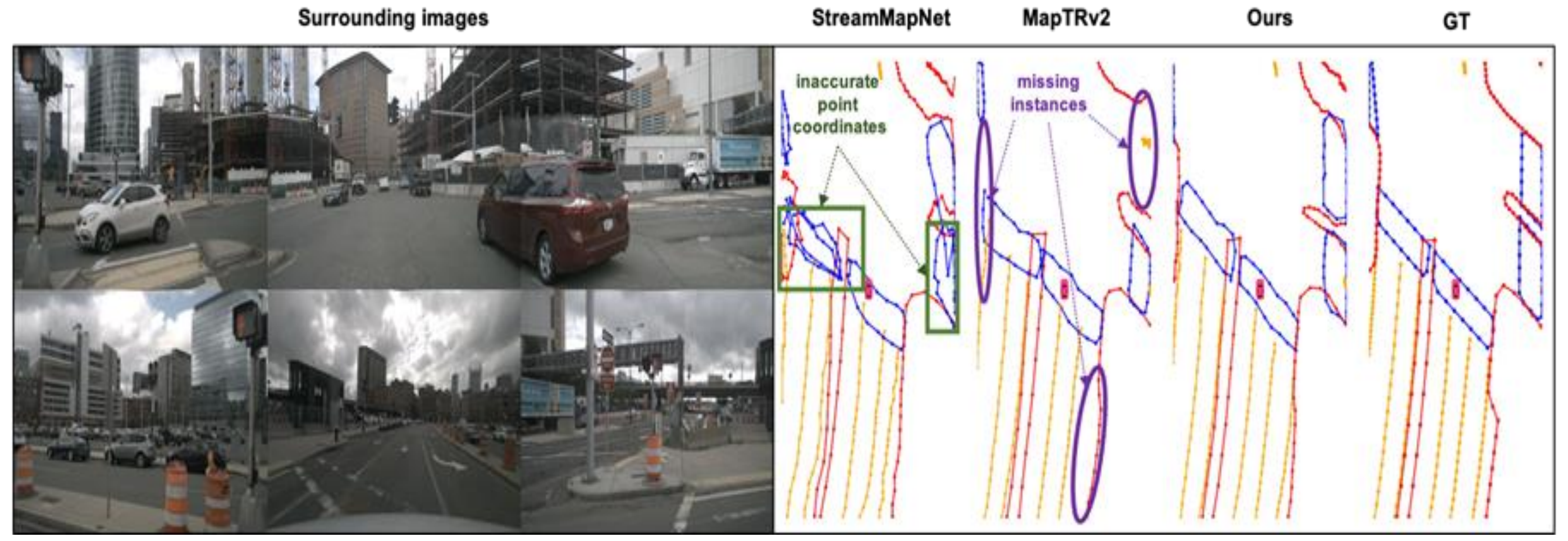
MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1]

[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025 | SINGAPORE EXPO

Challenge



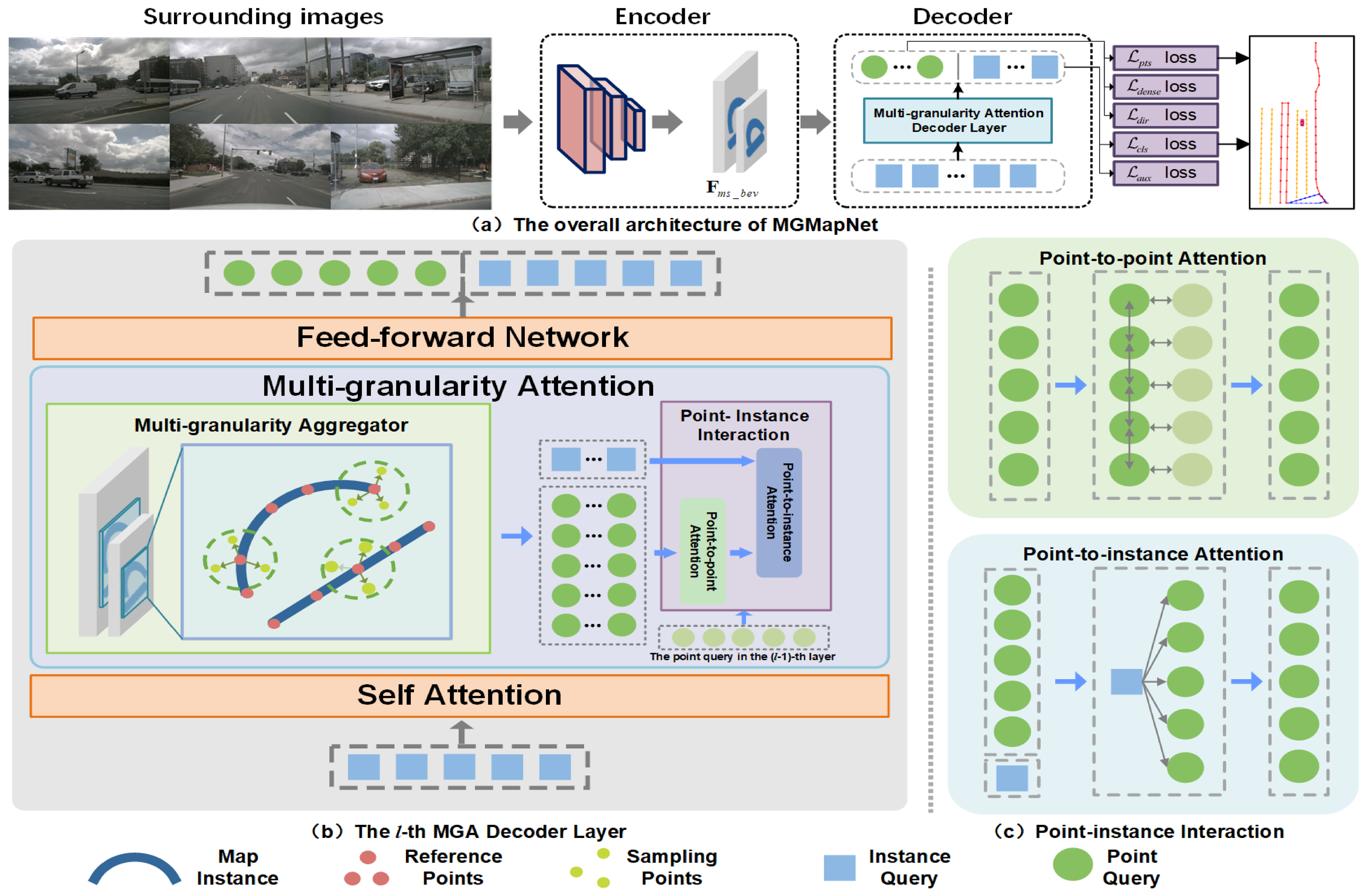
- In autonomous driving HD map reconstruction, point-level queries lack an overall semantic description of map elements and lane relationships (MapTRv2), while instance-level queries struggle with geometric details (StreamMapNet).
- Existing methods encounter difficulties in integrating fine-grained local positions with coarse-grained global classification, which constrains the holistic representation.

Innovation

- We propose a robust multigranularity representation, enabling the end-to-end construction of vectorized HD maps by employing coarse-grained instance-level and fine-grained point-level queries in one framework.
- The multi-granularity aggregator, combined with point-instance interaction, facilitates an efficient interaction between point-level and instance-level queries, effectively exchanging category and geometry information.
- We incorporate several strategy optimizations into model training, enabling our proposed MGMapNet to achieve state-of-the-art single-frame performances on both the nuScenes and Argoverse2.

