OptiMVMap: Offline Vectorized Map Construction via Optimal Multi-vehicle Perspectives

Supplementary Material

Table 1. Robustness of injected pose noise on nuScenes. We report mAP under varying standard deviations of Gaussian rotation/translation noise.

Rotation noise std. (rad) →					
Methods	0	0.005	0.01	0.02	0.05
MapTRv2	61.5	60.8	58.0	50.6	32.3
MapTRv2 + OptiMVMap	71.0	70.8	70.6	69.8	65.7
Translation noise std. (m) →					
Methods	0	0.05	0.1	0.5	1.0
MapTRv2	61.5	61.2	59.8	35.6	18.8
MapTRv2 + OptiMVMap	71.0	70.8	70.7	68.0	62.3

1. Implementation Details

We employ ResNet50 [1] as the backbone network for image processing and ResNet18 for uncertainty map feature extraction in OVS. We utilize InternImage [2] as the pretrained model for semantic prior generation. The default settings for instance queries, point queries, and decoder layers are 50, 20, and 6, respectively. We use the AdamW optimizer with a learning rate of 6×10^{-4} and a weight decay of 0.01. All models are trained using 8 80GB NVIDIA Tesla A100 GPUs, with a batch size of 4 per node, leading to a total batch size of 32. For the overall loss, we set $\beta_o = 1$, $\beta_m = 1$, $\beta_d = 1$, $\beta_c = 1$.

2. Noise robustness under pose perturbations.

Tab. 1 injects Gaussian rotation/translation noise into vehicle poses. From 0 to 0.05rad rotation noise, the MapTRv2 baseline drops -29.2 mAP, whereas CVA retains most accuracy (-5.3 mAP). At 1m translation std., the baseline collapses to 18.8 mAP while our model still achieves 62.3 mAP. These show that explicit cross-vehicle alignment via CVA is essential for making BEV fusion robust to extrinsic inaccuracies.

3. Performance on Geospatially Disjoint Split.

Tab. 2 shows results on StreamMapNet's disjoint split. OptiMVMap outperforms SOTA methods HRMapNet (+1.2) and MapExpert (+0.5 mAP), confirming generalization beyond training locations.

Table 2. **Performance on StreamMapNet Geo-Disjoint Split.**

Method	epoch	AP_{div}	AP_{ped}	AP_{bnd}	mAP
StreamMapNet	24	30.1	29.6	41.9	33.9
HRMapNet	24	30.3	36.9	44.0	37.1
Ours	24	**32.0**	**38.1**	**45.0**	**38.3**
MapTracker	72	30.0	45.9	45.1	40.3
MapExpert	100	34.1	46.7	45.1	42.0
Ours	72	**35.2**	**47.8**	**45.9**	**42.5**

Table 3. Computational Efficiency.

Map Construction Efficiency			
Baseline	**FPS**	**Ours**	**FPS**
HRMapNet	17.0	K=1	10.6
StreamMapNet	14.2	K=3	5.4
MapTRv2	14.1	K=5	4.8
Vehicle Selection Latency			
	Random	**Closest**	**OVS**
Latency (ms)	0.27	0.28	12.6

4. Computational Efficiency.

Tab. 3 shows our method at K=1 achieves near real-time inference comparable to MapTRv2. FPS decreases gradually with K, remaining acceptable for offline pipelines. OVS operates on low-resolution uncertainty maps, adding only 12.6ms/sample.

5. Extra Qualitative Results

We present 15 additional qualitative examples from the NuScenes and the Argoverse2 dataset using only 3 helper vehicles, covering challenging scenarios such as nighttime crossroads, severe occlusion in rainy day, and distant, curved lane-shape. These results reveal that OptiMVMap can select the right non-ego vehicles that provide complementary information for the uncertainty areas while not introducing too much noise. This enables the map decoder to restore missing map details due to occlusion or remote distance, and produce markedly higher-precision reconstructions. By contrast, HRMapNet occasionally underperforms even MapTRv2 in these challenging settings, indicating that the absence of explicit noise filtering and data-source selection in HRMapNet leads to error accumulation and ultimately degrades map quality.

