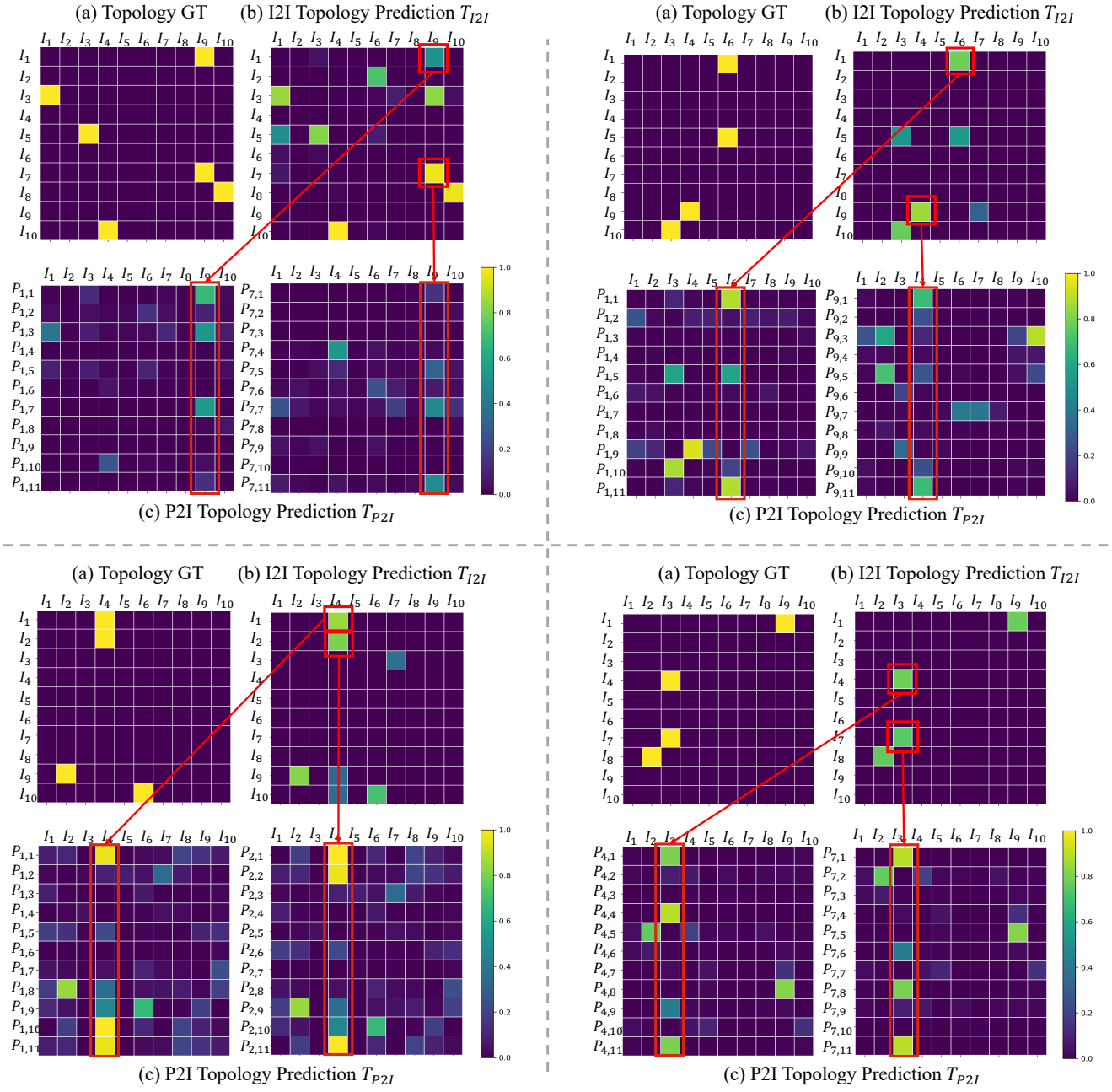

Figure 5. More instance-to-instance and point-to-instance topology reasoning results. (a) Groundtruth of centerline topology reasoning, where $(I_i, I_j) = 1$ denotes that the endpoint of the I_i-th centerline is connected to the start point of the I_j-th centerline. (b) Global instance-to-instance topology prediction. (c) Fine-grained point-to-instance topology prediction.

