Stereo Anywhere: Robust Zero-Shot Deep Stereo Matching Even Where Either Stereo or Mono Fail

Luca Bartolomei Fabio Tosi Matteo Poggi Stefano Mattoccia

University of Bologna

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Project Page

CVPR Nashville JUNE 11-15, 2025

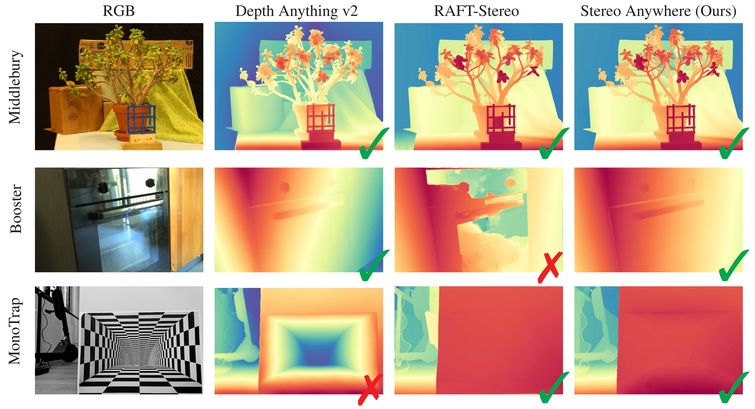
Introduction

Problems:

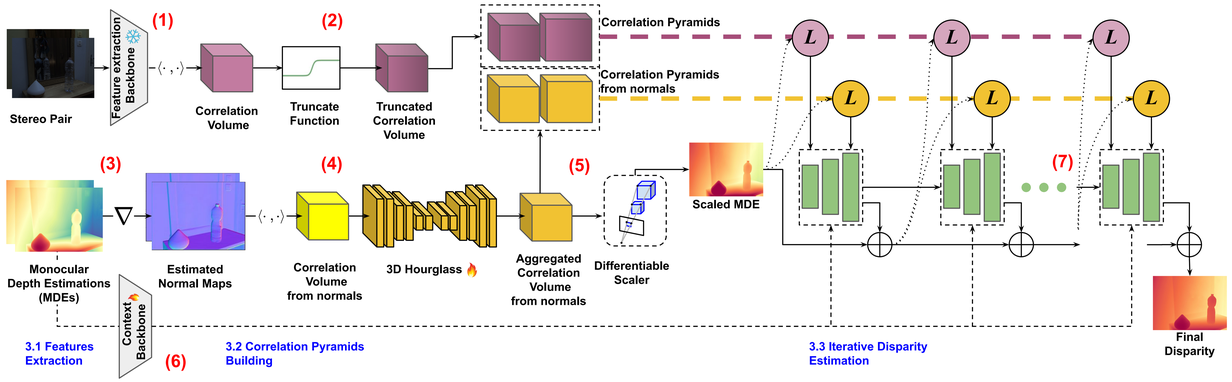
- Existing stereo matching techniques struggle on **non-Lambertian** and **textureless** surfaces.
- Monocular depth estimation is fooled by **perspective illusions**.
- **Stereo training data** for challenging materials is scarce.
- Neither modality alone achieves **robust depth**.

Key Contributions:

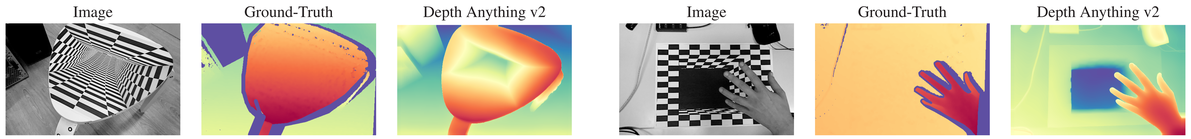
- **Stereo Anywhere**, a novel foundation model for stereo matching.
- Fuses **monocular cues** from single-view depth foundation models.
- **MonoTrap** dataset for benchmarking robustness to **optical illusions**.
- State-of-the-art **zero-shot generalization** on popular benchmarks.



Stereo Anywhere Architecture



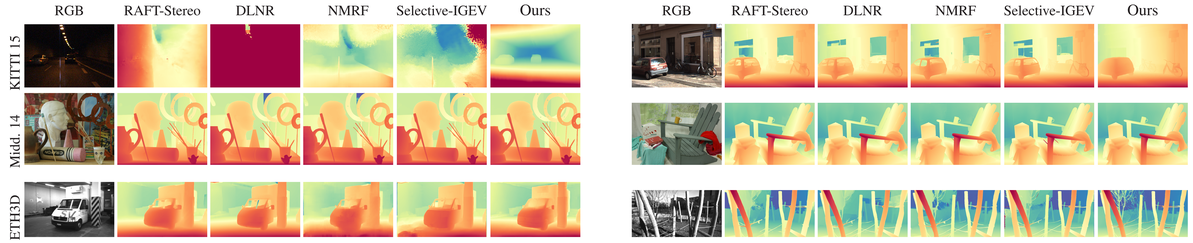
MonoTrap Dataset



Experimental Results

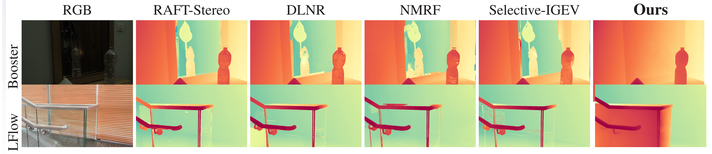
Zero-shot Generalization (Synth to Real)

