See through the Dark: Learning Illumination-affined Representations for Nighttime Occupancy Prediction

Yuan Wu, Zhiqiang Yan, Yigong Zhang, Xiang Li, Jian Yang

NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

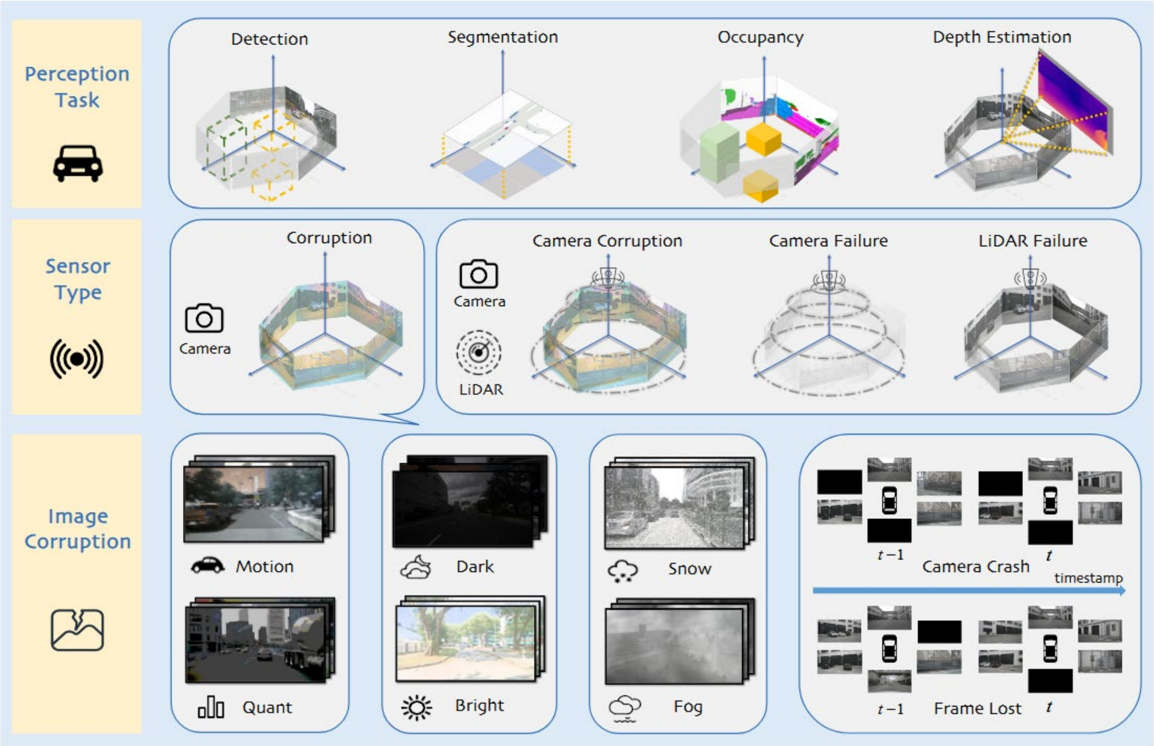
NUS National University of Singapore

NANKAI UNIVERSITY 1919

NEURAL INFORMATION PROCESSING SYSTEMS

Background

- Robust perception under extreme scenarios



RoboBEV benchmark

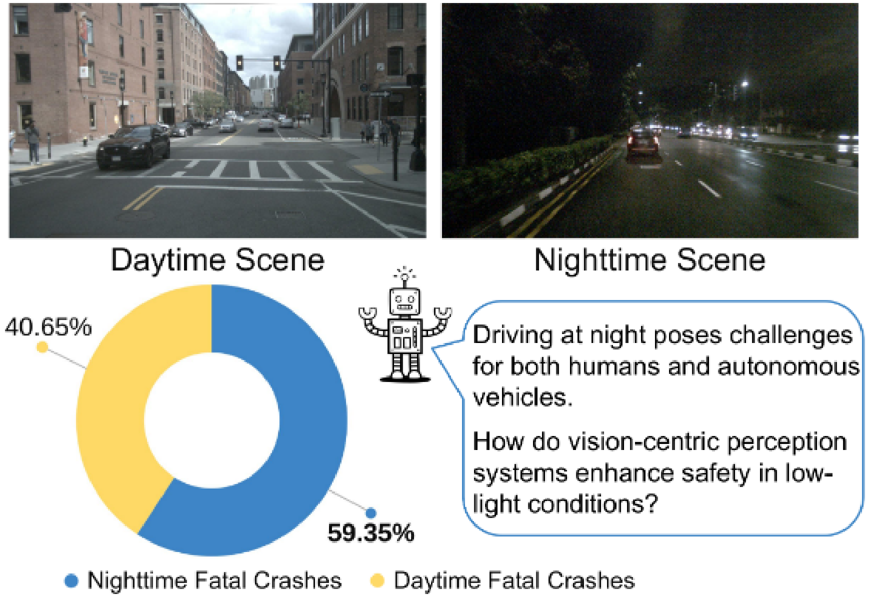


Figure 1. Nighttime driving scenarios pose a greater fatal threat than daytime. The fatal rate at night is much higher [4]. This paper aims to enhance nighttime images to improve the overall driving safety at night.

