GraphBEV: Towards Robust BEV Feature Alignment for Multi-Modal 3D Object Detection

Ziying Song[1,2], Lei Yang[3], Shaoqing Xu[4], Lin Liu[1,2], Dongyang Xu[3],
Caiyan Jia[1,2] *, Feiyang Jia[1,2], and Li Wang[5]

[1] School of Computer and Information Technology, Beijing Jiaotong University
[2] Beijing Key Lab of Traffic Data Analysis and Mining, China
[3] School of Vehicle and Mobility, Tsinghua University
[4] Department of Electrome chanical Engineering, University of Macau
[5] School of Mechanical Engineering, Beijing Institute of Technology
{songziying,cyjia}@bjtu.edu.cn

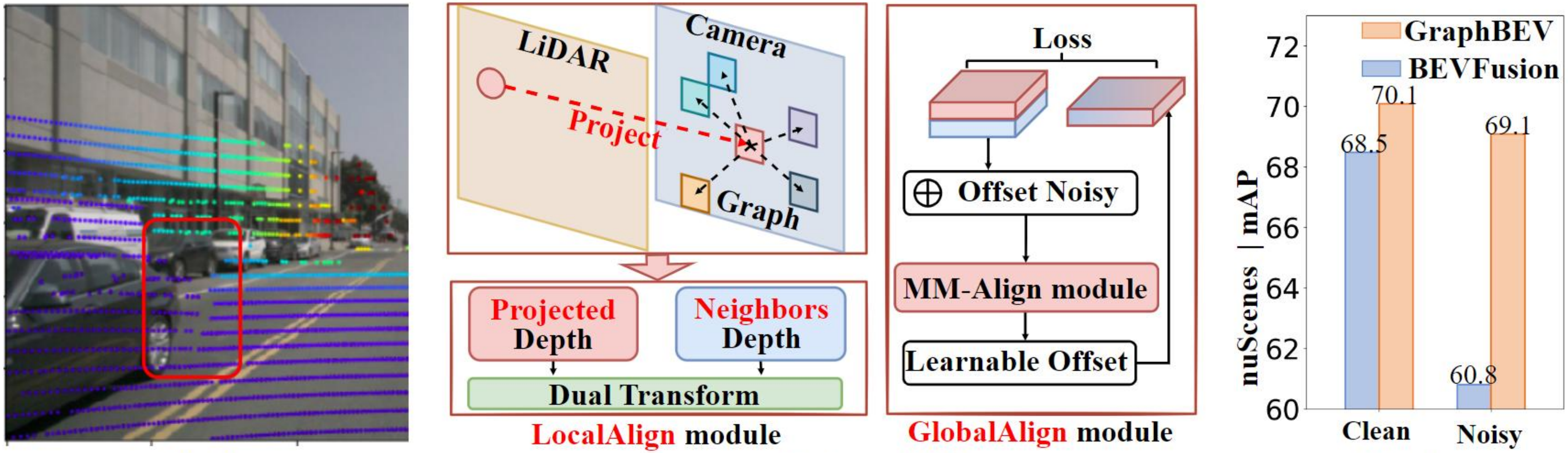
Our Code

Our Paper

Wechat

Other Paper

Problem Formulation

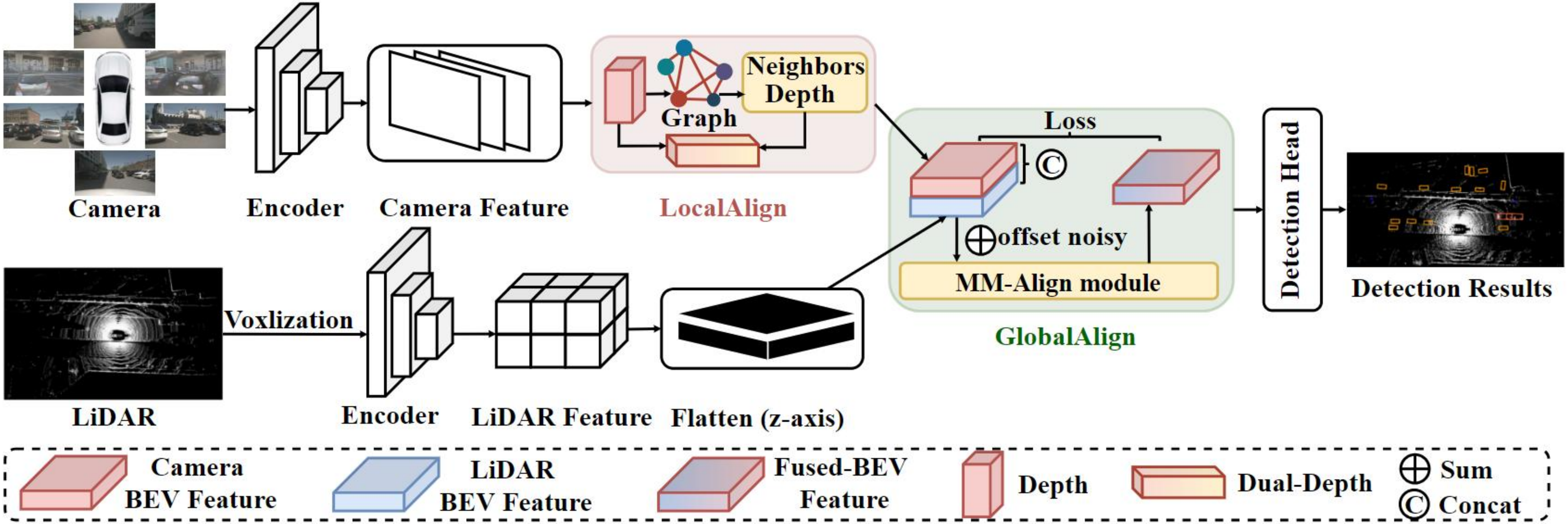


(a) Misalignment

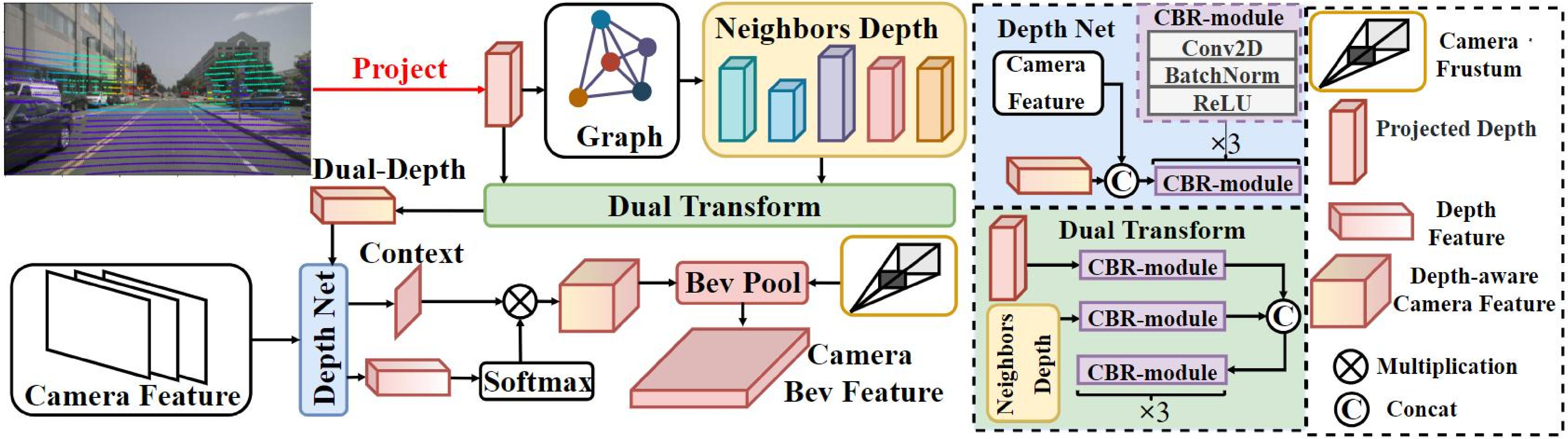
(b) GraphBEV

(c) Performance

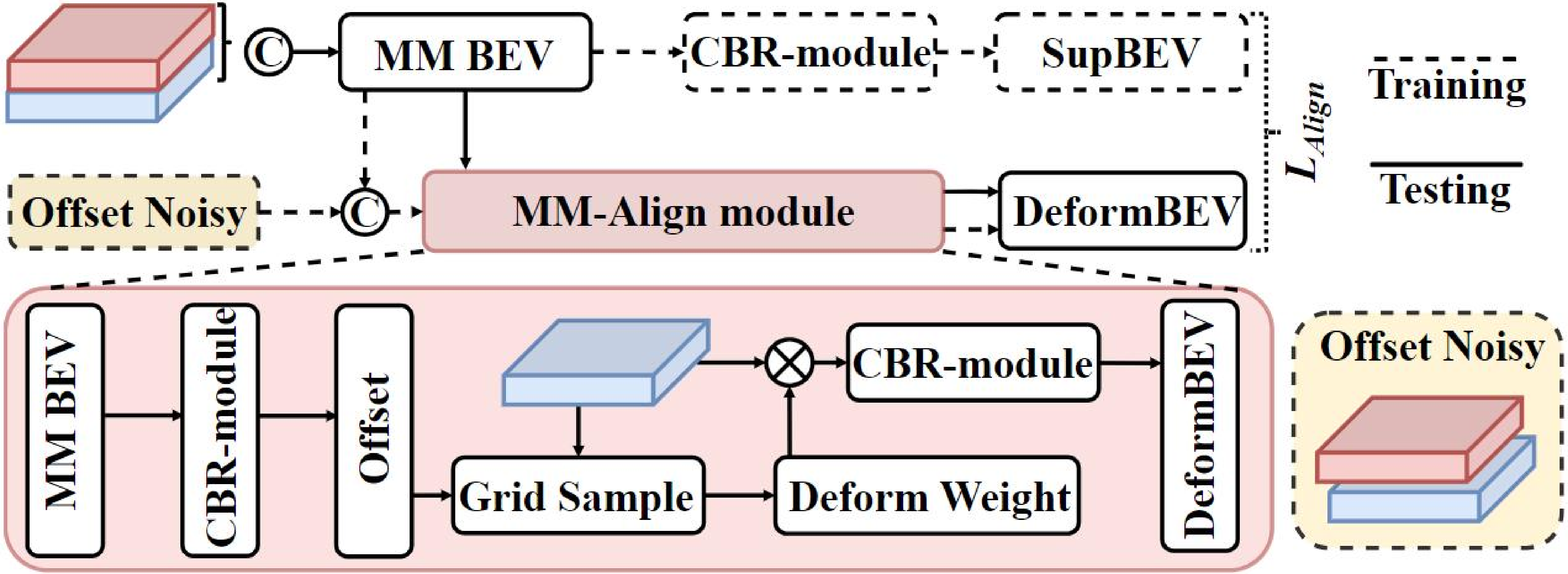
Method


Camera
Encoder
Camera Feature
Graph
Neighbors Depth
LocalAlign
Loss
©
⊕offset noisy
MM-Align module
GlobalAlign
Detection Head
Detection Results
LiDAR
Voxlization
Encoder
LiDAR Feature
Flatten (z-axis)
Camera BEV Feature
LiDAR BEV Feature
Fused-BEV Feature
Depth
Dual-Depth
⊕ Sum
© Concat

Method


Project
Graph
Neighbors Depth
Dual-Depth
Dual Transform
Context
Camera Feature
Depth Net
Bev Pool
Softmax
Camera
Bev Feature
Depth Net
Camera
Feature
CBR-module
Conv2D
BatchNorm
ReLU
×3
CBR-module
Dual Transform
CBR-module
CBR-module
CBR-module
Neighbors
Depth
×3
Camera
Frustum
Projected Depth
Depth
Feature
Depth-aware
Camera Feature
Multiplication
Concat

Method



MM BEV
CBR-module
SupBEV
L_{Align}
Training
Testing
Offset Noisy
MM-Align module
DeformBEV
MM BEV
CBR-module
Offset
Grid Sample
Deform Weight
CBR-module
DeformBEV
Offset Noisy