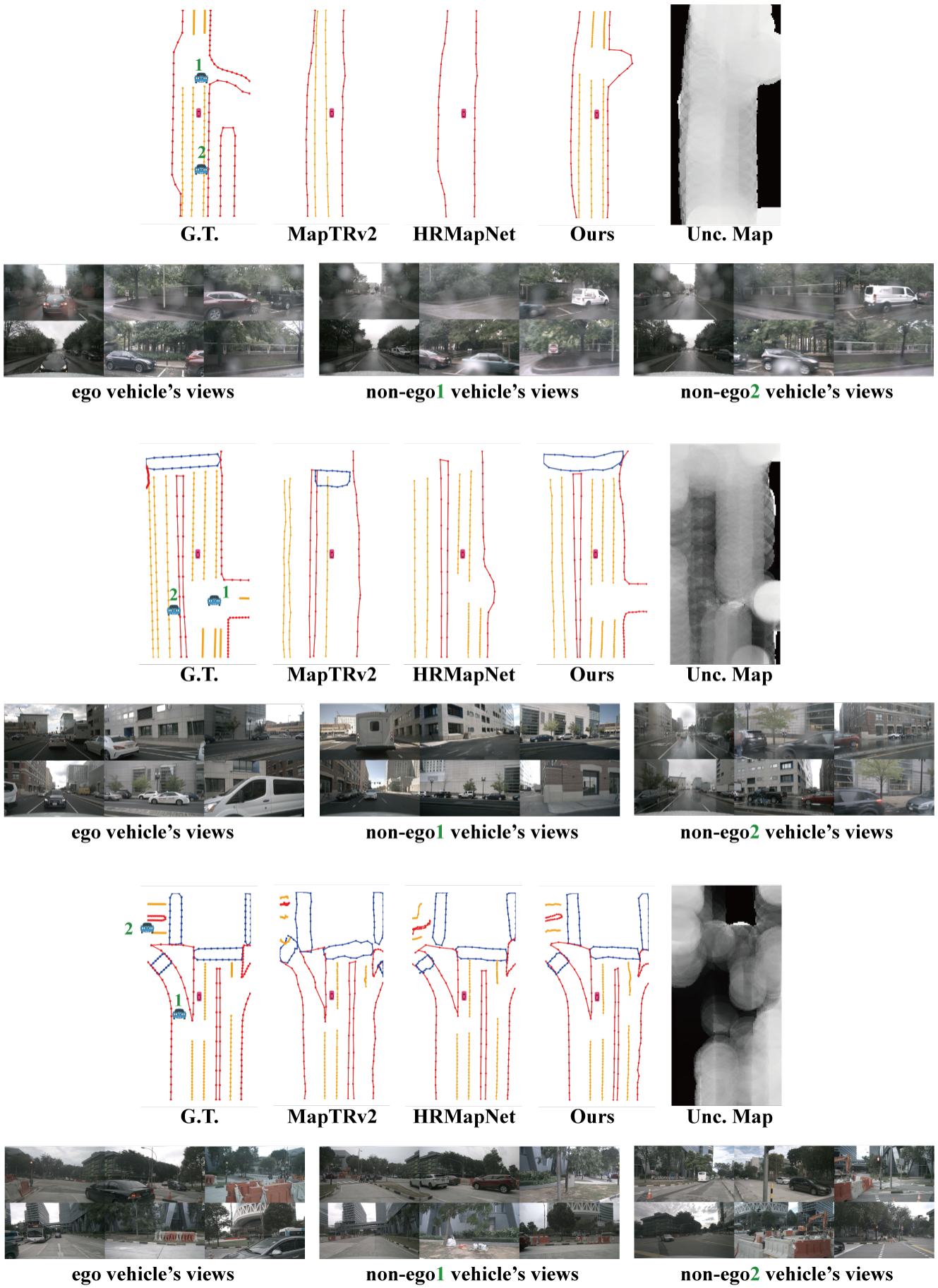

Figure 1. More qualitative results on the nuScenes dataset.

