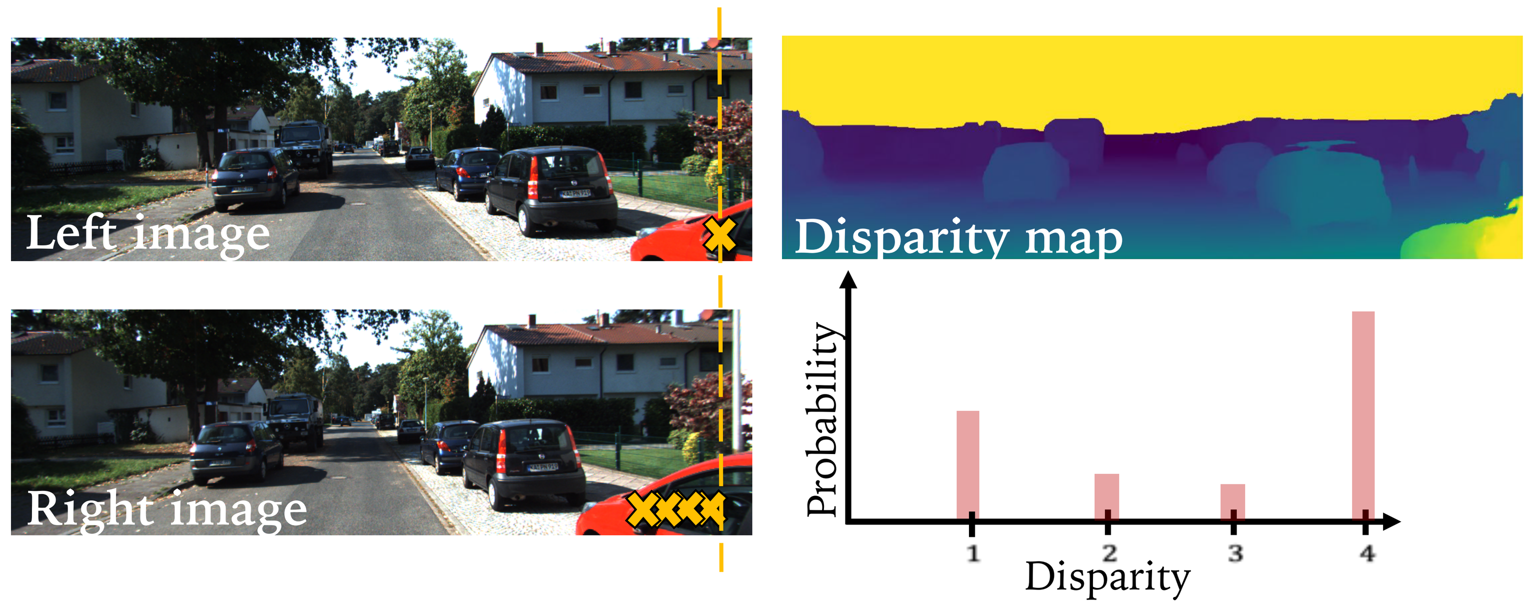
Wasserstein Distances for Stereo Disparity Estimation

Divyansh Garg[1], Yan Wang[1], Bharath Hariharan[1], Mark Campbell[1], Kilian Q. Weinberger[1], Wei-Lun Chao[2]