Experiments

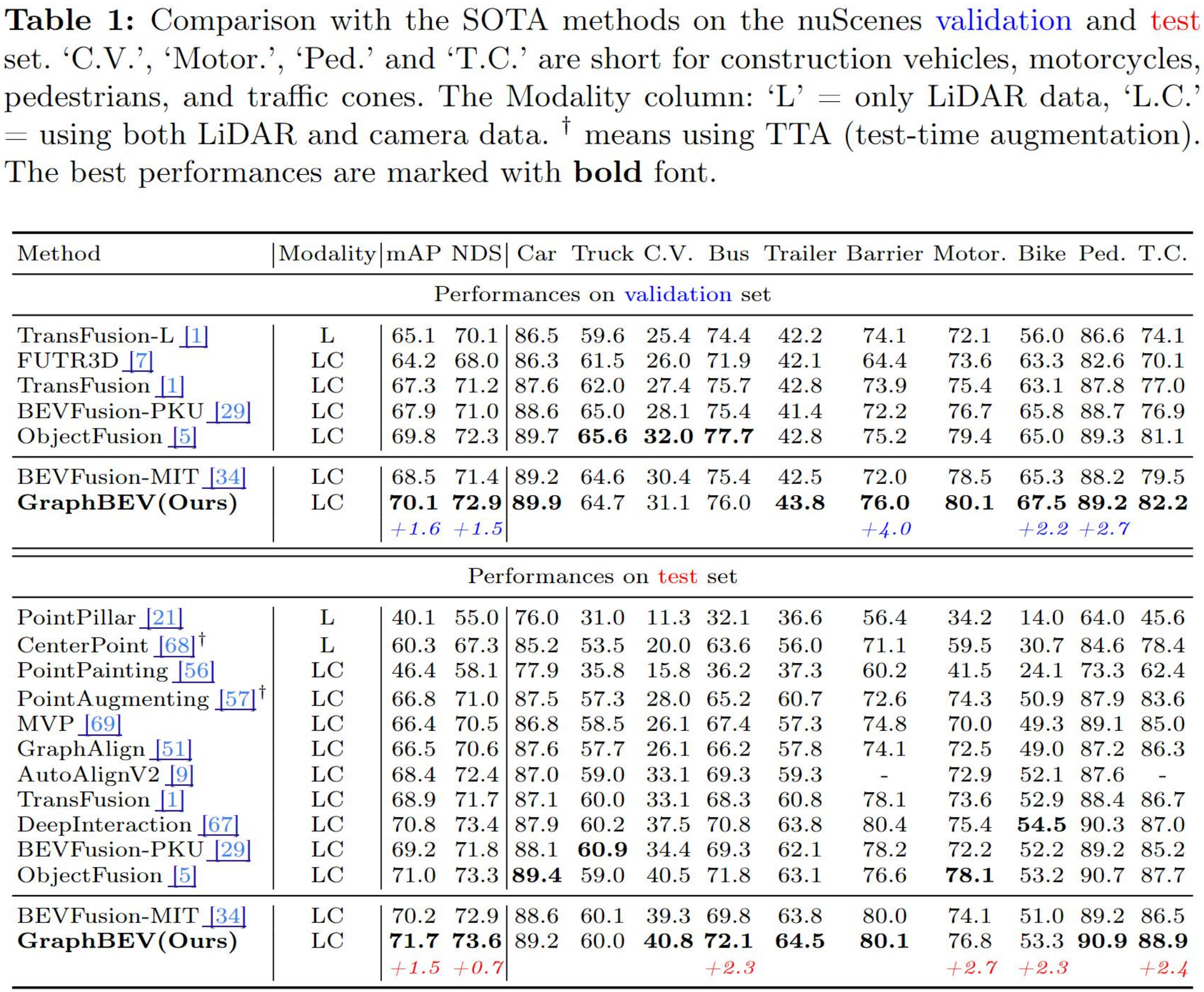


Table 1: Comparison with the SOTA methods on the nuScenes validation and test set. 'C.V.', 'Motor.', 'Ped.' and 'T.C.' are short for construction vehicles, motorcycles, pedestrians, and traffic cones. The Modality column: 'L' = only LiDAR data, 'L.C.' = using both LiDAR and camera data. † means using TTA (test-time augmentation). The best performances are marked with **bold** font.

Method	Modality	mAP	NDS	Car	Truck	C.V.	Bus	Trailer	Barrier	Motor.	Bike	Ped.	T.C.
Performances on validation set													
TransFusion-L [1]	L	65.1	70.1	86.5	59.6	25.4	74.4	42.2	74.1	72.1	56.0	86.6	74.1
FUTR3D [7]	LC	64.2	68.0	86.3	61.5	26.0	71.9	42.1	64.4	73.6	63.3	82.6	70.1
TransFusion [1]	LC	67.3	71.2	87.6	62.0	27.4	75.7	42.8	73.9	75.4	63.1	87.8	77.0
BEVFusion-PKU [29]	LC	67.9	71.0	88.6	65.0	28.1	75.4	41.4	72.2	76.7	65.8	88.7	76.9
ObjectFusion [5]	LC	69.8	72.3	89.7	**65.6**	**32.0**	**77.7**	42.8	75.2	79.4	65.0	89.3	81.1
BEVFusion-MIT [34]	LC	68.5	71.4	89.2	64.6	30.4	75.4	42.5	72.0	78.5	65.3	88.2	79.5
GraphBEV(Ours)	LC	**70.1**	**72.9**	**89.9**	64.7	31.1	76.0	**43.8**	**76.0**	**80.1**	**67.5**	**89.2**	**82.2**
