NEURAL INFORMATION PROCESSING SYSTEMS

TopoPoint: Enhance Topology Reasoning via Endpoint Detection in Autonomous Driving

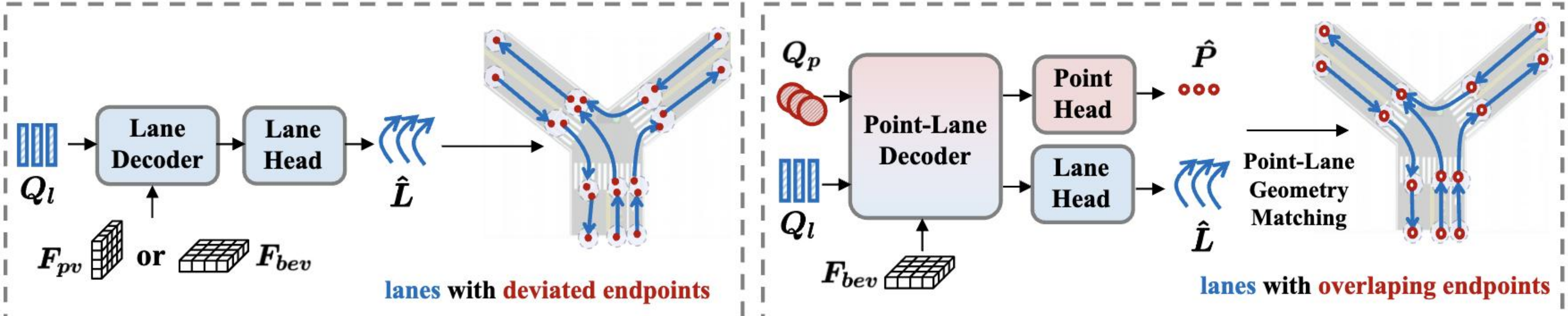
Yanping Fu[1,2,3], Xinyuan Liu[1,2], Tianyu Liu[3,4], Yike Ma[1], Yucheng Zhang[1], Feng Dai[1*]

[1]Institute of Computing Technology, Chinese Academy of Science;
[2]University of Chinese Academy of Sciences; [3]Shanghai AI Lab; [4]Shanghai Innovation Institute

fuyanping23s@ict.ac.cn

Motivation

- In the previous pipeline, lanes are predicted independently, which leads to obvious endpoint deviation.
- In our proposed pipeline, lane endpoints are explicitly modeled, and lanes with overlapping endpoints are obtained through point-lane geometry matching.



