TOPOHG: RELIABLE HYPERGRAPH REASONING FOR LANE TOPOLOGY

Anonymous authors
Paper under double-blind review

ABSTRACT

Inferring a structured road graph from surround-view images requires jointly detecting 3D lane centerlines and traffic elements, and predicting lane-to-lane and lane-to-traffic-element relations. Most query-based methods first detect individual road elements and then predict topological relations for candidate element pairs. Under this paradigm, false-positive lane queries may still receive high confidence, allowing detection errors to propagate into topology prediction and induce spurious links. Moreover, pairwise prediction does not explicitly capture group-level topology patterns spanning multiple road elements, such as lane splits, merges, and multi-lane control associations. To address these limitations, we propose TopoHG, a reliable-query-driven topology reasoning framework based on a structure-aware heterogeneous hypergraph. TopoHG combines contrastive query learning with online hard-negative mining to produce informative and discriminative lane-query representations. The resulting reliability estimates softly weight lane queries during hypergraph construction, limiting the propagation of unreliable candidates into topology reasoning. TopoHG then constructs structure-aware hypergraphs over lane and traffic-element queries, aligning with real-world split, merge, and control patterns. By integrating query semantics, lane geometry, and coarse pairwise predictions, cross-hyperedge interaction and gated bidirectional node–hyperedge message passing enable context sharing within and across structural groups. The resulting group-level context is propagated back to individual queries to refine L2L and L2T relation predictions. On OpenLane-V2, TopoHG achieves OLS scores of 50.71 on Subset-A and 48.20 on Subset-B, with consistent gains across element detection and topology tasks.

1 INTRODUCTION

Accurate perception of road elements and their topology is fundamental to motion planning, traffic-rule understanding, and safe decision making in autonomous driving. Unlike lane detection, which mainly focuses on geometric representation, lane topology reasoning must additionally recover connectivity between lanes and control relations between lanes and traffic elements. Benchmarks such as OpenLane-V2 (Wang et al., 2023a) have advanced this direction by enabling unified evaluation of element perception and structured relation reasoning directly from multi-view images.

Most recent lane-topology methods employ query-based detectors to extract lane and traffic elements, followed by graph neural networks, multilayer perceptrons, or relation-classification heads for topology prediction. TopoNet (Li et al., 2026) models element interactions through graph message passing, TopoMLP (Wu et al., 2024) performs lightweight pairwise relation classification, and TopoLogic (Fu et al., 2024) combines geometric distance with feature similarity for more interpretable connectivity estimation. TopoFG (Xu et al., 2026) further introduces fine-grained lane queries to better represent complex lane geometry. Despite this progress, existing approaches still face two key limitations: topology reasoning remains sensitive to unreliable lane queries, while pairwise relation modeling provides limited context for capturing higher-order group-level structures.

First, query-based detectors produce far more candidates than ground-truth lanes. Although Hungarian matching supervises unmatched queries as background, distractors such as stop lines, road boundaries, and ground markings may still receive high lane confidence. Since dense relation heads treat retained queries as candidate nodes, these high-confidence unmatched queries can induce spu-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

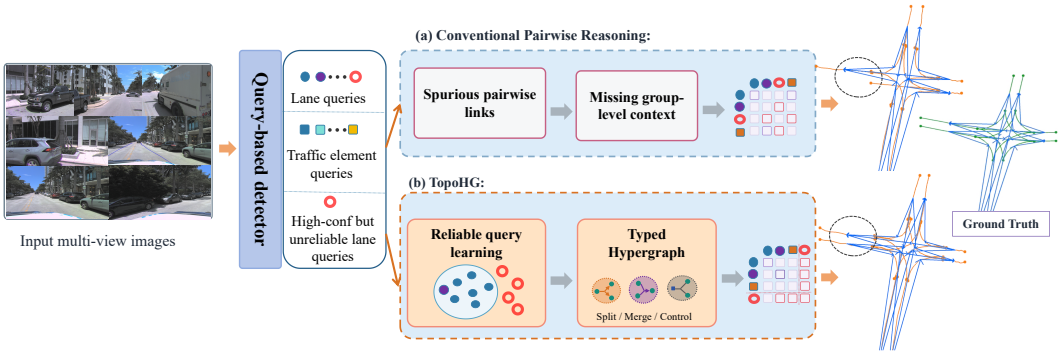

Figure 1: TopoHG: Beyond error-prone pairwise reasoning with reliable queries and heterogeneous hypergraphs. Conventional pairwise reasoning may propagate high-confidence false-positive lane queries into spurious topological links, while modeling split, merge, and traffic-control patterns only through isolated element pairs. TopoHG learns discriminative lane queries, softly weights their contributions to hypergraph construction by reliability, and reasons over typed structural groups to refine L2L and L2T relations.

rious links, while independent classification does not explicitly separate them from matched queries in feature space. CoPo (Luo et al., 2025) applies contrastive learning to lane-to-lane (L2L) and lane-to-traffic-element (L2T) relation embeddings, but distinguishes related from unrelated pairs after relation construction rather than suppressing unreliable queries beforehand. Direct auxiliary discrimination on shared decoder features may also interfere with the fine-grained geometric cues required for 3D lane regression. Therefore, improving query reliability without degrading geometry-sensitive representations remains challenging.

