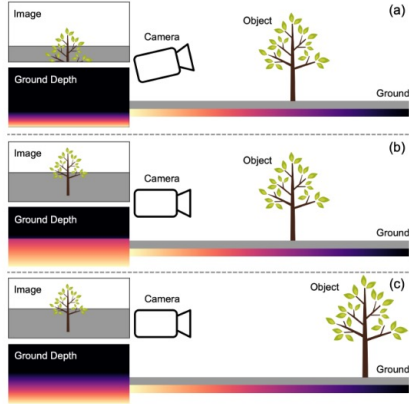


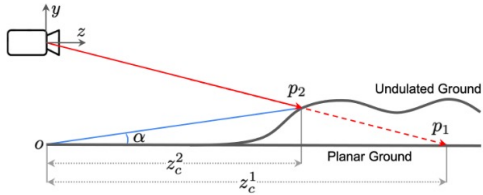
QCRAFT

Overview of Depth Estimation



Estimation depth from 2D image is a well-known ill-posed problem. Objects with different scales may generate the same image.

Ground As a Reference



Ground depth encodes the camera parameters naturally and it can be used as the reference by embedding.

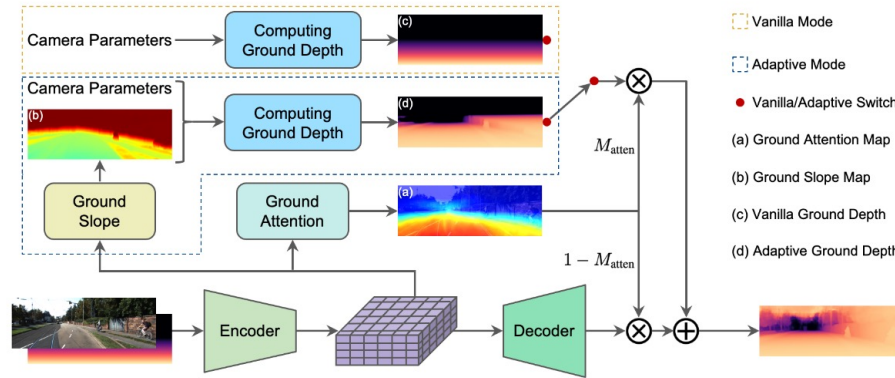


Code Available

GEDepth:Ground Embedding for Monocular Depth Estimation

Xiaodong Yang* Zhuang Ma* Zhiyu Ji* Zhe Ren

Network Architecture



Experimental Results

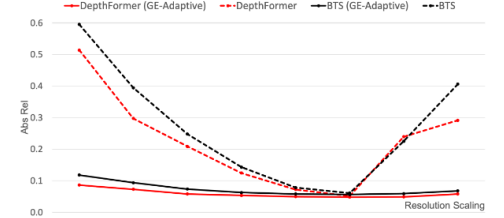
Method	Abs Rel ↓	Sq Rel ↓	RMSE ↓	Method	Abs Rel ↓	Sq Rel ↓	RMSE ↓
PackNet-SAN [10]	0.187	2.776	11.936	PackNet-SAN [10]	0.187	2.776	11.936
BTS [16]	0.162	2.492	11.466	BTS [16]	0.162	2.492	11.466
BTS (GE-Vanilla)	0.158	2.377	11.219	BTS (GE-Vanilla)	0.158	2.377	11.219
BTS (GE-Adaptive)	0.156	2.360	11.186	BTS (GE-Adaptive)	0.156	2.360	11.186
DepthFormer [20]	0.152	2.230	11.051	DepthFormer [20]	0.152	2.230	11.051
DepthFormer (GE-Vanilla)	0.149	2.121	10.790	DepthFormer (GE-Vanilla)	0.149	2.121	10.790
DepthFormer (GE-Adaptive)	0.145	2.119	10.596	DepthFormer (GE-Adaptive)	0.145	2.119	10.596
PixelFormer [1]	0.151	2.140	10.920	PixelFormer [1]	0.151	2.140	10.920
PixelFormer (GE-Vanilla)	0.148	2.123	10.848	PixelFormer (GE-Vanilla)	0.148	2.123	10.848
PixelFormer (GE-Adaptive)	0.145	2.122	10.803	PixelFormer (GE-Adaptive)	0.145	2.122	10.803
BinsFormer [21]	0.149	2.142	10.866	BinsFormer [21]	0.149	2.142	10.866
BinsFormer (GE-Vanilla)	0.146	2.109	10.561	BinsFormer (GE-Vanilla)	0.146	2.109	10.561
BinsFormer (GE-Adaptive)	0.145	2.101	10.459	BinsFormer (GE-Adaptive)	0.145	2.101	10.459

Comparison of GEDepth (in both vanilla and adaptive modes) and the state-of-the-art methods on KITTI (left) and DDAD (right). Groups 2-5 correspond to the four representative methods integrated with the proposed ground embedding modules.

ICCV23

PARIS

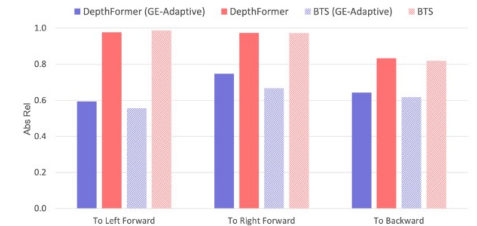
Generalization Enhancement



Comparison of the impact of image resolution change to DepthFormer and BTS on KITTI.

Dataset	Method	Abs Rel ↓	RMSE ↓
K → D	DepthFormer	0.644	17.083
	GE-Adaptive	0.261	16.132
D → K	DepthFormer	0.149	4.132
	GE-Adaptive	0.104	3.398

Comparison of the cross-dataset (KITTI and DDAD) performance based on DepthFormer.



Comparison of the view-transferring influence trained on the forward view and then tested on the left forward, right forward and backward view