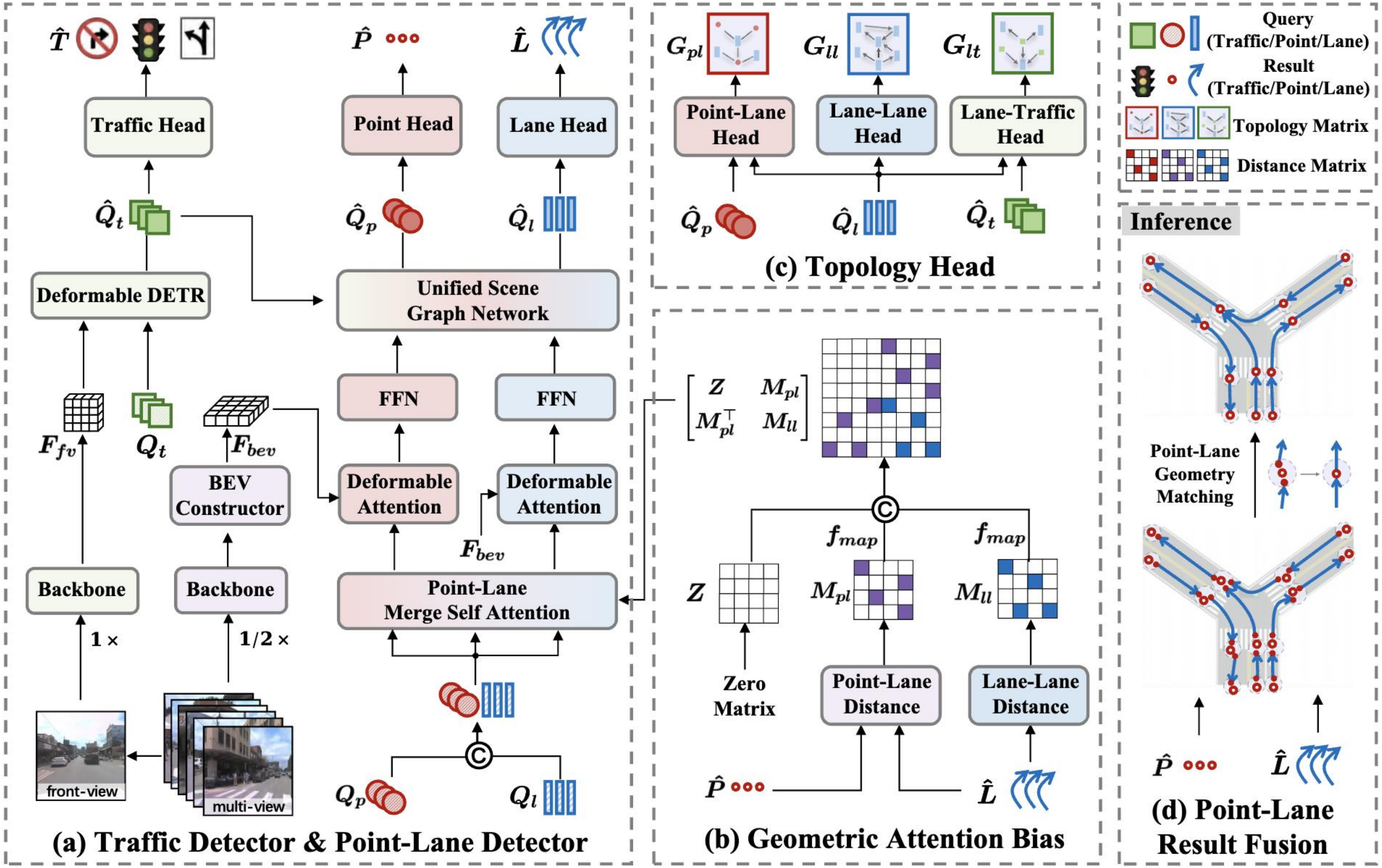
(a) The previous pipeline

(b) Our proposed pipeline

Overview

- **Traffic Detector**
- **Point-Lane Detector**
- **Topology Head**
- **Geometric Attention**
- **Point-Lane Fusion**

- The multi-view images are downsampled by a factor of 0.5, while keeping the front-view at its original resolution.
- All images are passed through ResNet50 with FPN. The features are then encoded into BEV representations using BevFormer encoder.



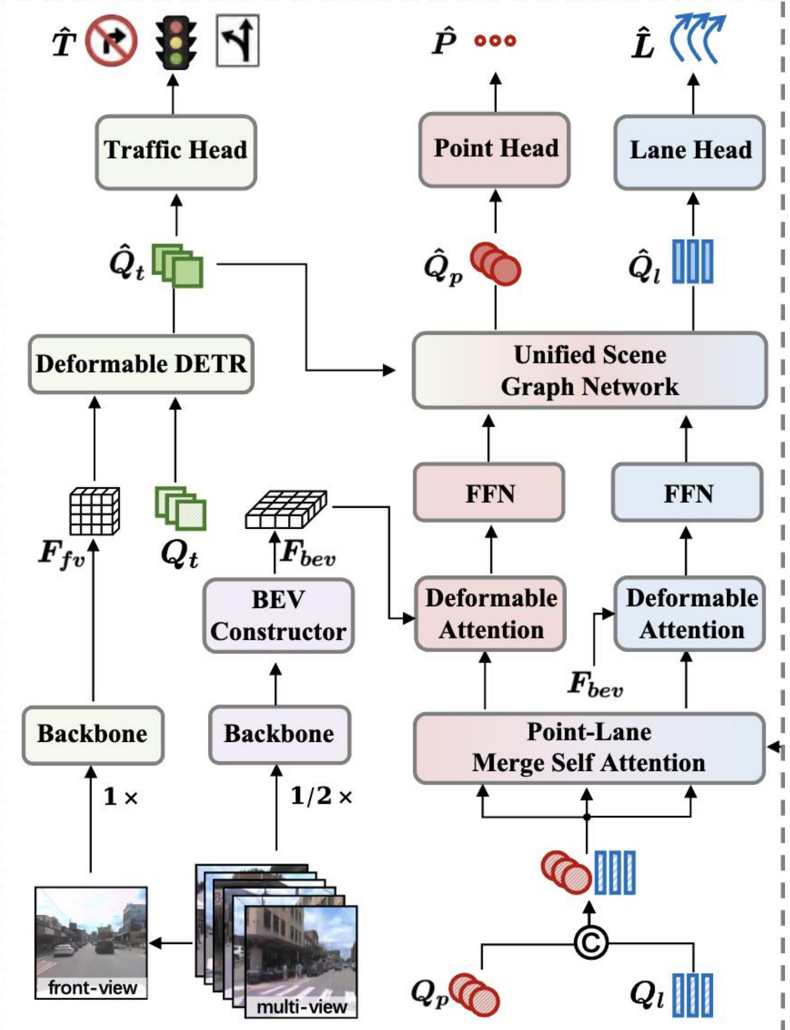
Pipeline

- **Traffic Detector:** In the traffic detector, front-view features are directly processed by Deformable DETR to produce traffic query.