Method

a. Multi-Granularity Attention



b. Multi-Granularity Aggregator

➢ ***Reference Point Update for Multi-Granularity Aggregator***

$$\begin{cases} \mathbf{RF}^l = \mathrm{MLP}(\mathbf{Q}^l_{ins}), l = 0, \\ \mathbf{RF}^l = \mathrm{sigmoid}(\mathrm{sigmoid}^{-1}(\mathbf{RF}^{l-1}) + \mathrm{MLP}(\mathbf{Q}^l_{pts})), l >= 1, \end{cases}$$

➢ ***Update Sampling Offset and Attention Weight***

$$\begin{aligned} \mathbf{PE}^{l-1}_{ref} &= \mathrm{MLP}^{l-1}_{ref}(\mathbf{RF}^{l-1}), \\ \Delta\mathbf{S}^l &= \mathrm{Sampling\_Offset}(\mathbf{Q}^{l-1}_{ins} + \mathbf{PE}^{l-1}_{ref}) \in \mathbf{R}^{N_q \times N_p \times N_{rep} \times 2}, \\ \mathbf{W}^l &= \mathrm{Weight\_Embed}(\mathbf{Q}^{l-1}_{ins} + \mathbf{PE}^{l-1}_{ref}) \in \mathbb{R}^{N_q \times N_p \times N_{rep}}, \\ \mathbf{S}^l &= (\mathbf{RF}^{l-1} + \Delta\mathbf{S}^l) \in \mathbb{R}^{N_q \times N_p \times N_{rep} \times 2}, \end{aligned}$$

➢ ***Generate Multi-Granularity Queries***

$$\begin{aligned} \mathbf{W}^l_{ins} &= \underset{(j,k)\in(N_p,N_{rep})}{\mathrm{softmax}} \left(\mathbf{W}^l_{j,k}\right) \in \mathbb{R}^{N_q \times (N_p \times N_{rep})}, \\ \mathbf{W}^l_{pts} &= \underset{k\in N_{rep}}{\mathrm{softmax}} \left(\mathbf{W}^l_{j,k}\right) \in \mathbb{R}^{N_q \times N_p \times N_{rep}}, \\ \mathbf{Q}^l_{ins} &= \sum_{j=1}^{N_p} \sum_{k=1}^{N_{rep}} \left[\mathbf{W}^l_{ins}\, \mathrm{sampling}(\mathbf{F}_{\mathrm{ms\_bev}}, \mathbf{S}^{\mathrm{l}}_{\mathrm{j,k}})\right] \in \mathbb{R}^{N_q \times C}, \\ \mathbf{Q}^l_{pts} &= \sum_{k=1}^{N_{rep}} \left[\mathbf{W}^l_{pts}\, \mathrm{sampling}(\mathbf{F}_{\mathrm{ms\_bev}}, \mathbf{S}^{\mathrm{l}}_{\mathrm{j,k}})\right] \in \mathbb{R}^{N_q \times N_p \times C}, \end{aligned}$$

c. Point Instance Interaction

➢ ***Positional Encoding of Different Granularities***

$$\begin{aligned} \mathbf{PE}^l_{ins} &= \mathrm{MLP}^l_{ins}(\mathbf{S}^l, \mathbf{W}^l_{ins}), \\ \mathbf{PE}^l_{pts} &= \mathrm{MLP}^l_{pts}(\mathbf{S}^l, \mathbf{W}^l_{pts}), \end{aligned}$$

➢ ***P2P Attention***

$$\begin{cases} \mathbf{Q}^{l'}_{pts} = \mathrm{SA}(\mathbf{Q}^l_{pts} + \mathbf{PE}^l_{pts}), l = 0, \\ \mathbf{Q}^{l'}_{pts} = \mathrm{CA}(\mathbf{Q}^l_{pts} + \mathbf{PE}^l_{pts}, \mathbf{Q}^{l-1}_{pts} + \mathbf{PE}^{l-1}_{pts}), l >= 1. \end{cases}$$

➢ ***P2I Attention***

$$\mathbf{Q}^{l''}_{pts} = \mathrm{CA}(\mathbf{Q}^{l'}_{pts} + \mathbf{PE}^l_{pts}, \mathbf{Q}^l_{ins} + \mathbf{PE}^l_{ins}).$$

➢ ***Output of Point Instance Interaction***

$$\mathbf{Q}^{l'}_{ins} = \mathrm{MLP}_{agg}(\sum_{j=1}^{N_p} \mathbf{Q}^{l''}_{pts,j}).$$

Experiments

a. results on nuScenes dataset

Method	Epoch	AP_{ped}	AP_{div}	AP_{bou}	mAP	FPS	Params (MB)
HDMapNet (Li et al., 2022a)	30	14.4	21.7	33.0	23.0	-	-
BeMapNet (Qiao et al., 2023a)	30	62.3	57.7	59.4	59.8	4.3	-
PivotNet (Ding et al., 2023)	24	56.5	56.2	60.1	57.6	9.2	-
MapTRv2 (Liao et al., 2023)	24	59.8	62.4	62.4	61.5	14.1	40.3
MGMap (Liu et al., 2024a)	24	61.8	65.0	67.5	64.8	12	55.9
MapQR (Liu et al., 2024b)	24	**68.0**	63.4	67.7	66.4	11.9	125.3
MGMapNet (Ours)	24	64.7	**66.1**	**69.4**	**66.8**	11.7	70.1
VectorMapNet (Liu et al., 2023a)	110	42.5	51.4	44.1	46.0	-	-
MapTRv2 (Liao et al., 2023)	110	68.1	68.3	69.7	68.7	14.1	40.3
MGMap (Liu et al., 2024a)	110	64.4	67.6	67.7	66.5	12	55.9
MapQR (Liu et al., 2024b)	110	**74.4**	70.1	73.2	72.6	11.9	125.3
MGMapNet (Ours)	110	74.3	**71.8**	**74.8**	**73.6**	11.7	70.1

b. results on Argoverse 2 dataset

Method	Map dim.	AP_{ped}	AP_{div}	AP_{bou}	mAP
HDMapNet (Li et al., 2022a)	2	13.1	5.7	37.6	18.8
VectorMapNet (Liu et al., 2023a)		38.3	36.1	39.2	37.9
MapTRv2 (Liao et al., 2023)		62.9	72.1	67.1	67.4
MapQR (Liu et al., 2024b)		64.3	72.3	68.1	68.2
HIMap (Zhou et al., 2024)		**69.0**	69.5	70.3	69.6
MGMapNet (Ours)		67.1	**74.6**	**71.7**	**71.2**
VectorMapNet (Liu et al., 2023a)	3	36.5	35.0	36.2	35.8
MapTRv2 (Liao et al., 2023)		60.7	68.9	64.5	64.7
MapQR (Liu et al., 2024b)		60.1	71.2	66.2	65.9
HIMap (Zhou et al., 2024)		**66.7**	68.3	70.3	68.4
MGMapNet (Ours)		64.7	**72.1**	**70.4**	**69.1**

