CVPR SEATTLE, WA JUNE 17-21, 2024

ZHEJIANG UNIVERSITY

MGMap: Mask-Guided Learning for Online Vectorized HD Map Construction

[1]Xiaolu Liu, [1]Song Wang, [1]Wentong Li, [1]Ruizi Yang, [2]Junbo Chen*, [1]Jianke Zhu*

[1]Zhejiang University, [2]Udeer.AI

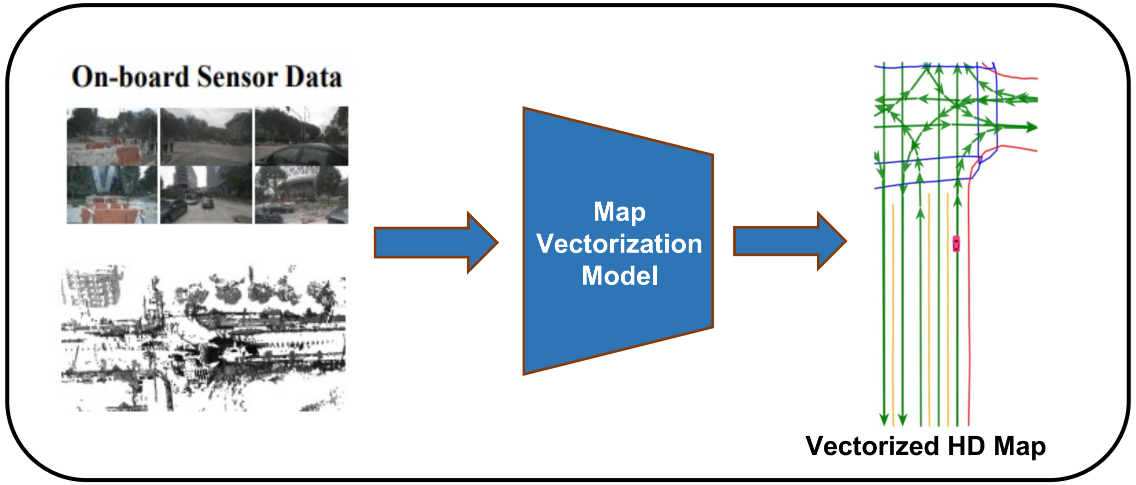
https://github.com/xiaolul2/MGMap

Poster: THU-PM-19

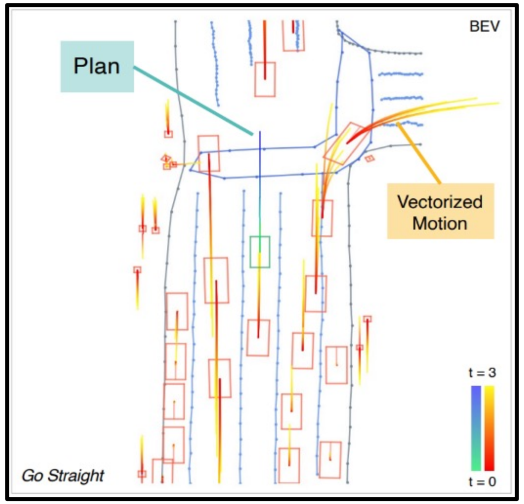
Contact: {xiaoluliu, songw, liwentong, ruiziyang, jkzhu}@zju.edu.cn, junbo@udeer.ai

Background

- Traditional offline annotation requires manual investments and lacks real-time update
- Light-weight online generation tendency for vectorized HD map construction