$$\hat{Q}_t = \text{DeformableDETR}(Q_t, F_{fv})$$
$$\hat{T} = \text{TrafficHead}(\hat{Q}_t)$$

- **Point-Lane Detector:** In the point-lane detector, point query and lane query interact via *Point-Lane Merge Self-Attention*, which computes geometric attention bias serving as an attention mask to enhance global information sharing. The resulting queries then perform cross-attention with BEV features. Then all queries are fed into *Unified Scene Graph Network*.

$$Q_{pl} = \text{Concat}\left(Q_p, Q_l\right)$$
$$Q_p, Q_l = \text{Point-LaneMergeSelfAttention}(Q_{pl})$$
$$Q_p, Q_l = \text{LN}(\text{DeformAttn}(Q_p, R_p, F_{bev})), \text{LN}(\text{DeformAttn}(Q_l, R_l, F_{bev}))$$
$$Q_p, Q_l = \text{LN}(\text{FFN}(Q_p)), \text{LN}(\text{FFN}(Q_l))$$
$$Q_p, Q_l = \text{UnifiedSceneGraphNetwork}(Q_p, Q_l, \hat{Q}_t)$$
$$\hat{P} = \text{PointHead}(\hat{Q}_p),\ \hat{L} = \text{LaneHead}(\hat{Q}_l)$$



(a) Traffic Detector & Point-Lane Detector

Pipeline

NEURAL INFORMATION PROCESSING SYSTEMS

- **Point-Lane Attention:** The geometric attention bias is also incorporated into the point-lane merge self attention module to exchange information.
- To incorporate the geometric relationships between points and lanes in the BEV space, we compute their pairwise geometric distances based on the predicted points and lanes from the previous decoder layer
- To compute self-attention, we concatenate distance matrixes to form geometric attention bias, which is added to the attention weights computed from original quries.

$$Q_{pl} = \text{Concat}\left(Q_p, Q_l\right)$$

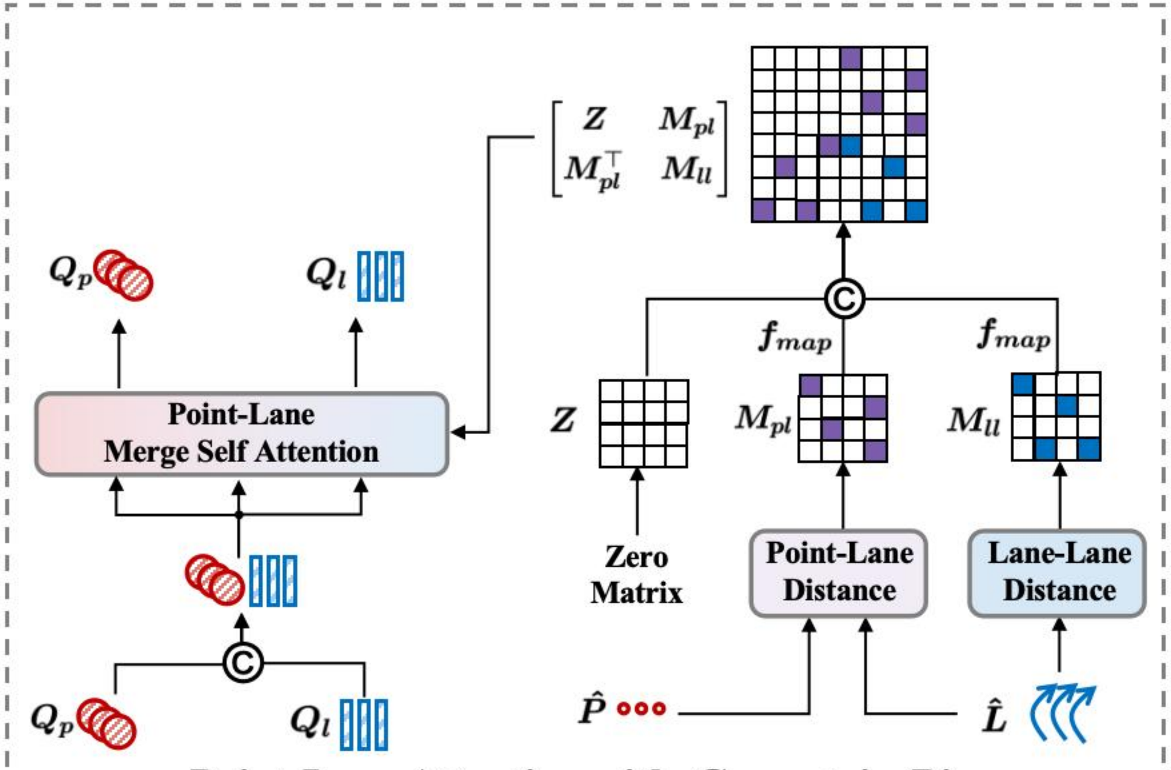
$$D_{ll} = \left\{ \sum |\hat{l}_i^e - \hat{l}_j^s| \,\middle|\, i = 1, 2, \ldots, N_p, j = 1, 2, \ldots, N_l \right\}$$