Challenges of nighttime scenes for autonomous driving

[1]. Benchmarking and Improving Bird's Eye View Perception Robustness in Autonomous Driving, TPAMI, 2025

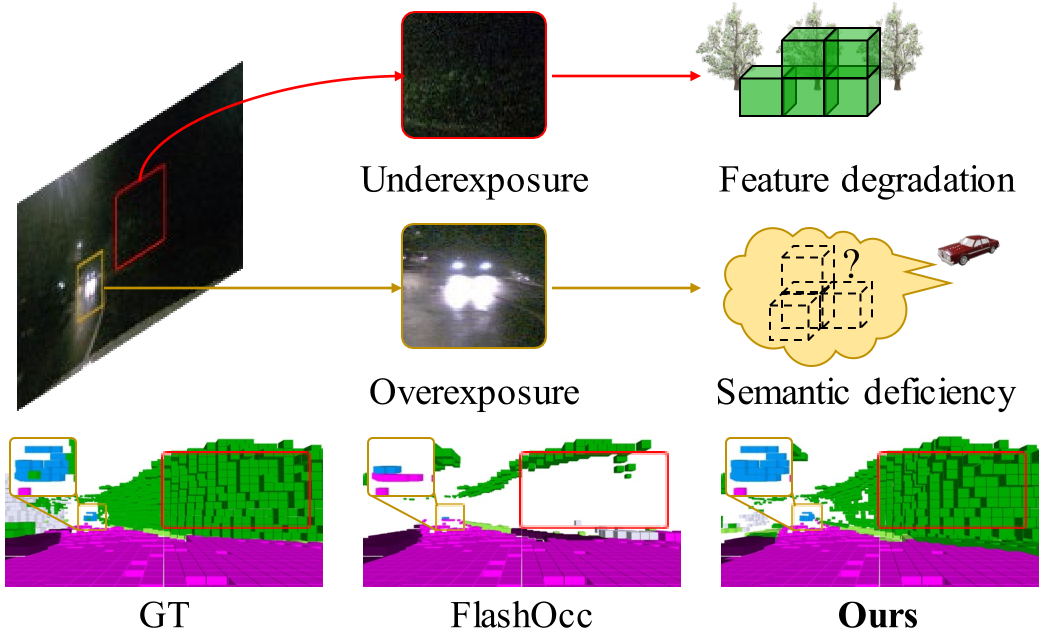
[2]. Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving, CVPR, 2025