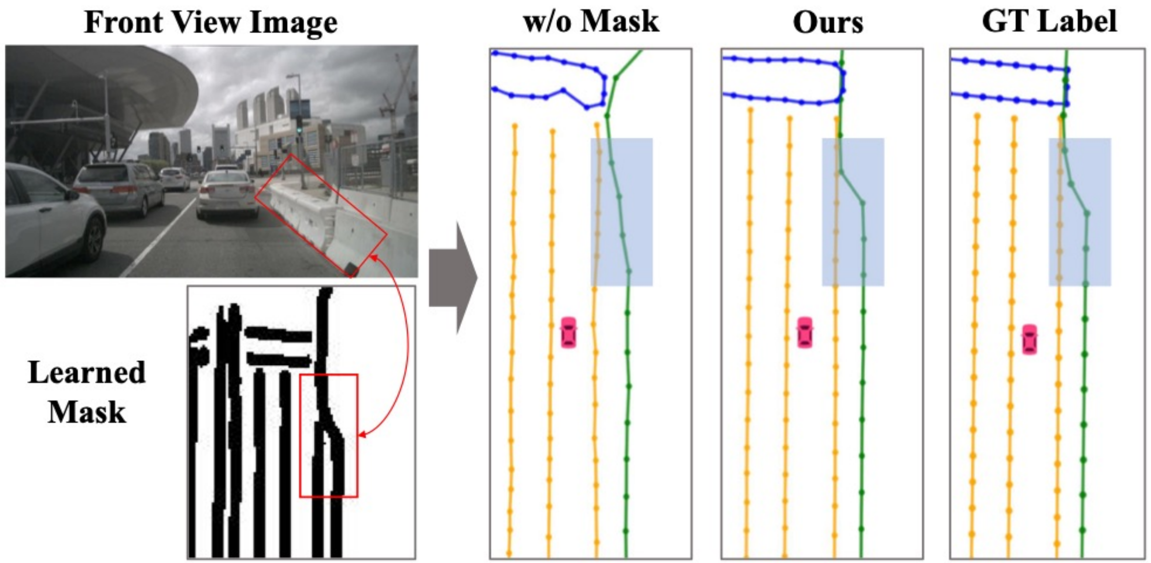


online HD map construction



map-based trajectory prediction

Motivation



☐ Current challenges

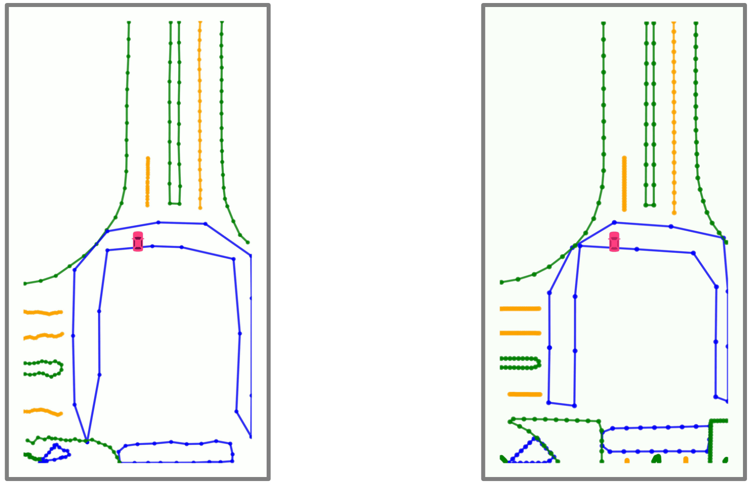
- Subtle annotations and strong shape priors for map
- Hard to capture specific structure features
- Lack detailed representation for road elements

MGMap: Improve localization and highlight specific features by

incorporating the guidance of mask information

Overview

- An efficient mask-guided learning framework for online vectorized HD map construction
- Instance and point-level mask information are introduced to highlight the effective features.
- MAI decoder and PG-MPR module are specifically designed for different level regressions.
- Our MGMap achieves superior performance and shows generalization capability.