$$D_{pl} = \left\{ \text{Min}\left(\sum |\hat{p}_i - \hat{l}_j^s|, \sum |\hat{p}_i - \hat{l}_j^e| \right) \,\middle|\, i = 1, 2, \ldots N_p, j = 1, 2, \ldots N_l \right\}$$

$$M_{pl} = f_{map}(D_{pl}),\ M_{ll} = f_{map}(D_{ll})$$

$$Q_p, Q_l = \text{Softmax}\left(\frac{Q_{pl} \cdot Q_{pl}^\top}{\sqrt{d}} + \begin{bmatrix} Z & M_{pl} \\ M_{pl}^\top & M_{ll} \end{bmatrix} \right) \cdot Q_{pl}$$

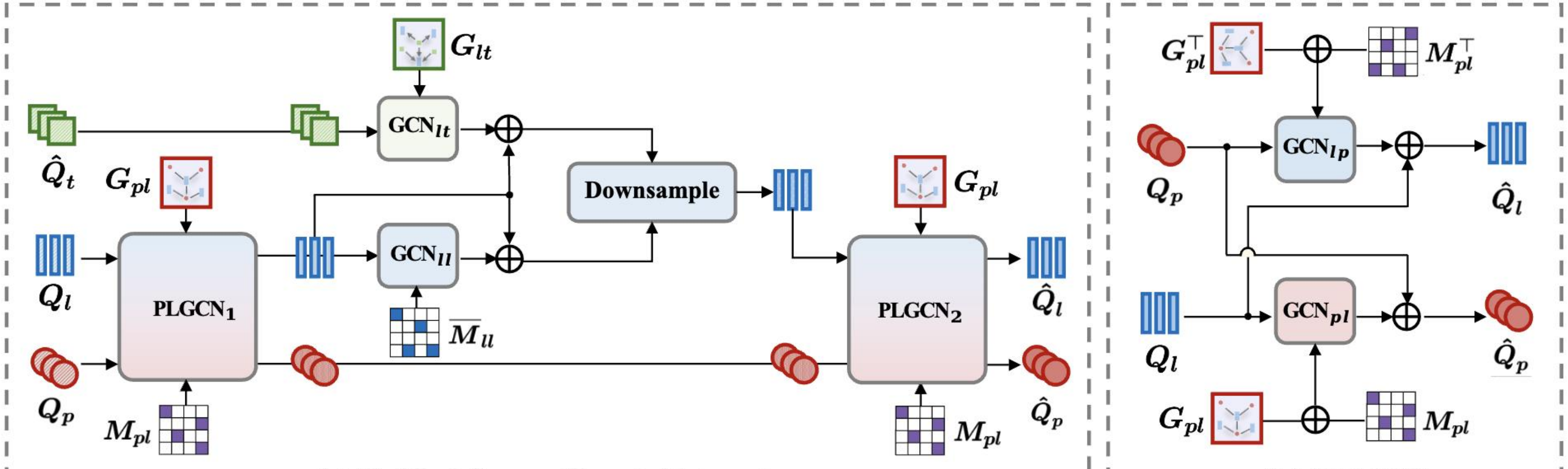
$$Q_p, Q_l = \text{LN}(Q_p), \text{LN}(Q_p)$$



Point-Lane Attention with Geometric Bias

Pipeline

- **Unified Scene Graph Network:** Based on geometric attention bias and reasoned topology, lane & point queries are enhanced from the associated traffic elements & lanes & points by the unified scene graph network.
- **PLGCN:** The submodule is designed to facilitate bidirectional feature aggregation between point and lane based on their geometric relationships.



(a) Unified Scene Graph Network (b) PLGCN

$$A_{pl} = \lambda_1 G_{pl} + \lambda_2 M_{pl}$$

$$Q_p = \mathrm{GCN}_{pl}\left(Q_l, A_{pl}\right) + Q_p,\ Q_l = \mathrm{GCN}_{lp}(Q_p, A_{pl}^\top) + Q_l$$

$$Q_p^1, Q_l^1 = \mathrm{PLGCN}_1(Q_p, Q_l, M_{pl}, G_{pl})$$

$$Q_l^2 = \mathrm{Downsample}\left(\mathrm{Concat}\left(\mathrm{GCN}_{ll}(Q_l^1, \overline{M}_{ll}) + Q_l^1,\ \mathrm{GCN}_{lt}(\hat{Q}_t, G_{lt}) + Q_l^1\right)\right)$$