Motivation

- Challenges

Low visibility

Complex lighting conditions

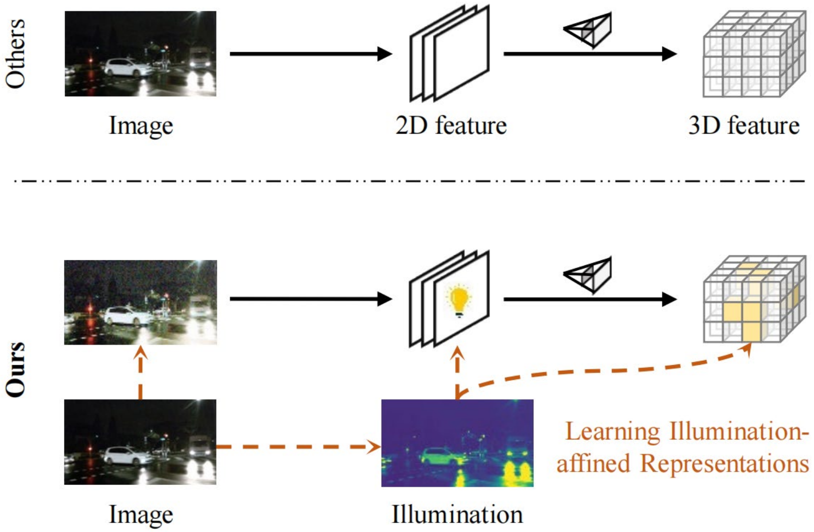


Motivation

- Learning illumination affined representation

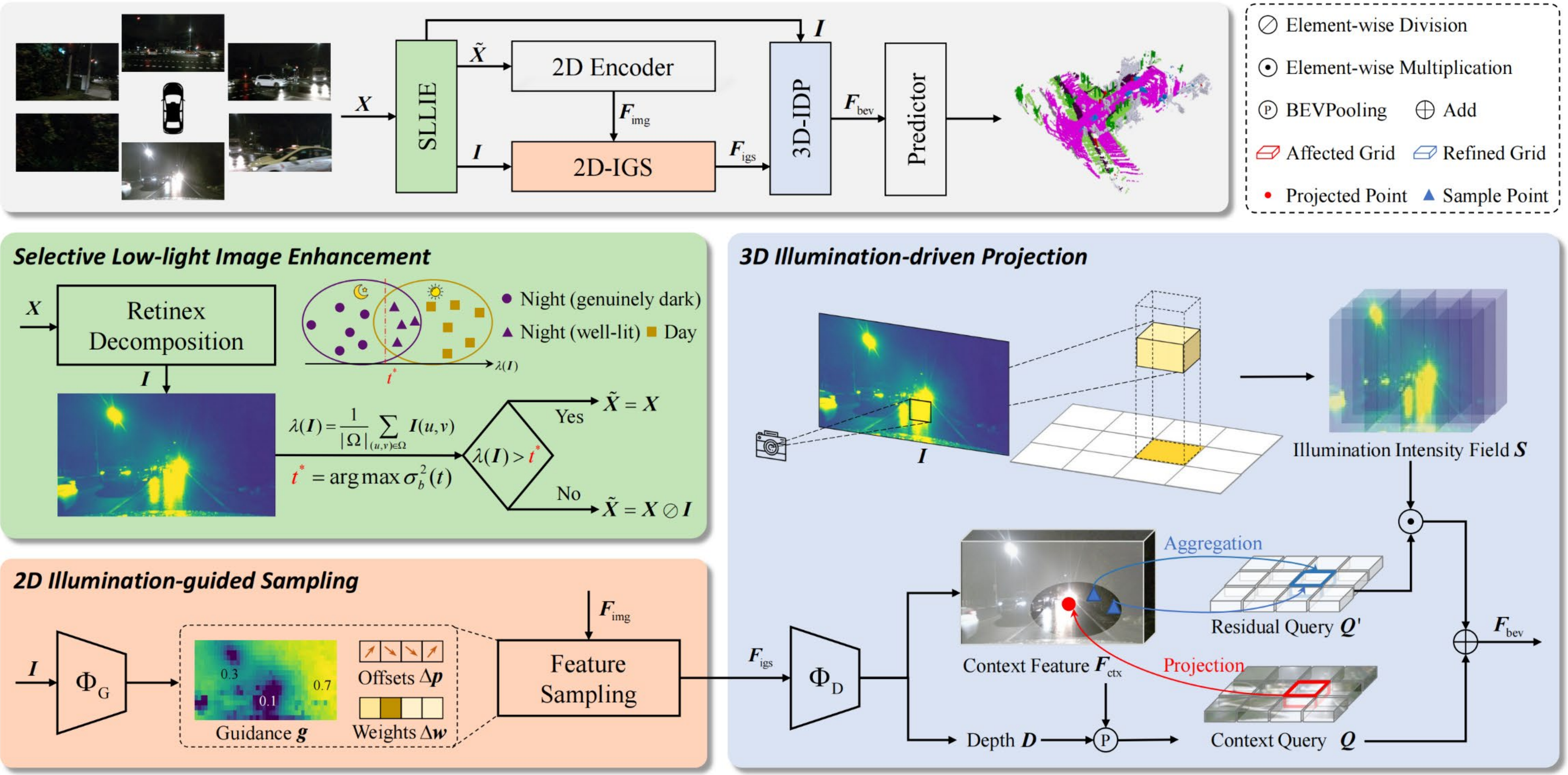
Illumination map as guidance

Pipeline comparison



Our approach learns illumination-aware representations to enhance both the fundamental 2D and 3D stages.

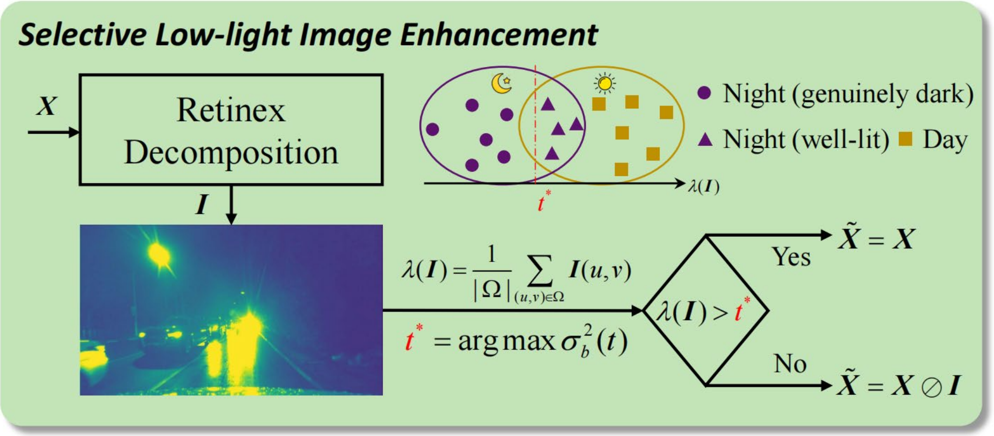
Method

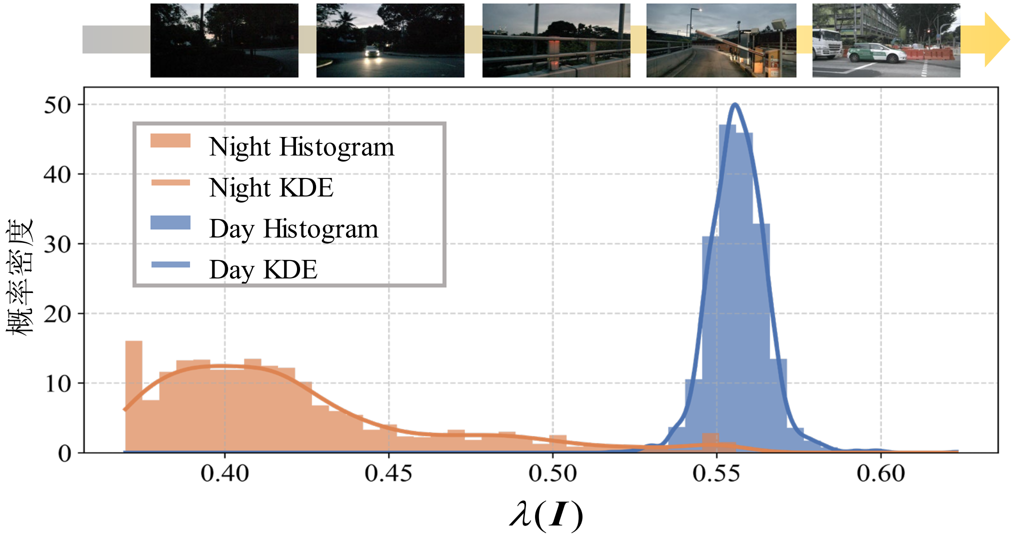


Element-wise Division
Element-wise Multiplication
BEVPooling
Add
Affected Grid
Refined Grid
Projected Point
Sample Point
SLLIE
2D Encoder
2D-IGS
3D-IDP
Predictor
X
$\tilde{X}$
I
F_{img}
F_{igs}
F_{bev}
Selective Low-light Image Enhancement
Retinex Decomposition
Night (genuinely dark)
Night (well-lit)
Day
$\lambda(I)$
t^*
$\lambda(I) = \frac{1}{|\Omega|} \sum_{(u,v)\in\Omega} I(u,v)$
$t^* = \arg\max \sigma_b^2(t)$
$\lambda(I) > t^*$
Yes
$\tilde{X} = X$
No
$\tilde{X} = X \oslash I$
2D Illumination-guided Sampling
Φ_G
0.3
0.7
0.1
Guidance g
Offsets Δp
Weights Δw
Feature Sampling
3D Illumination-driven Projection
Illumination Intensity Field S
Aggregation
Residual Query Q'
Context Feature F_{ctx}
Projection
Φ_D
Depth D
Context Query Q

Method

Selective Low-light Image Enhancement





1. Retinex Decomposition

$$\boldsymbol{X}_{\text{enh}} = \boldsymbol{X} \oslash \boldsymbol{I}, \quad \boldsymbol{I} \in (0,1],$$