Second, pairwise L2L and L2T prediction decomposes group-level structures into separate binary relations. A lane split shares one source across multiple successors, a merge shares one target across multiple predecessors, and a traffic-control association may involve several lanes. These relations are structurally coupled through a common anchor and shared topology pattern. Although graph message passing enables information exchange between adjacent elements, conventional pairwise graphs do not explicitly represent such groups as unified relational units, limiting direct group-level aggregation and consistency reasoning.

To address these limitations, we propose TopoHG, a unified framework that couples reliable lane-query learning with structure-aware heterogeneous hypergraph reasoning. TopoHG first contrasts a scene-conditioned prototype of Hungarian-matched lane queries against high-confidence unmatched candidates in a dedicated projection space. Controlled auxiliary-gradient propagation limits interference with geometry-sensitive representations required for 3D lane regression, while a reliability estimator provides a soft gate for subsequent hypergraph construction. TopoHG then organizes lane and traffic-element queries into typed soft hyperedges representing split, merge, and control structures. At the final lane-decoder layer, cross-type interaction and bidirectional node–hyperedge propagation aggregate group-level context and residually refine the base L2L and L2T relation logits.

Our main contributions are summarized as follows:

- We introduce reliable lane-query learning with gradient-controlled hard-negative contrastive supervision, improving query discrimination while limiting auxiliary-gradient interference with geometry-sensitive representations.
- We formulate topology reasoning as a structure-aware heterogeneous hypergraph that explicitly captures higher-order group-level split, merge, and traffic-control dependencies beyond isolated pairwise relations.
- Experiments on both OpenLane-V2 subsets evaluate TopoHG under the official protocol. Controlled ablations separately assess reliable query learning and group-level hypergraph reasoning and reveal distinct metric-level contributions.

2 Related Work

2.1 Lane Detection and Vectorized Map Construction

Lane detection estimates lane boundaries or centerlines for autonomous-driving perception. Early methods mainly operated in the perspective view using segmentation, anchors, or parametric curves, but their 2D outputs are limited by perspective distortion and missing depth information. Representative 3D lane detectors include 3D-LaneNet (Garnett et al., 2019), Gen-LaneNet (Guo et al., 2020), PersFormer (Chen et al., 2022), and BEV-LaneDet (Wang et al., 2023b). Recent methods increasingly perform 3D lane detection in bird's-eye view (BEV), using inverse perspective mapping or Transformer-based view transformation, such as BEVFormer (Li et al., 2022b), to aggregate multi-view features. Vectorized mapping further represents road elements as compact instances. HDMapNet (Li et al., 2022a) introduced an online HD-map construction and evaluation framework, while PivotNet (Ding et al., 2023), StreamMapNet (Yuan et al., 2024), and LaneSegNet (Li et al., 2024) improve vectorized map construction through pivot-based, temporal, and lane-segment representations, respectively. VectorMapNet (Liu et al., 2023) autoregressively generates vertices, whereas MapTR (Liao et al., 2023) models each element as a permutation-equivalent ordered point set. Finer geometric representations have also been explored. TopoMaskV2 (Kalfaoglu et al., 2024) combines instance masks with directional flow supervision, while RoadPainter (Ma et al., 2024a) jointly refines point and mask representations. These advances improve geometric reconstruction but do not explicitly recover lane connectivity, motivating joint topology reasoning.

2.2 Lane Topology Reasoning

Lane topology reasoning predicts lane-to-lane (L2L) connectivity and lane-to-traffic-element (L2T) associations. STSU (Can et al., 2021) jointly detects BEV lane centerlines and classifies lane-pair connectivity, while TopoNet (Li et al., 2026) applies graph message passing among sparse road elements. TopoMLP (Wu et al., 2024) instead uses lightweight MLP relation heads with positional encoding. Subsequent methods improve relation prediction through stronger geometric or structural cues. TopoLogic (Fu et al., 2024) combines endpoint distance with semantic similarity, and TopoPoint (Fu et al., 2025) corrects topology using explicit endpoint detection and point-to-lane matching. TopoFG (Xu et al., 2026) uses fine-grained lane queries and region-focused decoding to capture local structure, whereas TopoFormer (Lv et al., 2025) learns implicit structural rules through traffic-topology scene graphs and counterfactual intervention. Despite this progress, query reliability before relation prediction and explicit group-level topology modeling remain underexplored.

2.3 Hypergraph-Based Relation Reasoning

Hypergraphs generalize pairwise graphs by allowing each hyperedge to connect multiple nodes, providing an explicit representation of higher-order dependencies. Foundational spectral formulations establish hypergraph learning for clustering, classification, and embedding (Zhou et al., 2006), while AllSet (Chien et al., 2022) provides a multiset-function view of hypergraph neural networks. Existing methods investigate incidence-based convolution (Feng et al., 2019), graph approximations of hyperedges (Yadati et al., 2019), attention-based hyperedge modeling (Zhang et al., 2020), bidirectional node–hyperedge propagation (Huang & Yang, 2021), and directed hypergraphs that preserve asymmetric roles (Ma et al., 2024b). Hypergraphs have also been applied to regional road-network representation (Zhang & Long, 2023) and group interaction modeling for motion prediction (Wang et al., 2025). These studies, however, primarily address offline road networks or dynamic agents rather than lane topology inferred from multi-view images.

