

Adaptive Multi-Modal Cross-Entropy Loss for Stereo Matching

Peng Xu Zhiyu Xiang Chengyu Qiao Jingyun Fu Tianyu Pu

Motivation

- Stereo matching networks poorly estimate clear edges due to the over-smoothing problem, causing **bleeding artifacts** in reconstructed point clouds.



- Existing methods model the disparity ground-truth as the uni-modal distribution, but fail to suppress the multi-modal outputs at the edge. Meanwhile, the single-modal disparity estimator (SME) suffers from severe **misalignment artifacts**.

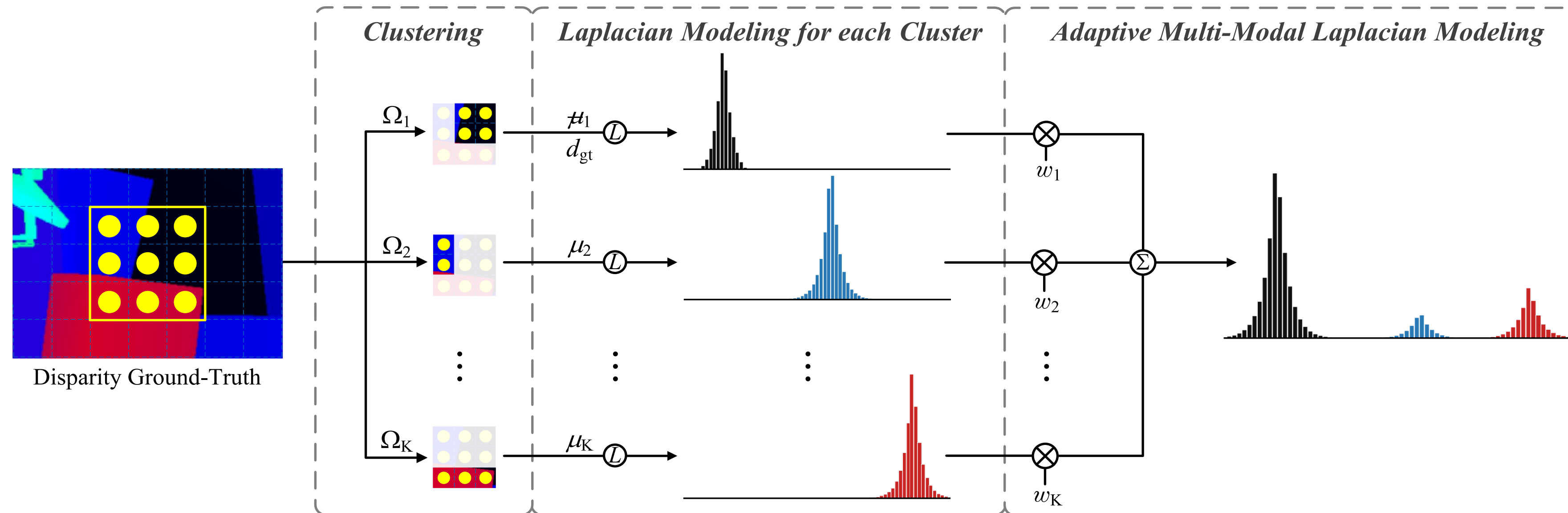


- Our work aims to explore a better modeling for the stereo ground-truth and improve the robustness of the disparity estimator.