2. Selective Mechanism

$$\lambda(\boldsymbol{I}) = \frac{1}{|\Omega|} \sum_{(u,v)\in\Omega} \boldsymbol{I}(u,v),$$

quantify brightness

$$\mu_T = \omega_0(t)\mu_0(t) + \omega_1(t)\mu_1(t),$$

optimal threshold

$$\sigma_b^2(t) = \omega_0(t)\left(\mu_0(t) - \mu_T\right)^2 + \omega_1(t)\left(\mu_1(t) - \mu_T\right)^2.$$

$$t^* = \arg\max_t \sigma_b^2(t).$$

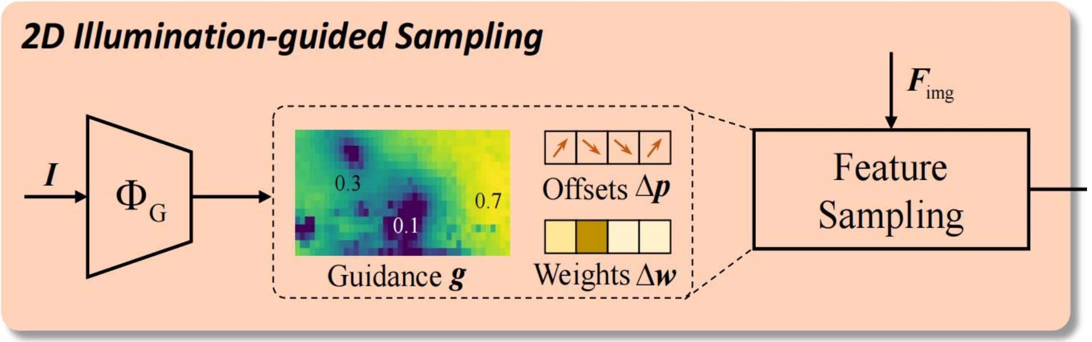
$$\tilde{\boldsymbol{X}} = \begin{cases} \boldsymbol{X}_{\text{enh}}, & \text{if } \lambda(\boldsymbol{I}) \le t^*, \\ \boldsymbol{X}, & \text{otherwise.} \end{cases}$$

selective enhancement

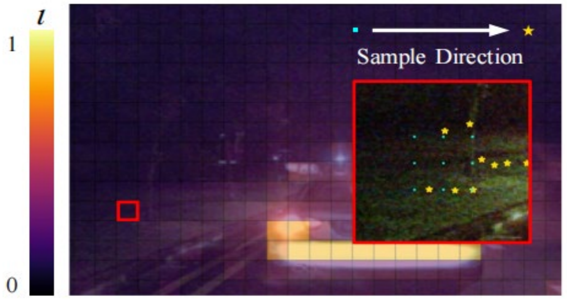
Method

2D Illumination-guided Sampling

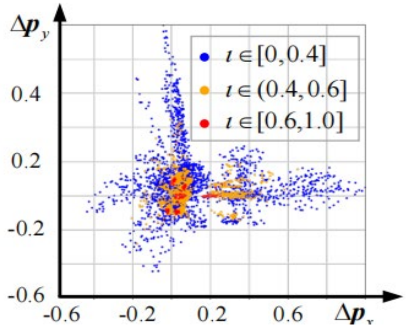
Enhance 2D image feature representations by learning adaptive sampling points from the illumination maps.



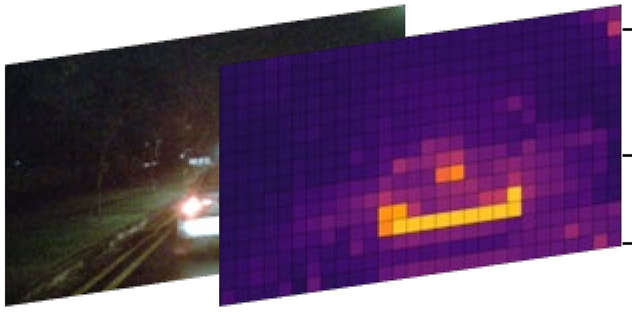
Aggregate semantic cues from adequately-exposed regions.



Visualization of sampling points



Offsets distribution



$$g = \frac{I'^{-1} - \min\left(I'^{-1}\right)}{\max\left(I'^{-1}\right) - \min\left(I'^{-1}\right)}$$

$$\Delta \boldsymbol{p} = \left(\Delta \boldsymbol{p}_x, \Delta \boldsymbol{p}_y\right)$$