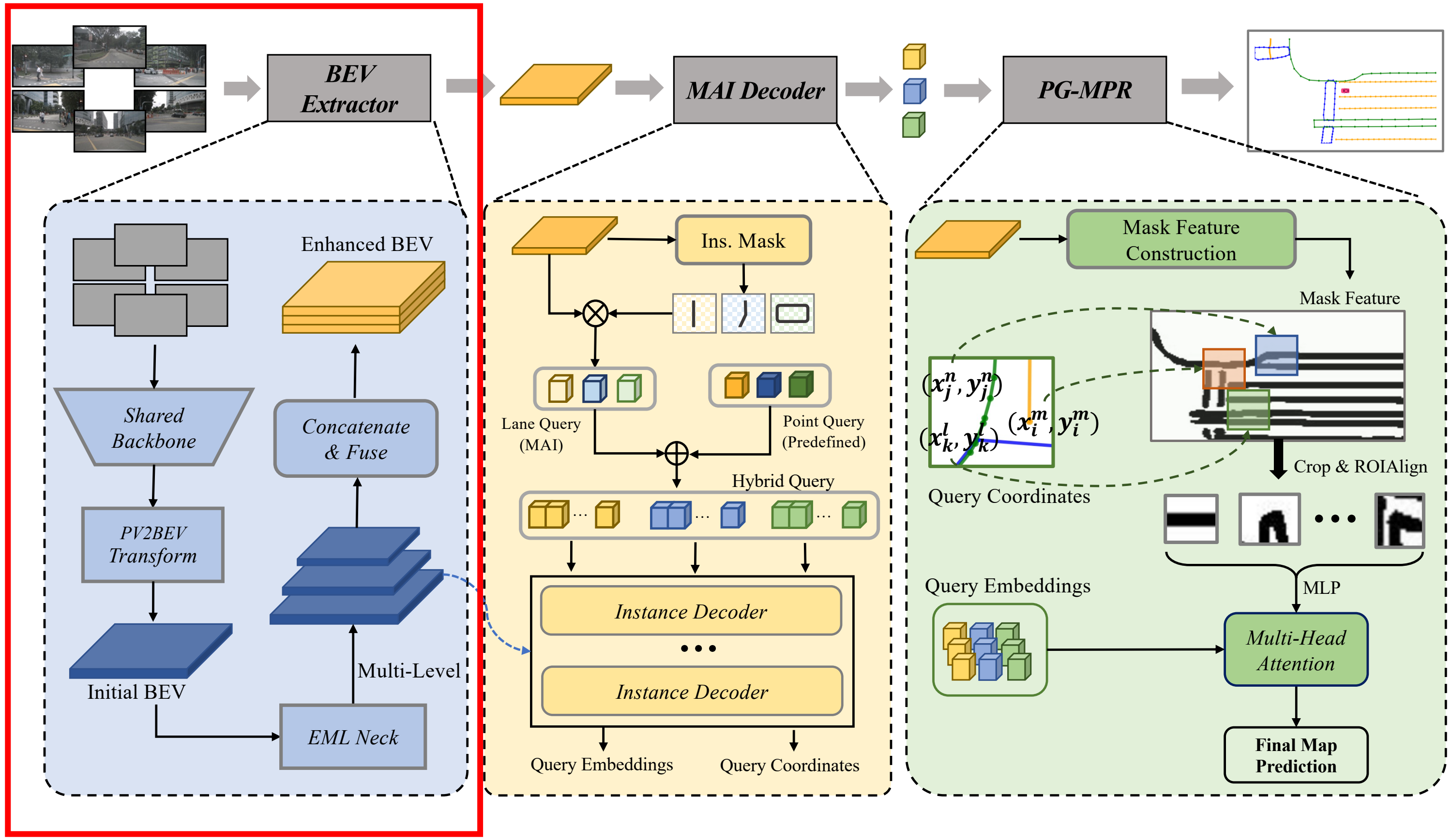


Pred. GT

Multi-View Surrounding Images

MGMap Design

➢ BEV Extractor

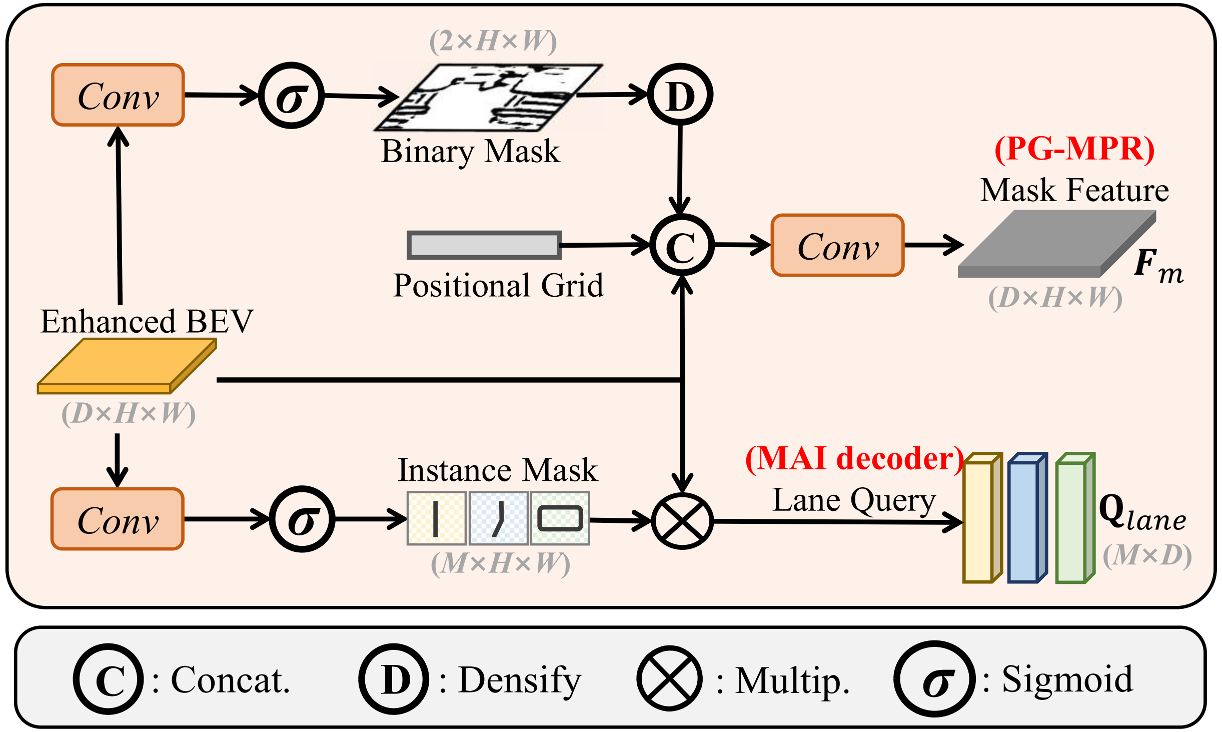


- Initial BEV feature extractor
- EML neck for multi-scale BEV with fused attention

$$\mathbf{F}_{i+1} = (\mathrm{CA}(\mathbf{F}_i) \times \mathbf{F}_i) \times \mathrm{SA}(\mathbf{F}_i).$$