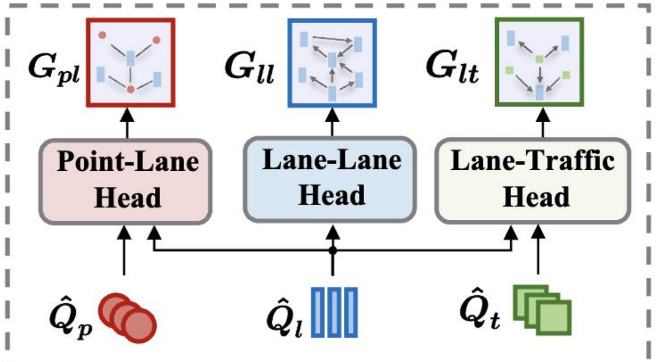
$$Q_p^3, Q_l^3 = \mathrm{PLGCN}_2(Q_p^1, Q_l^2, M_{pl}, G_{pl})$$

$$\hat{Q}_p, \hat{Q}_l = Q_p^3, Q_l^3$$

Pipeline

- **Topology Head :** The queries are used for topology reasoning, and the topology is also used for query enhancement in scene graph.

$$\hat{G}_{pl} = \mathrm{Sigmoid}(\mathrm{MLP}(\hat{Q}_p) \cdot \mathrm{MLP}(\hat{Q}_l)^\top)$$
$$\hat{G}_{ll} = \mathrm{Sigmoid}(\mathrm{MLP}(\hat{Q}_l) \cdot \mathrm{MLP}(\hat{Q}_l)^\top)$$
$$\hat{G}_{lt} = \mathrm{Sigmoid}(\mathrm{MLP}(\hat{Q}_l) \cdot \mathrm{MLP}(\hat{Q}_t)^\top)$$

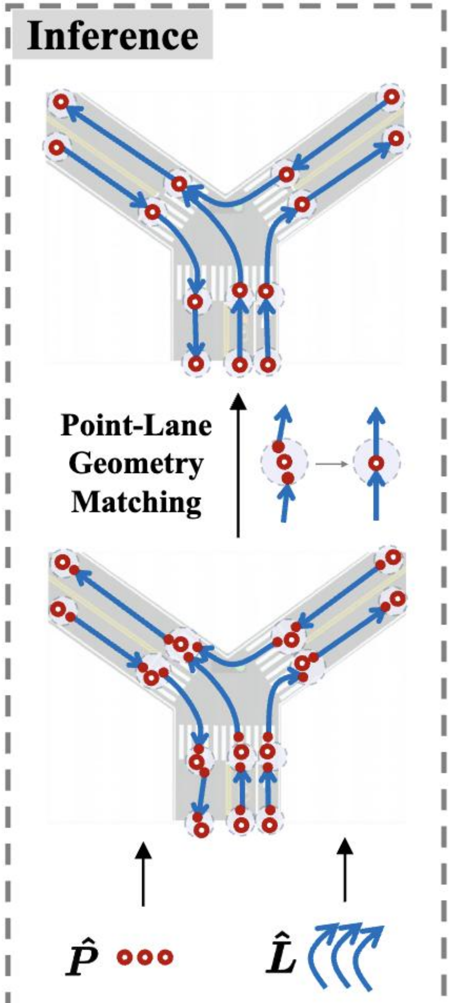


(c) Topology Head

- **PointLane Geometry Matching Algorithm:** During inference, predicted points and lanes are fused via Point-Lane Geometry Matching algorithm to refine lane endpoints and effectively mitigate the endpoint deviation problem.

Algorithm 1: Point-Lane Geometry Matching Algorithm

Input: Predicted points $\hat{P}_{reg}, \hat{P}_{cls}$; predicted lanes $\hat{L}_{reg}, \hat{L}_{cls}$; classification thresholds τ_p, τ_l; geometry distance threshold δ.
Output: Refined lanes $\hat{L}_{ref}$
Step 1: High-Confidence Filtering
Filter points with high classification scores: $\hat{P}_{select} = \{\hat{P}^i_{reg} \mid \hat{P}^i_{cls} > \tau_p\}$
Filter lanes with high classification scores: $\hat{L}_{select} = \{\hat{L}^j_{reg} \mid \hat{L}^j_{cls} > \tau_l\}$
Step 2: Geometry-Based Matching and Refinement
foreach *point* $\hat{P}_i \in \hat{P}_{select}$ **do**
  Initialize empty match set: $\mathcal{N}_i = \emptyset$;
  foreach *lane* $\hat{L}_j \in \hat{L}_{select}$ **do**
    if *distance*$(\hat{P}_i, \hat{L}_j^{endpoint}) < \delta$ **then**
      Add $\hat{L}_j$ to $\mathcal{N}_i$;
  if $\mathcal{N}_i \neq \emptyset$ **then**
    Compute refined endpoint:
    $\hat{E}_i = \frac{1}{|\mathcal{N}_i|+1}\left(\hat{P}_i + \sum_{\hat{L}_j \in \mathcal{N}_i} \hat{L}_j^{endpoint}\right)$;
    Update endpoints of all $\hat{L}_j \in \mathcal{N}_i$ with $\hat{E}_i$;
return $\hat{L}_{ref}$ *with refined endpoints*