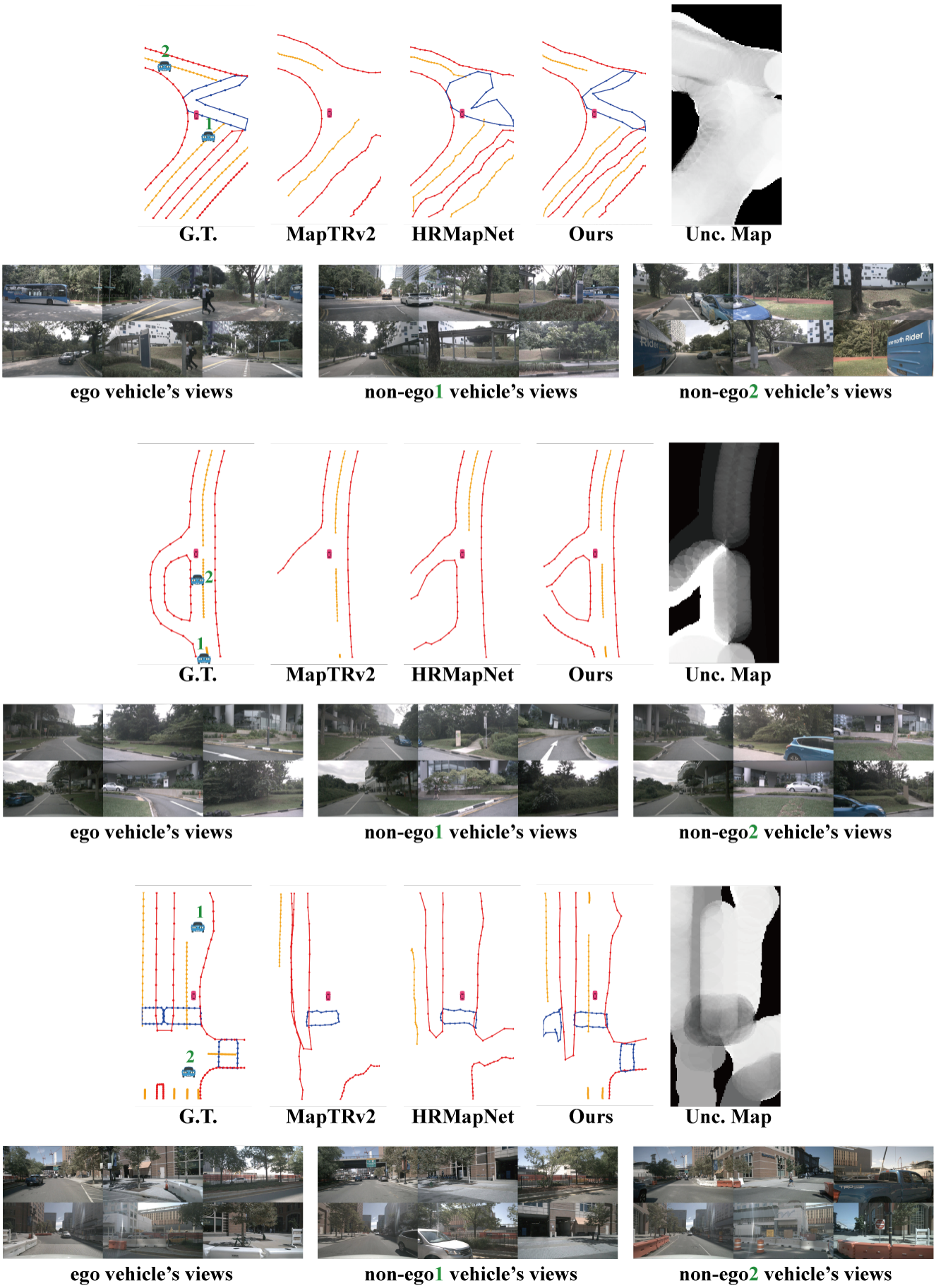

Figure 2. More qualitative results on the nuScenes dataset.