MGMap Design

➢ Mask Construction

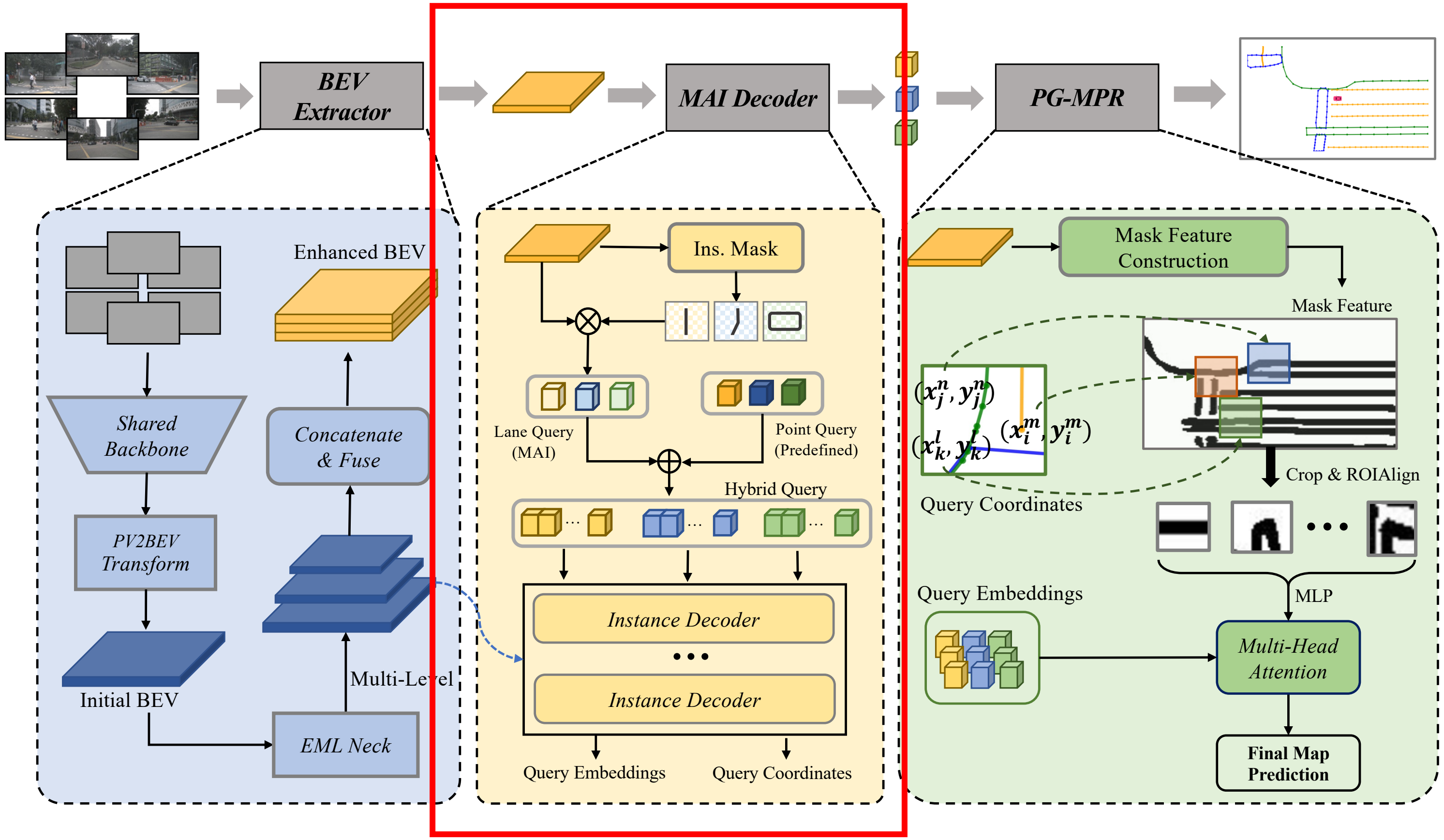


- **Instance level**: Instance mask to activate query
- **Point level**: Binary mask for patch extraction

$$\mathcal{L}_{mask} = \lambda_{ins}\mathcal{L}_{ins}(\hat{\mathbf{M}}_{ins}, \mathbf{M}_{ins}) + \lambda_{b}\mathcal{L}_{b}(\hat{\mathbf{M}}_{b}, \mathbf{M}_{b})$$

