Densely Constrained Depth Estimator for Monocular 3D Object Detection

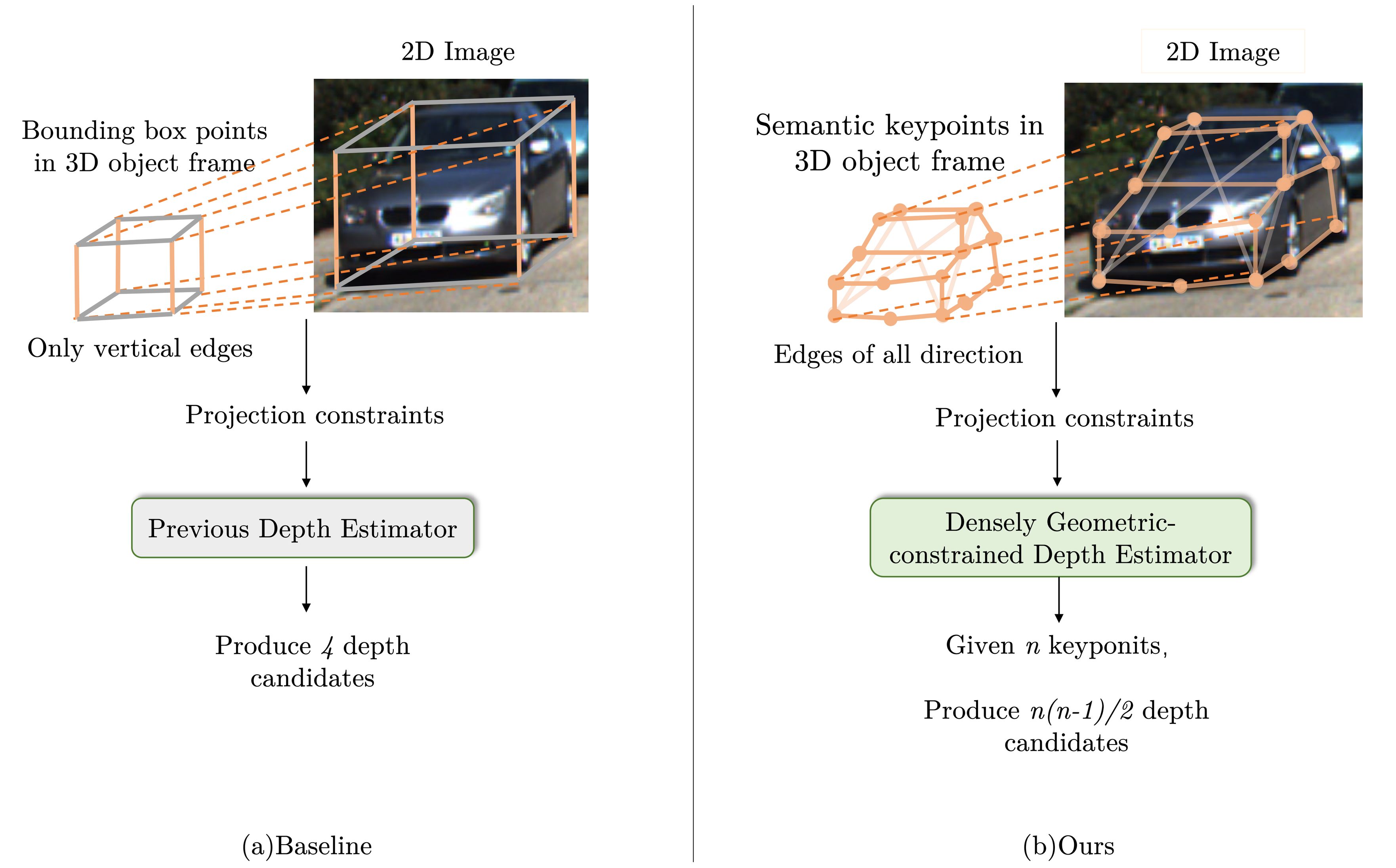
Yingyan Li, Yuntao Chen, Jiawei He, and Zhaoxiang Zhang
{liyingyan2021, hejiawei2019, zhaoxiang.zhang}@ia.ac.cn, chenyuntao08@gmail.com

