

MonoGRNet: A Geometric Reasoning Network for Monocular 3D Object Localization

Zengyi Qin^{1,2} Jinglu Wang¹ Yan Lu¹

¹Microsoft Research Asia ²Tsinghua University

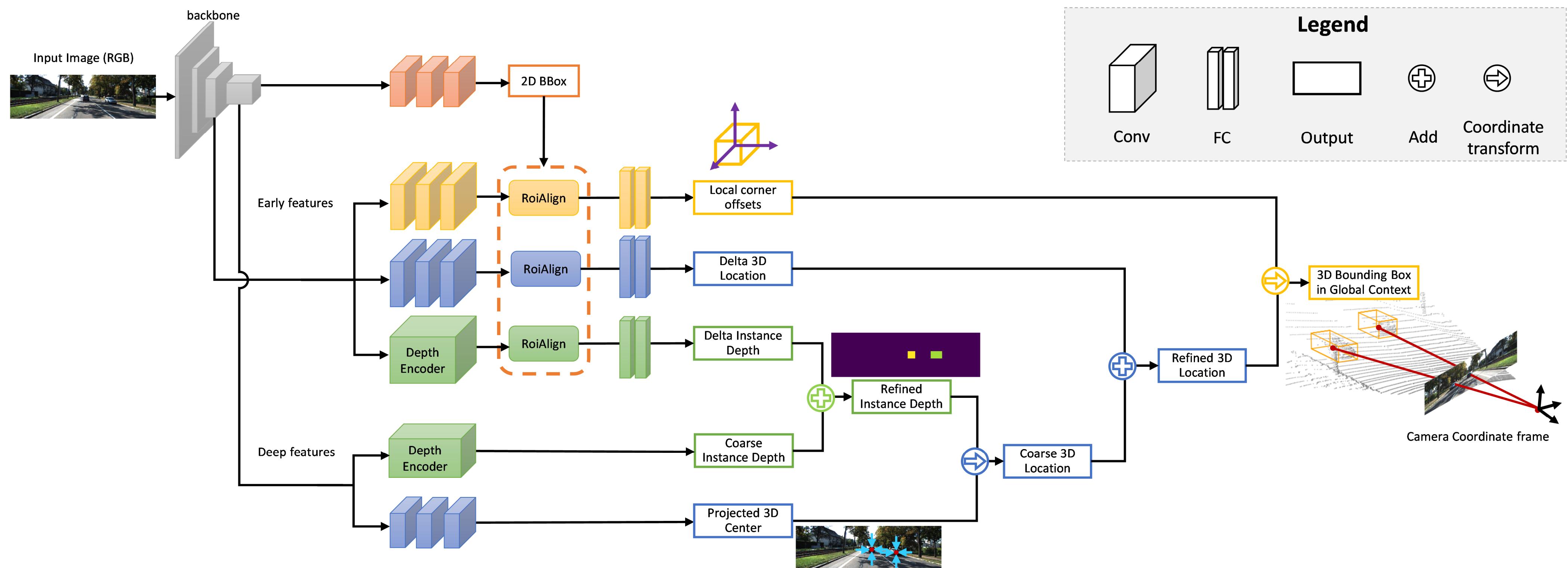


Figure 1. MonoGRNet

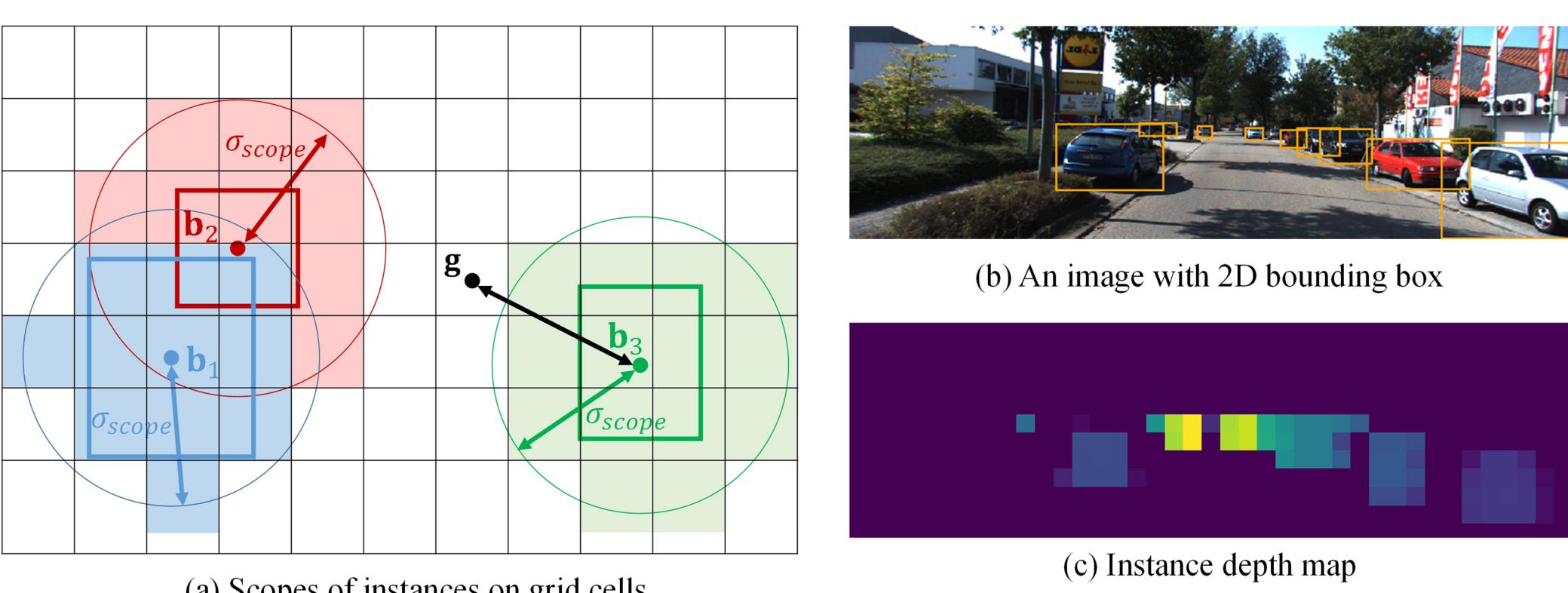
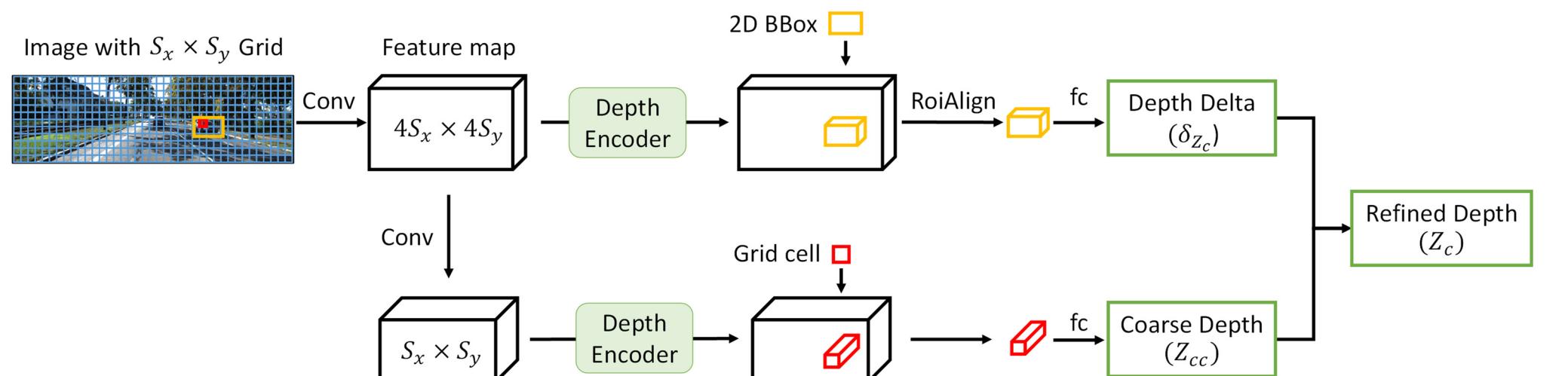
Detecting and localizing objects in 3D space is particularly challenging given only a monocular (single-view) RGB image, since the 3D geometric information is lost via imagery projection. To solve this problem, we propose a geometric reasoning network that outperforms state-of-the-art mono 3D object detectors on the challenging KITTI dataset by a significant margin across almost all evaluation metrics.