MGMap Design

➢ Instance-level: MAI Decoder



- Mask-activated instance decoder
- Activate query initialization
- Deformable Decoder to update instances

Initialize:

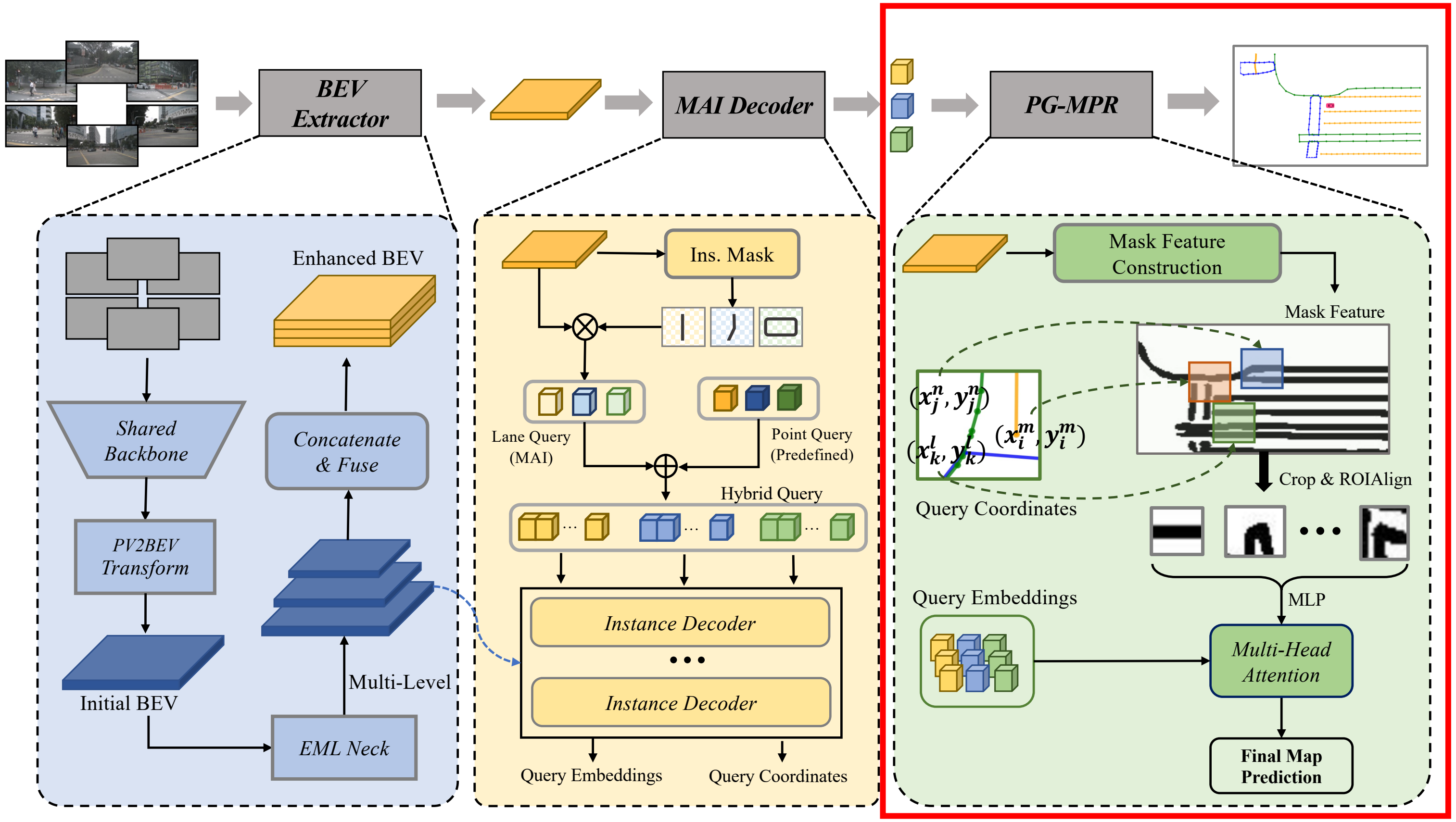
$$\mathbf{Q}_{lane} = (\sigma((\mathrm{Conv}(\mathbf{F}_c))) \times \mathbf{F}_c^{\top}$$

