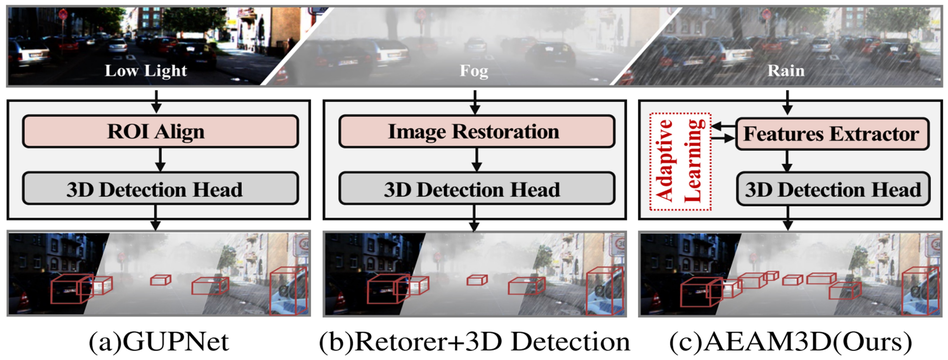
AEAM3D: Adverse Environment-adaptive Monocular 3d Object Detection Via Feature Extraction Regularization

Yixin Lei, Xingyuan Li, Zhiying Jiang, Xinrui Ju, Jinyuan Liu
Dalian University of Technology, China

DALIAN UNIVERSITY OF TECHNOLOGY

Contributions

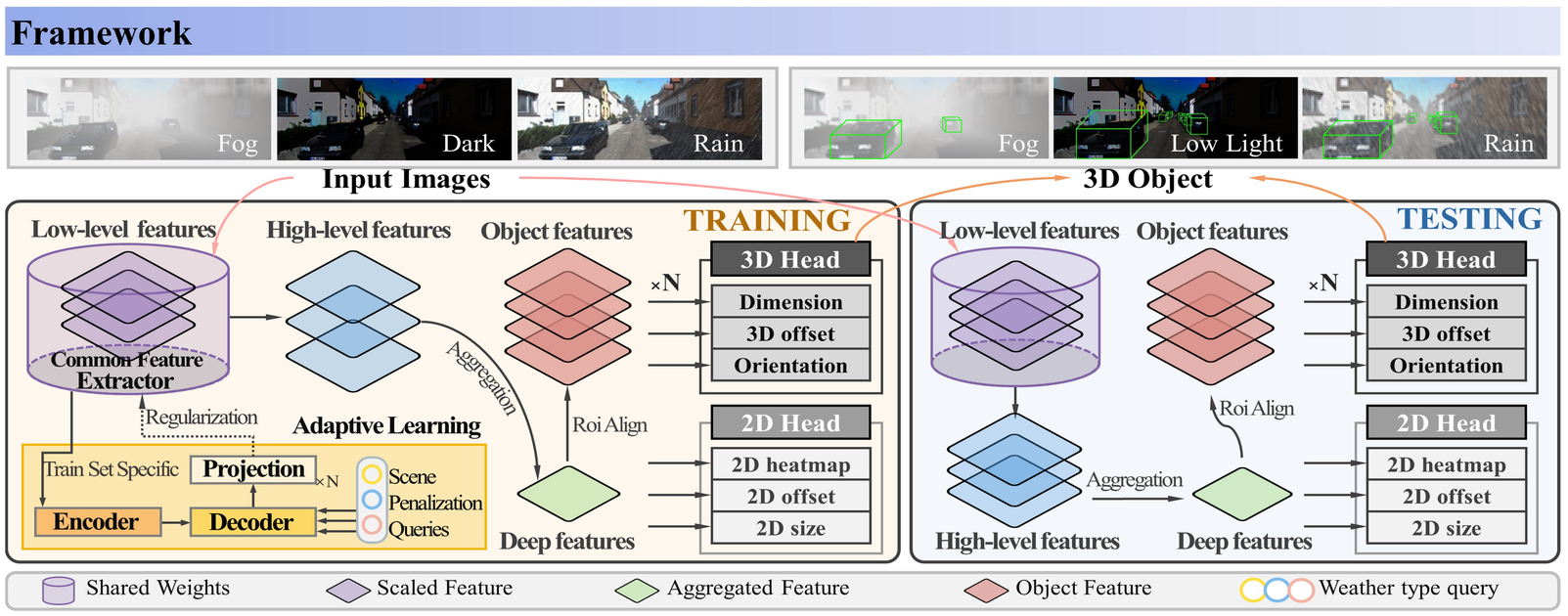


(a)GUPNet (b)Retorer+3D Detection (c)AEAM3D(Ours)

- We introduce a robust network specifically designed to handle adverse environments, significantly improving the performance of monocular 3D object detection models across various challenging real-world situations.
- We propose an adaptive learning strategy during the training process to extract resilient features that remain less susceptible to degrading factors, aiding the model in discerning various inclement environments.
- To support 3D object detection in harsh environments, we have compiled a comprehensive dataset comprising 7,481 images for seven demanding conditions.

Method

Framework



Adaptive Learning Strategy

We propose a novel adaptive learning strategy comprising an encoder and a decoder, which are specifically designed to act as a constraint, rather than focus on image restoration. Particularly the encoder assists the model in rectifying inaccurate feature perception under adverse conditions. The decoder employs learnable scene penalization queries to penalize incorrect perception by which the model can suppress potential errors. Notably, this learning strategy is only required during training.