$$\Delta \boldsymbol{w}$$

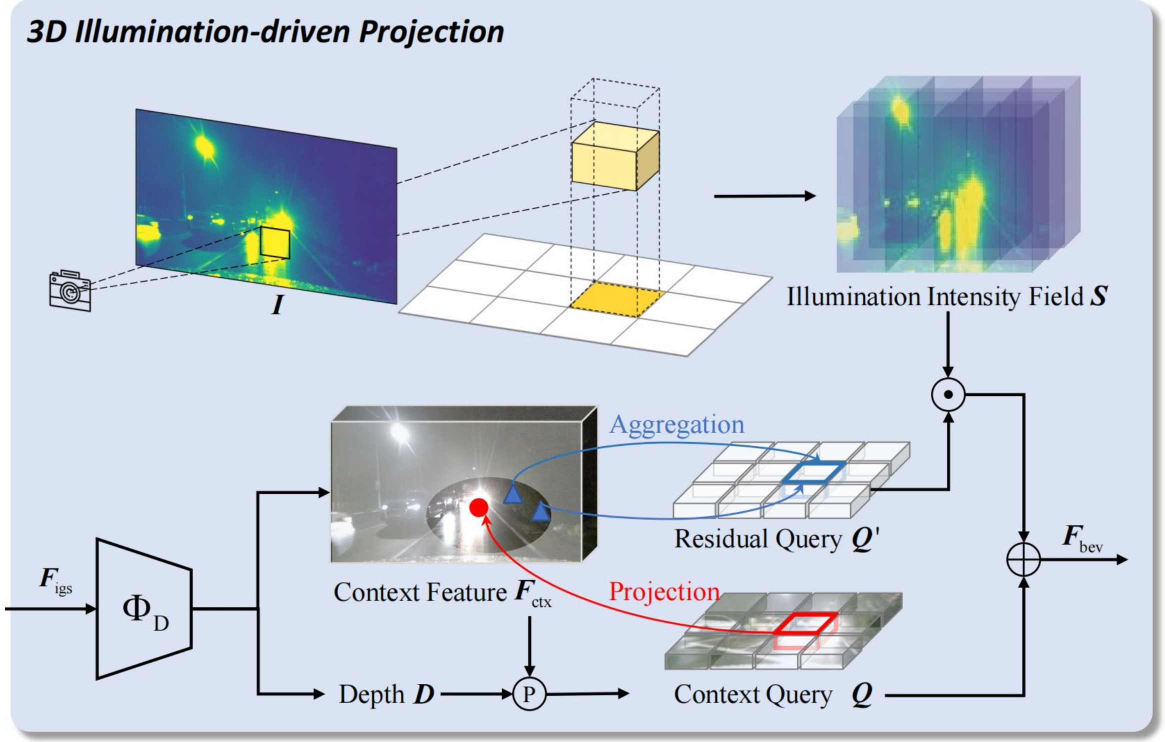
$$\Delta \tilde{p} = \Delta p \odot g$$

$$\boldsymbol{F}_{\text{igs}} = \mathcal{F}_{\text{dcn}}\left(\boldsymbol{F}_{\text{img}}, p + \Delta \tilde{\boldsymbol{p}}\right) \odot \Delta \boldsymbol{w} + \boldsymbol{F}_{\text{img}}$$

Method

3D Illumination-driven Projection

Mitigate the adverse effects of semantic deficiency caused by overexposure.



3D-IDP constructs 3D illumination intensity fields to refine the projection process.

1. Illumination Modeling

$$d_j \begin{bmatrix} u_j & v_j & 1 \end{bmatrix}^{\top} = M \begin{bmatrix} x & y & z_j & 1 \end{bmatrix}^{\top}, \forall j \in \{1, \ldots, N_z\}$$

$$\boldsymbol{S}(x,y) = \frac{1}{N_z} \sum_{j=1}^{N_z} \boldsymbol{I}\left(\lfloor v_j \rfloor, \lfloor u_j \rfloor \right)$$

2. BEV Context Refinement

$$\boldsymbol{Q}'(x,y) = \sum_{j=1}^{N_z} \mathcal{F}_{\text{dca}}\left(\boldsymbol{Q}(x,y), \mathcal{P}\left(x,y,z_j\right), \boldsymbol{F}_{\text{ctx}} \right)$$

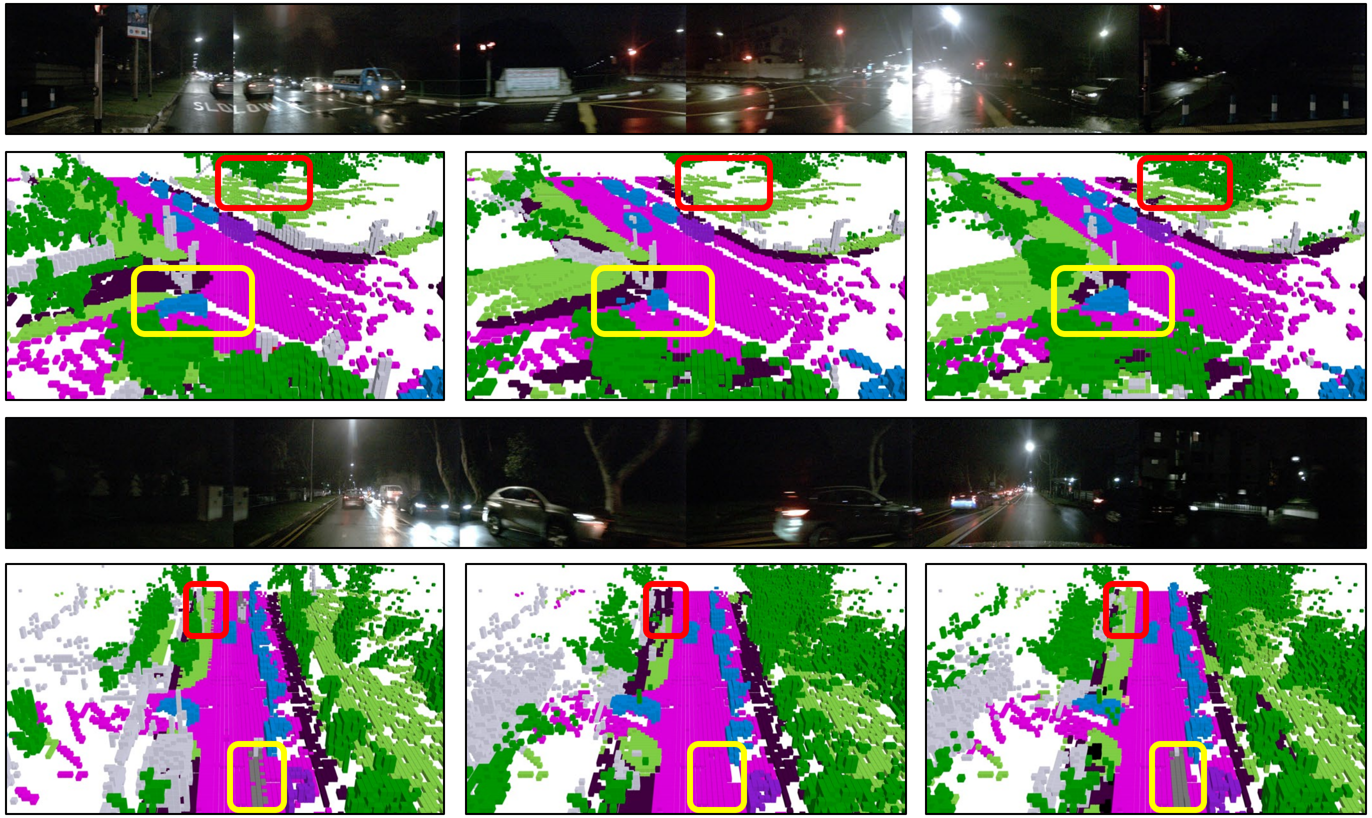
$$\boldsymbol{F}_{\text{bev}} = \boldsymbol{Q} + \boldsymbol{Q}' \odot \boldsymbol{S}$$

Results

Real: Occ3D-nuScenes

Table 1: Quantitative comparison on the Occ3D-nuScenes dataset. "1f" denotes single-frame method. "2f","4f" and "8f" denotes methods fusing temporal information from 2, 4 and 8 frames.