Update:

$$\mathbf{Q}^{l} = \mathrm{DeformAttn}(\mathbf{Q}^{l-1}, \mathbf{P}^{l}, \{\mathbf{F}_i\}_{i=1}^{3}, \mathbf{F}_c),$$
$$\mathbf{P}^{l+1} = \sigma(\sigma^{-1}(\mathbf{P}^{l}) + \mathrm{Reg}_{pos}(\mathbf{Q}^{l})).$$

MGMap Design

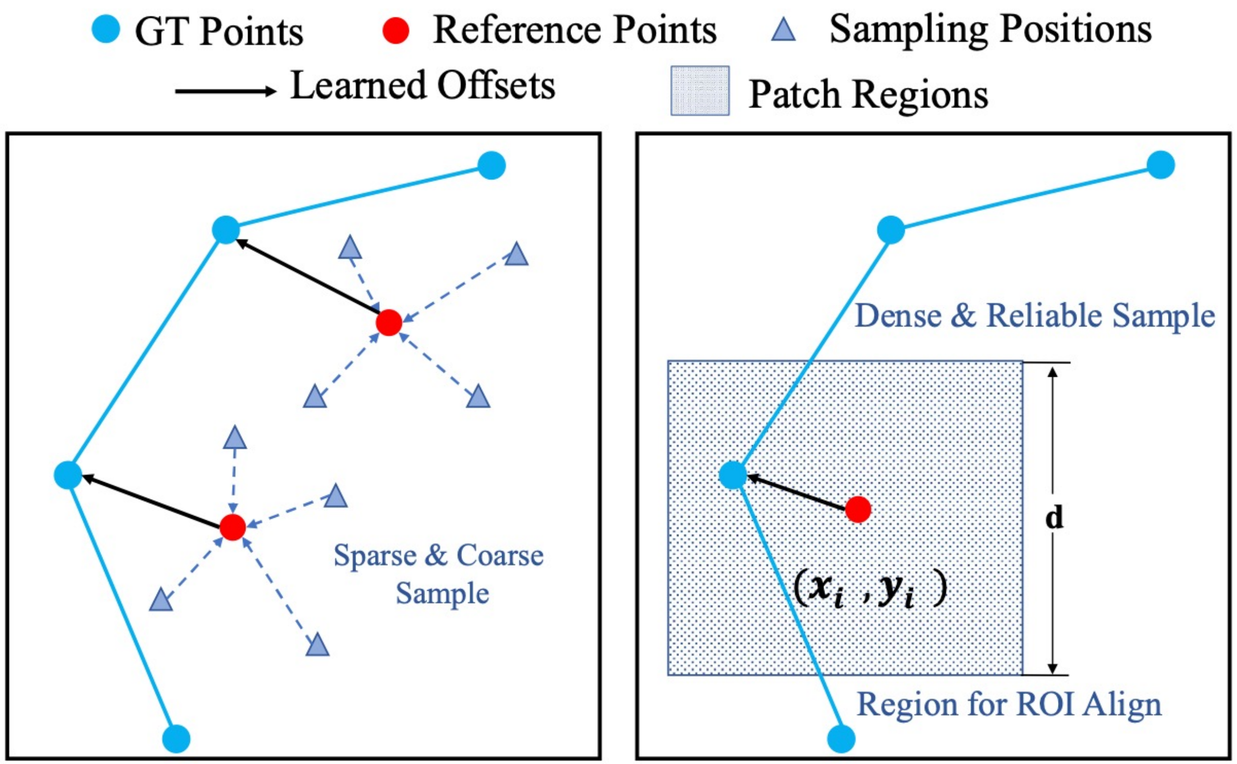
➢ **Point-level: PG-MPR**



- Position-guided mask patch refinement
- Construct binary mask feature
- ROI Align for patch feature extraction

MGMap Design

➢ Point-level: Mask Patch Extraction



(a) Original Deformable Attention (b) Mask Patch Refinement (Ours)

- dense & reliable patch feature

$$\mathbf{v}_i = f_{ext}\left(\mathbf{F}_m, \mathrm{R}(B_i), (x_i, y_i)\right)$$

$$\mathbf{Q}^{s+1} = \mathrm{MultiHeadAtt}\left(\frac{\mathbf{Q}^s\mathbf{V}^s}{\sqrt{D}}\right)$$