c. ablation study

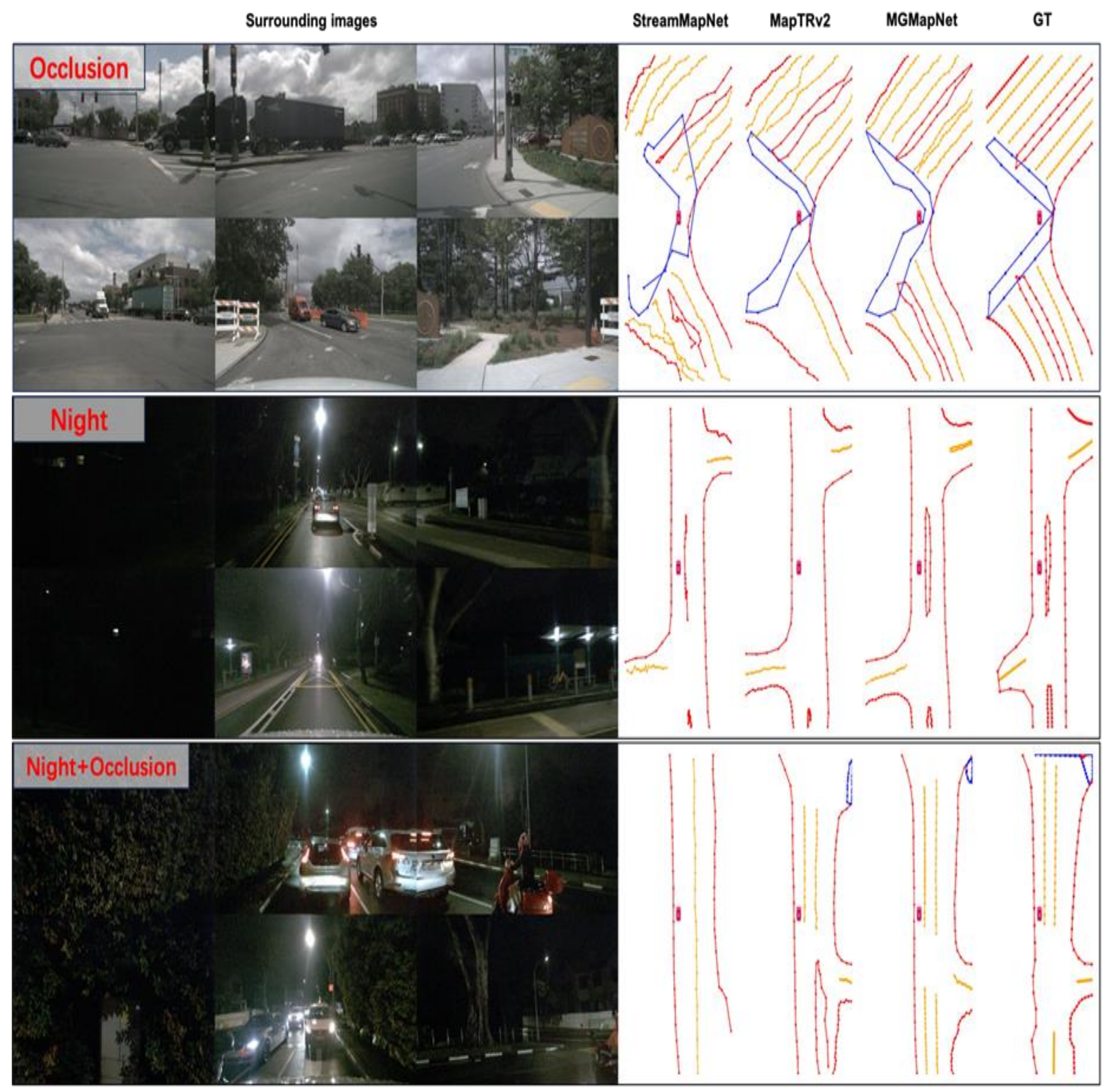
Method	Point-instance Interaction		mAP
	P2P Attention	P2I Attention	
Multi-point Attention (Yuan et al., 2024)	×	×	59.6
Multi-granularity Aggregator	×	×	62.7
	✓	×	64.8
	×	✓	65.0
	✓	✓	66.8

Effectiveness Validation
The proposed MGA includes a multi-granularity aggregator and point-instance interaction. Ablation studies demonstrate that the integration of both geometric and categorical information significantly enhances the quality of query representations

Experiment	Method	mAP
	Multi-point Attention	55.9
(a)	Multi-granularity Attention	63.6 (+7.7)
(b)	+ Auxiliary Loss	64.4 (+0.8)
(c)	+ Multi-scale BEV Feature	65.0 (+0.6)
(d)	+ Position Embeddings	66.2 (+1.2)
(e)	+ Increase Query Number	66.8 (+0.6)

Generality Validation
To verify generality of the proposed modules, we integrate several strategy into the MapTRv2 framework. Experiments (a-e) confirm the effectiveness of these designed techniques.

d. qualitative results



Conclusion

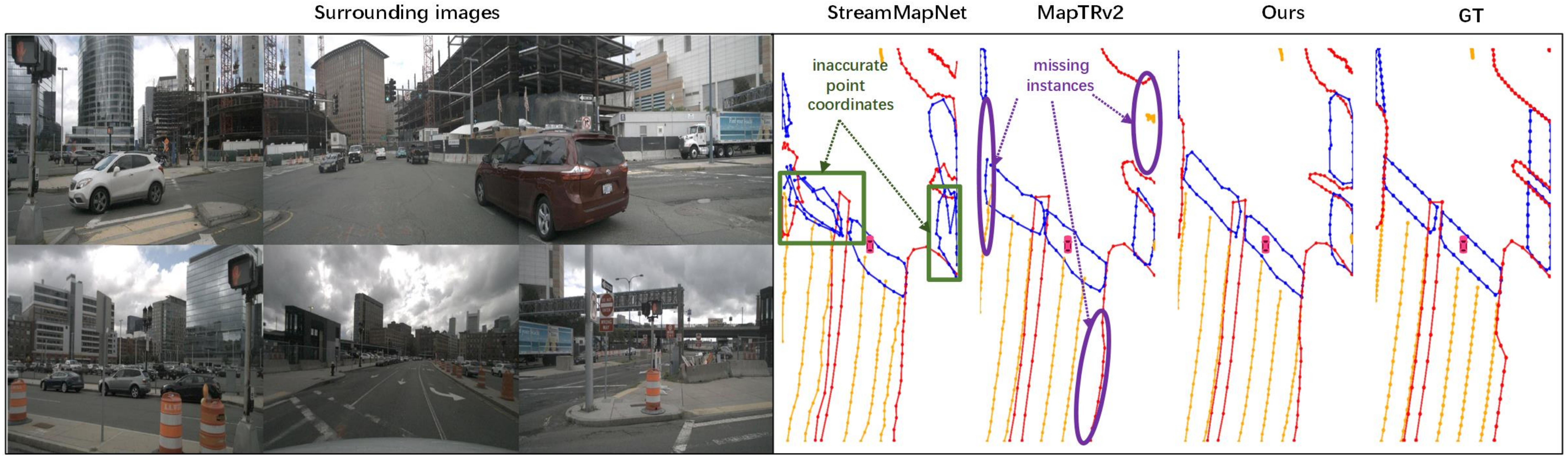
We propose a multi-granularity map network for end-to-end vectorized HD map construction. Through multi-granularity attention, coarse-grained instance queries represent fine-grained point queries, enabling category and geometric information capture via point-instance interaction. This approach improves map construction by leveraging multi-granularity queries for instance categories and polyline distributions.

MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]
[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025 | SINGAPORE EXPO

Challenge



Limitations of Existing Methods:

•**Point-level Queries** (e.g., MapTR series):

Strengths: Excel at capturing fine-grained geometric details.

Weaknesses: Lack holistic modeling of instance relationships, leading to missed detections in distant or complex scenarios (e.g., merging lanes).

•**Instance-level Queries** (e.g., StreamMapNet):

Strengths: Effective at global classification and capturing overall instance properties.

Weaknesses: Struggle with precise geometric representation, especially for irregular or elongated elements, resulting in inaccurate local coordinates.