		+1.6	*+1.5*						*+4.0*		*+2.2*	*+2.7*	
Performances on test set													
PointPillar [21]	L	40.1	55.0	76.0	31.0	11.3	32.1	36.6	56.4	34.2	14.0	64.0	45.6
CenterPoint [68]†	L	60.3	67.3	85.2	53.5	20.0	63.6	56.0	71.1	59.5	30.7	84.6	78.4
PointPainting [56]	LC	46.4	58.1	77.9	35.8	15.8	36.2	37.3	60.2	41.5	24.1	73.3	62.4
PointAugmenting [57]†	LC	66.8	71.0	87.5	57.3	28.0	65.2	60.7	72.6	74.3	50.9	87.9	83.6
MVP [69]	LC	66.4	70.5	86.8	58.5	26.1	67.4	57.3	74.8	70.0	49.3	89.1	85.0
GraphAlign [51]	LC	66.5	70.6	87.6	57.7	26.1	66.2	57.8	74.1	72.5	49.0	87.2	86.3
AutoAlignV2 [9]	LC	68.4	72.4	87.0	59.0	33.1	69.3	59.3	-	72.9	52.1	87.6	-
TransFusion [1]	LC	68.9	71.7	87.1	60.0	33.1	68.3	60.8	78.1	73.6	52.9	88.4	86.7
DeepInteraction [67]	LC	70.8	73.4	87.9	60.2	37.5	70.8	63.8	80.4	75.4	**54.5**	90.3	87.0
BEVFusion-PKU [29]	LC	69.2	71.8	88.1	**60.9**	34.4	69.3	62.1	78.2	72.2	52.2	89.2	85.2
ObjectFusion [5]	LC	71.0	73.3	**89.4**	59.0	40.5	71.8	63.1	76.6	**78.1**	53.2	90.7	87.7
BEVFusion-MIT [34]	LC	70.2	72.9	88.6	60.1	39.3	69.8	63.8	80.0	74.1	51.0	89.2	86.5
GraphBEV(Ours)	LC	**71.7**	**73.6**	89.2	60.0	**40.8**	**72.1**	**64.5**	**80.1**	76.8	53.3	**90.9**	**88.9**
		+1.5	*+0.7*				*+2.3*			*+2.7*	*+2.3*		*+2.4*

