SafeMap: Robust HD Map Construction from Incomplete Observations

Xiaoshuai Hao[1], Lingdong Kong[2], Rong Yin[3], Pengwei Wang[1], Jing Zhang[4], Yunfeng Diao[5 6 *], Shu Zhao[7]

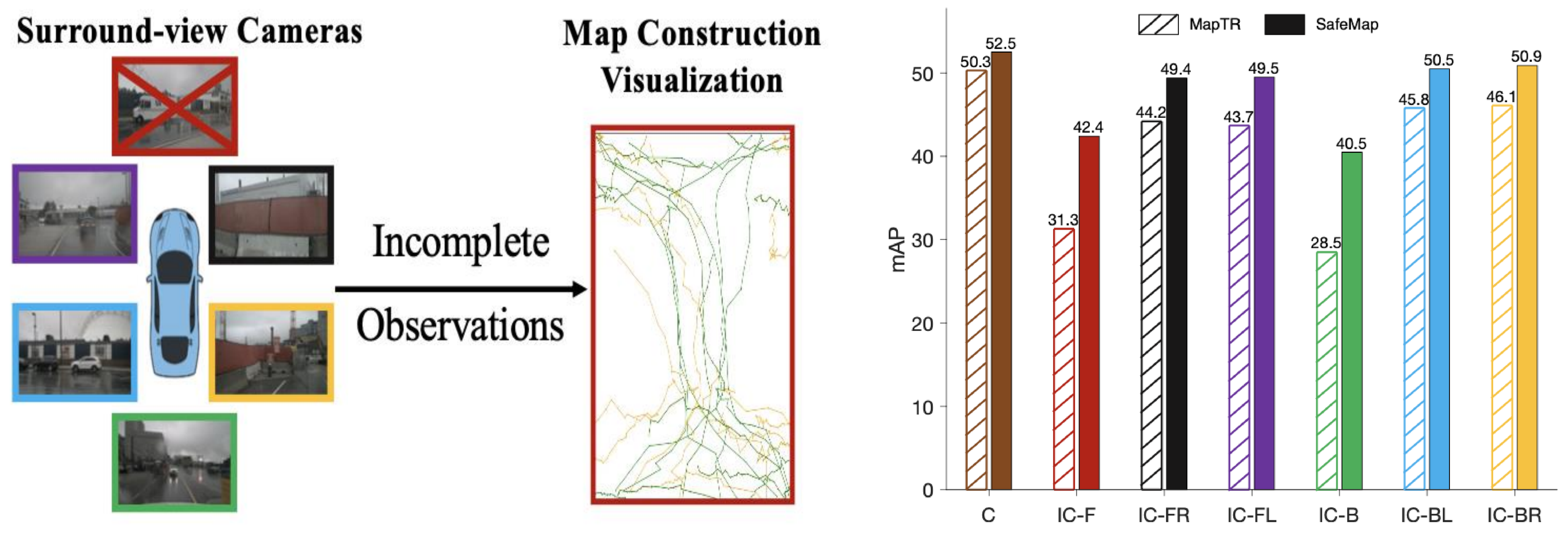
[1]Beijing Academy of Artificial Intelligence [2]National University of Singapore [3]Institute of Information Engineering, CAS [4]Wuhan University

[5]Hefei University of Technology [6]Intelligent Interconnected Systems Laboratory of Anhui Province [7]Independent Researcher

{xshao@baai.ac.cn,diaoyunfeng@hfut.edu.cn}

ICML International Conference On Machine Learning 2025

Section 1 :Motivation/Contribution



Map Construction Performance under Complete/Incomplete Observations

Motivation:

- Robust high-definition (HD) map construction is vital for autonomous driving, yet existing methods often struggle with incomplete multi-view camera data.
- As shown in Fig. 1, the absence of crucial visual information can significantly degrade overall map construction. Thus, it be- comes crucial to enhance the robustness of online vectorized HD map construction under incomplete visual observations. Overcoming these challenges is essential for ensuring safe navigation, particularly in complex and extreme driving scenarios, thereby significantly contributing to the overall reliability of autonomous systems. ***This paper presents SafeMap, a novel framework specifically designed to ensure accuracy even when certain camera views are missing.***

Contribution:

- We present SafeMap, a robust HD map construction framework that *ensures* high accuracy and reliability even in the presence of missing camera views.
- We introduce two innovative techniques in SafeMap: 1) the Gaussian-based Perspective View Reconstruction module, which utilizes relationships among available camera views to infer missing information through Gaussian-based reference point sampling, and 2) the Distillation-based BEV Correction module to further correct the BEV feature extracted from incomplete observations.
- SafeMap outperforms state-of-the-art methods in both complete and incomplete scenarios, demonstrating superior performance and robustness, thereby establishing a strong baseline for HD map construction research.