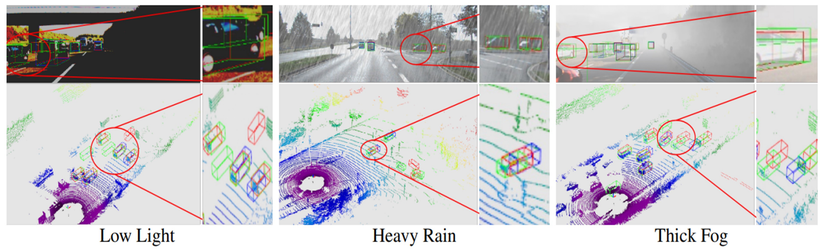
3D Object Detection in Adverse Scenes

Monocular 3D object detection takes an RGB image as input and constructs a 3D bounding box for the object in 3D space. Concretely, 2D detection backbone from low-level constraint features is applied to produce high-level deep features, and then these features are aggregated to get deep. Subsequently, we apply three 2D detection heads in deep features F to predict 2D heatmap H. Through using ROIAlign in deep feature map with 2D box information, the features are generated whose size is 7×7 and finally used in the 3D detection heads to predict the object 3D center offset O3d, 3D size S3d and direction Θ.

Experiments

Quantitative and Qualitative Comparisons: comparison of the latest 3D object detection methods on the moderate fog, thick fog, moderate rain, heavy rain, dense rain and low light dataset based on AP3D of car category. Our method achieves significant performance improvements across different weather conditions.

Methods	Venue	Mod. Fog			Thick Fog			Mod. Rain			Heavy Rain			Dense Rain			Low Light		
		Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard
SMOKE	*CVPR20*	8.86	5.98	4.53	5.10	3.31	2.28	7.33	5.24	4.03	5.97	3.78	2.77	5.64	3.88	3.21	5.48	4.03	3.49
MonoFLEX	*CVPR21*	19.97	14.11	11.86	18.37	13.28	10.57	17.21	12.94	11.55	16.99	11.83	10.12	15.35	12.14	10.38	10.43	8.32	7.75
MonoDLE	*CVPR21*	14.77	12.15	10.02	17.35	12.89	11.27	15.65	13.34	12.33	15.64	12.63	11.13	14.94	11.20	9.78	14.69	11.99	10.60
GUPNet	*ICCV21*	21.06	15.02	12.34	19.91	14.24	11.57	19.69	14.24	12.36	17.36	12.95	10.76	16.71	12.40	10.64	9.84	6.36	5.09
DID-M3D	*ECCV22*	22.75	15.52	12.61	22.19	15.96	12.86	22.42	15.30	12.43	21.40	14.79	12.05	20.56	14.07	11.88	21.92	14.79	12.10
DEVIANT	*ECCV22*	22.74	15.92	13.16	22.90	16.11	13.25	22.35	15.99	12.45	20.18	13.93	11.96	20.20	13.85	12.26	22.40	15.16	12.33
HomoLoss	*CVPR22*	14.31	12.27	11.12	19.32	13.26	11.51	18.23	13.19	12.56	17.69	13.01	12.23	16.33	13.40	10.76	15.88	13.89	11.42
CubeR-CNN	*CVPR23*	21.11	14.97	12.55	20.81	14.77	12.12	20.37	14.14	12.38	22.36	13.67	11.11	19.17	13.54	10.99	20.11	14.37	11.89
AEAM3D	-	23.13	16.03	13.19	23.24	16.28	13.35	23.08	16.01	12.98	23.06	15.77	12.92	21.31	15.40	12.52	22.55	15.70	12.80
Improvement	-	**+0.38**	**+0.11**	**+0.03**	**+0.34**	**+0.17**	**+0.10**	**+0.66**	**+0.02**	**+0.53**	**+1.66**	**+0.98**	**+0.87**	**+0.75**	**+1.33**	**+0.26**	**+0.15**	**+0.54**	**+0.49**



Comparison of our method with the combinations of our base 3D detection network and popular enhancement models under various challenging conditions.

Scene	Methods	Venue	Car 3D@IOU=0.7 Easy	Mod.	Hard
Thick Fog	Trans	*CVPR22*	22.95	16.03	13.21
	MSBDN	*CVPR20*	20.11	14.14	11.55
	GCA	*WACV19*	21.21	14.08	12.49
	DCPDN	*CVPR18*	19.97	13.25	11.34
	Ours	-	**23.13**	**16.03**	**12.95**
Heavy Rain	Trans	*CVPR22*	20.29	13.89	11.67
	RESCAN	*ECCV18*	20.06	13.81	10.99
	VRGNet	*CVPR21*	21.55	12.98	11.01
	PRENet	*CVPR19*	20.11	13.34	10.67
	Ours	-	**23.06**	**15.77**	**12.92**
Low Light	Trans	*CVPR22*	14.7	10.53	9.18
	SCI	*CVPR22*	19.88	14.12	10.68
	IAT	*BMVC22*	19.84	13.59	10.94
	SID	*CVPR18*	17.78	12.21	10.32
	Ours	-	**22.55**	**15.70**	**12.82**

Ablation study for the components of our method.

	Enc	Dec	3D@IoU=0.7 Easy↑	Mod.↑	Hard↑
(a)	✗	✗	18.53	13.09	10.89
(b)	✗	✓	19.12$_{\uparrow 0.59}$	14.43$_{\uparrow 1.34}$	11.74$_{\uparrow 0.85}$
(c)	✓	✗	20.35$_{\uparrow 1.82}$	14.86$_{\uparrow 1.77}$	12.11$_{\uparrow 1.22}$
(d)	✓	✓	23.22$_{\uparrow 4.69}$	15.55$_{\uparrow 2.46}$	12.31$_{\uparrow 1.42}$