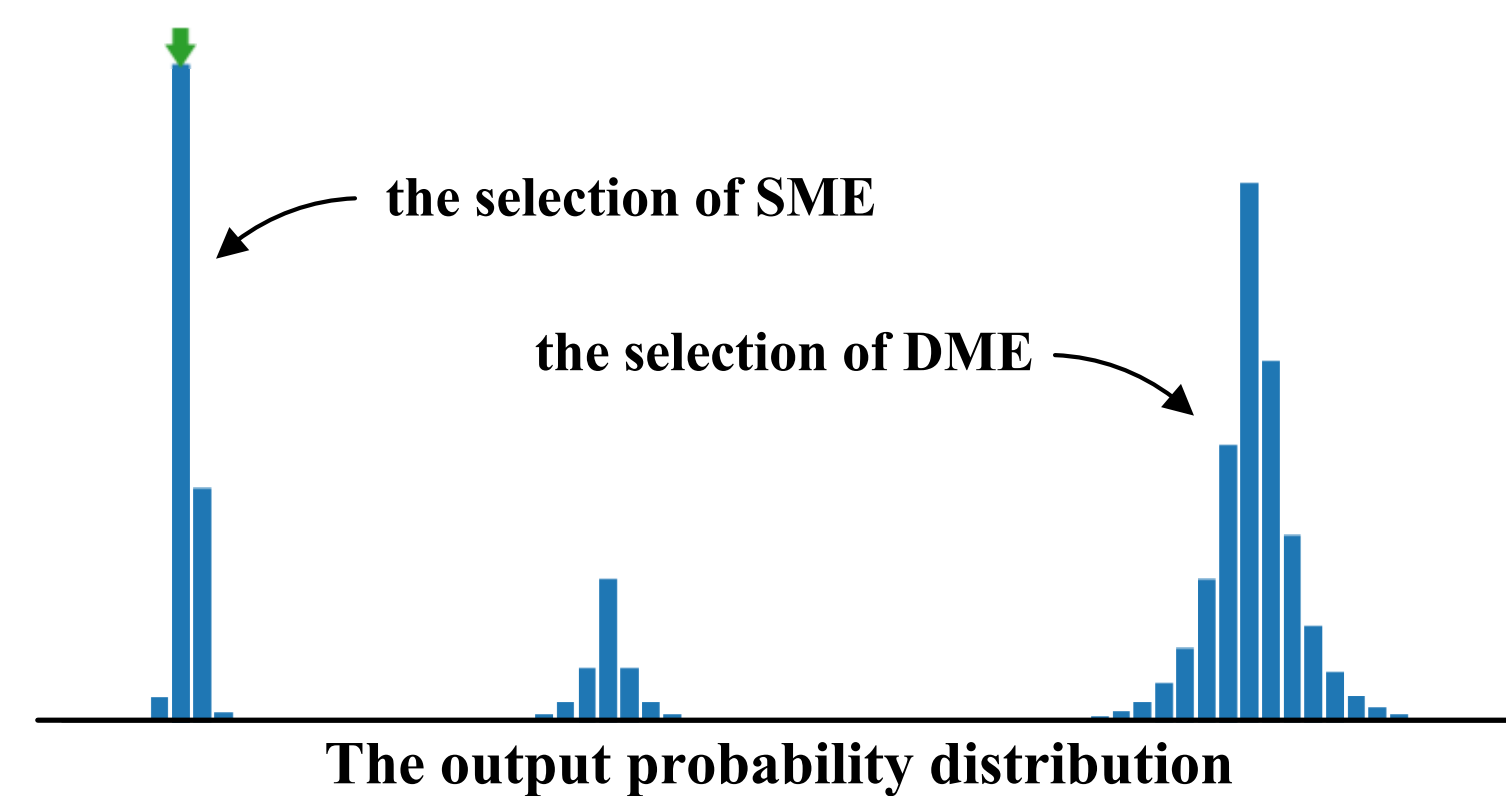
Contributions

- An **adaptive multi-modal probability modeling** for supervising stereo networks training, effectively guiding the networks to learn clear distribution patterns.
- A **dominant-modal disparity estimator (DME)** that can obtain accurate results upon multi-modal outputs.
- State-of-the-art performance** on both the KITTI 2015 and KITTI 2012 benchmarks.
- Excellent cross-domain generalization performance.**