Section 2: Method

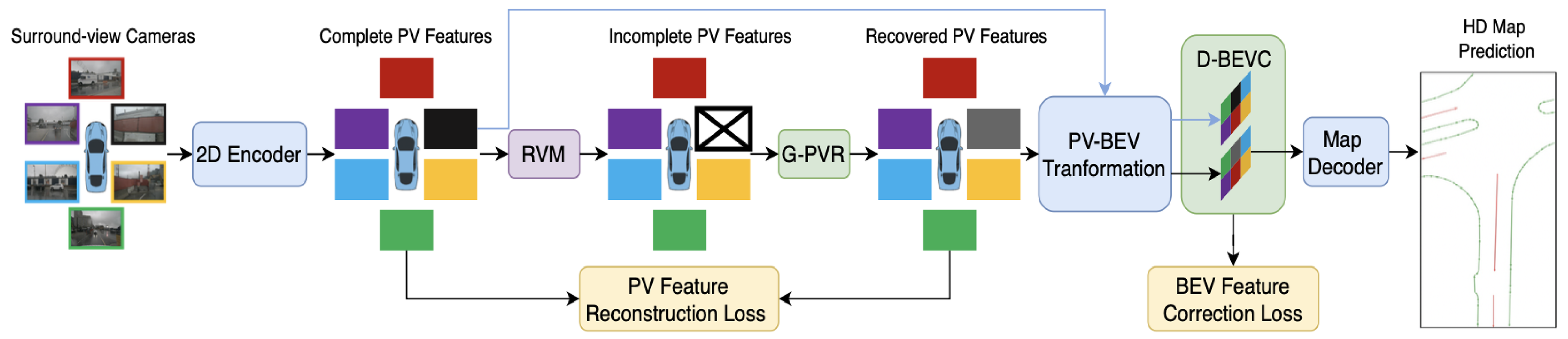


Figure 2. **Overview of the SafeMap Framework.** We first extract features from complete multi-view camera images and efficiently transform them into a unified BEV space using view transformations. To simulate emergency scenarios involving camera failures, we employ a Random View Masking (RVM) and recovery scheme. Specifically, we introduce a novel Gaussian-based Perspective View Reconstruction (G-PVR) module and a Distillation-based Bird's-Eye-View Correction (D-BEVC) module to reconstruct the missing view information. Finally, the reconstructed BEV features are processed by a map decoder and prediction heads for HD map construction.

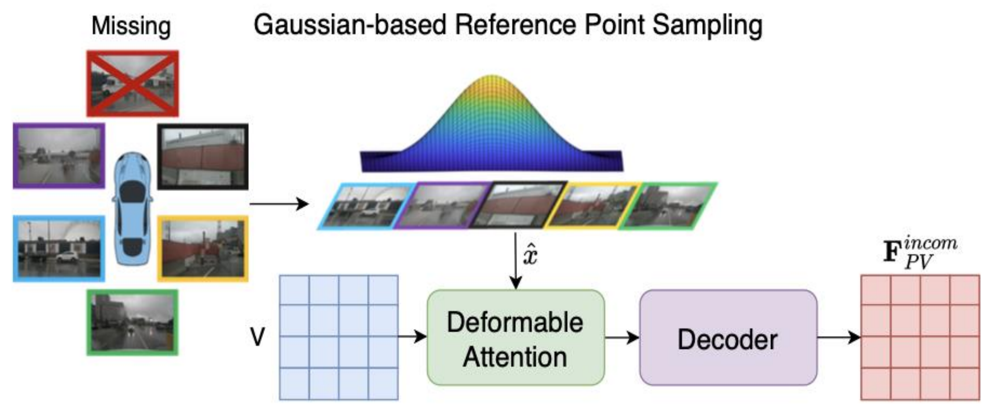


Figure 3. Illustration of the proposed Gaussian-based Perspective View Reconstruction (G-PVR) module.

