

# Benchmarking Stereo Geometry Estimation in the Wild

## Supplementary Material

	Real				Synthetic				
	Metric Scale		Abs. Depth		Metric Scale		Abs. Depth		
	Scale $\downarrow$	Rank $\downarrow$	$\delta_{1.25}^P \uparrow$	Rank $\downarrow$	Scale $\downarrow$	Rank $\downarrow$	$\delta_{1.25}^P \uparrow$	Rank $\downarrow$	
Stereo	FoundationStereo	1.18	2	<b>81.79</b>	<b>1</b>	1.03	3	<b>95.95</b>	<b>1</b>
	CREStereo	<b>1.10</b>	<b>1</b>	80.94	3	<b>1.02</b>	<b>1</b>	94.02	2
	MASt3R	2.07	15	7.08	17	1.96	11	38.41	11
	MapAnything-Stereo	1.23	5	54.00	8	1.90	10	38.75	9
	Pi3-Stereo	11.67	21	0.08	21	8.27	22	3.44	21
	VGGT-Stereo	13.35	23	0.03	24	9.46	24	3.13	24
Monocular	DepthPro	1.28	9	47.23	13	2.26	12	43.17	7
	Metric3Dv2	1.22	4	64.23	5	1.86	7	15.78	17
	UniDepthV2	1.30	11	64.21	6	2.26	13	52.28	6
	MoGeV2	1.19	3	52.82	9	1.59	5	59.22	5
Multi-view	Pi3Video	40.03	27	0.00	27	19.66	27	0.03	28
	VGGTVideo	92.84	28	0.00	31	35.51	28	0.02	29
	ViPE	1.30	10	50.79	12	2.61	14	33.31	13
	VideoDA-Metric	1.52	12	45.60	14	2.69	15	22.86	15

Table 8. **Absolute depth vs. metric scale.** We show that evaluating absolute depth (without scale alignment) produces similar results to the metric scale evaluation protocol in our benchmark.

## 6. Additional Details

### 6.1. Why for unified scale alignment?

As shown in Tab. 6 and Sec. 4.4, recovering absolute scale is a key advantage of stereo. However, our unified scale alignment aims to *decouple* metric scale evaluation from geometry quality in benchmark design, as otherwise metric scale errors would dominate absolute depth errors. In Tab. ??, we show that evaluating absolute depth (i.e. Tab. 5 w/o scale alignment) produces similar trends to evaluating metric scale (Tab. 6):

### 6.2. Impact of distant depth outliers

One difference between stereo metrics and depth metrics is that stereo metrics mainly focus on near-field errors where disparity error is high, while depth metrics mainly focus on far-field errors. In our paper, we use a max-depth threshold of 80m to clip extremely distant outliers. To analyze the impact of depth outliers on our evaluation, we repeat Tab. 3 with max-depth thresholds of 40m, 20m, and 10m. As shown in Tab. ??, the relative ranking of methods is stable, and the synthetic-to-real gap remains for stereo methods.

### 6.3. Accuracy of ground-truth datasets

To assess quality of ground-truth depths, we measure photometric consistency between depth and stereo pair by warping the right view to the left view (as in S2M2’s KITTI analysis) and measuring PSNR w/ the left view. Our real data has similar photometric consistency as our synthetic