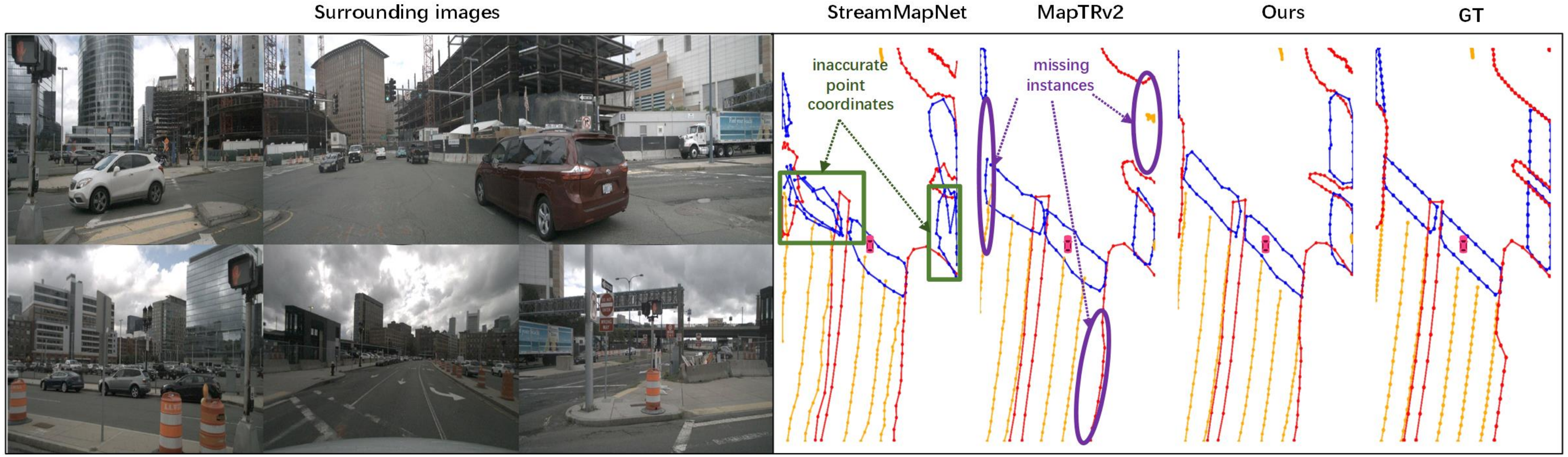
Challenge

Balancing local geometric accuracy with global instance-level understanding.

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]
[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025| SINGAPORE EXPO

Innovation



- We propose a robust multi-granularity representation, enabling the end-to-end construction of vectorized HD maps by employing coarse-grained instance-level and fine-grained point-level queries in one framework.
- The multi-granularity aggregator, combined with point-instance interaction, facilitates an efficient interaction between point-level and instance-level queries, effectively exchanging category and geometry information.
- We incorporate several strategy optimizations into model training, enabling our proposed MGMapNet to achieve state-of-the-art single-frame performances on both the nuScenes and Argoverse2.

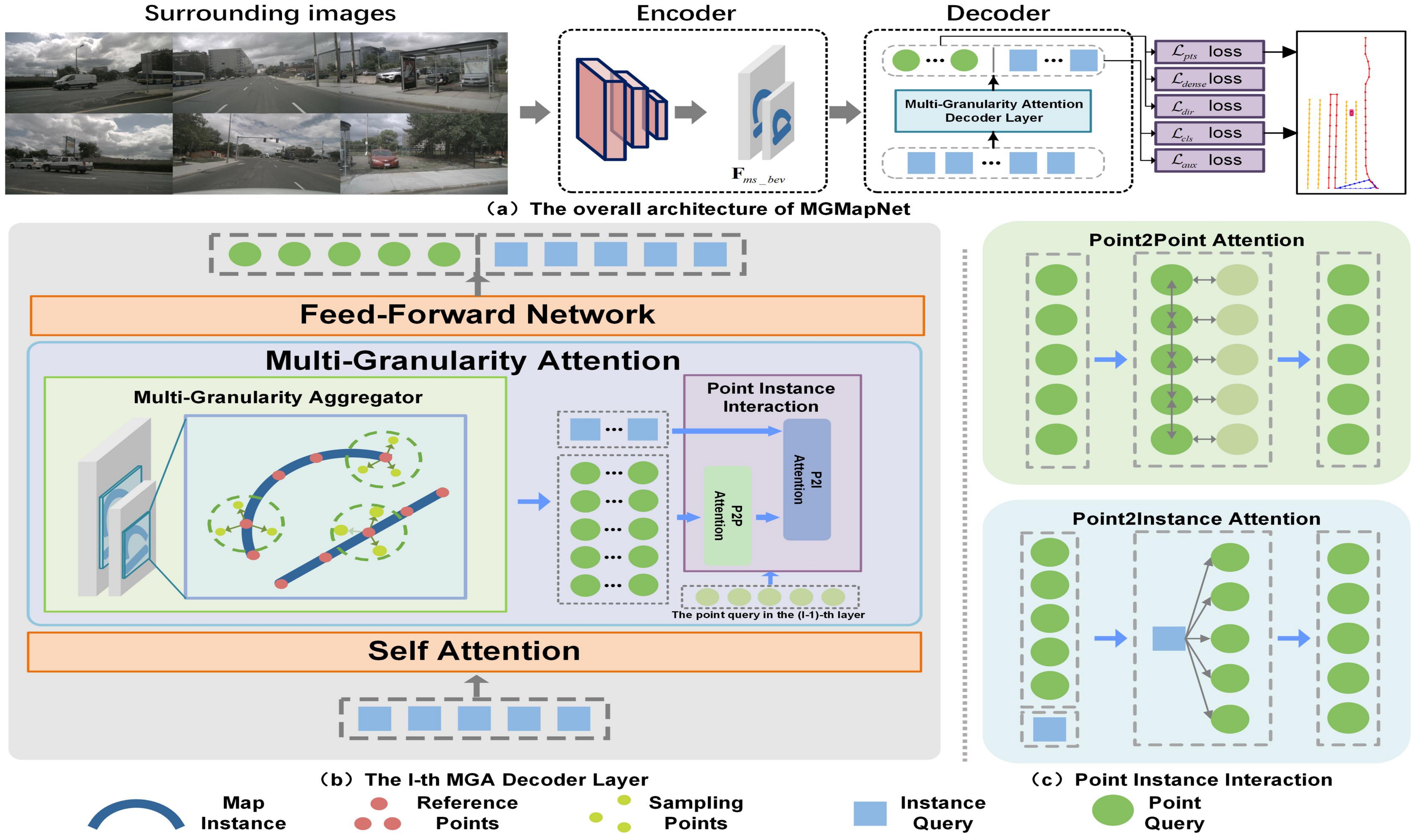
MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]

[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025| SINGAPORE EXPO

MGMapNet



(a) The overall architecture of MGMapNet

(b) The l-th MGA Decoder Layer

(c) Point Instance Interaction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1]
[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.
The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025| SINGAPORE EXPO

Multi-Granularity Aggregator

➢*Reference Point Update for Multi-Granularity Aggregator*

$$\begin{cases} \mathbf{RF}^l = \mathrm{MLP}(\mathbf{Q}_{ins}^l), l = 0, \\ \mathbf{RF}^l = \mathrm{sigmoid}(\mathrm{sigmoid}^{-1}(\mathbf{RF}^{l-1}) + \mathrm{MLP}(\mathbf{Q}_{pts}^l)), l >= 1, \end{cases}$$

➢*Update Sampling Offset and Attention Weight*

$$\begin{aligned} \mathbf{PE}_{ref}^{l-1} &= \mathrm{MLP}_{ref}^{l-1}(\mathbf{RF}^{l-1}), \\ \Delta\mathbf{S}^l &= \mathrm{Sampling\_Offset}(\mathbf{Q}_{ins}^{l-1} + \mathbf{PE}_{ref}^{l-1}) \in \mathbf{R}^{N_q \times N_p \times N_{rep} \times 2}, \\ \mathbf{W}^l &= \mathrm{Weight\_Embed}(\mathbf{Q}_{ins}^{l-1} + \mathbf{PE}_{ref}^{l-1}) \in \mathbb{R}^{N_q \times N_p \times N_{rep}}, \\ \mathbf{S}^l &= (\mathbf{RF}^{l-1} + \Delta\mathbf{S}^l) \in \mathbb{R}^{N_q \times N_p \times N_{rep} \times 2}, \end{aligned}$$