TEL AVIV 2022

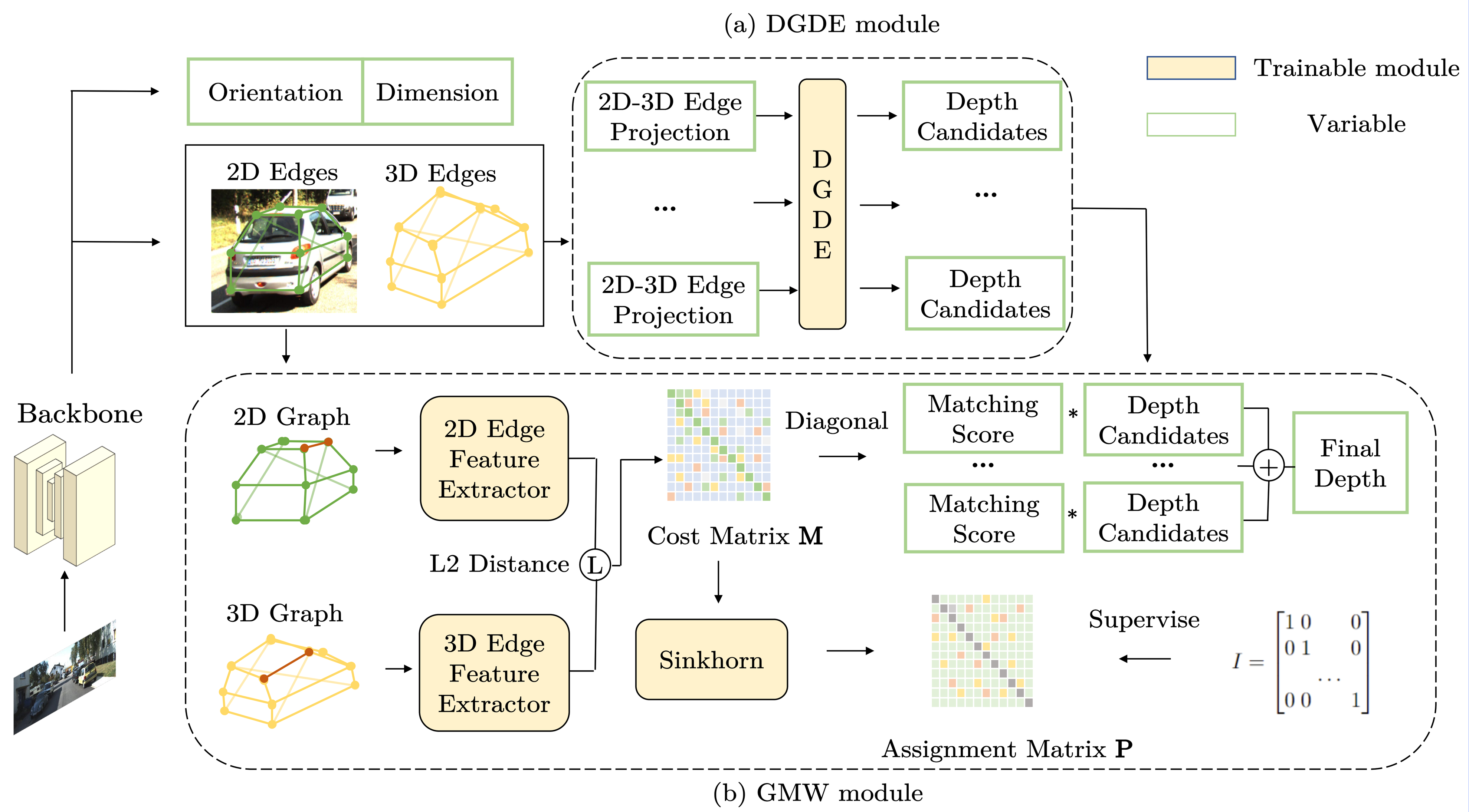
Introduction

- We propose a Dense Geometric-constrained Depth Estimator (DGDE). Different from the previous methods, DGDE estimates depth candidates utilizing projection constraints of edges of any direction. Therefore, considerable 2D-3D projection constraints are used, producing considerable depth candidates. We produce high-quality final depth based on these candidates.
- We propose an effective and interpretable Graph Matching Weighting module (GMW). We construct the 2D/3D graph from 2D/3D keypoints respectively. Then we regard the graph matching score of the 2D-3D edge as the weight of the corresponding depth candidate. This strategy utilizes all the keypoints' information and produces explicitly supervised weights.
- We localize each object more accurately by weighting the estimated depth candidates with corresponding matching scores. Our Densely Constrained Detector (DCD) achieves state-of-the-art performance on the KITTI and Waymo Open Dataset (WOD) benchmarks.

More Geometric Constraints



Framework



DGDE

DGDE can calculate the depth from 2D-3D edge of any direction.

$$z_c^{ij} = \begin{cases} \dfrac{l_i - l_j}{\tilde{u}_i - \tilde{u}_j}, & l_i = x_o^i \cos(r_y) + z_o^i \sin(r_y) + u^i(x_o^i \sin(r_y) - z_o^i \cos(r_y)) \\ \dfrac{h_i - h_j}{\tilde{v}_i - \tilde{v}_j}, & h_i = y_o^i + v^i(x_o^i \sin(r_y) - z_o^i \cos(r_y)) \end{cases}$$

WGM

GMW matches the 2D edges and 3D edges and produce the matching scores. The 2D-3D edge matching score is used as the weight of the corresponding depth candidate.