Experiments

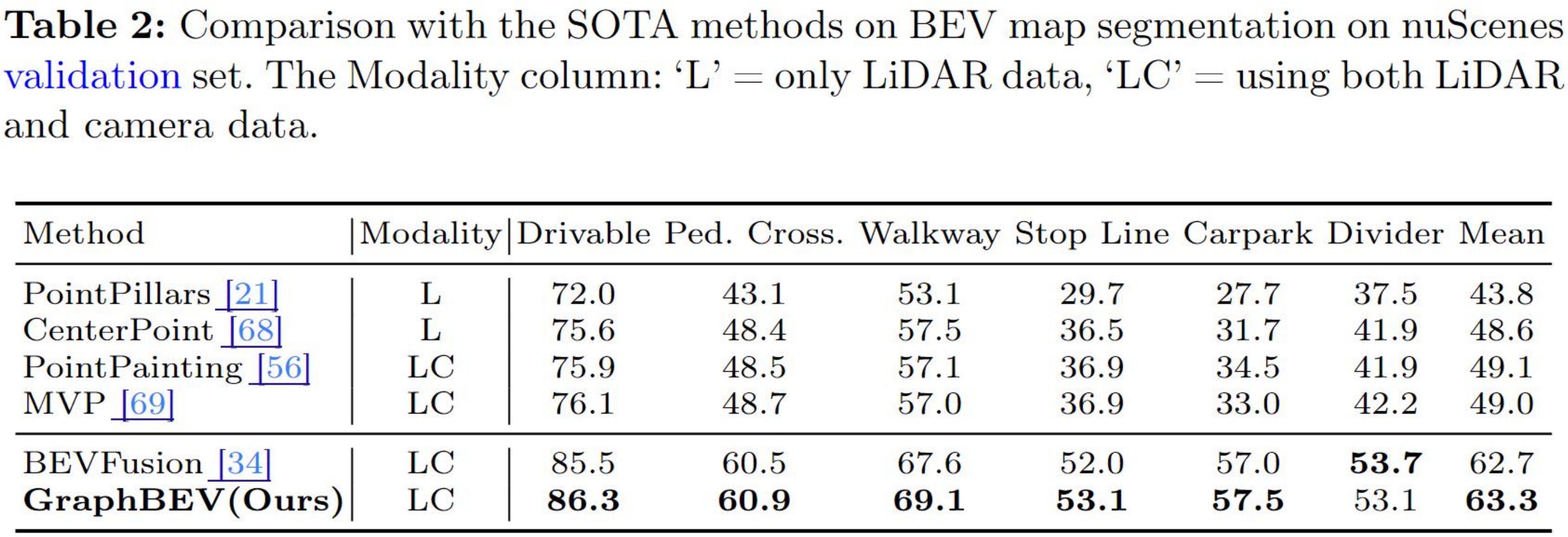


Table 2: Comparison with the SOTA methods on BEV map segmentation on nuScenes validation set. The Modality column: ‘L’ = only LiDAR data, ‘LC’ = using both LiDAR and camera data.

Method	Modality	Drivable	Ped. Cross.	Walkway	Stop Line	Carpark	Divider	Mean
PointPillars [21]	L	72.0	43.1	53.1	29.7	27.7	37.5	43.8
CenterPoint [68]	L	75.6	48.4	57.5	36.5	31.7	41.9	48.6
PointPainting [56]	LC	75.9	48.5	57.1	36.9	34.5	41.9	49.1
MVP [69]	LC	76.1	48.7	57.0	36.9	33.0	42.2	49.0
BEVFusion [34]	LC	85.5	60.5	67.6	52.0	57.0	**53.7**	62.7
GraphBEV(Ours)	LC	**86.3**	**60.9**	**69.1**	**53.1**	**57.5**	53.1	**63.3**