Our **key idea** is to decouple the monocular 3D detection task into four sub-tasks that are solvable using only a single-view image:

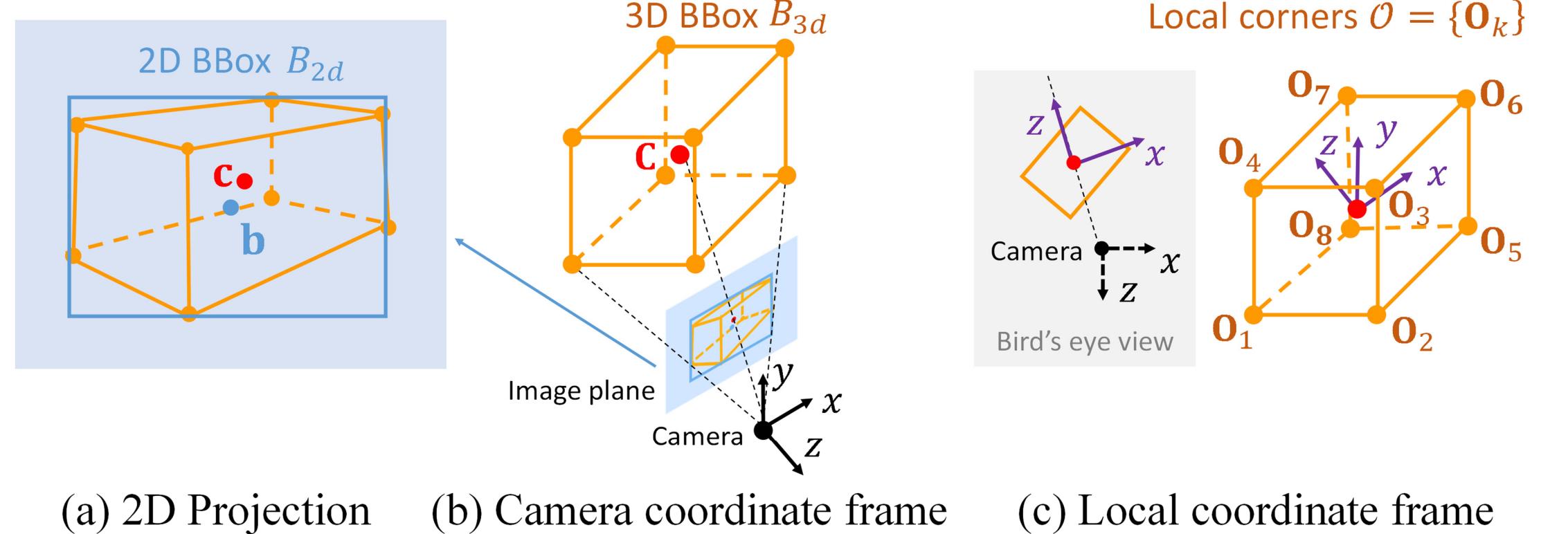
- **2D detection**
- **Instance-level depth estimation**
- **Projected 3D center estimation**
- **Local corner regression**

The four colors correspond to the four task-specific branches of MonoGRNet in Figure 1. By such decomposition, the 3D detection problem can be easily solved using only a monocular image, without any 3D data, e.g., LiDAR point clouds, as input.

Instance-level depth estimation



Projected 3D center estimation



Results



Method	Type	Time (s)	AP _{3D} (IoU=0.3)			AP _{3D} (IoU=0.5)			AP _{3D} (IoU=0.7)		
			Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
3DOP	Stereo	4.2	69.79	52.22	49.64	46.04	34.63	30.09	6.55	5.07	4.10
Mono3D	Mono	3	28.29	23.21	19.49	25.19	18.20	15.22	2.53	2.31	2.31
MF3D	Mono	0.12	/	/	/	47.88	29.48	26.44	10.53	5.69	5.39
Ours	Mono	0.06	72.17	59.57	46.08	50.51	36.97	30.82	13.88	10.19	7.62

Microsoft
Research
微软亚洲研究院