Our Method

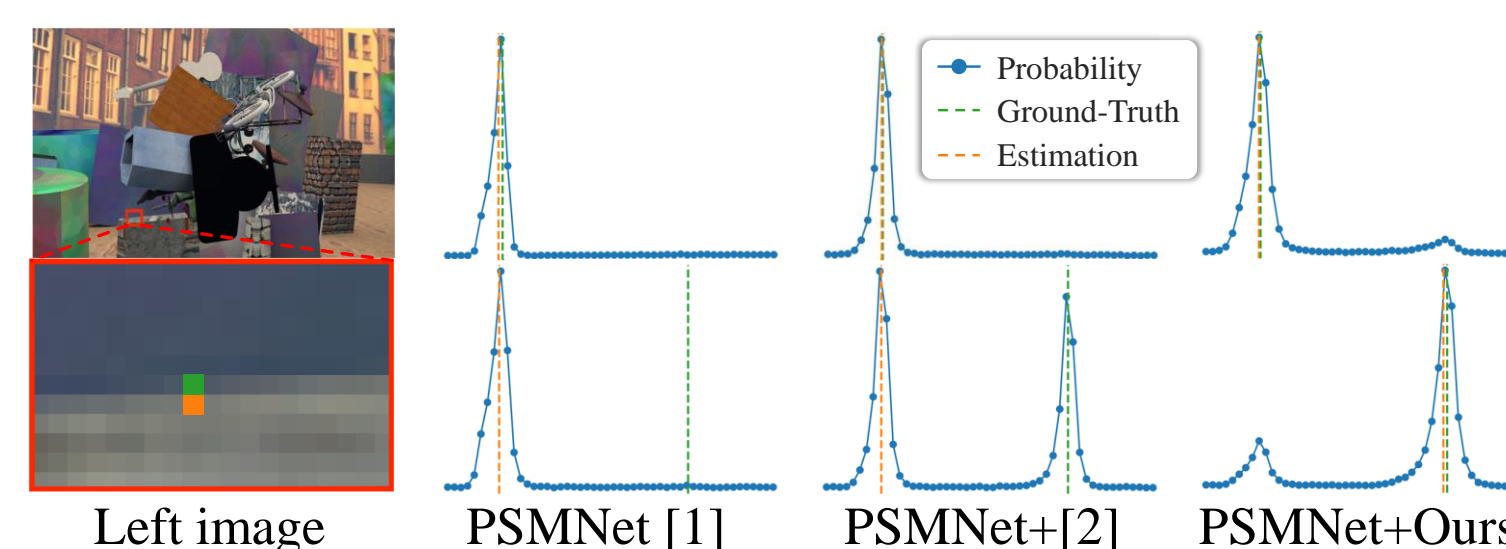


- Clustering is applied within the local window to separate different modals.
- Laplacian distribution is then employed for modeling each cluster.
- Local structural information is used to fuse the generated uni-modal distributions.



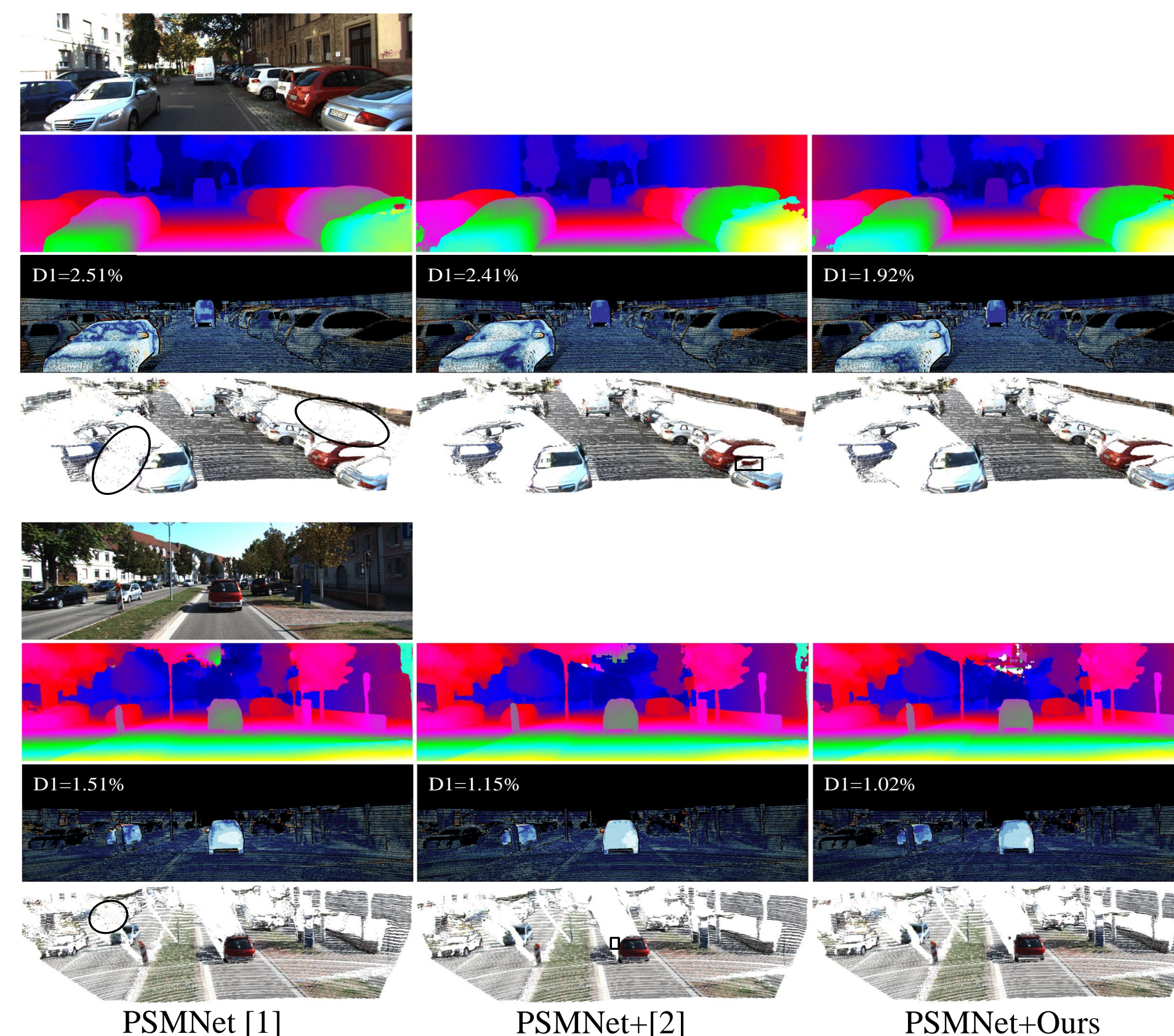
- We propose to select the modal with the **maximum cumulative probability** as the dominant modal.