[1]Cornell University, [2]The Ohio State University

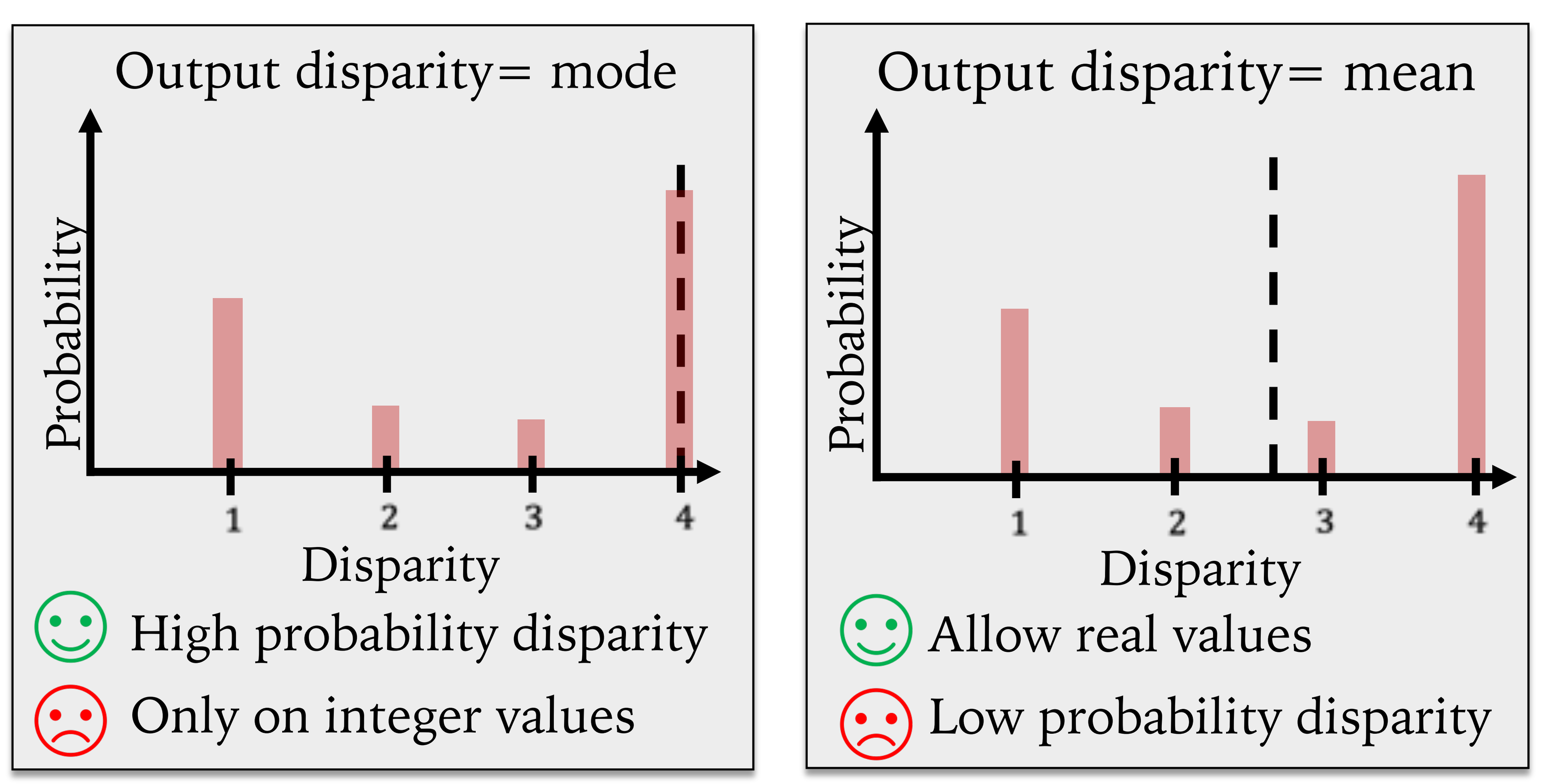
NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