(d) Point-Lane Result Fusion

Experiments Setup

- **Dataset:** OpenLaneV2, which is constructed based on Argoverse2 and nuScenes. OpenLane-V2 is divided into two subsets: subset_A and subset_B, each containing 1,000 scenes captured at 2 Hz with multi-view images and corresponding annotations.
- **Metric:** We adopt the evaluation metrics defined by OpenLane-V2, including DETl, DETt, TOPll, and TOPlt, all of which are computed based on mean Average Precision (mAP)

$$\mathrm{OLS} = \frac{1}{4}[\mathrm{DET}_l + \mathrm{DET}_t + \sqrt{\mathrm{TOP}_{ll}} + \sqrt{\mathrm{TOP}_{lt}}]$$

- **Point Metric:** In addition, to evaluate the performance of endpoint detection, we define a custom metric DETp, which is computed as the average over match thresholds T = {1.0, 2.0, 3.0} based on the point-wise Fréchet distance, as follows:

$$\mathrm{DET}_p = \frac{1}{|\mathbb{T}|}\sum_{t\in\mathbb{T}} AP_t$$

Main Results

NEURAL INFORMATION PROCESSING SYSTEMS

- **Comparison on OpenLane-v2 Benchmark:** New SOTA results and more precise endpoints.

Data	Method	Conference	$DET_l\uparrow$	$DET_t\uparrow$	$TOP_{ll}\uparrow$	$TOP_{lt}\uparrow$	OLS↑	$DET_p\uparrow$
subset_A	STSU[13]	*ICCV2021*	12.7	43.0	2.9	19.8	29.3	-
	VectorMapNet[10]	*ICML2023*	11.1	41.7	2.7	9.2	24.9	-
	MapTR[48]	*ICLR2023*	17.7	43.5	5.9	15.1	31.0	-
	TopoNet[26]	*Arxiv2023*	28.6	48.6	10.9	23.8	39.8	43.8
	TopoMLP[29]	*ICLR2024*	28.3	49.5	21.6	26.9	44.1	43.4
	TopoLogic[15]	*NeurIPS2024*	29.9	47.2	23.9	25.4	44.1	45.2
	TopoFormer*[31]	*CVPR2025*	**34.7**	48.2	24.1	29.5	46.3	-
	TopoPoint (Ours)	-	31.4	**55.3**	**28.7**	**30.0**	**48.8**	**52.6**
subset_B	STSU[13]	*ICCV2021*	8.2	43.9	-	-	-	-
	VectorMapNet[10]	*ICML2023*	3.5	49.1	-	-	-	-
	MapTR[48]	*ICLR2023*	15.2	54.0	-	-	-	-
	TopoNet[26]	*Arxiv2023*	24.3	55.0	6.7	16.7	36.8	38.5
	TopoMLP[29]	*ICLR2024*	26.6	58.3	21.0	19.8	43.8	39.6
	TopoLogic[15]	*NeurIPS2024*	25.9	54.7	21.6	17.9	42.3	39.2
	TopoFormer*[31]	*CVPR2025*	**34.8**	58.9	23.2	23.3	47.5	-
	TopoPoint (Ours)	-	31.2	**60.2**	**28.3**	**27.1**	**49.2**	**45.1**

Ablation Studies

- **Impact of each module:**

Module	$DET_l\uparrow$	$DET_t\uparrow$	$TOP_{ll}\uparrow$	$TOP_{lt}\uparrow$	OLS↑	$DET_p\uparrow$
Baseline	29.2	46.8	23.4	24.3	43.4	44.5
+ FVScale	29.4	53.8	23.8	27.0	46.0	44.8
+ PLMSA	30.2	54.8	27.2	28.5	47.6	49.8
+ PLGCN	30.8	55.3	28.0	29.2	48.3	51.8
+ PLGM	**31.4**	**55.3**	**28.7**	**30.0**	**48.8**	**52.6**