➢*Generate Multi-Granularity Queries*

$$\mathbf{W}_{ins}^l = \underset{(j,k)\in(N_p,N_{rep})}{\mathrm{softmax}} \left(\mathbf{W}_{j,k}^l\right) \in \mathbb{R}^{N_q \times (N_p \times N_{rep})},$$

$$\mathbf{W}_{pts}^l = \underset{k\in N_{rep}}{\mathrm{softmax}} \left(\mathbf{W}_{j,k}^l\right) \in \mathbb{R}^{N_q \times N_p \times N_{rep}},$$

$$\mathbf{Q}_{ins}^l = \sum_{j=1}^{N_p} \sum_{k=1}^{N_{rep}} \left[\mathbf{W}_{ins}^l \, \mathrm{sampling}(\mathbf{F}_{\mathrm{ms\_bev}}, \mathbf{S}_{\mathrm{j,k}}^{\mathrm{l}})\right] \in \mathbb{R}^{N_q \times C},$$

$$\mathbf{Q}_{pts}^l = \sum_{k=1}^{N_{rep}} \left[\mathbf{W}_{pts}^l \, \mathrm{sampling}(\mathbf{F}_{\mathrm{ms\_bev}}, \mathbf{S}_{\mathrm{j,k}}^{\mathrm{l}})\right] \in \mathbb{R}^{N_q \times N_p \times C},$$

MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]

[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025 | SINGAPORE EXPO

. Point Instance Interaction

➢ ***Positional Encoding of Different Granularities***

$$\mathbf{PE}^l_{ins} = \mathrm{MLP}^l_{ins}(\mathbf{S}^l, \mathbf{W}^l_{ins}),$$
$$\mathbf{PE}^l_{pts} = \mathrm{MLP}^l_{pts}(\mathbf{S}^l, \mathbf{W}^l_{pts}),$$

➢ ***P2P Attention***

$$\begin{cases} \mathbf{Q}^{l'}_{pts} = \mathrm{SA}(\mathbf{Q}^l_{pts} + \mathbf{PE}^l_{pts}), l = 0, \\ \mathbf{Q}^{l'}_{pts} = \mathrm{CA}(\mathbf{Q}^l_{pts} + \mathbf{PE}^l_{pts}, \mathbf{Q}^{l-1}_{pts} + \mathbf{PE}^{l-1}_{pts}), l >= 1. \end{cases}$$

➢ ***P2I Attention***

$$\mathbf{Q}^{l''}_{pts} = \mathrm{CA}(\mathbf{Q}^{l'}_{pts} + \mathbf{PE}^l_{pts}, \mathbf{Q}^l_{ins} + \mathbf{PE}^l_{ins}).$$

➢ ***Output of Point Instance Interaction***

$$\mathbf{Q}^{l'}_{ins} = \mathrm{MLP}_{agg}(\sum_{j=1}^{N_p} \mathbf{Q}^{l''}_{pts,j}).$$

MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]

[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025| SINGAPORE EXPO

Experiments

a. results on nuScenes dataset

Method	Epoch	AP_{ped}	AP_{div}	AP_{bou}	mAP	FPS	Params (MB)
HDMapNet (Li et al., 2022a)	30	14.4	21.7	33.0	23.0	-	-
BeMapNet (Qiao et al., 2023a)	30	62.3	57.7	59.4	59.8	4.3	-
PivotNet (Ding et al., 2023)	24	56.5	56.2	60.1	57.6	9.2	-
MapTRv2 (Liao et al., 2023)	24	59.8	62.4	62.4	61.5	14.1	40.3
MGMap (Liu et al., 2024a)	24	61.8	65.0	67.5	64.8	12	55.9
MapQR (Liu et al., 2024b)	24	**68.0**	63.4	67.7	66.4	11.9	125.3
MGMapNet (Ours)	24	64.7	**66.1**	**69.4**	**66.8**	11.7	70.1
VectorMapNet (Liu et al., 2023a)	110	42.5	51.4	44.1	46.0	-	-
MapTRv2 (Liao et al., 2023)	110	68.1	68.3	69.7	68.7	14.1	40.3
MGMap (Liu et al., 2024a)	110	64.4	67.6	67.7	66.5	12	55.9
MapQR (Liu et al., 2024b)	110	**74.4**	70.1	73.2	72.6	11.9	125.3
MGMapNet (Ours)	110	74.3	**71.8**	**74.8**	**73.6**	11.7	70.1

b. results on Argoverse 2 dataset

Method	Map dim.	AP_{ped}	AP_{div}	AP_{bou}	mAP
HDMapNet (Li et al., 2022a)	2	13.1	5.7	37.6	18.8
VectorMapNet (Liu et al., 2023a)		38.3	36.1	39.2	37.9
MapTRv2 (Liao et al., 2023)		62.9	72.1	67.1	67.4
MapQR (Liu et al., 2024b)		64.3	72.3	68.1	68.2
HIMap (Zhou et al., 2024)		**69.0**	69.5	70.3	69.6
MGMapNet (Ours)		67.1	**74.6**	**71.7**	**71.2**
VectorMapNet (Liu et al., 2023a)	3	36.5	35.0	36.2	35.8
MapTRv2 (Liao et al., 2023)		60.7	68.9	64.5	64.7
MapQR (Liu et al., 2024b)		60.1	71.2	66.2	65.9
HIMap (Zhou et al., 2024)		**66.7**	68.3	70.3	68.4
MGMapNet (Ours)		64.7	**72.1**	**70.4**	**69.1**

MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]

[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025 | SINGAPORE EXPO

Ablation study

Method	Point-instance Interaction		mAP
	P2P Attention	P2I Attention	
Multi-point Attention (Yuan et al., 2024)	×	×	59.6
Multi-granularity Aggregator	×	×	62.7
	✓	×	64.8
	×	✓	65.0
	✓	✓	66.8

Effectiveness Validation

The proposed MGA includes a multi-granularity aggregator and point-instance interaction. Ablation studies demonstrate that the integration of both geometric and categorical information significantly enhances the quality of query representations

Experiment	Method	mAP
	Multi-point Attention	55.9
(a)	Multi-granularity Attention	63.6 (+7.7)
(b)	+ Auxiliary Loss	64.4 (+0.8)
(c)	+ Multi-scale BEV Feature	65.0 (+0.6)
(d)	+ Position Embeddings	66.2 (+1.2)
(e)	+ Increase Query Number	66.8 (+0.6)

Generality Validation

To verify generality of the proposed modules, we integrate several strategy into the MapTRv2 framework. Experiments (a-e) confirm the effectiveness of these designed techniques.