Vis. of Output Distributions



- Top row: **background** pixel
- Bottom row: **foreground** pixel

Qualitative Results



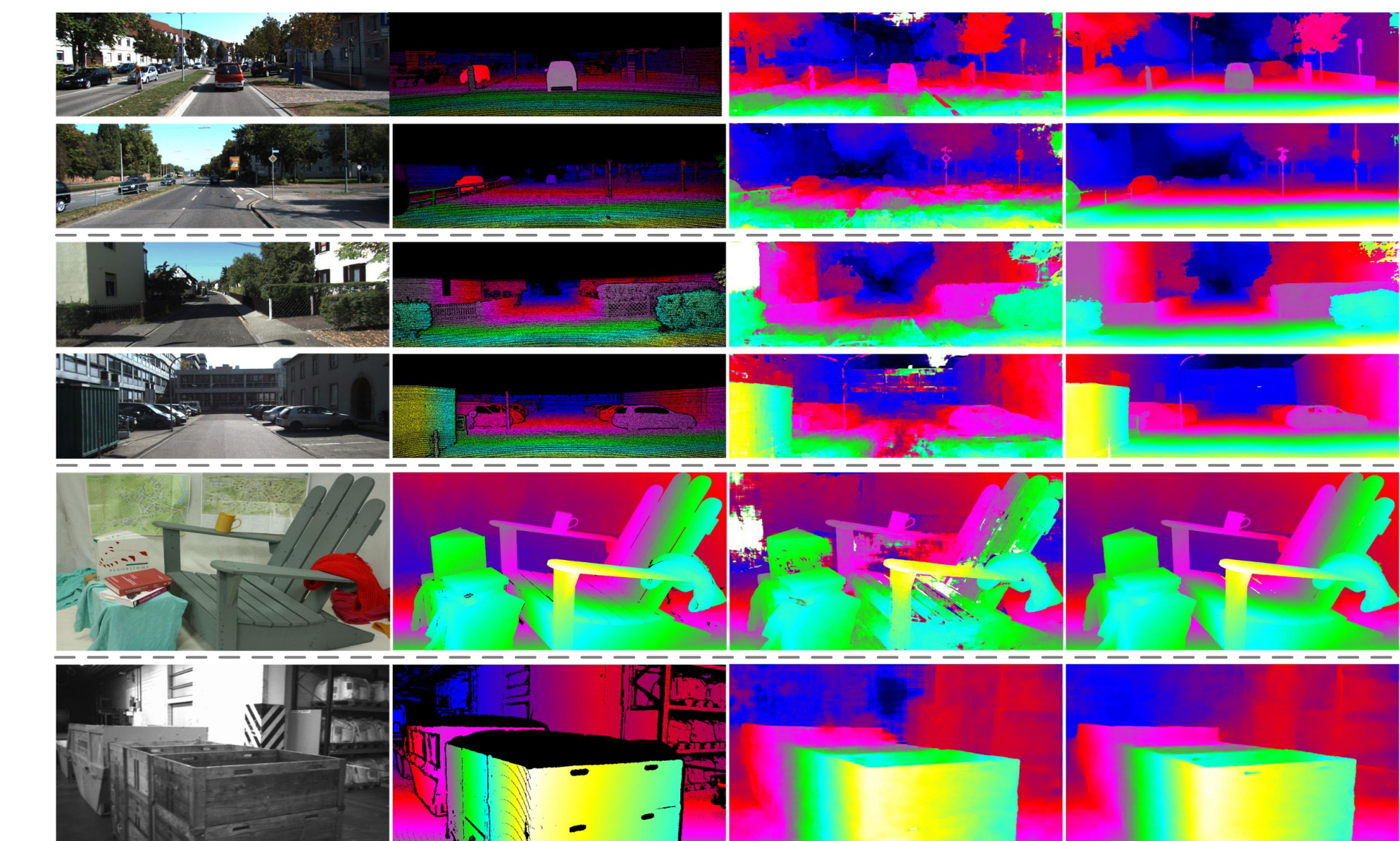
- From top to bottom: left images, disparity maps, error maps, and reconstructed point clouds.