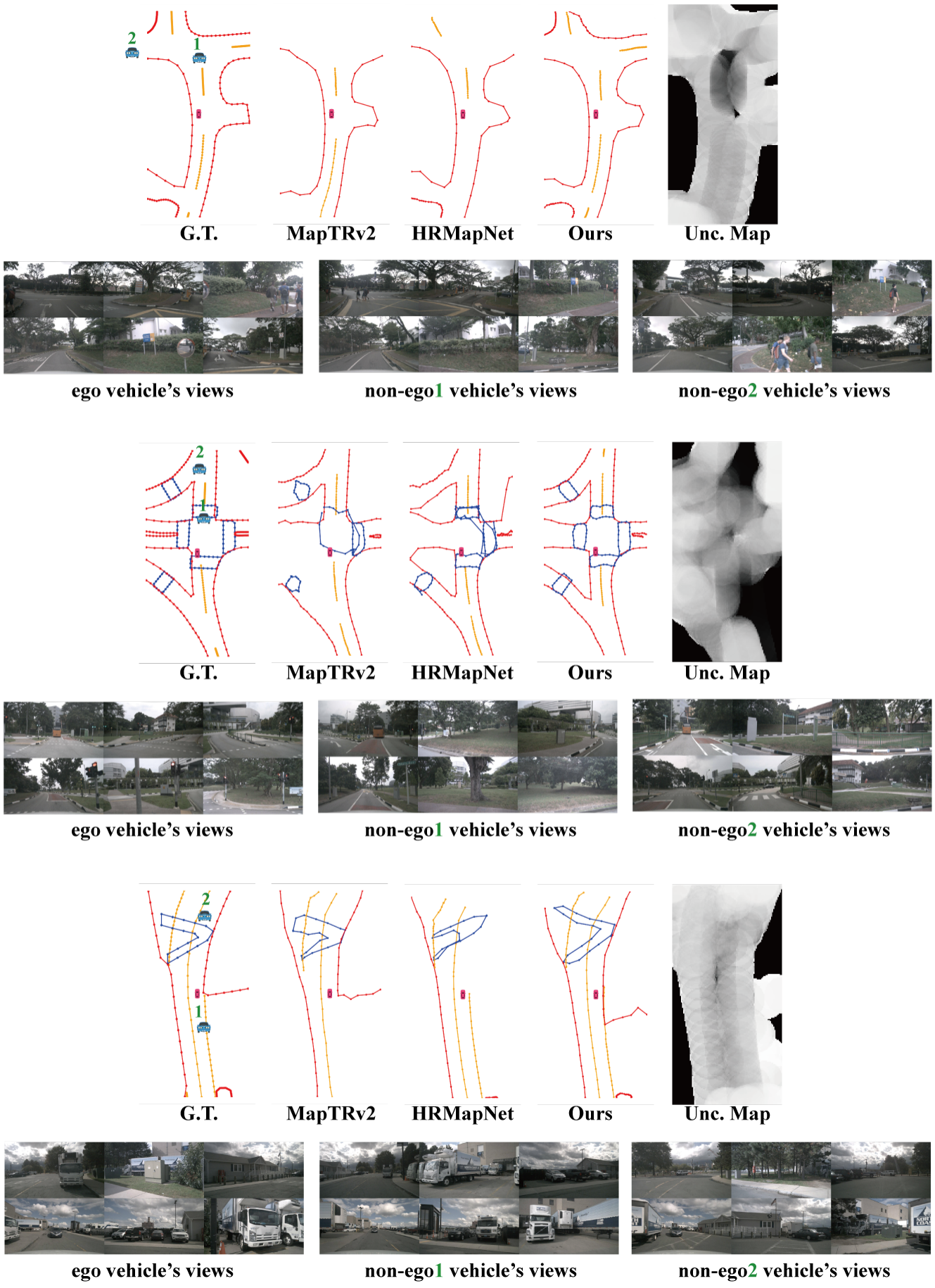

Figure 3. More qualitative results on the nuScenes dataset.