Existing lane-topology methods mainly predict L2L and L2T relations from independent element pairs, without explicitly representing the group structures shared by splits, merges, and traffic-control associations. We instead construct a type-aware heterogeneous hypergraph with source-lane-anchored split, target-lane-anchored merge, and traffic-element-anchored control hyperedges. Cross-hyperedge interaction and bidirectional node–hyperedge propagation aggregate pairwise evidence into group-level context for topology refinement.

3 METHOD

3.1 OVERVIEW

As illustrated in Fig. 2, TopoHG is an end-to-end query-to-scene-graph framework that jointly predicts 3D lane centerlines, traffic elements, and L2L and L2T relations from synchronized surround-view images.

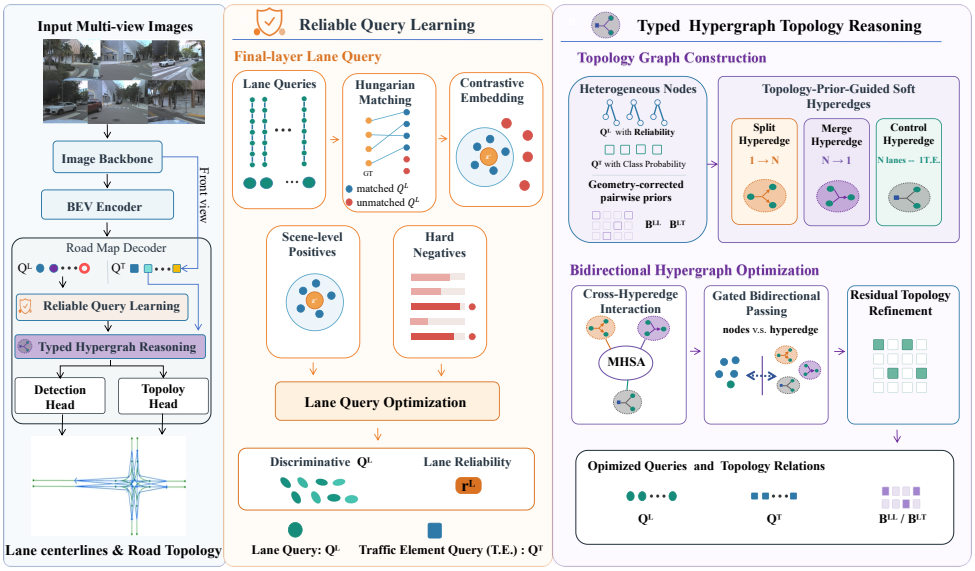

Figure 2: Overview of TopoHG. Reliable-query learning improves lane-query discrimination and supplies reliability weights for typed split, merge, and control hyperedges. Bidirectional node–hyperedge reasoning then uses group-level context to refine L2L and L2T relations.

Given multi-view images $\mathcal{I} = \{I_v\}_{v=1}^{N_{\text{cam}}}$, a shared backbone and feature pyramid extract multi-scale features. A BEV encoder produces $\mathbf{F}_{\text{B}}$, while front-view features $\mathbf{F}_{\text{F}}$ are retained for traffic-element perception. Following TopoFG (Xu et al., 2026), each lane is represented by a lane-level query and ordered point-level queries interacting with $\mathbf{F}_{\text{B}}$, while traffic-element queries interact with $\mathbf{F}_{\text{F}}$. TopoHG learns lane-query reliability and constructs typed split, merge, and control hyperedges at the final decoder layer to refine the L2L and L2T relation logits.

3.2 RELIABLE LANE QUERY LEARNING

Following the set-prediction formulation of DETR (Carion et al., 2020), Hungarian matching assigns predicted lane candidates to ground-truth lanes during training. However, unmatched queries remain in relation prediction, so high-confidence false positives may induce spurious L2L or L2T relations at inference. We therefore improve lane-query discrimination and estimate candidate reliability before topology reasoning.

3.2.1 SCENE-CONDITIONED LANE QUERY OPTIMIZATION

The i-th regular lane candidate contains ordered point-query features $\mathcal{H}_i = \{\mathbf{h}_{i,k}\}_{k=1}^{N_p}$, where $\mathbf{h}_{i,k} \in \mathbb{R}^d$ represents the k-th centerline point. We average them into $\mathbf{q}_i = N_p^{-1} \sum_{k=1}^{N_p} \mathbf{h}_{i,k}$ for reliability learning, while retaining the ordered features for fine-grained centerline regression. To decouple reliability learning from geometry decoding, we project $\mathbf{q}_i$ into a normalized reliability space:

$$\mathbf{z}_i = \text{Norm}(g(G_\alpha(\mathbf{q}_i))), \tag{1}$$

where $g(\cdot)$ is a learnable projection head and $\text{Norm}(\cdot)$ denotes ℓ_2 normalization.