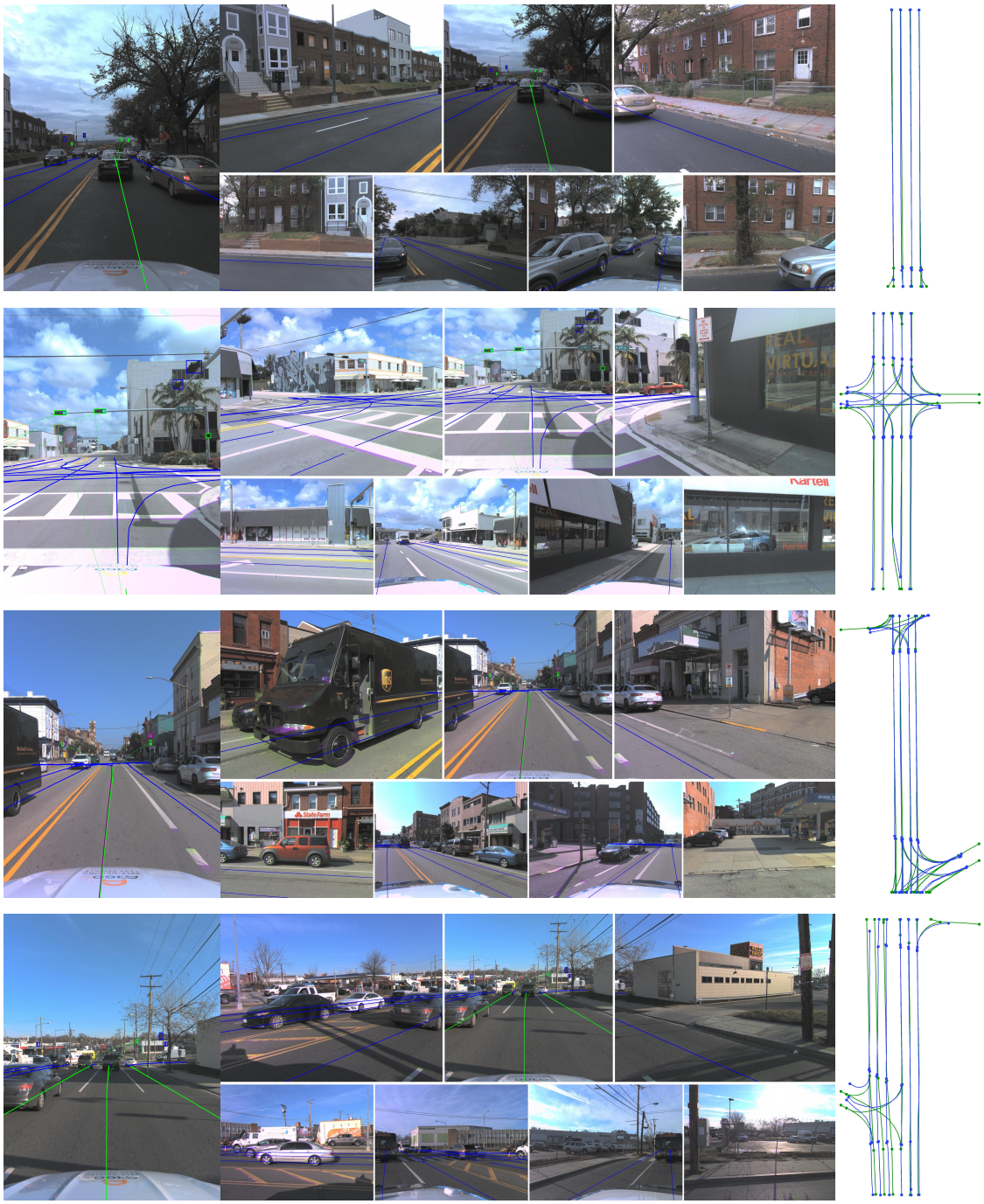

Figure 6. Illustration of detection and topology reasoning results on OpenLane-V2 validation dataset.