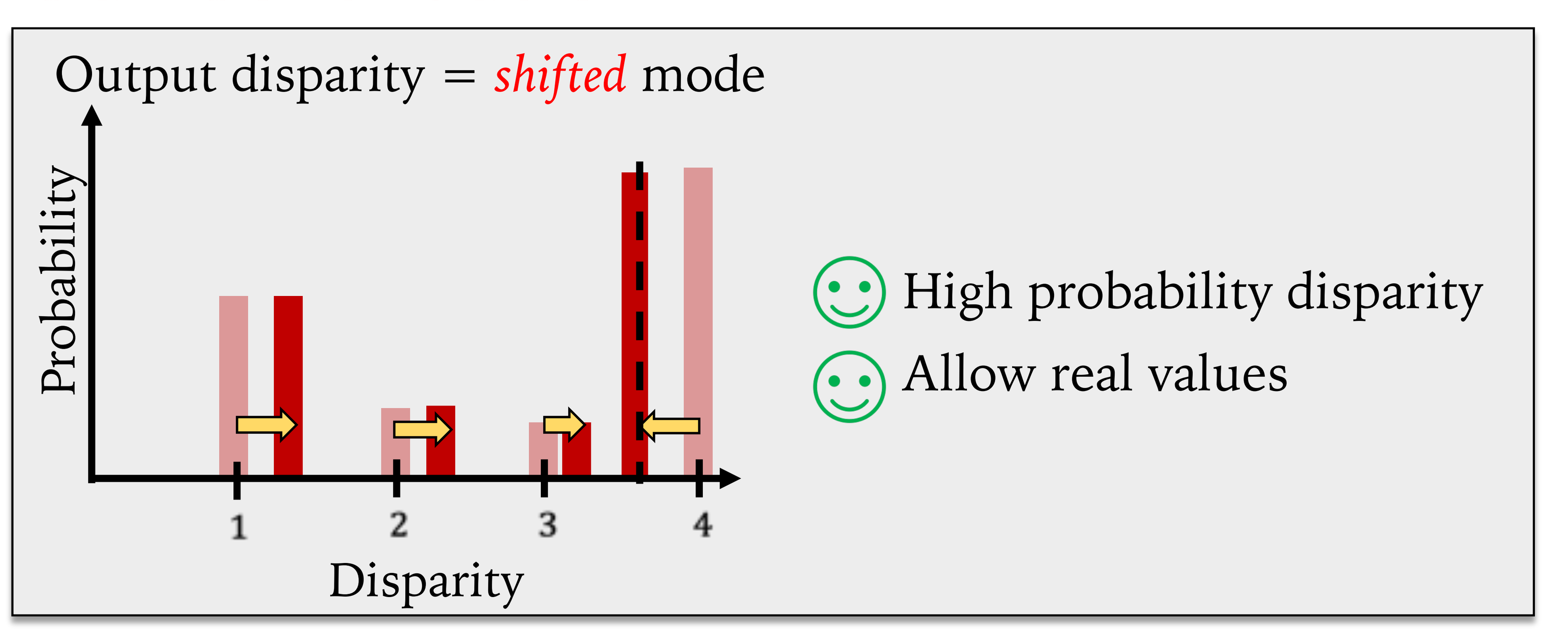


Existing approaches: generate a distribution over a set of integral disparity values. Problems:

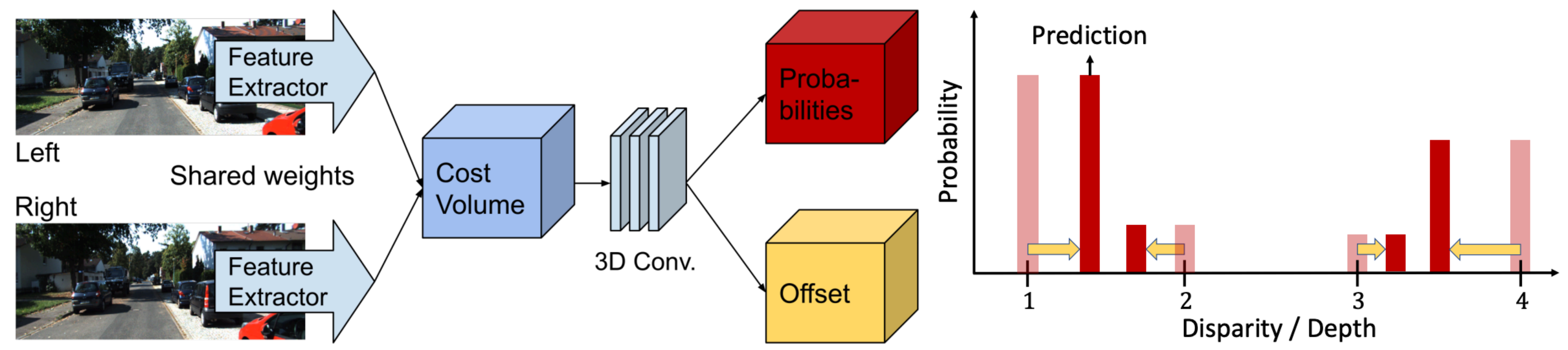
- Inaccurate estimates:
  (a) output the mode but true disparity is sub-pixel.
  (b) output the mean but true distribution is multi-modal.
- Sub-optimal learning: using regression loss, not a distribution matching loss.