Let $\mathcal{Q}_M$ and $\mathcal{Q}_U$ denote the matched and unmatched lane-query sets. The matched embeddings form a scene-conditioned valid-lane prototype $\bar{\mathbf{z}}^+$. Lane-like unmatched queries with high detached

classification confidence are retained as hard negatives, with at most the top-K candidates forming $\mathcal{N}_{\text{hard}}$. Motivated by contrastive predictive coding (van den Oord et al., 2018) and supervised contrastive learning (Khosla et al., 2020), we contrast each matched query with the scene-conditioned prototype and the selected hard negatives:

$$\mathcal{L}_{\text{r}} = -\frac{1}{|\mathcal{Q}_M|} \sum_{i \in \mathcal{Q}_M} \log \frac{\exp(\mathbf{z}_i^{\mathsf{T}} \bar{\mathbf{z}}^{+}/\tau)}{\exp(\mathbf{z}_i^{\mathsf{T}} \bar{\mathbf{z}}^{+}/\tau) + \sum_{j \in \mathcal{N}_{\text{hard}}} \exp(\mathbf{z}_i^{\mathsf{T}} \mathbf{z}_j/\tau)}. \tag{2}$$

This objective pulls matched queries toward the scene-conditioned prototype and separates them from high-confidence unmatched candidates. If no valid hard negative is available, the loss is omitted for that scene.

3.2.2 Geometry-Preserving Gradient Propagation

Although reliability learning is performed in a dedicated projection space, its gradients still propagate to the shared point-query features. Applying the full auxiliary gradient may interfere with the fine-grained information required for lane-geometry regression. We therefore use

$$G_\alpha(\mathbf{q}) = \text{sg}(\mathbf{q}) + \alpha\,(\mathbf{q} - \text{sg}(\mathbf{q}))\,. \tag{3}$$

The operator preserves the forward representation while scaling the auxiliary gradient by α.

A detached reliability head predicts p_i^{rel} and is trained with the focal-loss objective $\mathcal{L}_{\text{rel}}$ to distinguish matched from unmatched regular queries. We combine the prediction with detached lane-classification confidence to obtain the soft graph-construction gate

$$\rho_i^L = \rho_{\min} + (1 - \rho_{\min}) \sqrt{p_i^{\text{rel}} \max_{c \in \mathcal{C}_{\text{L}}} \sigma(\text{sg}(a_{i,c}))}. \tag{4}$$

A query receives a large hypergraph weight only when its reliability prediction and lane-classification confidence agree. The gate is detached before entering the hypergraph branch, preventing topology losses from modifying the contrastive and reliability representations through this side path. Detailed prototype construction, hard-negative selection, gradient analysis, and reliability supervision are provided in Appendix A.1.

3.3 Typed Hypergraph Topology Reasoning

Pairwise predictors provide local evidence but do not explicitly model the dependencies shared by lane splits, lane merges, and traffic-control associations. We therefore organize lane and traffic-element queries into a full-scene topology graph represented by typed soft hyperedges. The hypergraph is applied only at the final decoder layer, while earlier layers retain their original pairwise predictions.

3.3.1 Full-Scene Topology Graph Construction

At the final decoder layer, regular lane and traffic-element queries form the heterogeneous node set $\mathcal{V} = \mathcal{V}_L \cup \mathcal{V}_T$; denoising queries are excluded. Each lane node combines its pooled query representation, predicted centerline geometry, and lane-type embedding, while each traffic node combines its query representation, detached class probabilities, and traffic-element type embedding. Features entering the hypergraph branch are detached to protect the shared detection and geometry representations. The resulting features $\mathbf{V}_L$ and $\mathbf{V}_T$, together with the base logits $\mathbf{B}^{LL}$ and $\mathbf{B}^{LT}$, provide local evidence for typed hyperedge construction.

3.3.2 Structure-Aware Hypergraph Construction.

We construct three families of candidate hyperedges for split, merge, and traffic-control structures. Each lane anchors one split and one merge hyperedge, while each traffic element anchors one control hyperedge.

For each directed lane pair (i, j), a boundary-geometric feature $\mathbf{g}_{ij}$ encodes endpoint displacement, boundary distance, and tangent alignment. It produces the geometry-corrected prior $\widetilde{B}_{ij}^{LL} = B_{ij}^{LL} + \gamma_g \phi_R(\mathbf{g}_{ij})$.

We estimate the memberships of lane j in the split hyperedge anchored by lane i, lane i in the merge hyperedge anchored by lane j, and lane i in the control hyperedge anchored by traffic element k as

$$\begin{aligned} S^s_{ij} &= \phi_s([\mathbf{v}^L_i \| \mathbf{v}^L_j \| \mathbf{g}_{ij}]) + \beta\, \mathrm{sg}(\widetilde{B}^{LL}_{ij}), \\ S^m_{ij} &= \phi_m([\mathbf{v}^L_i \| \mathbf{v}^L_j \| \mathbf{g}_{ij}]) + \beta\, \mathrm{sg}(\widetilde{B}^{LL}_{ij}), \\ S^c_{ik} &= \phi_c([\mathbf{v}^L_i \| \mathbf{v}^T_k]) + \beta\, \mathrm{sg}(B^{LT}_{ik}). \end{aligned} \tag{5}$$

The detached pairwise logits provide topology priors without receiving gradients through the hyperedge-association terms.

For each traffic-element candidate, we derive a soft reliability weight ρ^T_k from its detached maximum classification confidence using the same reliability-floor principle as for lane queries.