- **Effect of different GCNs:**

Module	$DET_l\uparrow$	$DET_t\uparrow$	$TOP_{ll}\uparrow$	$TOP_{lt}\uparrow$	OLS↑	$DET_p\uparrow$
w/o GCN	28.9	53.9	25.6	26.4	46.2	48.6
+ GCN_{ll}	29.8	54.2	26.9	27.1	47.0	49.8
+ GCN_{lt}	30.6	54.5	27.4	28.8	47.8	50.5
+ $PLGCN_1$	30.9	55.0	28.2	29.5	48.3	51.9
+ $PLGCN_2$	**31.4**	**55.3**	**28.7**	**30.0**	**48.8**	**52.6**

Ablation Studies

- **Image scales set up:**

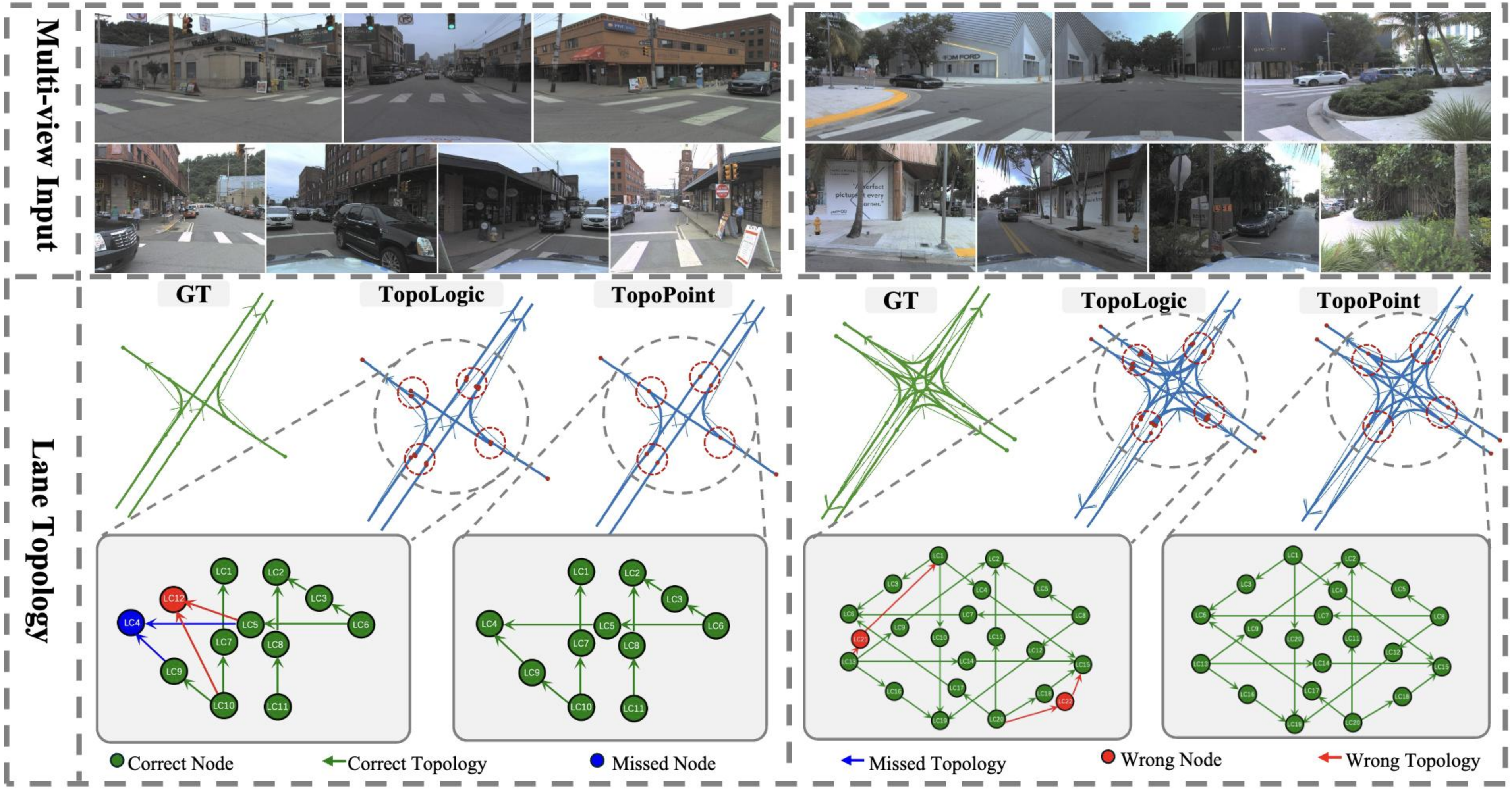
S_{fv}	S_{mv}	DET$_l$↑	DET$_t$↑	TOP$_{ll}$↑	TOP$_{lt}$↑	OLS↑	DET$_p$↑
0.5	0.5	31.2	48.6	28.5	28.4	46.6	52.3
0.5	1.0	30.5	48.3	28.0	27.9	46.1	51.5
1.0	0.5	**31.4**	**55.3**	**28.7**	**30.0**	**48.8**	**52.6**
1.0	1.0	30.8	54.7	28.3	28.9	48.1	51.8