Method	mIoU	others	barrier	bicycle	car ↑	motorcycle	pedestrian	traffic cone	truck	drive. surf.	other flat	sidewalk	terrain	manmade	vegetation	Venue
Train: night *Test: night*																
BEVDetOcc (1f) [11]	13.22	0.0	0.0	0.0	34.5	0.0	0.0	0.0	0.1	**58.7**	1.0	24.1	26.0	19.7	21.1	arXiv22
FlashOcc (1f) [50]	13.42	0.0	0.0	0.0	35.3	0.0	0.6	0.0	3.6	57.2	2.3	21.8	25.9	19.6	**21.7**	arXiv23
SparseOcc (1f) [50]	13.32	**10.4**	0.5	6.4	37.3	10.9	8.2	**1.4**	**13.1**	46.4	4.4	19.7	9.6	8.8	9.3	ECCV24
OPUS (1f) [40]	11.32	0.2	0.0	0.0	33.2	4.6	0.1	0.0	13.0	50.9	1.0	15.0	13.5	11.4	15.8	NIPS24
LIAR (1f)	**19.27**	9.7	**11.2**	**8.5**	**37.7**	**14.8**	**12.2**	1.1	12.5	**58.7**	**8.5**	**27.2**	**27.2**	**20.1**	20.6	-
BEVDetOcc (2f) [11]	15.86	0.6	0.0	0.0	40.3	0.0	6.0	0.0	4.5	62.0	3.0	28.3	28.8	23.7	24.8	arXiv21
BEVFormer (2f) [25]	16.57	3.6	0.0	0.0	40.3	16.1	9.9	0.0	10.1	62.1	4.8	19.9	24.4	18.8	22.2	ECCV22
FlashOcc (2f) [50]	18.15	5.4	0.0	0.0	41.5	9.5	10.5	0.0	17.7	63.5	1.7	26.8	28.0	25.0	24.6	arXiv23
FBOcc (2f) [24]	19.79	9.3	**16.3**	5.4	40.0	13.6	12.5	0.1	17.6	59.4	6.8	24.5	29.7	20.0	21.9	CVPR23
OPUS (2f) [40]	12.77	1.5	0.0	0.0	35.4	8.5	1.1	0.0	17.0	52.9	2.2	16.6	14.6	13.0	16.1	NIPS24
SparseOcc (2f) [27]	14.29	12.0	1.9	7.9	37.9	9.8	9.3	0.4	16.6	46.9	4.8	21.8	11.4	9.7	9.9	ECCV24
COTR (2f) [31]	20.01	**15.3**	0.3	1.5	**44.0**	18.1	10.1	0.0	7.7	63.2	**7.0**	**30.3**	**31.1**	25.1	**26.6**	CVPR24
LIAR (2f)	**22.09**	13.0	5.3	**13.5**	42.8	**19.3**	**18.6**	**1.1**	**20.2**	**64.3**	2.8	29.0	29.3	**25.4**	24.7	-
Train: day & night *Test: night*																
BEVDetOcc (1f) [11]	18.96	4.7	22.5	2.6	38.5	6.6	6.5	0.0	12.5	**63.6**	5.7	**29.0**	28.8	21.3	**23.1**	arXiv22
FlashOcc (1f) [50]	18.93	4.3	1.0	0.0	38.7	7.7	5.8	0.0	13.6	60.6	2.1	27.1	30.0	**21.7**	22.5	arXiv23
LIAR (1f)	**23.67**	**12.0**	**36.1**	**14.5**	**40.7**	**19.2**	**14.0**	**1.3**	**19.3**	60.5	**11.1**	28.9	**30.5**	21.4	21.8	-
BEVDetOcc (2f) [11]	21.98	9.1	16.5	2.4	44.4	8.2	11.4	0.0	28.9	64.6	8.6	29.9	31.4	26.1	**26.3**	arXiv21
BEVFormer (4f) [25]	13.77	3.1	19.8	0.7	44.1	16.6	14.5	0.0	22.3	35.8	5.6	13.9	8.4	3.1	5.0	ECCV22
FlashOcc (2f) [50]	23.40	13.4	18.5	2.3	**46.5**	10.7	14.3	0.0	30.0	66.5	5.9	**32.5**	32.9	**28.1**	26.1	arXiv23
OPUS (8f) [40]	20.28	14.0	10.4	12.1	40.4	15.2	13.0	0.0	27.9	62.9	5.0	25.5	19.8	17.1	20.6	NIPS24
SparseOcc (8f) [27]	22.79	15.8	40.0	**23.0**	43.5	15.3	13.6	0.4	28.9	58.8	10.2	26.2	16.7	13.2	13.5	ECCV24
COTR (2f)[31]	25.17	**16.0**	**41.5**	8.9	42.1	17.4	12.5	0.0	26.6	65.2	**11.2**	27.0	**33.7**	24.8	25.6	CVPR24
LIAR (2f)	**27.33**	15.9	37.7	19.0	45.4	**19.1**	**17.8**	**1.9**	**33.6**	**67.1**	8.1	31.2	**33.7**	27.4	24.7	-