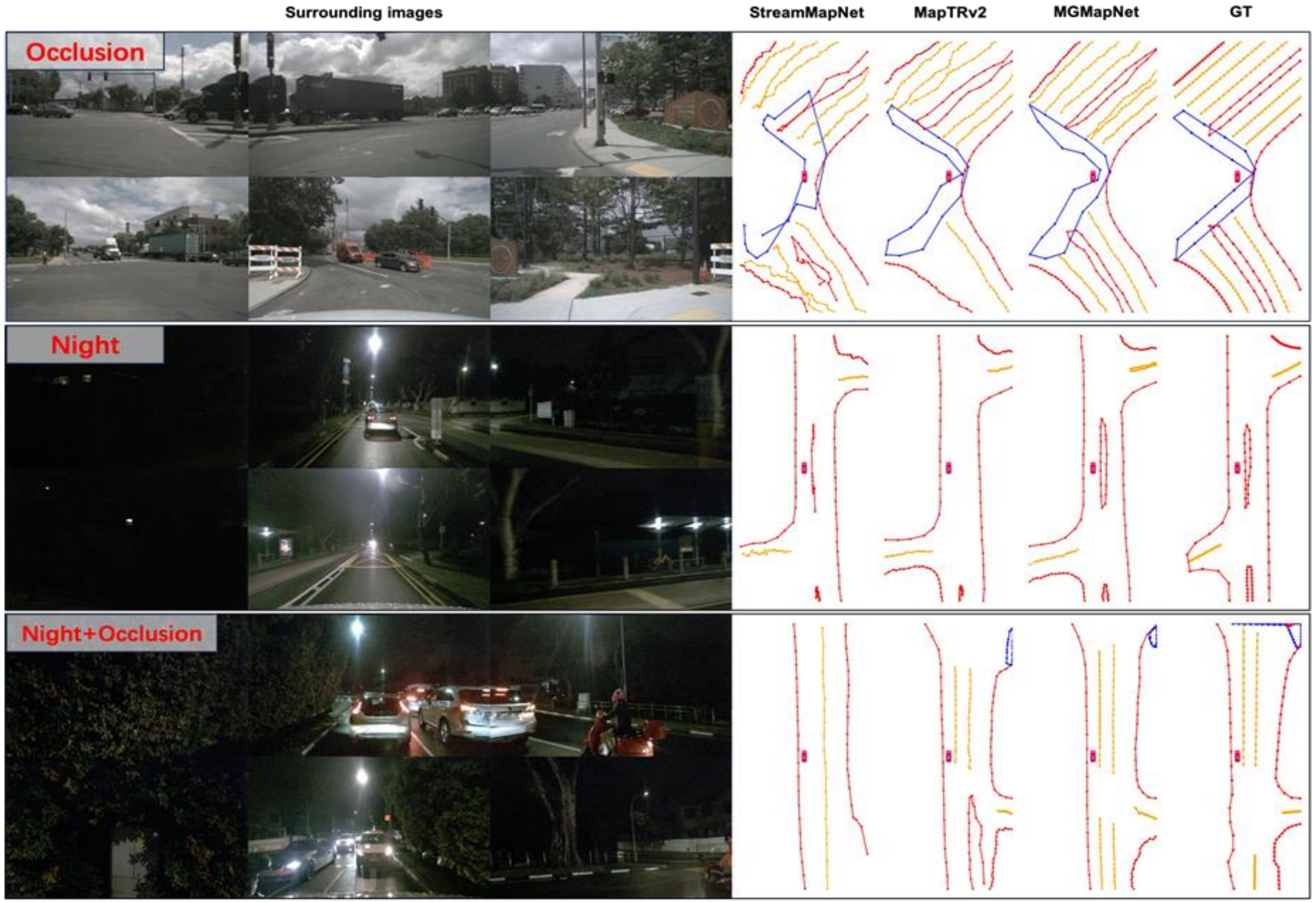
MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]

[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025|
SINGAPORE EXPO

Qualitative Results



MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

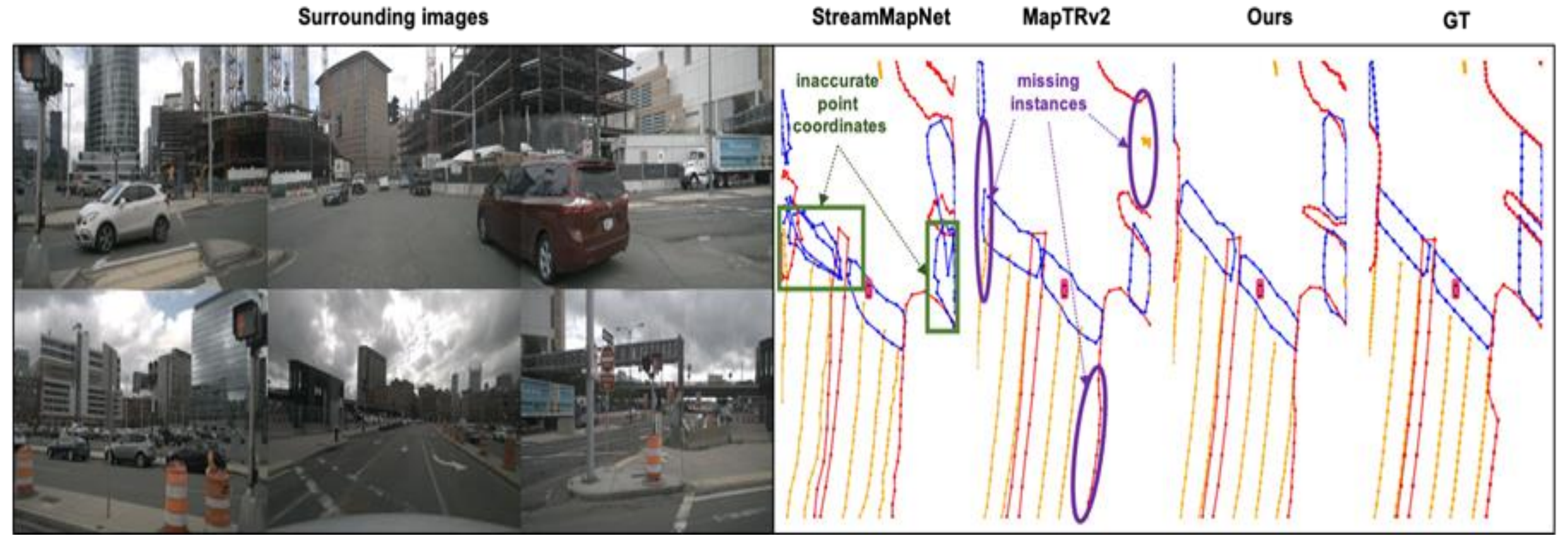
Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]