Method	Gap (Rel <sup>P</sup> )	Real			Synthetic			
		Rel <sup>P</sup> $\downarrow$	$\delta_{1.25}^P \uparrow$	Rank $\downarrow$	Rel <sup>P</sup> $\downarrow$	$\delta_{1.25}^P \uparrow$	Rank $\downarrow$	
Stereo (40m)	FoundationStereo	-14.46	21.13	75.59	25	6.67	<b>95.53</b>	<b>1</b>
	CREStereo	-12.34	19.27	77.48	24	6.93	93.67	6
	MASt3R	<b>-3.87</b>	13.18	86.30	3	9.31	90.95	9
	MapAnything-Stereo	-4.71	15.38	83.59	10	10.67	87.55	15
	Pi3-Stereo	-5.92	<b>12.40</b>	<b>86.78</b>	<b>1</b>	<b>6.48</b>	94.08	3
	VGGT-Stereo	-6.48	13.04	83.16	13	6.56	93.88	5
Stereo (20m)	FoundationStereo	-14.25	18.89	79.08	26	<b>4.65</b>	<b>95.71</b>	<b>1</b>
	CREStereo	-11.16	17.21	80.66	24	6.05	93.98	6
	MASt3R	<b>-3.63</b>	12.21	88.16	4	8.59	92.27	8
	MapAnything-Stereo	-4.40	14.57	85.48	14	10.17	88.79	15
	Pi3-Stereo	-5.40	11.55	<b>88.81</b>	<b>1</b>	6.15	94.60	3
	VGGT-Stereo	-4.90	<b>11.12</b>	87.67	7	6.22	94.54	4
Stereo (10m)	FoundationStereo	-12.64	17.11	81.90	26	<b>4.47</b>	<b>95.05</b>	<b>1</b>
	CREStereo	-9.55	15.37	83.43	25	5.82	93.34	6
	MASt3R	<b>-2.61</b>	10.80	89.91	8	8.18	92.01	8
	MapAnything-Stereo	-3.48	13.25	87.31	17	9.76	88.65	15
	Pi3-Stereo	-4.09	9.98	91.10	2	5.89	94.19	4
	VGGT-Stereo	-3.19	<b>9.10</b>	<b>91.88</b>	<b>1</b>	5.90	94.28	3

Table 9. **Impact of far-field depth threshold.** Although evaluating stereo methods with depth metrics places additional emphasis on far-field errors, we show that our findings remain stable under different choices of far-field depth threshold.

data (which is expected to have perfect GT). KITTI’s consistency score is the worst among real datasets.

PSNR	Eden	IRS	MidAir	TartanAirV2	Botanic	KITTI	MS2	VB-Rome
	28.08	32.50	26.06	30.71	25.91	24.97	30.24	26.64

### 6.4. Potential training-set overlap risk

To our knowledge, data overlaps do not impact our key conclusion of a synth-to-real gap in stereo geometry estimation. No evaluated method trains on our real-world datasets due to the need for dense ground-truth. We also find no overlap for stereo (i.e. FoundationStereo, CREStereo) and multi-view methods (i.e. VGGT, Pi3, MapAnything), except Pi3 on MidAir. Our TartanAirV2 subset only includes scenes not seen in TartanAir. Some monodepth models (e.g. DepthPro, MoGeV2) train on synthetic datasets such as Eden, but there are limited new synthetic datasets that have truly no overlap with their training set.

### 6.5. Choice of ranking metric

We use  $\delta_1$  as the ranking criterion because AbsRel can be skewed by large outliers. We re-rank all tables using AbsRel and show that rank correlations between AbsRel and  $\delta_1$  are above  $r = 0.9$ :

Tab. 3 Real	Tab. 3 Synth	Tab. 5 Real	Tab. 5 Synth	Tab. 7 Real	Tab. 7 Online
0.906	0.960	0.925	0.928	0.900	0.950

## 6.6. Multiview-to-stereo repurposing procedure

Given a  $T$ -frame stereo video, we repurpose multi-view models into “-Stereo” variants by running  $T$  inferences, each taking the left and right image as input. We take the left image’s depth as output. As most multi-view models (e.g. VGGT, Pi3) cannot take calibration as input, we assume calibration is not available for all repurposed multi-view models (including MapAnything which can take optional calibration inputs).

## 6.7. Comparability of scale error across modalities

Our scale comparisons are intended to measure absolute scale recovery by decoupling it from relative depth quality. As shown on this page, absolute depth and metric scale evaluation are highly correlated. Our intent is that monocular models can be fairly compared against monocular (e.g. VGGT-Mono), while multi-view models can be fairly compared against multi-view (and a naive way to repurpose monocular models for multi-view is to run it per frame).

## 6.8. Preprocessing and depth-range details

In general, we use the default input image resizing procedure in the public codebase of each model (e.g. padding to multiples of 32 for stereo models, resizing to 518px longest side for VGGT-like models, and resizing to the range of valid image areas for monocular models). We use the original aspect ratio for all images. For depth-range handling, we clip all datasets at a max-depth of 80m and only evaluate on depth pixels that are valid in ground-truth.