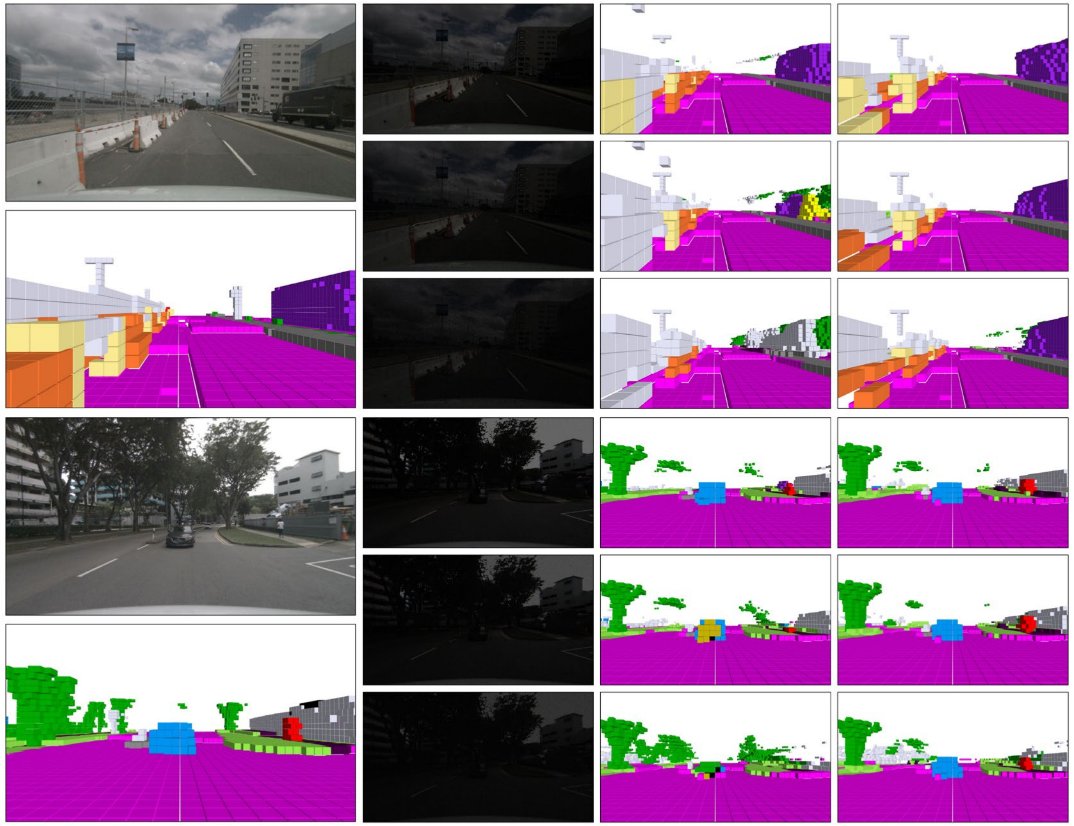
GT FlashOcc **Ours**

Results

Synthetic: nuScenes-C

Table 2: Quantitative comparison on the nuScenes-C dataset under three severity levels.

Method	mIoU	others	barrier	bicycle	bus	car	const. veh.	motorcycle	pedestrian	traffic cone	trailer	truck	drive. surf.	other flat	sidewalk	terrain	manmade	vegetation
Severity: easy																		
BEVDetOcc (1f) [11]	15.08	0.5	12.4	1.3	14.1	24.9	7.7	3.8	7.8	4.5	2.1	10.1	62.5	12.0	25.5	32.0	17.1	18.2
FlashOcc (1f) [50]	12.47	0.2	7.5	1.9	16.1	19.1	6.6	3.0	5.6	3.4	4.6	9.1	49.7	9.9	19.7	25.8	14.9	14.7
LIAR (1f)	**22.52**	**3.7**	**19.4**	**12.1**	**24.4**	**31.8**	**11.0**	**14.8**	**16.3**	**14.4**	**11.6**	**17.9**	**67.9**	**21.5**	**34.1**	**38.2**	**21.8**	**22.2**
BEVDetOcc (2f) [11]	20.22	0.9	18.2	4.3	20.4	35.6	13.2	8.3	12.4	9.3	3.4	18.3	67.2	14.4	29.2	35.6	28.0	24.9
BEVFormer (4f) [25]	14.13	0.7	20.4	3.8	28.9	33.6	4.4	8.3	11.3	7.8	10.2	18.0	34.4	14.0	17.4	11.2	7.2	8.6
FlashOcc (2f) [50]	21.83	1.7	19.5	7.7	24.0	35.0	14.6	11.6	14.3	14.3	4.4	19.0	63.3	21.3	33.2	33.3	28.9	25.1
OPUS (8f) [40]	23.63	4.0	22.2	14.0	31.7	37.5	16.3	15.5	13.3	13.8	12.0	24.0	65.6	23.1	34.4	29.4	21.2	23.7
SparseOcc (8f) [27]	21.98	3.9	23.0	16.6	28.7	37.5	13.2	18.6	18.9	**24.7**	7.4	21.8	58.0	18.9	26.9	23.3	16.5	15.8
COTR (2f) [31]	21.36	1.3	23.6	10.2	16.0	31.4	8.2	11.1	17.1	18.5	3.3	15.0	67.6	19.0	33.0	35.9	29.4	22.5
LIAR (2f)	**30.66**	**5.1**	**31.6**	**17.2**	**32.2**	**43.8**	**16.7**	**19.8**	**21.5**	23.9	**20.0**	**27.6**	**75.5**	**33.1**	**42.5**	**45.0**	**35.7**	**30.1**
Severity: moderate																		
BEVDetOcc (1f) [11]	11.50	0.3	6.6	0.8	10.0	20.7	6.1	2.8	6.0	2.2	1.0	5.5	56.5	5.7	18.2	24.8	13.6	14.8
FlashOcc (1f) [50]	9.05	0.0	3.1	0.7	13.4	15.4	5.9	1.7	4.3	1.7	1.5	5.5	42.0	3.3	13.2	18.4	11.9	11.9
LIAR (1f)	**18.44**	**2.2**	**11.1**	**11.1**	**23.0**	**28.0**	**8.5**	**13.8**	**14.0**	**12.5**	**5.0**	**13.7**	**62.8**	**12.0**	**27.5**	**31.4**	**17.8**	**19.0**
BEVDetOcc (2f) [11]	15.95	0.4	10.3	3.2	14.9	31.8	10.1	5.7	9.7	4.9	1.7	12.6	62.4	8.5	22.2	28.8	23.6	20.5
BEVFormer (4f) [25]	9.98	0.4	11.6	2.9	25.2	29.5	1.5	5.4	8.6	2.7	3.8	11.9	27.3	8.1	11.9	7.5	4.8	6.8
FlashOcc (2f) [50]	17.34	0.9	12.6	5.4	17.6	31.4	10.7	8.6	11.8	9.2	2.2	12.7	60.6	14.0	26.0	28.1	23.2	19.9
OPUS (8f) [40]	17.50	2.0	11.6	9.6	25.9	32.9	**13.4**	12.1	10.7	8.2	3.3	17.3	59.4	13.6	24.8	18.4	15.4	18.9
SparseOcc (8f) [27]	16.81	1.9	13.6	11.7	22.8	33.5	9.2	15.5	16.0	19.2	2.0	16.2	53.2	8.3	19.3	16.3	13.9	13.1
COTR (2f) [31]	17.77	0.8	18.2	8.6	11.4	28.4	6.8	9.7	14.1	15.0	1.3	11.3	62.9	13.2	27.1	29.9	25.1	18.3
LIAR (2f)	**25.67**	**2.7**	**20.0**	**13.2**	**28.7**	**39.8**	12.8	**17.6**	**18.5**	**20.2**	**11.9**	**22.8**	**72.0**	**25.6**	**37.0**	**39.6**	**29.6**	**24.7**
Severity: hard																		
BEVDetOcc (1f) [11]	7.81	0.0	2.1	0.3	5.2	14.6	3.1	1.4	3.2	0.8	0.3	2.1	49.4	1.4	11.2	16.0	10.2	11.6
FlashOcc (1f) [50]	5.82	0.0	0.7	0.1	7.9	11.0	1.3	0.5	2.4	0.4	0.5	2.2	34.3	0.2	7.3	11.0	9.2	9.9
LIAR (1f)	**12.21**	**0.8**	**4.7**	**7.4**	**15.4**	**20.9**	**3.7**	**9.1**	**8.8**	**7.5**	**2.2**	**6.8**	**53.3**	**2.2**	**16.1**	**21.9**	**12.3**	**14.5**
BEVDetOcc (2f) [11]	10.46	0.1	4.2	1.6	7.1	23.9	3.8	2.5	5.5	1.4	0.6	5.1	53.1	2.8	14.1	19.5	17.5	14.9
BEVFormer (4f) [25]	5.71	0.2	5.1	0.9	13.1	22.3	0.2	2.4	4.8	0.6	0.5	5.7	19.4	2.8	6.3	5.1	3.0	4.7
FlashOcc (2f) [50]	11.76	0.2	4.8	2.0	9.7	25.0	5.0	5.2	7.2	3.9	0.8	5.3	56.3	4.9	16.5	22.3	16.4	14.3
OPUS (8f) [40]	10.80	0.5	3.1	5.8	**14.7**	25.0	**10.6**	5.3	6.4	2.9	0.5	8.6	49.7	3.9	13.6	10.3	9.0	13.7
SparseOcc (8f) [27]	10.77	**0.8**	5.1	7.9	10.8	25.9	4.2	8.2	11.3	11.0	0.3	8.3	45.2	1.0	11.6	10.6	10.9	10.0
COTR (2f) [31]	12.01	0.4	**9.1**	6.0	3.9	23.4	1.7	6.4	8.3	9.2	0.4	5.6	53.9	5.3	18.6	21.0	18.2	13.0
LIAR (2f)	**17.31**	**0.8**	**9.1**	**8.3**	14.2	**31.5**	9.0	**10.2**	**12.3**	**13.2**	**2.3**	**13.1**	**63.0**	**14.4**	**26.0**	**30.1**	**19.5**	**17.6**