The association probabilities are modulated by node reliability:

$$\begin{aligned} P^s_{ij} &= \sigma(S^s_{ij}/\tau_h)\rho^L_i \rho^L_j, \\ P^m_{ij} &= \sigma(S^m_{ij}/\tau_h)\rho^L_i \rho^L_j, \\ P^c_{ik} &= \sigma(S^c_{ik}/\tau_h)\rho^L_i \rho^T_k. \end{aligned} \tag{6}$$

Reliability is applied only at this graph-construction stage and is not multiplied into the final topology residual a second time.

To avoid dense and overly similar hyperedges, we retain the top-K_h non-anchor memberships to form the typed incidence matrices $\mathbf{H}^s$, $\mathbf{H}^m$, and $\mathbf{H}^c$. Reliability-weighted diagonal entries serve only as structural anchors and are excluded from split and merge supervision. The complete orientation and sparsification rules are given in Appendix A.3.

3.3.3 Bidirectional Hypergraph Optimization.

Given the typed incidence matrices, TopoHG aggregates incident nodes into type-specific hyperedges, exchanges context within each structural type and across their summary tokens, and propagates the updated context back to lane and traffic-element nodes through gated residual updates. The resulting representations predict residual corrections ΔB^{LL}_{ij} and ΔB^{LT}_{ik}. A curriculum coefficient η_e and bounded learnable scales progressively add these corrections to the base logits. The residual heads are zero-initialized to preserve the base-relation path at initialization, and reliability is not reapplied because it has already filtered hypergraph membership. Complete node-construction and bidirectional-propagation equations are provided in Appendix A.

3.4 Training Objective

The contrastive query objective $\mathcal{L}_{\mathrm{r}}$ is the batch-averaged form of Eq. (2), with samples lacking valid hard negatives contributing zero. The query-reliability objective $\mathcal{L}_{\mathrm{rel}}$ is defined in Eq. (12). Both objectives are applied only to the final lane-decoder layer.

Ground-truth L2L and L2T relations are mapped to the regular-query space through Hungarian assignment to supervise the split, merge, and control association heads. Self-membership is retained only as a structural anchor and is excluded from the split and merge supervision targets. We further assign larger weights to positive memberships belonging to genuine one-to-many structures. Additional details on target construction, group-aware weighting, loss coefficients, and the curriculum schedule are included in Appendix A.5.

The overall training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{base}} + \lambda_{\mathrm{r}}\mathcal{L}_{\mathrm{r}} + \lambda_{\mathrm{rel}}\mathcal{L}_{\mathrm{rel}} + \eta_e \sum_{r\in\{s,m,c\}} \lambda_r\, \mathrm{FL}\left(\mathbf{S}^r, \mathbf{Y}^r; \mathbf{W}^r\right), \tag{7}$$

where FL denotes the sigmoid focal loss introduced by Lin et al. (2017b), and $\mathcal{L}_{\mathrm{base}}$ contains the perception, geometry, denoising, and pairwise-topology objectives of the base framework.

4 EXPERIMENTS

4.1 DATASETS AND EVALUATION METRICS

Datasets. We evaluate our method on the lane-centerline topology task of OpenLane-V2 (Wang et al., 2023a). The benchmark contains 2,000 scene segments and is divided into Subset-A and Subset-B. Subset-A is collected from Argoverse 2 and provides seven surround-view cameras,whereas Subset-B is derived from nuScenes and contains six camera views. Each subset consists of 700 training, 150 validation, and 150 test scenes sampled at 2 Hz. Main results are evaluated on the validation splits of both Subset-A and Subset-B, while all ablation studies are conducted on the Subset-A validation split.

Evaluation metrics. Following the official OpenLane-V2 protocol, we report lane detection (DET_l), traffic-element detection (DET_t), lane-to-lane topology (TOP_{ll}), lane-to-traffic-element topology (TOP_{lt}), and the aggregate OpenLane-V2 Score (OLS):

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

Here, DET_l evaluates 3D lane-centerline detection using Fréchet-distance-based matching, while DET_t measures traffic-element detection. The two topology metrics evaluate L2L and L2T relation prediction on matched road elements. We use the official OpenLane-V2 v2.1.0 evaluator for all reported results.

Table 1: Performance comparison with existing lane-topology methods on OpenLane-V2 Subset-A and Subset-B. "–" denotes that the corresponding result is not reported. Higher values indicate better performance.

Dataset	Method	Venue	OLS↑	DET_l ↑	DET_t ↑	TOP_{ll} ↑	TOP_{lt} ↑
Subset-A	STSU	ICCV 2021	29.3	12.7	43.0	2.9	19.8
	VectorMapNet	ICML 2023	24.9	11.1	41.7	2.7	9.2
	MapTR	ICLR 2023	31.0	17.7	43.5	5.9	15.1
	TopoNet	SCIS 2026	39.8	28.6	48.6	10.9	23.8
	TopoMLP	ICLR 2024	44.1	28.5	49.5	21.7	26.9
	TopoLogic	NeurIPS 2024	44.1	29.9	47.2	23.9	25.4
	TopoFG	AAAI 2026	48.0	33.8	47.2	30.8	30.9
	TopoHG (ours)	–	**50.71**	**36.42**	**53.31**	**31.56**	**32.40**
Subset-B	STSU	ICCV 2021	–	8.2	43.9	–	–
	VectorMapNet	ICML 2023	–	3.5	49.1	–	–
	MapTR	ICLR 2023	–	15.2	54.0	–	–
	TopoNet	SCIS 2026	36.8	24.3	55.0	6.7	16.7
	TopoLogic	NeurIPS 2024	42.3	25.9	54.7	21.6	17.9
	TopoFG	AAAI 2026	45.4	30.0	53.0	27.2	21.7
	TopoHG (ours)	–	**48.20**	**32.83**	**56.05**	**29.12**	**24.96**