## 7. Full Results

We include the full results of our benchmarking below. For each setting, we report the AbsRel ( $\text{Rel}^p$ , ↓),  $\delta_1^p$  (↑), and scale ratio (↓) separately for each dataset.

Table 10. Relative single-frame geometry estimation on synthetic data. We report Rel<sup>P</sup> (↓), δ<sub>1</sub><sup>P</sup> (↑), and Scale Ratio (↓), where Rel<sup>P</sup> and δ<sub>1</sub><sup>P</sup> are percentages.

Method	Average			Eden			IRS			MidAir			TartanAirV2			
	Rank↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓
DepthAnything	13.00	13.92	87.72	471.64	20.12	82.22	139.88	9.38	93.48	205.02	11.77	88.70	1124.30	14.42	86.48	417.36
DepthAnythingV2	14.00	13.48	87.60	981.04	18.92	82.26	303.18	8.82	93.63	474.08	12.03	87.99	2291.80	14.13	86.50	855.10
Depth Pro	20.00	13.75	82.22	2.15	15.50	76.61	1.97	10.41	90.96	1.72	13.58	80.41	3.09	15.52	80.91	1.80
Lotus	21.00	17.28	81.24	2.52	21.94	79.27	1.20	9.69	90.45	1.32	16.56	79.07	5.11	20.94	76.18	2.44
Metric3Dv2	23.00	14.02	80.38	1.85	16.30	76.35	2.03	8.14	91.53	1.70	15.91	76.09	2.13	15.75	77.54	1.55
MoGeV1	11.00	9.58	88.65	6.30	14.81	78.59	1.25	6.98	93.66	1.87	5.15	95.98	17.60	11.38	86.37	4.50
MoGeV2	7.00	8.01	92.06	1.59	10.15	90.09	2.89	5.48	<b>95.25</b>	1.13	5.70	95.61	1.05	10.73	87.29	1.29
UniDepth	24.00	16.37	76.70	2.64	19.73	69.57	3.27	11.51	86.37	1.62	15.71	78.20	3.96	18.53	72.65	1.72
UniDepthV2	18.00	14.52	84.85	2.24	8.77	92.22	4.10	8.72	92.38	1.32	24.80	73.37	2.10	15.79	81.44	1.43
DUST3R-Mono	30.00	20.07	68.86	14.77	23.30	63.09	3.16	11.30	85.74	5.15	22.52	62.59	38.77	23.19	64.03	12.00
MASt3R-Mono	25.00	16.61	76.09	1.99	19.81	69.58	1.21	11.07	87.16	1.22	17.73	72.99	3.96	17.84	74.65	1.56
MapAnything-Mono	19.00	12.24	84.32	1.84	15.85	77.20	3.30	9.20	90.61	1.32	12.77	82.63	1.63	11.14	86.84	1.13
Pi3-Mono	12.00	10.07	87.83	9.13	13.95	80.29	1.97	7.83	91.38	2.31	7.56	93.19	26.30	10.94	86.48	5.93