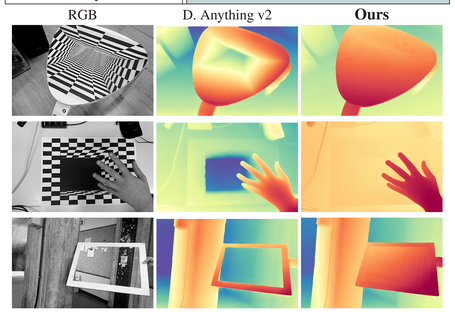
Model	Middlebury 2014 (H) bad > 2 All	Noc	Occ	Avg. (px)	Middlebury 2021 bad > 2 All	Noc	Occ	Avg. (px)	ETH3D bad > 1 All	Noc	Occ	Avg. (px)	KITTI 2012 bad > 3 All	Noc	Occ	Avg. (px)	KITTI 2015 bad > 3 All	Noc	Occ	Avg. (px)
RAFT-Stereo	11.15	8.06	29.06	1.55	12.05	9.38	37.89	1.81	2.59	2.24	8.78	0.25	4.80	4.23	29.21	0.89	5.44	5.21	14.09	1.16
PSMNet	18.79	13.80	53.22	4.63	23.67	20.61	53.75	5.70	19.75	18.62	42.05	0.94	6.73	5.81	46.24	1.22	6.78	6.40	24.85	1.38
GMStereo	15.63	10.98	46.04	1.87	25.43	22.43	54.70	2.86	6.22	5.58	19.97	0.42	5.68	4.87	38.84	1.10	5.72	5.44	17.33	1.21
ELFNet	24.48	16.94	77.06	8.61	27.08	21.77	85.56	11.01	25.61	24.50	46.06	5.65	10.52	8.67	88.21	2.30	9.61	8.22	85.64	2.16
PCVNet	16.79	13.54	35.66	2.96	12.92	10.19	40.23	2.18	4.24	3.61	14.01	0.41	4.44	3.92	27.70	0.89	5.08	4.88	13.72	1.24
DLNR	9.46	6.20	28.75	1.45	8.44	5.88	32.71	1.24	23.12	22.94	26.93	9.89	9.45	8.83	36.75	1.59	15.74	15.41	34.32	2.83
Selective-RAFT	12.05	9.46	27.42	2.35	15.69	13.86	36.32	5.92	4.36	3.81	10.23	0.34	5.71	5.16	30.54	1.08	6.50	6.22	18.44	1.27
Selective-IGEV	9.98	7.09	27.62	1.60	8.89	6.34	32.88	1.60	6.42	5.71	18.71	1.73	6.22	5.54	34.78	1.09	5.87	5.66	14.99	1.42
IGEV-Stereo	9.91	7.08	26.26	1.84	9.15	6.43	34.88	1.53	4.30	3.86	12.65	0.38	5.65	4.43	33.38	1.03	5.87	5.13	14.31	1.34
NMRF	14.08	10.87	34.62	2.91	23.36	21.69	42.51	8.57	4.34	3.66	17.15	0.42	4.62	4.05	30.65	0.92	5.24	5.07	12.28	1.16
Stereo Anywhere (ours)	**6.96**	**4.75**	**20.34**	**0.94**	**7.97**	**5.71**	**29.52**	**1.08**	**1.66**	**1.43**	**5.29**	**0.24**	**3.90**	**3.52**	**21.65**	**0.83**	**3.93**	**3.79**	**11.01**	**0.97**



Zero-shot Non-Lambertian Generalization

Model	Booster (Q) Error Rate (%) > 2	> 4	> 6	> 8	Avg. (px)	LayeredFlow (E) Error Rate (%) > 1	> 3	> 5	Avg. (px)
RAFT-Stereo	17.84	13.06	10.76	9.24	3.59	89.21	79.02	71.61	19.27
PSMNet	34.47	24.83	20.46	17.77	7.26	91.85	79.84	70.04	21.18
GMStereo	32.44	22.52	17.96	15.02	5.29	92.95	83.68	74.76	20.91
ELFNet	45.52	35.79	30.72	27.33	14.04	93.08	82.24	70.41	20.19
PCVNet	22.63	16.51	13.81	12.08	4.70	88.27	76.65	66.79	18.19
DLNR	18.56	14.55	12.61	11.22	3.97	89.90	79.46	72.72	18.97
Selective-RAFT	20.01	15.08	12.52	10.88	4.12	92.69	86.32	78.82	20.18
Selective-IGEV	18.52	14.24	12.14	10.77	4.38	91.31	81.72	74.74	19.65
IGEV-Stereo	16.90	13.23	11.40	10.20	3.94	87.28	80.07	72.91	19.07
NMRF	27.08	19.06	15.43	13.21	5.02	89.08	79.13	70.51	20.17
Stereo Anywhere (ours)	**9.01**	**5.40**	**4.12**	**3.34**	**1.21**	**81.83**	**57.66**	**45.12**	**11.20**
	Booster (Q) Online Benchmark					LayeredFlow (E)			
DKT-RAFT	10.32	7.13	5.65	4.36	1.70	66.05	46.95	37.77	8.72
Stereo Anywhere (ours) (*)	**6.52**	**2.82**	**1.77**	**1.27**	**0.73**	**51.24**	**25.63**	**15.65**	**4.84**



Results on MonoTrap

Model	MonoTrap AbsRel (%)↓	RMSE (m)↓	$\sigma < 1.05$ (%)↑
Depth Anything v2	53.46	0.36	15.21
Depth Anything v2 †	27.92	0.27	19.43
DepthPro	47.77	0.32	21.90
DepthPro †	20.82	0.22	22.88
RAFT-Stereo	5.01	0.09	77.05
Stereo Anywhere	**3.50**	**0.06**	**80.27**