Performance Comparison

- Nuscenes dataset

Method	Backbone	Epochs	$AP_{chamfer}$ ped.	$AP_{chamfer}$ div.	$AP_{chamfer}$ bou.	$AP_{chamfer}$ *avg.*	AP_{raster} ped.	AP_{raster} div.	AP_{raster} bou.	AP_{raster} *avg.*	FPS
Camera-Based Methods											
HDMapNet [ICRA22] [19]	EB0	30	14.4	21.7	33.0	23.0	-	-	-	-	-
InstaGraM [CVPRW23] [39]	EB4	30	33.8	47.2	44.0	41.7	-	-	-	-	-
MapTR [ICLR23] [23]	R50	24	46.3	51.5	53.1	50.3	32.4	23.5	17.1	24.3	15.7
MapTR [ICLR23] [23]	R50	30	45.2	53.8	54.3	51.1	32.9	24.9	18.9	25.6	15.7
MapVR [NeurIPS23] [49]	R50	24	47.7	54.4	51.4	51.2	37.5	33.1	23.0	31.2	15.7
PivotNet [ICCV23] [9]	R50	30	53.8	55.8	59.6	57.4	-	-	-	-	9.6
BeMapNet [CVPR23] [32]	R50	30	**57.7**	62.3	59.4	59.8	-	-	-	-	4.4
MGMap(Ours)	R50	30	57.4	**63.5**	**63.3**	**61.4**	**46.5**	**36.5**	**28.7**	**37.2**	11.6
MapTRv2 [arxiv23] [24]	R50	24	59.8	62.4	62.4	61.5	-	-	-	-	-
MGMap*(Ours)	R50	24	**61.8**	**65.0**	**67.5**	**64.8**	-	-	-	-	-
VectorMapNet [ICML23] [26]	R50	110+ft	42.5	51.4	44.1	46.0	-	-	-	-	-
MapTR [ICLR23] [23]	R50	110	56.2	59.8	60.1	58.7	43.6	35.7	25.8	35.0	15.7
MapVR [NeurIPS23] [49]	R50	110	55.0	61.8	59.4	58.8	46.0	39.7	29.9	38.5	15.7
BeMapNet [CVPR23] [32]	R50	110	62.6	66.7	65.1	64.8	-	-	-	-	4.4
MGMap(Ours)	R50	110	**64.4**	**67.6**	**67.7**	**66.5**	**54.5**	**42.1**	**37.4**	**44.7**	11.6
LiDAR-Based Methods											
HDMapNet [ICRA22] [19]	PP	30	10.4	24.1	37.9	24.1	-	-	-	-	-
VectorMapNet [ICML23] [26]	PP	110	42.5	51.4	44.1	34.0	-	-	-	-	-
MapTR [ICLR23] [23]	Sec	24	48.5	53.7	64.7	55.6	38.9	30.1	41.1	36.7	6.0
MGMap(Ours)	Sec	24	**63.5**	**66.7**	**73.6**	**67.9**	**53.4**	**44.4**	**52.6**	**50.1**	5.5
Camera-LiDAR Fusion Methods											
HDMapNet [ICRA22] [19]	EB0 & PP	30	16.3	29.6	46.7	31.0	-	-	-	-	-
VectorMapNet [ICML23] [26]	EB0 & PP	110+ft	48.2	60.1	53.0	53.7	-	-	-	-	-
MapTR [ICLR23] [23]	R50 & Sec	24	55.9	62.3	69.3	62.5	46.4	38.4	49.2	44.7	5.2
MapVR [NeurIPS23] [49]	R50 & Sec	24	60.4	62.7	67.2	63.5	52.4	46.4	54.4	51.1	5.2
MGMap(Ours)	R50 & Sec	24	**67.7**	**71.1**	**76.2**	**71.7**	**59.6**	**47.3**	**54.6**	**53.8**	4.8

Performance Comparison

- Argoverse dataset

Method	Backbone	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
HDMapNet [19]	EB0	5.7	13.1	37.6	18.8
VectorMapNet [26]	R50	37.2	50.4	40.6	42.7
MapTR [23]	R50	50.6	60.7	61.2	57.4
MGMap(Ours)	R50	**52.8**	**67.5**	**68.1**	**62.8**

- Enlarged perception range

BEV range	Method	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
$60m \times 60m$	MapTR [23]	37.9	44.4	41.6	41.3
	MGMap(Ours)	48.7	53.4	50.0	50.8
	ΔmAP	**+10.8**	**+9.0**	**+8.4**	**+9.5**
$90m \times 30m$	MapTR [23]	36.5	43.7	39.1	39.8
	MGMap(Ours)	46.4	56.1	49.1	50.5
	ΔmAP	**+9.9**	**+12.4**	**+10.0**	**+10.2**

Ablation Study

Mask-guided design

Ins.	Point	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
		53.1	60.1	59.5	57.6
✓		55.4	61.7	61.5	59.5
	✓	54.9	63.4	62.3	60.2
✓	✓	**57.4**	**63.5**	**63.3**	**61.4**