VGGT-Mono	8.00	8.36	90.93	10.56	8.72	91.57	2.03	7.42	91.63	2.81	5.21	95.99	29.08	12.10	84.51	8.30
CREStereo	6.00	7.43	93.20	1.03	9.41	93.67	1.01	7.74	91.94	1.08	5.57	94.53	1.01	7.00	92.64	1.03
FoundationStereo	<b>1.00</b>	7.05	<b>95.19</b>	1.04	12.95	<b>95.60</b>	<b>1.00</b>	6.65	93.19	1.15	<b>3.73</b>	<b>96.63</b>	<b>1.00</b>	<b>4.88</b>	<b>95.33</b>	<b>1.02</b>
DUST3R-Stereo	22.00	14.01	80.83	17.49	16.49	76.60	3.69	8.58	90.91	5.33	13.77	80.80	47.53	17.19	75.02	13.43
MASt3R-Stereo	9.00	9.92	89.69	2.12	9.80	90.77	1.14	8.30	91.75	1.22	9.82	89.42	4.52	11.75	86.80	1.61
UFM-Base	16.00	15.08	86.07	<b>1.03</b>	13.40	89.36	1.02	12.02	88.52	<b>1.07</b>	16.90	83.90	1.01	17.99	82.52	1.03
UniMatchFlow	10.00	10.64	89.44	1.17	8.74	93.96	1.01	17.46	81.85	1.61	8.09	90.71	1.02	8.28	91.24	1.06
MapAnything-Stereo	17.00	11.31	86.01	1.85	14.87	78.84	3.32	8.71	91.48	1.32	11.22	85.53	1.64	10.46	88.19	1.12
Pi3-Stereo	4.00	6.84	93.50	9.15	7.71	92.93	2.03	6.32	93.71	2.29	5.57	95.33	26.31	7.74	92.06	5.98
VGGT-Stereo	5.00	6.88	93.36	10.61	7.18	94.06	2.05	6.13	93.64	2.82	4.79	96.24	29.08	9.41	89.50	8.51
MapAnything-Video	15.00	11.30	86.10	1.84	15.06	78.89	3.28	7.93	92.58	1.30	12.45	83.46	1.64	9.76	89.47	1.13
Pi3-Video	3.00	6.54	93.64	19.68	9.17	89.86	4.56	<b>5.40</b>	94.62	3.30	5.61	95.40	60.66	5.95	94.67	10.22
VGGT-Video	2.00	<b>6.33</b>	94.12	35.51	<b>6.73</b>	94.45	5.73	5.62	94.36	5.37	4.61	96.21	117.29	8.35	91.48	13.65
MegaSAM	26.00	19.41	72.98	7.13	24.32	61.06	7.41	10.40	88.78	1.45	26.82	61.48	17.85	16.10	80.59	1.81
ViPE	29.00	19.59	72.15	2.27	24.46	63.17	4.78	14.63	82.18	1.34	16.39	75.89	1.30	22.88	67.34	1.67
VideoDA-Metric	28.00	23.08	72.77	2.54	48.57	47.69	5.48	8.68	92.59	1.34	16.73	76.73	1.67	18.34	74.08	1.68
Aether	31.00	43.47	49.19	180.01	88.72	6.61	702.21	15.70	80.74	4.39	21.85	66.82	3.08	47.61	42.58	10.36
GeometryCrafter	27.00	25.37	72.87	4.59	15.94	77.45	2.31	14.84	86.70	2.20	54.67	45.10	7.17	16.05	82.21	6.66