[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations

THU APR 24 – MON APR 28TH, 2025| SINGAPORE EXPO

Challenge



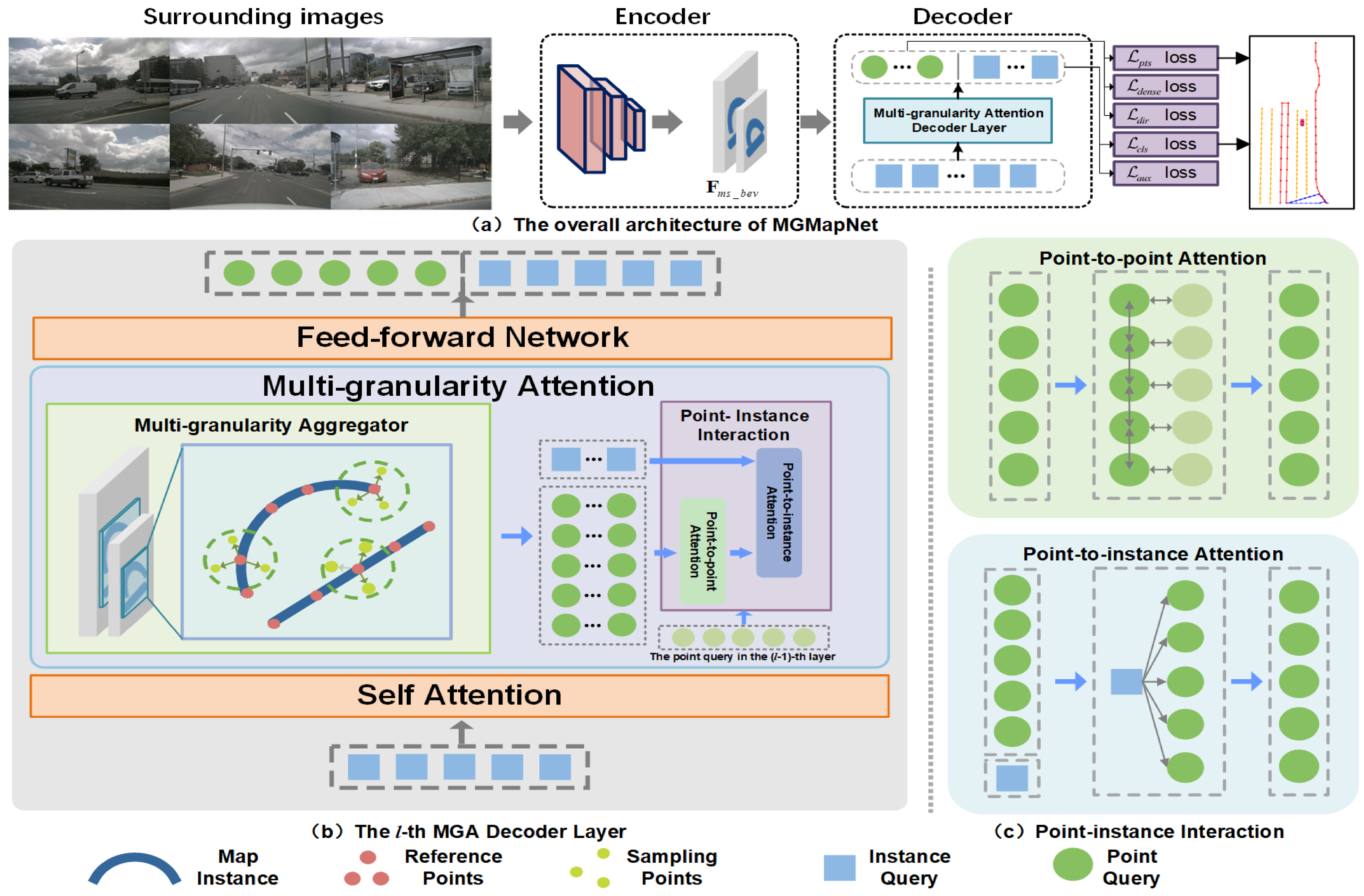
- In autonomous driving HD map reconstruction, point-level queries lack an overall semantic description of map elements and lane relationships (MapTRv2), while instance-level queries struggle with geometric details (StreamMapNet).
- Existing methods encounter difficulties in integrating fine-grained local positions with coarse-grained global classification, which constrains the holistic representation.

Innovation

- We propose a robust multigranularity representation, enabling the end-to-end construction of vectorized HD maps by employing coarse-grained instance-level and fine-grained point-level queries in one framework.
- The multi-granularity aggregator, combined with point-instance interaction, facilitates an efficient interaction between point-level and instance-level queries, effectively exchanging category and geometry information.
- We incorporate several strategy optimizations into model training, enabling our proposed MGMapNet to achieve state-of-the-art single-frame performances on both the nuScenes and Argoverse2.

Method

a. Multi-Granularity Attention



b. Multi-Granularity Aggregator

➢ ***Reference Point Update for Multi-Granularity Aggregator***

$$\begin{cases} \mathbf{RF}^l = \mathrm{MLP}(\mathbf{Q}^l_{ins}), l = 0, \\ \mathbf{RF}^l = \mathrm{sigmoid}(\mathrm{sigmoid}^{-1}(\mathbf{RF}^{l-1}) + \mathrm{MLP}(\mathbf{Q}^l_{pts})), l >= 1, \end{cases}$$

➢ ***Update Sampling Offset and Attention Weight***

$$\mathbf{PE}^{l-1}_{ref} = \mathrm{MLP}^{l-1}_{ref}(\mathbf{RF}^{l-1}),$$
$$\Delta\mathbf{S}^l = \mathrm{Sampling\_Offset}(\mathbf{Q}^{l-1}_{ins} + \mathbf{PE}^{l-1}_{ref}) \in \mathbf{R}^{N_q \times N_p \times N_{rep} \times 2},$$
$$\mathbf{W}^l = \mathrm{Weight\_Embed}(\mathbf{Q}^{l-1}_{ins} + \mathbf{PE}^{l-1}_{ref}) \in \mathbb{R}^{N_q \times N_p \times N_{rep}},$$
$$\mathbf{S}^l = (\mathbf{RF}^{l-1} + \Delta\mathbf{S}^l) \in \mathbb{R}^{N_q \times N_p \times N_{rep} \times 2},$$

➢ ***Generate Multi-Granularity Queries***

$$\mathbf{W}^l_{ins} = \underset{(j,k)\in(N_p, N_{rep})}{\mathrm{softmax}} \left(\mathbf{W}^l_{j,k}\right) \in \mathbb{R}^{N_q \times (N_p \times N_{rep})},$$
$$\mathbf{W}^l_{pts} = \underset{k \in N_{rep}}{\mathrm{softmax}} \left(\mathbf{W}^l_{j,k}\right) \in \mathbb{R}^{N_q \times N_p \times N_{rep}},$$
$$\mathbf{Q}^l_{ins} = \sum_{j=1}^{N_p} \sum_{k=1}^{N_{rep}} \left[\mathbf{W}^l_{ins}\, \mathrm{sampling}(\mathbf{F}_{\mathrm{ms\_bev}}, \mathbf{S}^{\mathrm{l}}_{\mathrm{j,k}})\right] \in \mathbb{R}^{N_q \times C},$$
$$\mathbf{Q}^l_{pts} = \sum_{k=1}^{N_{rep}} \left[\mathbf{W}^l_{pts}\, \mathrm{sampling}(\mathbf{F}_{\mathrm{ms\_bev}}, \mathbf{S}^{\mathrm{l}}_{\mathrm{j,k}})\right] \in \mathbb{R}^{N_q \times N_p \times C},$$

c. Point Instance Interaction

➢ ***Positional Encoding of Different Granularities***

$$\mathbf{PE}^l_{ins} = \mathrm{MLP}^l_{ins}(\mathbf{S}^l, \mathbf{W}^l_{ins}),$$
$$\mathbf{PE}^l_{pts} = \mathrm{MLP}^l_{pts}(\mathbf{S}^l, \mathbf{W}^l_{pts}),$$

➢ ***P2P Attention***

$$\begin{cases} \mathbf{Q}^{l'}_{pts} = \mathrm{SA}(\mathbf{Q}^l_{pts} + \mathbf{PE}^l_{pts}), l = 0, \\ \mathbf{Q}^{l'}_{pts} = \mathrm{CA}(\mathbf{Q}^l_{pts} + \mathbf{PE}^l_{pts}, \mathbf{Q}^{l-1}_{pts} + \mathbf{PE}^{l-1}_{pts}), l >= 1. \end{cases}$$

➢ ***P2I Attention***

$$\mathbf{Q}^{l''}_{pts} = \mathrm{CA}(\mathbf{Q}^{l'}_{pts} + \mathbf{PE}^l_{pts}, \mathbf{Q}^l_{ins} + \mathbf{PE}^l_{ins}).$$

➢ ***Output of Point Instance Interaction***

$$\mathbf{Q}^{l'}_{ins} = \mathrm{MLP}_{agg}(\sum_{j=1}^{N_p} \mathbf{Q}^{l''}_{pts,j}).$$