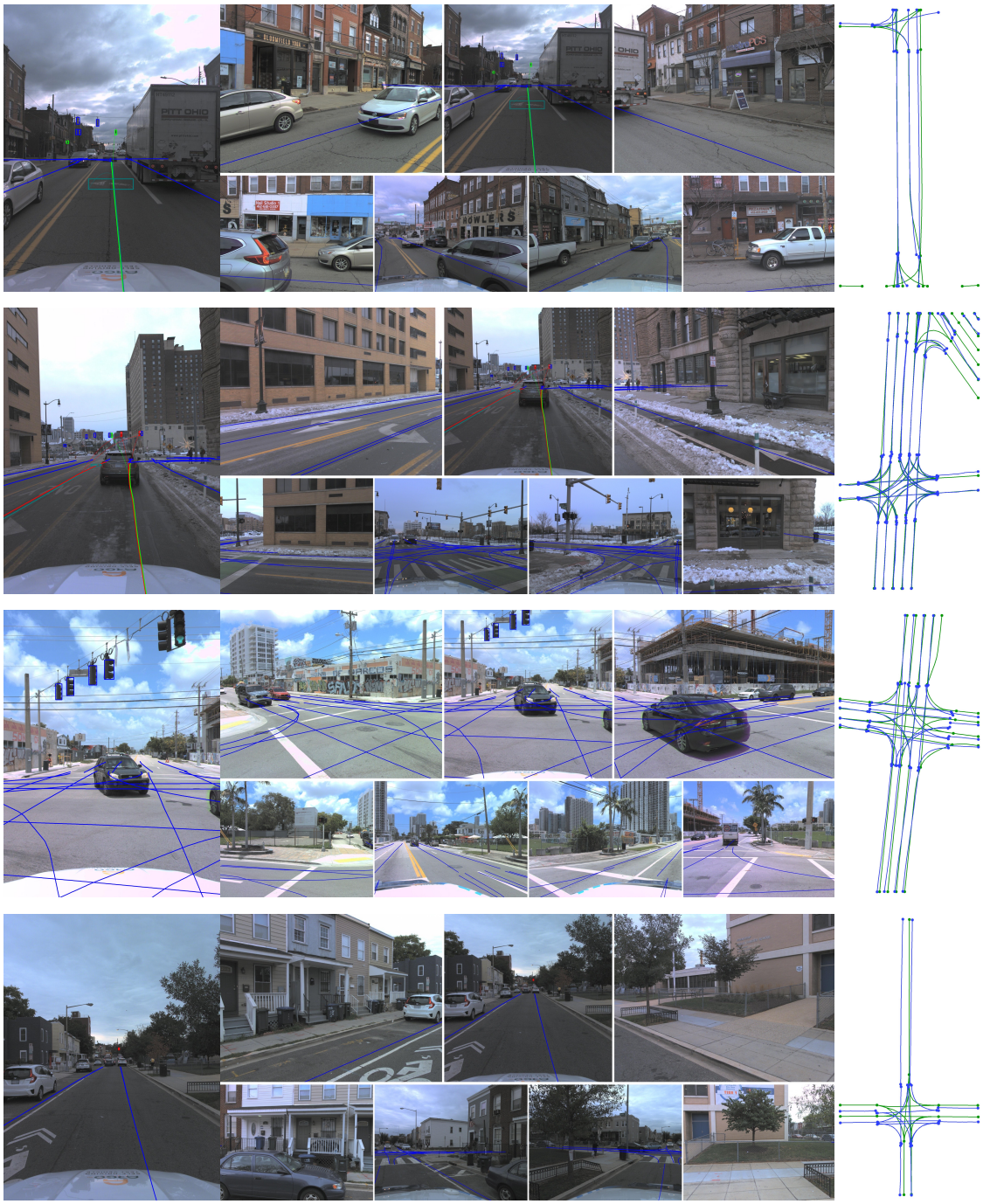

Figure 7. More qualitative results. With our proposed designs, TopoHR achieves more accurate reasoning of both centerline-to-centerline and centerline-to-traffic topological relationships.