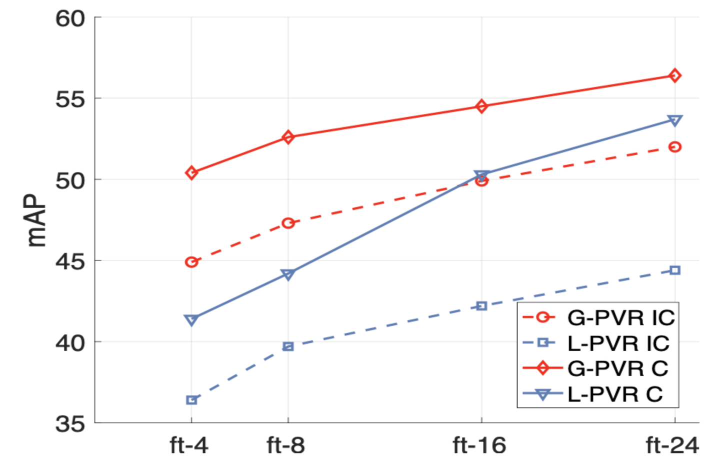


Figure 4. Comparisons of the Gaussian-based PVR *vs.* Local PVR. ft-N denotes fine-tuning N epochs.

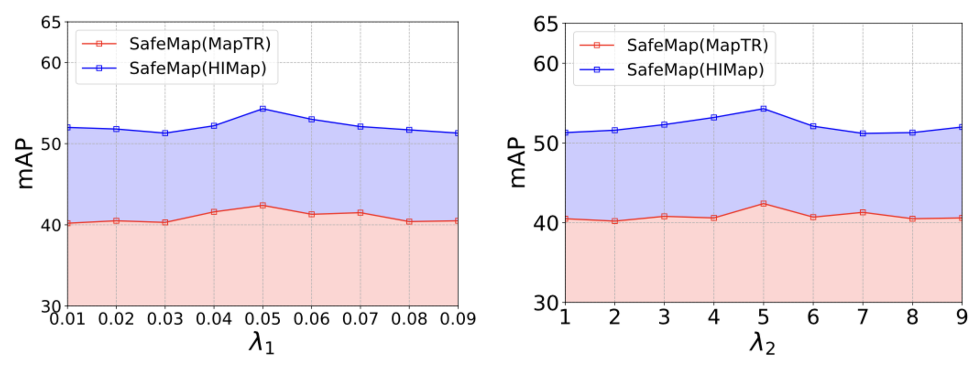


Figure 5. **Sensitivity Analysis of Hyperparameters λ_1 and λ_2.**

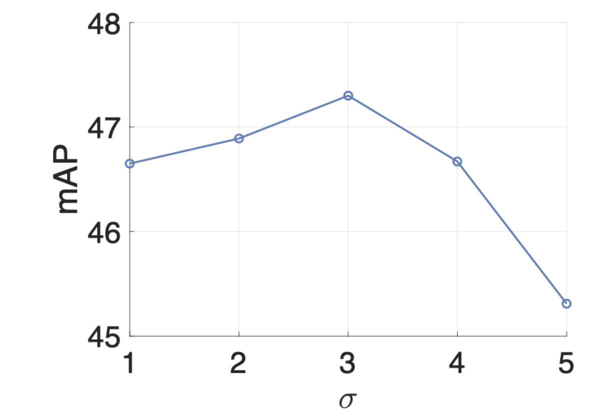


Figure 6. **Sensitivity Analysis of Hyperparameter Variance σ.**

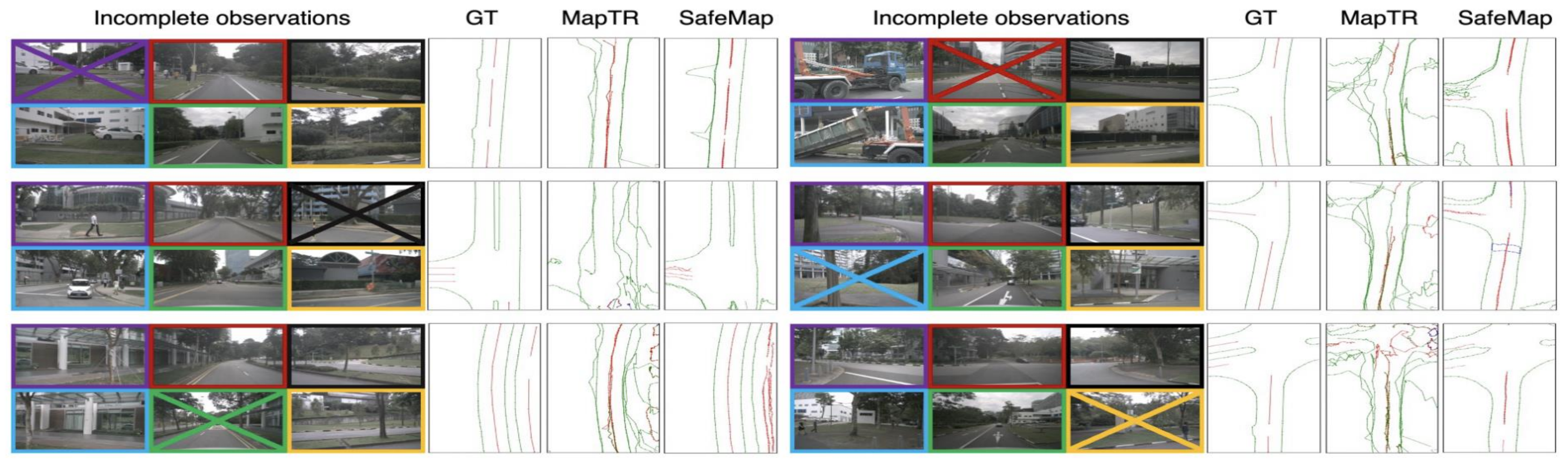


Figure 7. Qualitative Comparisons. The camera view marked with the symbol ✕ indicates the absence of this perspective.

Section 3:Experiments

Table 1. Performance comparisons with (Liao et al., 2023a) when losing each of six camera views on the nuScenes validation set.

Standard	Method	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
All Views	MapTR	46.3	51.5	53.1	50.3
	Ours	**48.1**	**54.3**	**55.3**	**52.5**
View Missing	**Method**	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	**mAP**
Front View (Center)	MapTR	25.7	34.5	33.6	31.3
	Ours	**36.6**	**45.0**	**45.8**	**42.4**
Front Left View (Left)	MapTR	37.9	47.8	45.6	43.7
	Ours	**44.3**	**52.0**	**52.4**	**49.5**
Front Right View (Right)	MapTR	38.8	46.6	47.1	44.2
	Ours	**45.3**	**52.3**	**51.7**	**49.4**
Back View (Center)	MapTR	33.4	25.2	27.0	28.5