EML Neck design

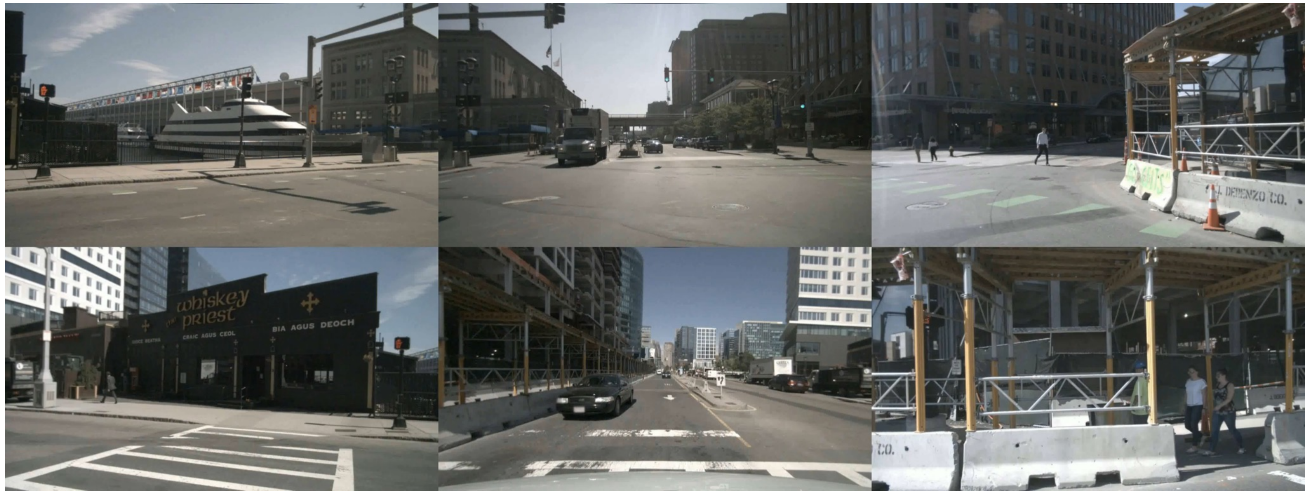
Baseline	PV	BEV	MG	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
✓				45.2	53.8	54.3	51.1
✓	✓			41.7	49.9	52.6	48.1
✓	✓		✓	47.3	54.1	55.1	52.2
✓		✓		53.1	60.1	59.5	57.6
✓		✓	✓	**57.4**	**63.5**	**63.3**	**61.4**

PG-MPR design

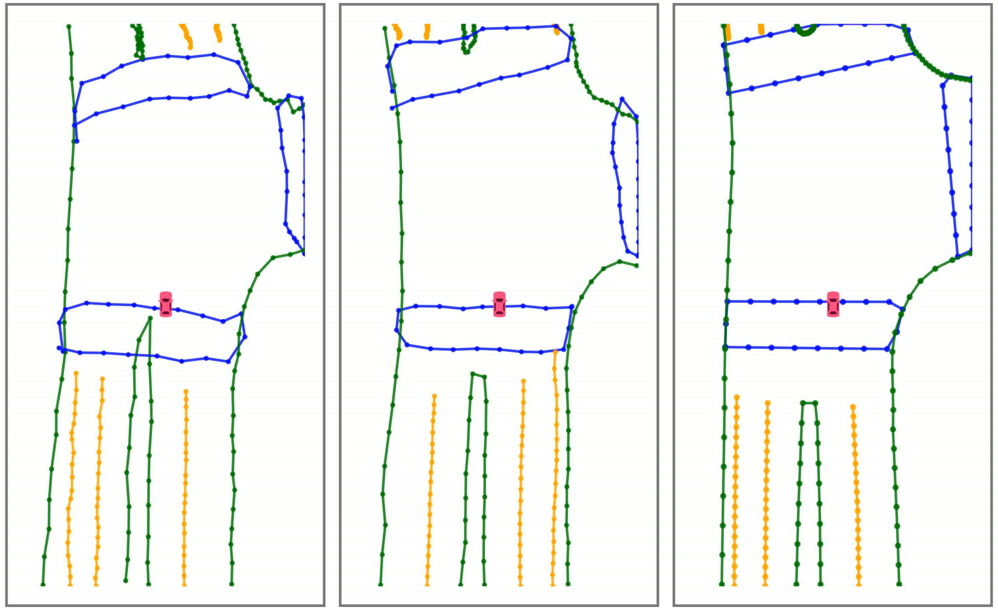
d	s	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
/	/	55.4	61.7	61.5	59.5
0.08	2	55.8	64.3	62.3	60.8
0.1	1	56.3	62.6	62.5	60.5
0.1	**2**	**57.4**	**63.5**	**63.3**	**61.4**
0.1	3	56.0	62.9	62.9	60.6
0.12	2	56.0	62.7	61.5	60.1

Visualization

☐ Sunny Condition



Multi-View Surrounding Images



MapTR   Ours   GT

Thanks!