Experiments

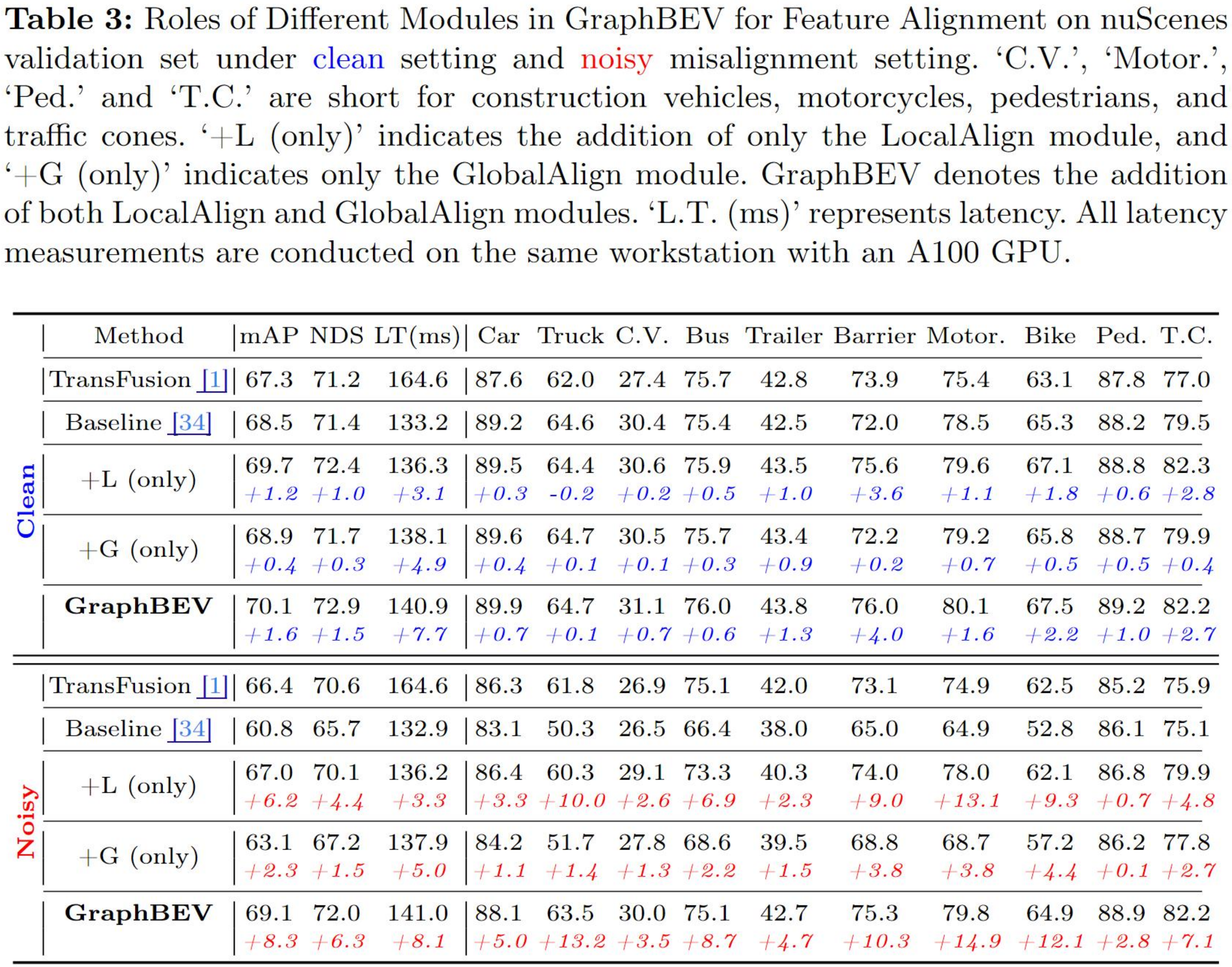


Table 3: Roles of Different Modules in GraphBEV for Feature Alignment on nuScenes validation set under clean setting and noisy misalignment setting. 'C.V.', 'Motor.', 'Ped.' and 'T.C.' are short for construction vehicles, motorcycles, pedestrians, and traffic cones. '+L (only)' indicates the addition of only the LocalAlign module, and '+G (only)' indicates only the GlobalAlign module. GraphBEV denotes the addition of both LocalAlign and GlobalAlign modules. 'L.T. (ms)' represents latency. All latency measurements are conducted on the same workstation with an A100 GPU.