4.2 IMPLEMENTATION DETAILS

We use ResNet-50 (He et al., 2016) with a feature pyramid network (FPN) (Lin et al., 2017a) as the image encoder and construct a 200×100 BEV grid. The model contains 200 lane queries and 100 traffic-element queries. Each lane is represented by 11 point queries, and the lane decoder has six layers. The front-view image is scaled by 0.75, while the remaining views are scaled by 0.5. Reliable-query learning is applied to the final decoder layer with temperature $\tau = 0.5$ and gradient scale $\alpha = 0.1$. Unmatched queries with a maximum lane-class confidence above $\theta = 0.6$ are treated as hard-negative candidates, from which at most $K = 16$ are retained per image. The typed heterogeneous hypergraph head is also applied only to the final lane-decoder layer, where typed-hyperedge supervision and residual L2L/L2T refinement are performed. We train the model for 24 epochs using the AdamW optimizer (Loshchilov & Hutter, 2019) with an initial learning rate of 2×10^{-4} and a 0.1 learning-rate multiplier for the image backbone. Experiments are conducted on four NVIDIA RTX 3090 GPUs with one sample per GPU.

4.3 PERFORMANCE COMPARISON

Table 1 compares representative methods using version 2.1.0 of the official protocol, with published baselines reported under their corresponding settings. TopoHG obtains OLS scores of 50.71 on Subset-A and 48.20 on Subset-B. These results summarize the performance of the complete model, while the contributions of its individual components are examined through the controlled ablations below.

4.4 ABLATION STUDY

Progressive component ablation. Table 1 reports the published TopoFG result (48.0 OLS), whereas all ablations use our reproduced baseline (47.44 OLS) under a common training and evaluation setting. Starting from the reproduced TopoFG baseline, we sequentially introduce the front-view resolution adjustment, Reliable Query Learning (RQL), and Typed Heterogeneous Hypergraph (THG) while keeping all other settings unchanged. Table 2 therefore reports their cumulative contributions. The progressive ablation shows increasing OLS across the cumulative configurations.

Table 2: Progressive component ablation on OpenLane-V2 Subset-A. Each row cumulatively adds the corresponding component to the preceding setting. RQL and THG denote reliable query learning and the typed heterogeneous hypergraph, respectively.

Setting	DET_l	DET_t	TOP_{ll}	TOP_{lt}	OLS
TopoFG (reproduced)	*34.54*	*47.18*	*29.73*	*28.63*	*47.44*
+ Front-view scale	*34.82*	*52.74*	*30.80*	*29.99*	*49.45*
+ RQL	*36.08*	*51.86*	*31.66*	*31.69*	*50.13*
+ THG	*36.42*	*53.31*	*31.56*	*32.40*	*50.71*

Table 3: Independent component ablation on OpenLane-V2 Subset-A. All settings are constructed from the same baseline model. RQL and THG are introduced individually to evaluate their respective contributions.

Setting	DET_l	DET_t	TOP_{ll}	TOP_{lt}	OLS
TopoFG (reproduced)	*34.54*	*47.18*	*29.73*	*28.63*	*47.44*
+ RQL only	*35.87*	*47.91*	*31.07*	*29.85*	*48.54*
+ THG only	*34.32*	*46.95*	*31.11*	*29.85*	*47.92*

Starting from the reproduced TopoFG baseline with an OLS of 47.44, adjusting the front-view resolution raises OLS to 49.45. Adding RQL on top of this setting further increases OLS to 50.13, a gain of 0.68 points. The gains in DET_l, TOP_{ll}, and TOP_{lt} are consistent with the intended role of RQL in improving the lane queries used by downstream topology prediction. Adding THG further increases OLS from 50.13 to 50.71. In this configuration, DET_t and TOP_{lt} increase to 53.31 and 32.40, respectively, while TOP_{ll} changes only slightly. This incremental result suggests that THG provides additional structural information beyond RQL under the evaluated configuration. Because the components are added cumulatively, Table 2 measures incremental gains under this order; their isolated effects are examined in Table 3. Overall, the complete configuration improves OLS by 3.27 points over our reproduced TopoFG baseline.

Independent contribution of the proposed modules. To further separate the contributions of the two proposed modules, we conduct controlled ablations using our reproduced TopoFG baseline. RQL and THG are introduced individually, while the remaining training and evaluation settings are kept unchanged. As shown in Table 3, RQL alone increases OLS from 47.44 to 48.54 and improves all four reported detection and topology metrics. In particular, TOP_{ll} and TOP_{lt} increase from 29.73 and 28.63 to 31.07 and 29.85, respectively. This metric-wide improvement is consistent with the intended role of RQL in providing more reliable lane queries for downstream topology prediction.