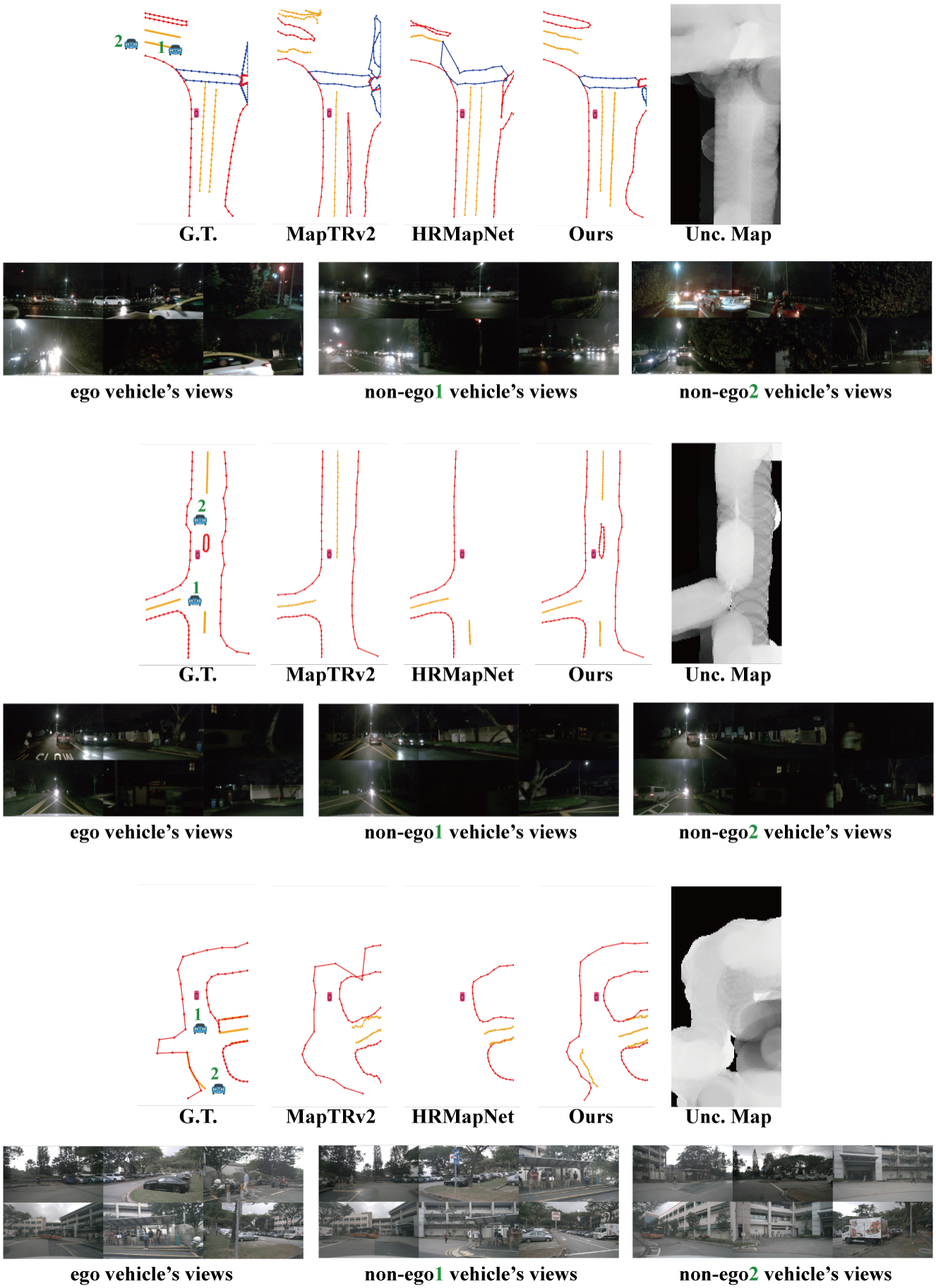

Figure 4. More qualitative results on the nuScenes dataset.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[2] Wenhai Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, Xiaogang Wang, and Yu Qiao. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1