- **Effect of point and lane query numbers:**

N_p	N_l	DET$_l$↑	DET$_t$↑	TOP$_{ll}$↑	TOP$_{lt}$↑	OLS↑	DET$_p$↑
100	200	29.5	54.3	25.6	27.0	46.5	49.7
200	200	30.7	53.7	27.4	28.2	47.5	51.8
200	300	**31.4**	**55.3**	**28.7**	**30.0**	**48.8**	**52.6**
300	300	30.8	54.6	28.2	29.8	48.3	51.4

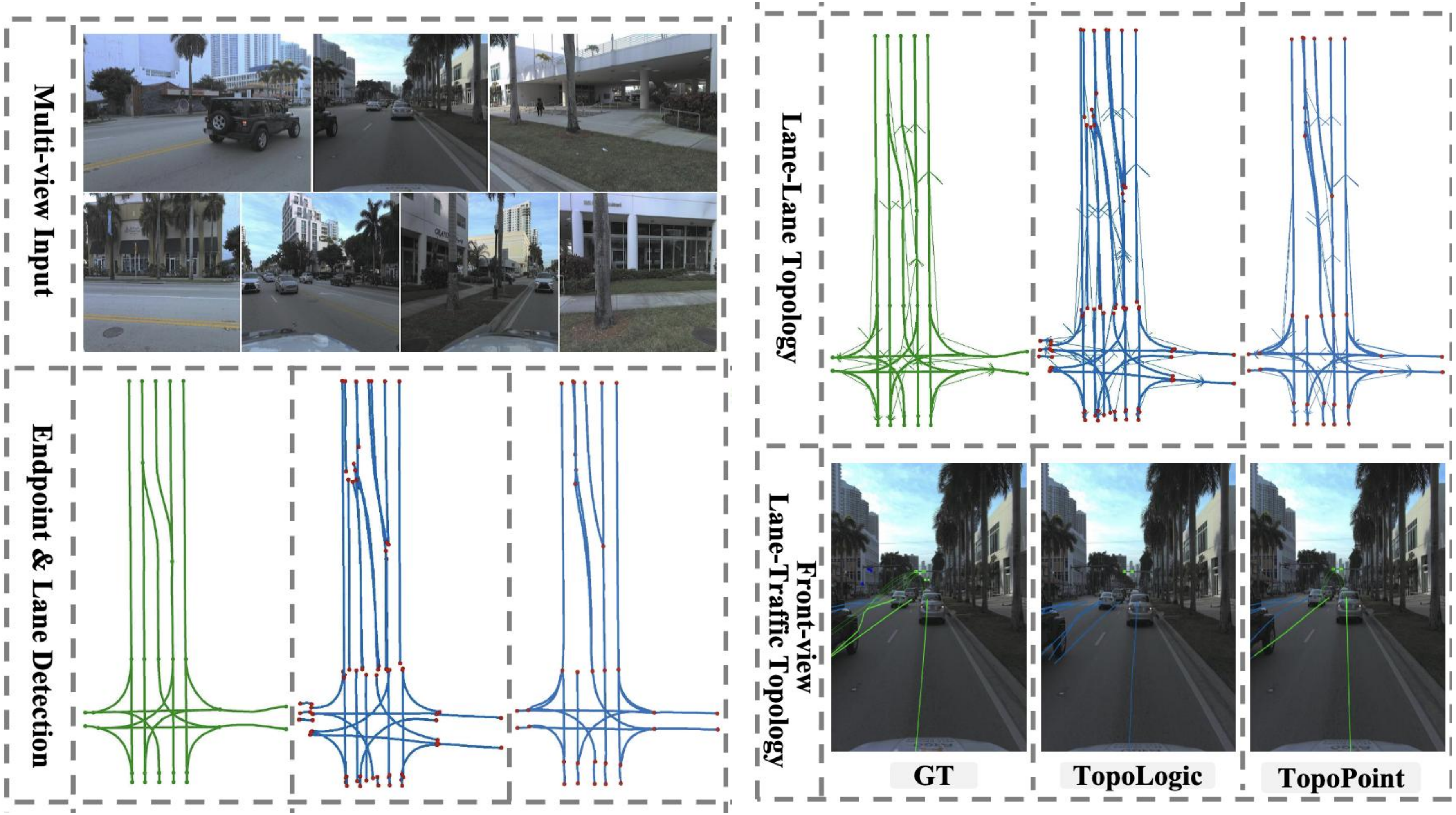
Qualitative Analysis

- **Comparison of TopoLogic and our TopoPoint:**



Qualitative Analysis

- Comparison of TopoLogic and our TopoPoint:



THANK YOU!