When THG is introduced independently, OLS increases from 47.44 to 47.92. Although DET_l and DET_t change only slightly, TOP_{ll} and TOP_{lt} increase to 31.11 and 29.85, respectively. The concentration of the gains in the topology metrics is consistent with the intended structural-reasoning role of THG. Taken together, the independent ablations reveal different metric patterns for RQL and THG, while the progressive experiment shows that combining them yields a higher OLS under the evaluated setting.

4.5 Qualitative Results

We qualitatively compare TopoHG with several representative lane topology reasoning methods, including TopoNet, TopoLogic, and TopoFG. As shown in Fig. 3, the columns from left to right present the multi-view input images, predictions produced by the compared methods, our TopoHG results, and the ground truth.

In complex scenes involving lane splits, merges, and intersecting road structures, existing methods may produce incomplete lane centerlines, miss valid branches, or introduce spurious topological connections. In the shown scenes, TopoHG recovers more complete lane structures and produces fewer missing or spurious connections. Together with Table 3, these results support distinct roles: RQL improves both detection and topology metrics, consistent with enhanced lane-query reliability, whereas THG mainly improves the topology metrics, consistent with strengthened structural reasoning.

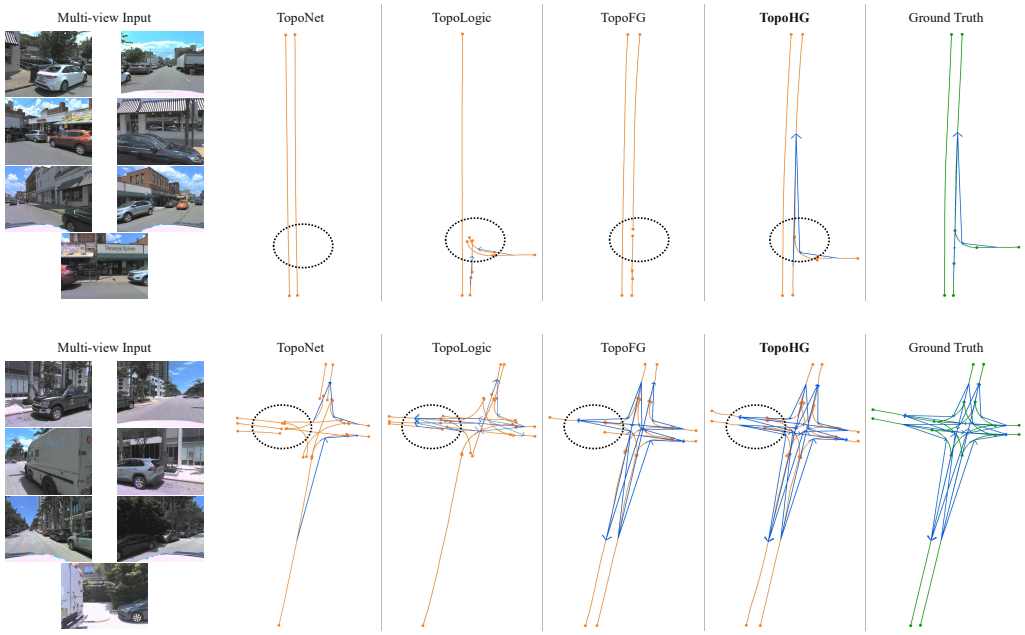

Figure 3: Qualitative comparison on OpenLane-V2. From left to right, the columns show the multi-view inputs, results of TopoNet, TopoLogic, TopoFG, our TopoHG, and the ground truth. Orange curves denote lane centerlines, and blue arrows denote lane-to-lane topological connections.

5 Conclusion

We presented TopoHG, a unified framework that improves lane-query reliability and captures group-level lane-topology dependencies through typed heterogeneous hypergraph reasoning. Experiments on both OpenLane-V2 subsets demonstrate its effectiveness, achieving an OLS of 50.71 on Subset-A and consistent improvements on Subset-B. Ablation studies further validate the complementary contributions of reliable query learning and heterogeneous hypergraph reasoning.

AI USE STATEMENT

In this work, generative AI tools were mainly used to assist with literature search and background learning. Specifically, they were used to help identify relevant prior work, locate potentially related studies, and support the authors in understanding background concepts, technical terminology, and related methods. Generative AI tools were also used for LaTeX formatting assistance, including checking syntax, refining table and equation formatting, and improving the consistency and presentation of the manuscript.

REFERENCES

Yigit Baran Can, Alexander Liniger, Danda Pani Paudel, and Luc Van Gool. Structured bird's-eye-view traffic scene understanding from onboard images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15661–15670, 2021.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *Computer Vision–ECCV 2020*, pp. 213–229, 2020.

Li Chen, Chonghao Sima, Yang Li, Zehan Zheng, Jiajie Xu, Xiangwei Geng, Hongyang Li, Conghui He, Jianping Shi, Yu Qiao, and Junchi Yan. PersFormer: 3d lane detection via perspective transformer and the OpenLane benchmark. In *European Conference on Computer Vision*, 2022.

Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are allset: A multiset function framework for hypergraph neural networks. In *International Conference on Learning Representations*, 2022.

Wenjie Ding, Limeng Qiao, Xi Qiu, and Chi Zhang. PivotNet: Vectorized pivot learning for end-to-end hd map construction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3672–3682, 2023.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3558–3565, 2019. doi: 10.1609/aaai.v33i01.33013558.