	Ours	**39.6**	**40.8**	**41.2**	**40.5**
Back Left View (Left)	MapTR	41.3	48.3	47.8	45.8
	Ours	**45.5**	**53.1**	**52.9**	**50.5**
Back Right View (Right)	MapTR	41.2	49.5	47.8	46.1
	Ours	**46.4**	**53.1**	**53.2**	**50.9**

Table 2. Performance comparisons with (Zhou et al., 2024) when losing each of six camera views on the nuScenes validation set.

Standard	Method	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
All Views	HIMap	62.2	66.5	67.9	65.5
	Ours	**62.6**	**66.7**	**68.7**	**66.0**
View Missing	**Method**	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	**mAP**
Front View (Center)	HIMap	39.2	41.6	33.1	38.0
	Ours	**50.7**	**55.7**	**56.3**	**54.3**
Front Left View (Left)	HIMap	51.5	59.6	60.0	57.0
	Ours	**59.2**	**64.2**	**65.7**	**63.0**
Front Right View (Right)	HIMap	57.0	62.1	62.6	60.6
	Ours	**60.0**	**64.2**	**66.1**	**63.4**
Back View (Center)	HIMap	46.3	31.4	21.7	33.1
	Ours	**51.4**	**50.8**	**51.6**	**51.3**
Back Left View (Left)	HIMap	58.1	63.8	64.2	62.0
	Ours	**60.7**	**65.4**	**67.0**	**64.4**
Back Right View (Right)	HIMap	58.5	62.8	63.4	61.6
	Ours	**60.8**	**65.1**	**66.5**	**64.2**

Table 3. Performance comparison on MapTR (Liao et al., 2023a) when losing each of seven camera views on Argoverse2 val set.