RGB&GT | Synthetic | FlashOcc | **Ours**

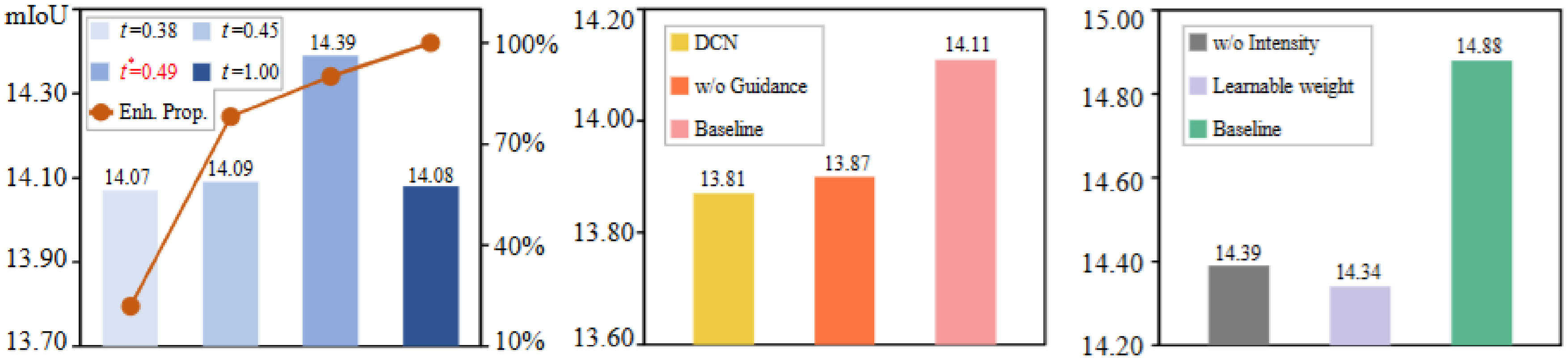
Ablations

Table 3: Ablation study of LIAR on the Occ3D-nuScenes.

LIAR	SLLIE	2D Feature Enh.			View Transformation		mIoU
		Add	Concat	2D-IGS	BEVPooling	3D-IDP	
baseline					✓		13.42
i	✓				✓		14.39
ii		✓			✓		13.86
iii			✓		✓		13.81
iv				✓	✓		14.11
v						✓	14.88
vi	✓			✓		✓	15.31

Input image w/o sel. mech.

Adverse effect of indiscriminate enhancement.



(a) Threshold in SLLIE (b) Sampling strategy in 2D IGS (c) Illumination intensity in 3D IDP

Ablation study on the impact of illumination in our SLLIE, 2D-IGS, and 3D-IDP.

Thanks!