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. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-1970.

Yanping Fu, Xinyuan Liu, Tianyu Li, Yike Ma, Yucheng Zhang, and Feng Dai. TopoPoint: Enhance topology reasoning via endpoint detection in autonomous driving. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Noa Garnett, Rafi Cohen, Tomer Pe'er, Roee Lahav, and Dan Levi. 3D-LaneNet: End-to-end 3d multiple lane detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2921–2930, 2019.

Yuliang Guo, Guang Chen, Peitao Zhao, Weide Zhang, Jinghao Miao, Jingao Wang, and Tae Eun Choe. Gen-LaneNet: A generalized and scalable approach for 3d lane detection. In *European Conference on Computer Vision*, 2020.

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*, pp. 770–778, 2016.

Jing Huang and Jie Yang. UniGNN: A unified framework for graph and hypergraph neural networks. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 2563–2569, 2021. doi: 10.24963/ijcai.2021/353.

M. Esat 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.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 18661–18673, 2020.

Qi Li, Yue Wang, Yilun Wang, and Hang Zhao. HDMapNet: An online hd map construction and evaluation framework. In *2022 IEEE International Conference on Robotics and Automation*, pp. 4628–4634, 2022a.

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 *International Conference on Learning Representations*, 2024.

Tianyu Li, Li Chen, Huijie Wang, Yang Li, Jiazhi Yang, Xiangwei Geng, Hang Xu, Chunjing Xu, Junchi Yan, Ping Luo, and Hongyang Li. Graph-based topology reasoning for driving scenes. *Science China Information Sciences*, 69(5):152103, 2026. doi: 10.1007/s11432-025-4815-9.

Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Yu Qiao, and Jifeng Dai. BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. In *Computer Vision–ECCV 2022*, pp. 1–18, 2022b. doi: 10.1007/978-3-031-20077-9_1.

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 *International Conference on Learning Representations*, 2023.

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*, pp. 2117–2125, 2017a.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017b.

Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. VectorMapNet: End-to-end vectorized hd map learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22352–22369, 2023.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Yueru Luo, Changqing Zhou, Yiming Yang, Erlong Li, Chao Zheng, Shuguang Cui, and Zhen Li. Geometry-guided representations for coherent lane and traffic topology reasoning in driving scenes. *arXiv preprint arXiv:2506.13553*, 2025. Accepted to the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

Changsheng Lv, Mengshi Qi, Liang Liu, and Huadong Ma. T2SG: Traffic topology scene graph for topology reasoning in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17197–17206, 2025.

Zhongxing Ma, Shuang Liang, Yongkun Wen, Weixin Lu, and Guowei Wan. RoadPainter: Points are ideal navigators for topology transformer. In *Computer Vision–ECCV 2024*, volume 15070 of *Lecture Notes in Computer Science*, pp. 179–195, 2024a. doi: 10.1007/978-3-031-73254-6_11.

Zitong Ma, Wenbo Zhao, and Zhe Yang. Directed hypergraph representation learning for link prediction. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3268–3276, 2024b.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Chengyue Wang, Haicheng Liao, Bonan Wang, Yanchen Guan, Bin Rao, Ziyuan Pu, Zhiyong Cui, Cheng-Zhong Xu, and Zhenning Li. NEST: A neuromodulated small-world hypergraph trajectory prediction model for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 808–816, 2025. doi: 10.1609/aaai.v39i1.32064.

Huijie Wang, Tianyu Li, Yang Li, Li Chen, Chonghao Sima, Zhenbo Liu, Bangjun Wang, Peijin Jia, Yuting Wang, Shengyin Jiang, Feng Wen, Hang Xu, Ping Luo, Junchi Yan, Wei Zhang, and Hongyang Li. OpenLane-V2: A topology reasoning benchmark for unified 3d hd mapping. In *Advances in Neural Information Processing Systems*, volume 36, 2023a.

Ruihao Wang, Jian Qin, Kaiying Li, Yaochen Li, Dong Cao, and Jintao Xu. 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*, pp. 1002–1011, 2023b.

Dongming Wu, Jiahao Chang, Fan Jia, Yingfei Liu, Tiancai Wang, and Jianbing Shen. TopoMLP: A simple yet strong pipeline for driving topology reasoning. In *International Conference on Learning Representations*, 2024.

Guoqing Xu, Yiheng Li, and Yang Yang. Fine-grained representation for lane topology reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11214–11222, 2026. doi: 10.1609/aaai.v40i13.38101.

Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. HyperGCN: A new method for training graph convolutional networks on hypergraphs. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Tianyuan Yuan, Yicheng Liu, Yue Wang, Yilun Wang, and Hang Zhao. StreamMapNet: Streaming mapping network for vectorized online hd map construction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 7356–7365, 2024.

Liang Zhang and Cheng Long. Road network representation learning: A dual graph-based approach. *ACM Transactions on Knowledge Discovery from Data*, 17(9):121:1–121:25, 2023. doi: 10.1145/3592859.

Ruochi Zhang, Yuesong Zou, and Jian Ma. Hyper-SAGNN: A self-attention based graph neural network for hypergraphs. In *International Conference on Learning Representations*, 2020.

Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems*, volume 19, pp. 1601–1608, 2006.