	Method	mAP	NDS	LT(ms)	Car	Truck	C.V.	Bus	Trailer	Barrier	Motor.	Bike	Ped.	T.C.
Clean	TransFusion [1]	67.3	71.2	164.6	87.6	62.0	27.4	75.7	42.8	73.9	75.4	63.1	87.8	77.0
	Baseline [34]	68.5	71.4	133.2	89.2	64.6	30.4	75.4	42.5	72.0	78.5	65.3	88.2	79.5
	+L (only)	69.7 *+1.2*	72.4 *+1.0*	136.3 *+3.1*	89.5 *+0.3*	64.4 *-0.2*	30.6 *+0.2*	75.9 *+0.5*	43.5 *+1.0*	75.6 *+3.6*	79.6 *+1.1*	67.1 *+1.8*	88.8 *+0.6*	82.3 *+2.8*
	+G (only)	68.9 *+0.4*	71.7 *+0.3*	138.1 *+4.9*	89.6 *+0.4*	64.7 *+0.1*	30.5 *+0.1*	75.7 *+0.3*	43.4 *+0.9*	72.2 *+0.2*	79.2 *+0.7*	65.8 *+0.5*	88.7 *+0.5*	79.9 *+0.4*
	GraphBEV	70.1 *+1.6*	72.9 *+1.5*	140.9 *+7.7*	89.9 *+0.7*	64.7 *+0.1*	31.1 *+0.7*	76.0 *+0.6*	43.8 *+1.3*	76.0 *+4.0*	80.1 *+1.6*	67.5 *+2.2*	89.2 *+1.0*	82.2 *+2.7*
Noisy	TransFusion [1]	66.4	70.6	164.6	86.3	61.8	26.9	75.1	42.0	73.1	74.9	62.5	85.2	75.9
	Baseline [34]	60.8	65.7	132.9	83.1	50.3	26.5	66.4	38.0	65.0	64.9	52.8	86.1	75.1
	+L (only)	67.0 *+6.2*	70.1 *+4.4*	136.2 *+3.3*	86.4 *+3.3*	60.3 *+10.0*	29.1 *+2.6*	73.3 *+6.9*	40.3 *+2.3*	74.0 *+9.0*	78.0 *+13.1*	62.1 *+9.3*	86.8 *+0.7*	79.9 *+4.8*
	+G (only)	63.1 *+2.3*	67.2 *+1.5*	137.9 *+5.0*	84.2 *+1.1*	51.7 *+1.4*	27.8 *+1.3*	68.6 *+2.2*	39.5 *+1.5*	68.8 *+3.8*	68.7 *+3.8*	57.2 *+4.4*	86.2 *+0.1*	77.8 *+2.7*
	GraphBEV	69.1 *+8.3*	72.0 *+6.3*	141.0 *+8.1*	88.1 *+5.0*	63.5 *+13.2*	30.0 *+3.5*	75.1 *+8.7*	42.7 *+4.7*	75.3 *+10.3*	79.8 *+14.9*	64.9 *+12.1*	88.9 *+2.8*	82.2 *+7.1*

Experiments

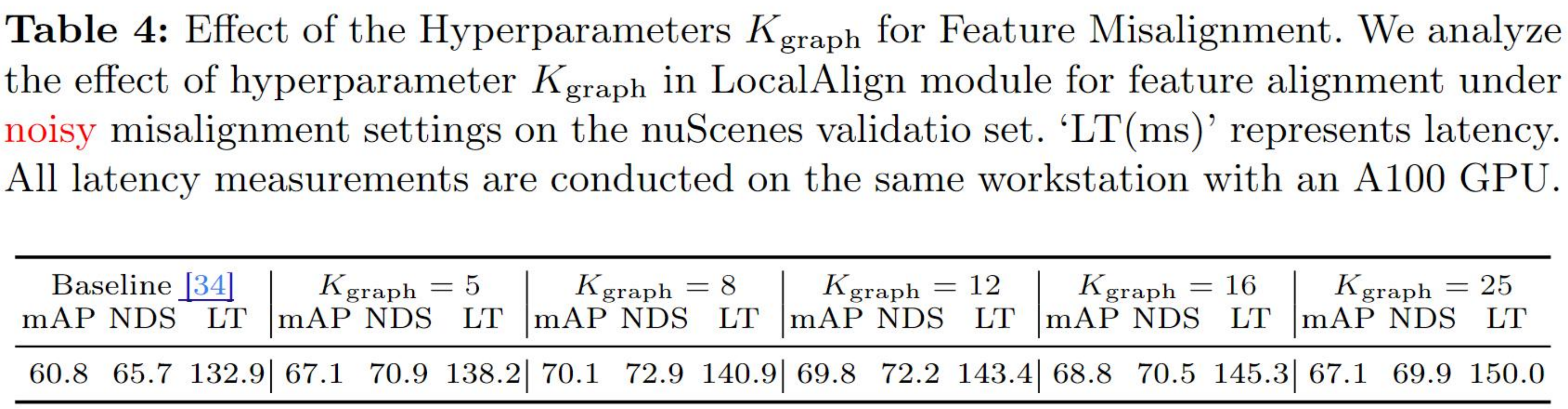


Table 4: Effect of the Hyperparameters K_{graph} for Feature Misalignment. We analyze the effect of hyperparameter K_{graph} in LocalAlign module for feature alignment under noisy misalignment settings on the nuScenes validatio set. 'LT(ms)' represents latency. All latency measurements are conducted on the same workstation with an A100 GPU.