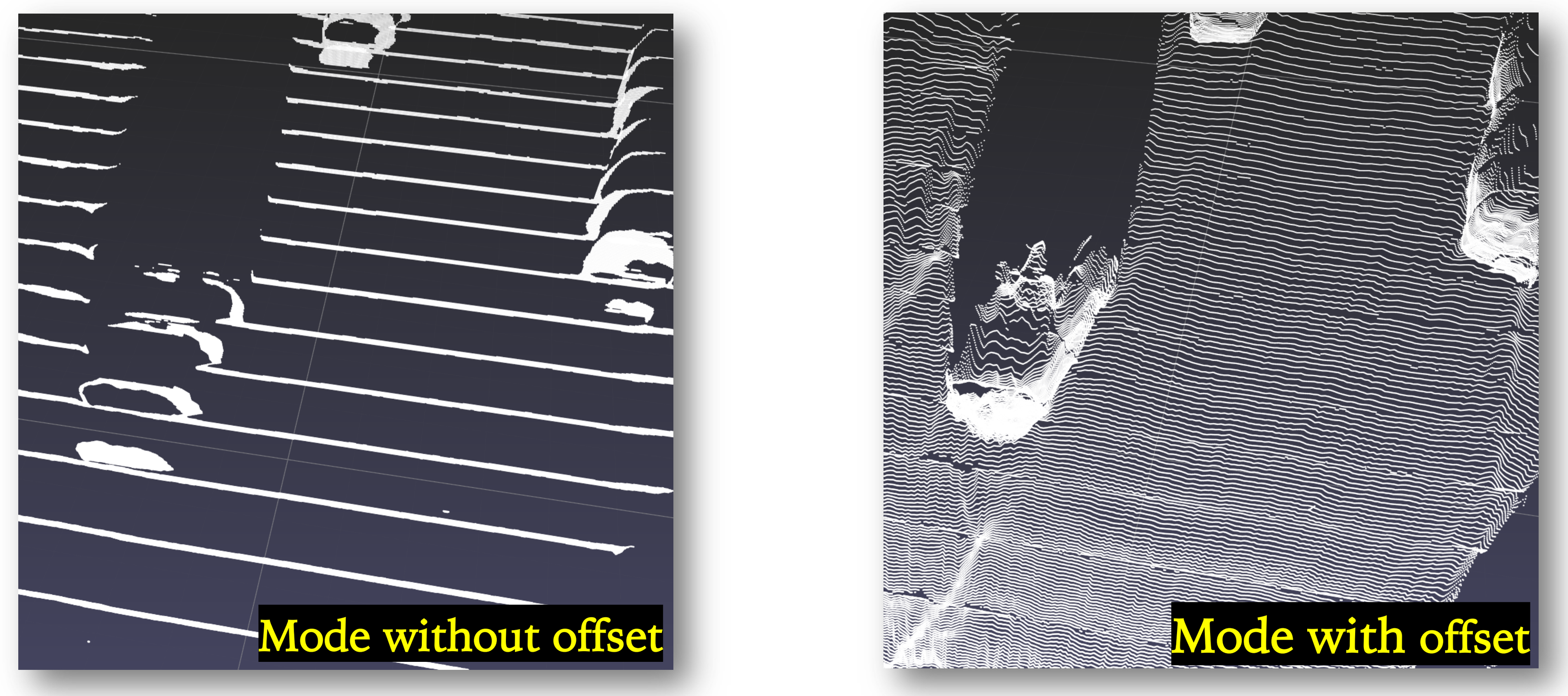
Our Contribution



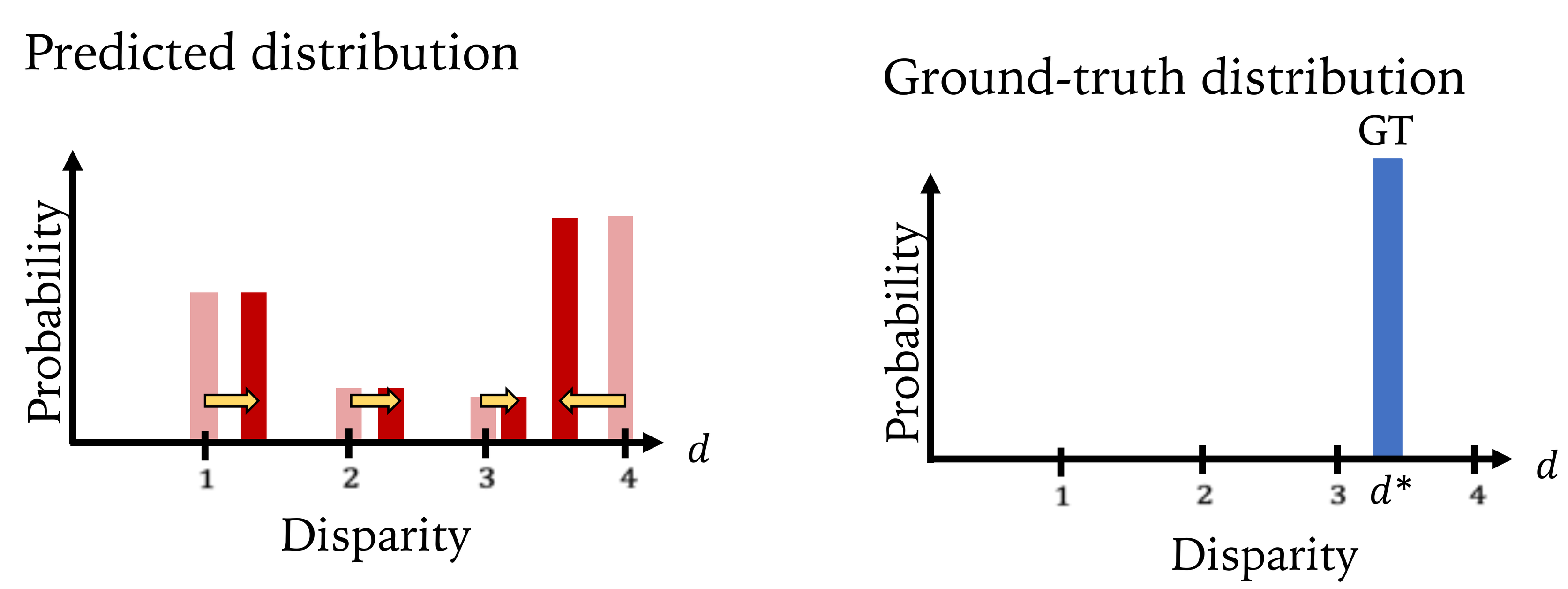
Approach



Visualization of offsets



Learning with Wasserstein Distances



$$\text{Loss} = \sum_d \textbf{probability}(d) \| d + \textbf{offset}(d) - d^* \|$$

Benefit of Wasserstein Distances

1. directly learn the distribution
2. suitable for distributions without common supports
3. efficient implementation for distribution of one variable
4. can extend to multi-modal ground truths (see paper)

Experiments

Disparity estimation results

Method	Scene Flow			KITTI 2015			
				Non Occlusion 3PE		All Areas 3PE	
	EPE	1PE	3PE	Foreground	All	Foreground	All
GC-Net	2.51	16.9	9.34	5.58	2.61	6.16	2.87
PSMNet	1.09	12.1	4.56	4.31	2.14	4.62	2.32
GANet	0.84	9.9	-	3.37	1.73	3.82	1.93
GANet Deep	0.78	8.7	-	3.11	**1.63**	3.46	**1.81**
CDN-PSMNet	0.98	9.1	3.99	4.01	2.12	4.34	2.29
CDN-GANet Deep	**0.70**	**7.7**	**2.98**	**2.79**	1.72	**3.20**	1.92