$$z_c = \sum_{i<j} w_{i,j} z_c^{ij}$$

Experimental Results

Ablation Study

	Weighting Method	DGDE	#Keypoints	#Depth Candidates	$AP_{3D\|R40\|IoU@0.7}$ Easy	Mod	Hard
(a)	Uncertainty (Baseline) [51]		10	5	21.63	15.87	13.38
(b)	Uncertainty	✓	10	45	21.72	16.09	13.35
(c)	Uncertainty	✓	73	1500	22.84	16.53	13.77
(d)	*GMW*		10	5	22.58	16.14	13.63
(e)	*GMW*	✓	10	45	23.30	16.91	14.93
(f)	*GMW*	✓	73	1500	**23.94**	**17.38**	**15.32**

Comparison with state-of-the-art methods on KITTI test server

Methods	Reference	Category	$AP_{3D\|R40\|IoU@0.7}$ Easy	Mod.	Hard	$AP_{BEV\|R40\|IoU@0.7}$ Easy	Mod.	Hard
PatchNet [26]	ECCV20	Pretrained Depth	15.68	11.12	10.17	22.97	16.86	14.97
D4LCN [8]	CVPR20		16.65	11.72	9.51	22.51	16.02	12.55
DDMP-3D [40]	CVPR21		19.71	12.78	9.80	28.08	17.89	13.44
CaDDN [31]	CVPR21	LiDAR Auxiliary	19.17	13.41	11.46	27.94	18.91	17.19
RTM3D [18]	ECCV20	Directly Regress	14.41	10.34	8.77	19.17	14.20	11.99
Movi3D [36]	ECCV20		15.19	10.90	9.26	22.76	17.03	14.85
Ground-Aware [21]	RAL21		21.65	13.25	9.91	29.81	17.98	13.08
MonoDLE [28]	CVPR21		17.23	12.26	10.29	24.79	18.89	16.00
MonoRCNN [35]	ICCV21		18.36	12.65	10.03	25.48	18.11	14.10
MonoEF [56]	CVPR21		21.29	13.87	11.71	29.03	19.70	17.26
MonoRUn [6]	CVPR21	Geometric-based	19.65	12.30	10.58	27.94	17.34	15.24
AutoShape [23]	ICCV21		22.47	14.17	11.36	30.66	20.08	15.59
GUPNet [25]	ICCV21		22.20	15.02	13.12	30.29	21.19	18.20
MonoFlex(Baseline) [51]	CVPR21		19.94	13.89	12.07	28.23	19.75	16.89
DCD(Ours)	ECCV22		**23.81**	**15.90**	**13.21**	**32.55**	**21.50**	**18.25**

Visualization