Quantitative Results

Method	KITTI 2015						KITTI 2012			
	All	All	All	Noc	Noc	Noc	>2px	>2px	>3px	>3px
	D1-bg	D1-fg	D1-all	D1-bg	D1-fg	D1-all	Out-Noc	Out-All	Out-Noc	Out-All
PDSNet	2.29	4.05	2.58	2.09	3.68	2.36	3.82	4.65	1.92	2.53
PSMNet	1.86	4.62	2.32	1.71	4.31	2.14	2.44	3.01	1.49	1.89
PSMNet + [2]	1.54	4.33	2.14	1.70	3.90	1.93	2.17	2.81	1.35	1.81
GwcNet	1.74	3.93	2.11	1.61	3.49	1.92	2.16	2.71	1.32	1.70
PSMNet+SMDNet	1.69	4.01	2.08	1.54	3.70	1.89	—	—	—	—
CDN	1.66	3.20	1.92	1.50	2.79	1.72	—	—	—	—
AcfNet	1.51	3.80	1.89	1.43	3.25	1.73	1.83	2.35	1.17	1.54
GANet	1.48	3.46	1.81	1.34	3.11	1.63	1.89	2.50	1.19	1.60
GANet + LaC	1.44	2.83	1.67	1.26	2.64	1.49	1.72	2.26	1.05	1.42
ACVNet	1.37	3.07	1.65	1.26	2.84	1.52	1.83	2.34	1.13	1.47
LEAStereo	1.40	2.91	1.65	1.29	2.65	1.51	1.90	2.39	1.13	1.45
IGEVStereo	1.38	2.67	1.59	1.27	2.62	1.49	1.71	2.17	1.12	1.44
CroCoStereo	1.38	2.65	1.59	1.30	2.56	1.51	—	—	—	—
PSMNet + Ours	1.44	3.25	1.74	1.30	3.04	1.59	1.80	2.32	1.14	1.50
GwcNet + Ours	1.42	3.01	1.68	1.30	2.76	1.54	1.65	2.17	1.05	1.42
GANet + Ours	1.38	2.38	1.55	1.24	2.18	1.40	1.52	2.01	0.98	1.29

- All of the three baselines are lifted to a highly competitive level by our method.
- GANet with our method achieves new state-of-the-art results on both KITTI 2015 and KITTI 2012 benchmarks.

Generalization Performance

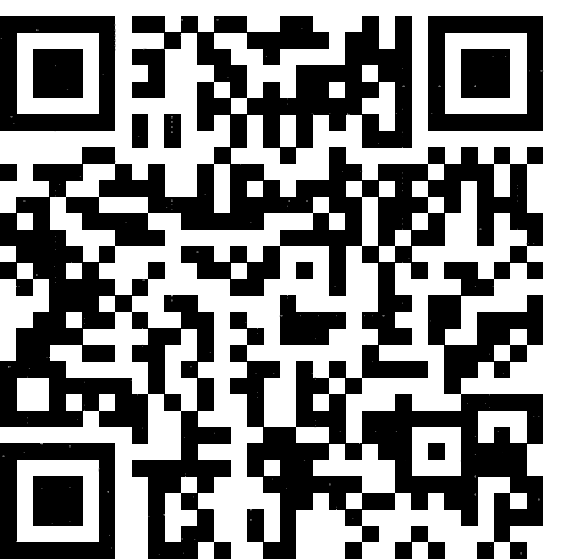


- From top to bottom: KITTI 2015, KITTI 2012, Middlebury, and ETH3D.

References

- [1] Chang and Chen. Pyramid stereo matching network. CVPR 2018.
- [2] Chen, Chen, and Cheng. On the over-smoothing problem of cnn based disparity estimation. ICCV 2019.

Links



paper



demo