Lite Any Stereo: Efficient Zero-Shot Stereo Matching

Supplementary Material

6. Perturbations

In **Stage ②**, we apply strong spatial and photometric perturbations to the training stereo pairs. Photometric augmentations include random `ColorJitter` with large variations in brightness, contrast, saturation, and hue; optional application of a 5×5 Gaussian blur ($\sigma \in [0.1, 2.0]$); and random gamma correction ($\gamma \in [0.7, 1.5]$). These transformations are applied either symmetrically to both views or asymmetrically with a small probability.

To simulate occlusions, we use an "eraser" augmentation that replaces 1–2 rectangular regions (typically 50–100 pixels wide) in the right image with its mean color. Spatial perturbations randomly rescale the images (and dense/sparse flow fields) using scale factors sampled in log-space, followed by up to 20% anisotropic stretching and a final random crop to the target resolution.

7. More Results on Real-world Images

We further assess the zero-shot generalization ability of Lite Any Stereo on diverse real-world stereo images captured under challenging indoor and outdoor scenes. Figure 8 presents representative qualitative results, including the left/right RGB inputs, predicted disparity maps, and reconstructed metric point clouds.

Even without exposure to these environments during training, our method produces clean and geometrically coherent disparity estimates, enabling accurate 3D reconstruction. Indoors, Lite Any Stereo preserves sharp boundaries and maintains stable depth gradients even under low lighting or textureless regions. Outdoors, it handles complex geometry, shadows, and natural clutter while producing reliable disparities and faithful 3D structure.

These results highlight the strong real-world robustness of our method and its ability to generalize beyond curated benchmarks without fine-tuning or domain adaptation.

8. More Results on Middlebury Dataset

We additionally report zero-shot generalization results on the Middlebury dataset across its three official resolution settings—Full (F), Half (H), and Quarter (Q). Table 8 compares Lite Any Stereo with a wide range of efficient stereo matching models, including those trained solely on Scene Flow and those trained with million-scale data or large pseudo-labeled sets.

Across all resolutions, Lite Any Stereo consistently ranks among the top-performing methods in both Bad 2.0 and EPE metrics. These results further validate the effectiveness of our lightweight design and the strong domain generalization offered by our stereo priors.

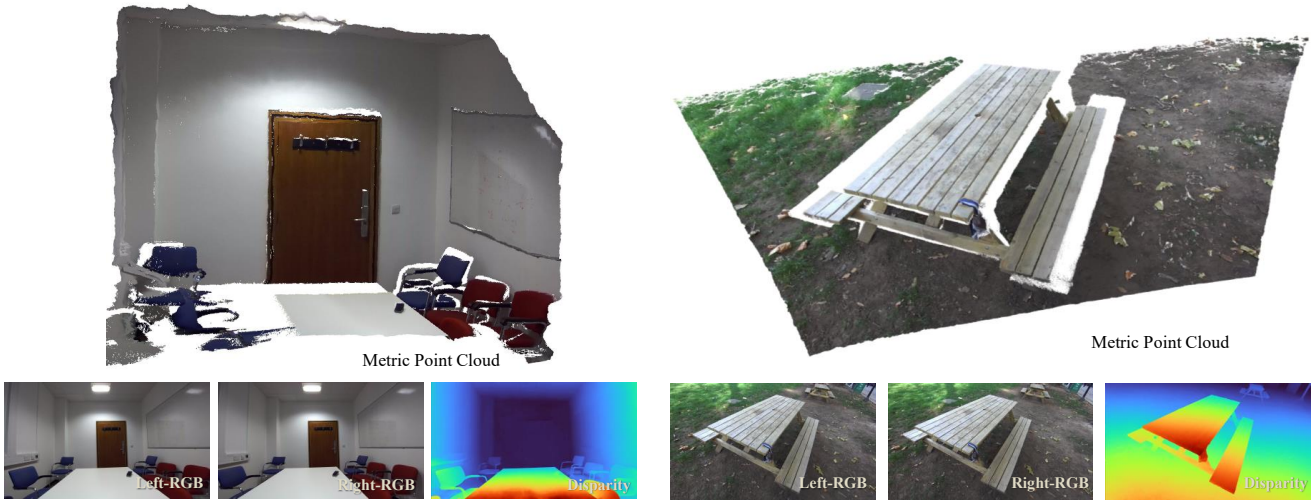

Figure 8. Zero-shot prediction on in-the-wild stereo images.

Method	Middlebury (F)		Middlebury (H)		Middlebury (Q)	
	Bad 2.0	EPE	Bad 2.0	EPE	Bad 2.0	EPE
Efficient methods: SceneFlow						
CoEX [1]	46.88	12.81	26.42	4.90	18.28	2.24
MobileStereoNet-2D [48]	66.02	15.26	37.98	7.54	20.63	3.34
FastACV [65]	37.80	14.23	19.61	4.66	13.50	2.20
FastACV+ [67]	51.87	23.05	27.34	7.16	16.45	2.73
Lite-CREStereo++ [25]	25.09	6.70	14.91	3.32	12.46	1.85
LightStereo-M [20]	31.55	5.61	16.99	2.06	11.64	**1.25**
LightStereo-L [20]	31.17	6.00	17.23	2.88	13.18	1.32
BANet-2D [68]	52.24	16.28	26.79	6.96	23.51	6.81
BANet-3D [68]	52.44	21.97	28.79	8.05	18.69	3.80
Lite Any Stereo	**21.34**	**4.92**	**13.13**	**1.60**	**10.73**	1.28
Efficient methods: Million-scale						
LightStereo-M* [20]	18.60	3.23	10.85	1.51	9.33	1.04
BANet-2D* [68]	18.12	2.37	10.05	1.34	8.99	1.05
StereoAnything-L† [19]	25.24	3.81	9.82	1.21	**5.85**	**0.77**
Lite Any Stereo	**14.81**	**2.95**	**7.51**	**0.94**	7.02	0.85

Table 8. Zero-shot generalization results on Middlebury [46] with different resolution settings. The most commonly used metrics are adopted. In the first block, all efficient methods are trained only on Scene Flow [38]. In the second blocks, methods are allowed to train on any existing datasets excluding the four target domains. The weights and parameters are fixed for evaluation. * indicates models retrained with official code on the same synthetic data as ours. † denotes results trained on 30M pseudo-labeled samples using the strategy in [19]. The **best** and second best are marked with colors.