Standard	Method	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
All Views	MapTR	57.7	58.9	59.4	58.7
	Ours	**58.7**	**59.7**	**60.6**	**59.7**
View Missing	**Method**	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	**mAP**
Front View (Center)	MapTR	50.8	49.8	53.9	51.5
	Ours	**53.8**	**54.9**	**58.1**	**55.6**
Front Left View (Left)	MapTR	51.8	55.4	53.4	53.5
	Ours	**54.6**	**58.6**	**57.9**	**57.0**
Front Right View (Right)	MapTR	52.7	57.3	54.2	54.7
	Ours	**55.6**	**58.9**	**57.3**	**57.3**
Rear Left View (Left)	MapTR	50.5	48.1	50.0	49.5
	Ours	**54.6**	**53.9**	**55.5**	**54.7**
Rear Right View (Right)	MapTR	49.1	52.6	47.0	49.6
	Ours	**54.8**	**57.2**	**54.0**	**55.3**
Side Left View (Left)	MapTR	55.0	57.9	57.2	56.7
	Ours	**57.4**	**59.5**	**59.7**	**58.9**
Side Right View (Right)	MapTR	56.0	58.3	57.8	57.4
	Ours	**58.0**	**59.3**	**59.8**	**59.1**

Table 4. Ablation study of components on the Gaussian-based Perspective View Reconstruction module (G-PVR) and the Distill-based BEV Correction module (D-BEVC).

G-PVR	D-BEVC	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
✗	✗	36.4	42.0	41.5	39.9
✓	✗	42.7	47.7	49.2	46.5
✗	✓	42.4	47.7	49.4	46.5
✓	✓	**42.9**	**49.4**	**49.5**	**47.3**

Table 5. Ablation study on the use of the G-PVR module.

Setting	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
SafeMap (*w/o* G-PVR)	42.4	47.7	49.4	46.5
SafeMap (*w/* Mean-PVR)	35.8	41.0	42.8	39.9
SafeMap (*w/* MAE-PVR)	42.3	47.9	49.2	46.5
SafeMap (*w/* Standard-PVR)	42.6	48.9	48.7	46.8
SafeMap (*w/* Gaussian-PVR)	**42.9**	**49.4**	**49.5**	**47.3**

Table 6. Ablation study on the use of the distillation loss.

Method	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
SafeMap (*w/o* D-BEVC)	42.7	47.7	49.2	46.5
SafeMap (*w/* KL)	42.3	48.6	49.5	46.8
SafeMap (*w/* L_1)	42.8	49.2	49.4	47.1
SafeMap (*w/* L_2)	**42.9**	**49.4**	**49.5**	**47.3**

Table 7. Impact of different numbers of missing views.

Method	#View	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
MapTR	1×	36.4	42.0	41.5	39.9
	2×	27.5	31.3	29.9	29.6
	3×	18.8	20.9	18.8	19.5
	4×	11.0	11.6	9.4	10.6
	5×	4.6	4.3	3.0	4.0
SafeMap	1×	42.9	49.4	49.5	47.3
	2×	33.5	37.0	35.0	35.2
	3×	23.7	24.4	20.2	22.8
	4×	15.6	16.2	12.6	14.8
	5×	6.7	7.0	4.3	6.0

Table 8. Accuracy-computation analysis. We report the mAP performance under the "complete" / "incomplete" observations.

Method	mAP	GPU Mem	Param	FPS
MapTR	50.3 / 39.9	2298 MB	39.1 M	21.5
SafeMap	**52.5 / 47.3**	**2300 MB**	**39.5 M**	**21.4**
HIMap	65.5 / 52.1	4091 MB	68.1 M	9.7
SafeMap	**66.0 / 60.1**	**4155 MB**	**71.7 M**	**9.2**

Table 9. Experimental results on the robustness of HD map construction under camera sensor corruptions.

Method	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP	mRR↑	mCE↓
MapTR	46.3	51.5	53.1	50.3	49.3	100.0
SafeMap	**48.1**	**54.3**	**55.3**	**52.5**	**51.2**	**90.6**
HIMap	62.2	66.5	67.9	65.5	56.6	100.0
SafeMap	**62.6**	**66.7**	**68.7**	**66.0**	**62.8**	**83.2**