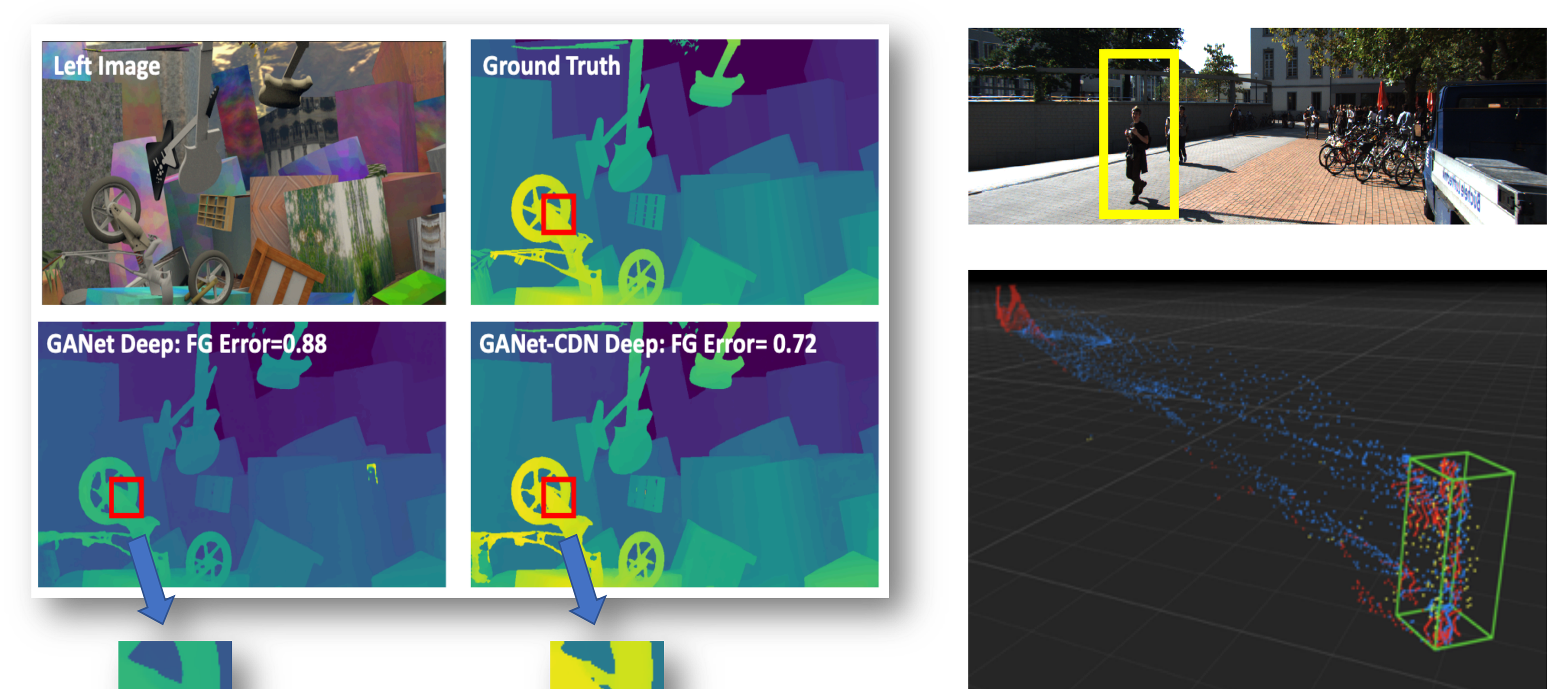
3D object prediction results

Method	BEV Detection AP (AP_{BEV})			3D Detection AP (AP_{3D})		
	Easy	Moderate	Hard	Easy	Moderate	Hard
S-RCNN	61.9	41.3	33.4	47.6	30.2	23.7
OC-STEREO	68.9	51.5	43.0	55.2	37.6	30.3
DISP R-CNN	74.1	52.4	43.8	59.6	39.4	32.0
PSEUDO-LIDAR	67.3	45.0	38.4	54.5	34.1	28.3
PSEUDO-LIDAR ++	78.3	58.0	51.3	61.1	42.4	37.0
PSEUDO-LIDAR E2E	79.6	58.8	52.1	64.8	43.9	38.1
CDN-PSEUDO-LIDAR ++	81.3	61.0	52.8	64.3	44.9	38.1
DSGN	82.9	65.0	56.6	73.5	52.2	45.1
CDN-DSGN	**83.3**	**66.2**	**57.7**	**74.5**	**54.2**	**46.4**

Object boundaries disparity results

Method	EPE	1PE	3PE
PSMNet	3.10	20.1	11.33
CDN-PSMNet	2.10	15.3	8.92
CDN-PSMNet Multi-Modal	**2.08**	**13.2**	**8.65**

Qualitative results



CDN, Baseline, Ground truth bbox