Baseline [34]			$K_{\text{graph}} = 5$			$K_{\text{graph}} = 8$			$K_{\text{graph}} = 12$			$K_{\text{graph}} = 16$			$K_{\text{graph}} = 25$		
mAP	NDS	LT	mAP	NDS	LT	mAP	NDS	LT	mAP	NDS	LT	mAP	NDS	LT	mAP	NDS	LT
60.8	65.7	132.9	67.1	70.9	138.2	70.1	72.9	140.9	69.8	72.2	143.4	68.8	70.5	145.3	67.1	69.9	150.0

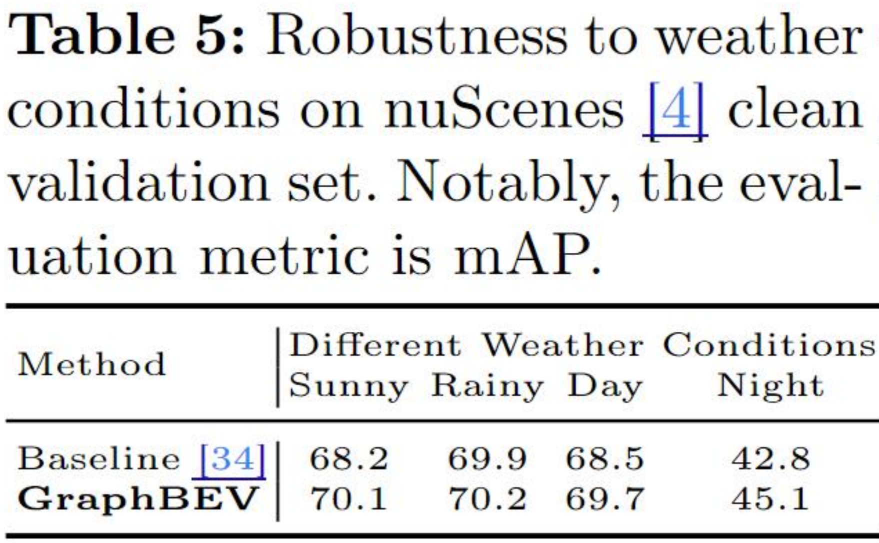


Table 5: Robustness to weather conditions on nuScenes [4] clean validation set. Notably, the evaluation metric is mAP.

Method	Different Weather Conditions			
	Sunny	Rainy	Day	Night
Baseline [34]	68.2	69.9	68.5	42.8
GraphBEV	70.1	70.2	69.7	45.1

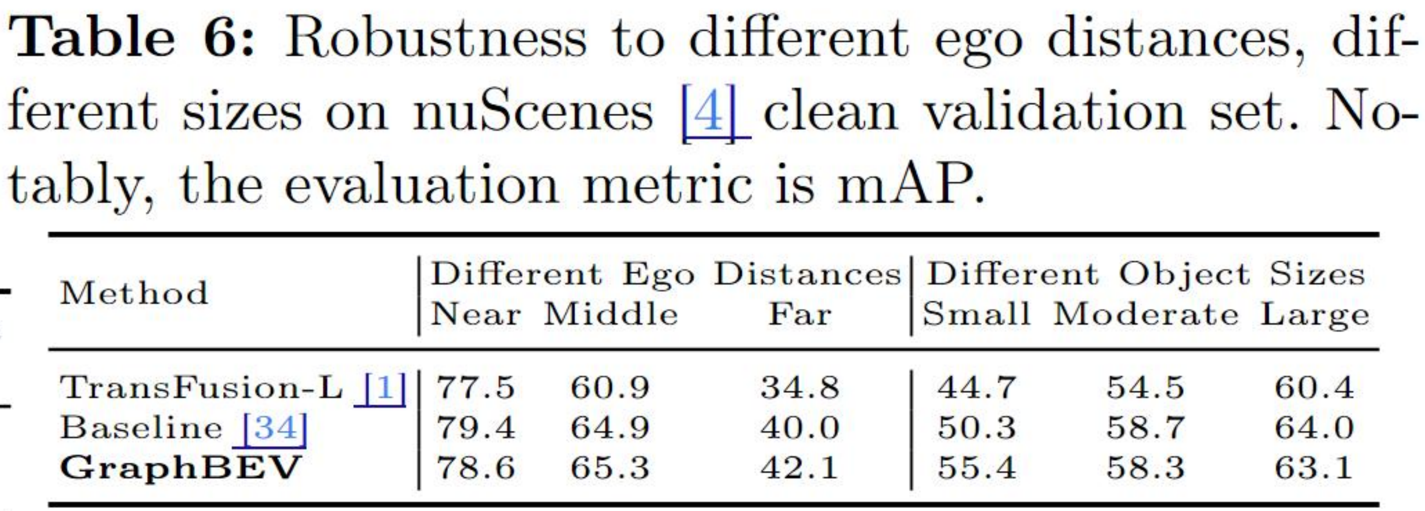


Table 6: Robustness to different ego distances, different sizes on nuScenes [4] clean validation set. Notably, the evaluation metric is mAP.

Method	Different Ego Distances			Different Object Sizes		
	Near	Middle	Far	Small	Moderate	Large
TransFusion-L [1]	77.5	60.9	34.8	44.7	54.5	60.4
Baseline [34]	79.4	64.9	40.0	50.3	58.7	64.0
GraphBEV	78.6	65.3	42.1	55.4	58.3	63.1

Thanks !

Our Code

Our Paper

Wechat

Other Paper