Experiments

a. results on nuScenes dataset

Method	Epoch	AP_{ped}	AP_{div}	AP_{bou}	mAP	FPS	Params (MB)
HDMapNet (Li et al., 2022a)	30	14.4	21.7	33.0	23.0	-	-
BeMapNet (Qiao et al., 2023a)	30	62.3	57.7	59.4	59.8	4.3	-
PivotNet (Ding et al., 2023)	24	56.5	56.2	60.1	57.6	9.2	-
MapTRv2 (Liao et al., 2023)	24	59.8	62.4	62.4	61.5	14.1	40.3
MGMap (Liu et al., 2024a)	24	61.8	65.0	67.5	64.8	12	55.9
MapQR (Liu et al., 2024b)	24	**68.0**	63.4	67.7	66.4	11.9	125.3
MGMapNet (Ours)	24	64.7	**66.1**	**69.4**	**66.8**	11.7	70.1
VectorMapNet (Liu et al., 2023a)	110	42.5	51.4	44.1	46.0	-	-
MapTRv2 (Liao et al., 2023)	110	68.1	68.3	69.7	68.7	14.1	40.3
MGMap (Liu et al., 2024a)	110	64.4	67.6	67.7	66.5	12	55.9
MapQR (Liu et al., 2024b)	110	**74.4**	70.1	73.2	72.6	11.9	125.3
MGMapNet (Ours)	110	74.3	**71.8**	**74.8**	**73.6**	11.7	70.1

b. results on Argoverse 2 dataset

Method	Map dim.	AP_{ped}	AP_{div}	AP_{bou}	mAP
HDMapNet (Li et al., 2022a)	2	13.1	5.7	37.6	18.8
VectorMapNet (Liu et al., 2023a)		38.3	36.1	39.2	37.9
MapTRv2 (Liao et al., 2023)		62.9	72.1	67.1	67.4
MapQR (Liu et al., 2024b)		64.3	72.3	68.1	68.2
HIMap (Zhou et al., 2024)		**69.0**	69.5	70.3	69.6
MGMapNet (Ours)		67.1	**74.6**	**71.7**	**71.2**
VectorMapNet (Liu et al., 2023a)	3	36.5	35.0	36.2	35.8
MapTRv2 (Liao et al., 2023)		60.7	68.9	64.5	64.7
MapQR (Liu et al., 2024b)		60.1	71.2	66.2	65.9
HIMap (Zhou et al., 2024)		**66.7**	68.3	70.3	68.4
MGMapNet (Ours)		64.7	**72.1**	**70.4**	**69.1**

c. ablation study

Method	Point-instance Interaction		mAP
	P2P Attention	P2I Attention	
Multi-point Attention (Yuan et al., 2024)	×	×	59.6
Multi-granularity Aggregator	×	×	62.7
	✓	×	64.8
	×	✓	65.0
	✓	✓	66.8

Effectiveness Validation

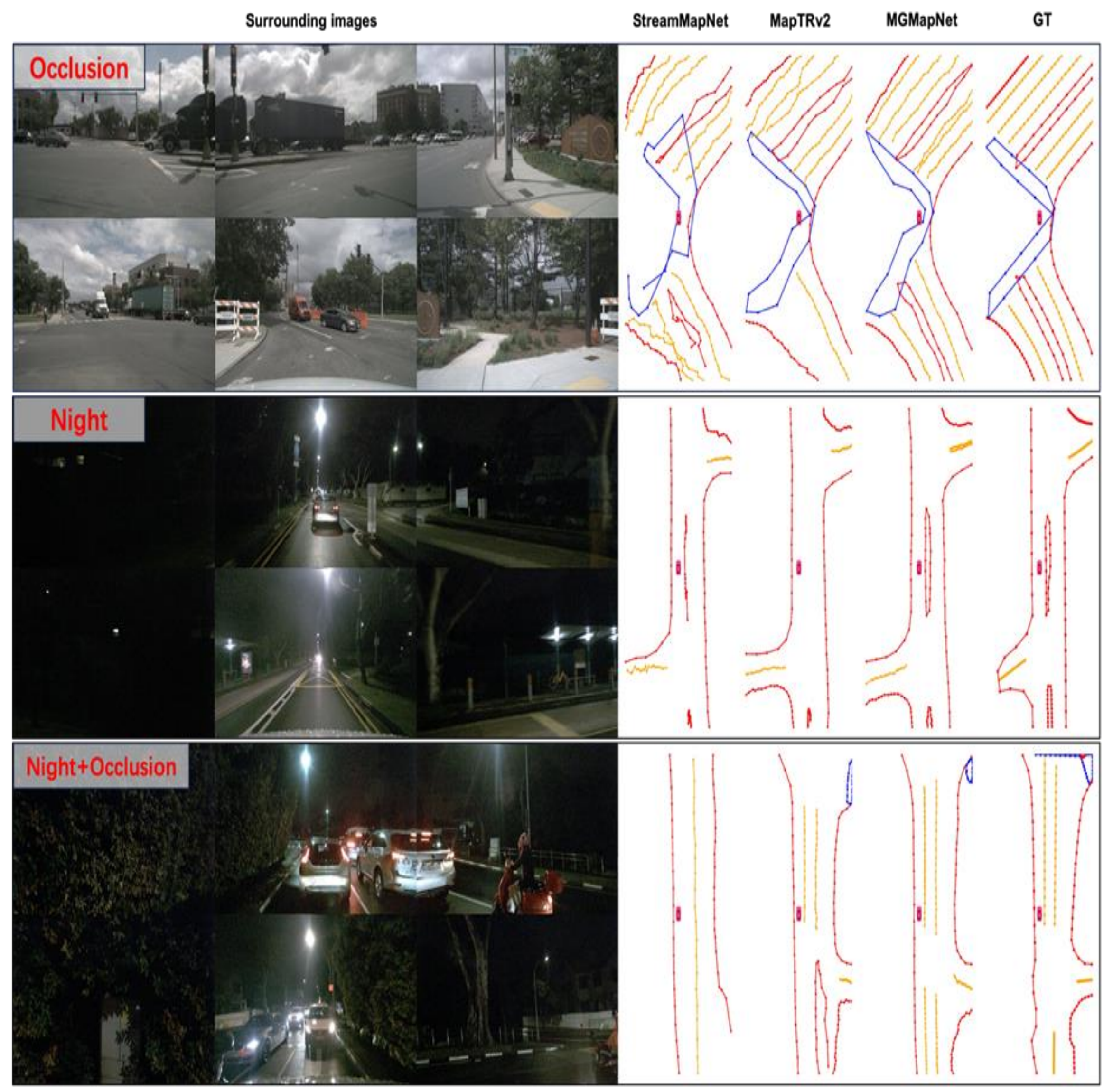
The proposed MGA includes a multi-granularity aggregator and point-instance interaction. Ablation studies demonstrate that the integration of both geometric and categorical information significantly enhances the quality of query representations

Experiment	Method	mAP
	Multi-point Attention	55.9
(a)	Multi-granularity Attention	63.6 (+7.7)
(b)	+ Auxiliary Loss	64.4 (+0.8)
(c)	+ Multi-scale BEV Feature	65.0 (+0.6)
(d)	+ Position Embeddings	66.2 (+1.2)
(e)	+ Increase Query Number	66.8 (+0.6)

Generality Validation

To verify generality of the proposed modules, we integrate several strategy into the MapTRv2 framework. Experiments (a-e) confirm the effectiveness of these designed techniques.

d. qualitative results



Conclusion

We propose a multi-granularity network for end-to-end vectorized HD map construction. Through multi-granularity attention, coarse-grained instance queries represent fine-grained point queries, enabling category and geometric information capture via point-instance interaction. This approach improves map construction by leveraging multi-granularity queries for instance categories and polyline distributions.

MGMapNet: Multi-Granularity Representation Learning for End-to-End Vectorized HD Map Construction

Jing Yang[1,*], Minyue Jiang[2,*], Sen Yang[2,*], Xiao Tan[2], Yingying Li[2], Erri Ding[2], Jingdong Wang[2], Hanli Wang[1,✉]
[1]College of Electronic and Information Engineering, Tongji University; [2]Baidu Inc.

The Thirteenth International Conference on Learning Representations
THU APR 24 – MON APR 28TH, 2025|
SINGAPORE EXPO

Thanks