Table 11. Relative single-frame geometry estimation on real data. We report  $\text{Rel}^P$  ( $\downarrow$ ),  $\delta_1^P$  ( $\uparrow$ ), and Scale Ratio ( $\downarrow$ ), where  $\text{Rel}^P$  and  $\delta_1^P$  are percentages.

Method	Rank $\downarrow$	Average			BotanicGarden			KITTI			MS2			VB-Rome		
		$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$
DepthAnything	16.00	18.25	80.16	632.21	27.71	71.15	164.32	10.86	89.56	1408.60	20.03	72.79	478.60	14.41	87.15	477.31
DepthAnythingV2	17.00	18.77	79.89	1279.71	29.63	70.25	260.77	10.85	89.55	2992.20	20.16	73.55	933.04	14.43	86.20	932.83
Depth Pro	19.00	16.50	78.41	1.39	22.55	68.17	1.77	8.55	91.24	1.11	18.33	73.80	1.17	16.57	80.45	1.52
Lotus	22.00	18.03	77.76	3.82	25.00	66.96	1.87	12.07	86.42	5.41	20.36	72.64	4.08	14.69	85.01	3.91
Metric3Dv2	4.00	13.55	84.64	1.24	18.21	76.46	1.54	6.82	93.67	1.09	<b>14.57 82.35</b>	1.11	14.60	86.07	1.21	
MoGeV1	6.00	13.57	83.85	7.74	17.95	75.50	4.04	7.08	93.57	10.14	16.96	78.33	8.41	12.30	87.98	8.37
MoGeV2	7.00	13.87	83.37	<b>1.19</b>	19.51	72.50	1.13	7.08	93.69	1.29	16.58	78.97	1.17	12.33	88.32	1.17
UniDepth	8.00	13.69	83.02	1.26	18.96	72.60	1.87	7.04	92.43	1.02	15.19	80.25	1.05	13.57	86.79	<b>1.10</b>
UniDepthV2	5.00	14.94	84.06	1.28	22.48	73.73	1.60	7.36	94.08	1.05	15.12	82.18	1.09	14.81	86.26	1.38
DUST3R-Mono	27.00	19.36	73.16	24.36	27.28	60.07	12.50	11.97	83.77	27.13	20.55	70.02	29.24	17.63	78.78	28.56
MASt3R-Mono	18.00	16.34	79.11	2.26	22.41	69.70	1.77	10.36	87.95	2.42	17.61	74.32	2.23	15.00	84.46	2.63
MapAnything-Mono	13.00	16.80	81.26	1.26	24.69	70.55	1.55	9.05	91.30	1.08	18.51	77.11	1.08	14.94	86.08	1.31
Pi3-Mono	11.00	14.94	82.20	12.10	19.17	73.86	4.93	8.80	91.23	14.27	18.88	75.96	15.30	12.91	87.78	13.91
VGGT-Mono	21.00	15.20	77.94	14.04	20.34	67.35	6.22	9.06	89.37	19.05	17.39	73.65	16.56	14.03	81.38	14.35
CREStereo	24.00	20.85	76.16	1.22	34.02	60.49	1.51	6.68	94.05	1.02	18.70	78.12	<b>1.04</b>	24.02	71.99	1.33
FoundationStereo	25.00	22.32	74.46	1.38	37.73	56.55	2.02	5.94	<b>94.90</b>	1.01	20.34	76.33	1.09	25.27	70.05	1.42
DUST3R-Stereo	23.00	17.61	77.45	25.13	23.60	68.52	13.18	10.27	87.11	27.76	19.24	73.04	30.22	17.32	81.14	29.34
MASt3R-Stereo	3.00	14.04	84.91	2.31	19.49	79.46	1.90	7.36	92.53	2.62	15.06	80.80	2.24	14.26	86.87	2.49
UFM-Base	29.00	31.56	66.98	1.58	47.49	52.76	<b>1.05</b>	14.36	86.82	1.01	23.19	73.59	1.11	41.21	54.76	3.14
UniMatchFlow	20.00	19.64	78.10	1.43	21.13	78.03	1.09	<b>5.56</b>	94.51	<b>1.01</b>	18.55	78.16	1.06	33.32	61.70	2.57
MapAnything-Stereo	10.00	16.08	82.60	1.25	23.43	73.32	1.53	8.56	91.98	1.07	18.08	77.99	1.07	14.24	87.12	1.31
Pi3-Stereo	<b>1.00</b>	13.17	<b>85.81</b>	12.46	<b>16.12 80.41</b>	4.97	7.17	93.51	14.19	16.81	80.19	15.42	12.57	89.15	15.27	
VGGT-Stereo	15.00	14.07	80.66	14.18	17.76	73.03	6.30	8.30	90.88	19.05	16.42	75.89	16.64	13.78	82.84	14.73
MapAnything-Video	12.00	16.26	81.98	1.24	23.71	71.34	1.48	8.46	92.22	1.06	18.53	76.98	1.12	14.32	87.39	1.32
Pi3-Video	2.00	<b>12.93</b>	85.68	39.65	17.13	78.36	21.96	6.30	94.10	28.60	16.23	80.46	69.63	<b>12.05 89.80</b>	38.40	
VGGT-Video	14.00	13.62	81.20	93.07	17.73	72.51	40.82	7.59	92.10	55.59	16.26	75.93	208.92	12.90	84.24	66.95
MegaSAM	26.00	19.55	73.46	6.22	27.09	62.75	3.14	11.83	84.99	9.15	22.35	64.92	5.36	16.95	81.20	7.23
ViPE	30.00	22.18	65.66	1.26	27.92	56.28	1.40	15.33	76.67	1.17	26.39	54.27	1.19	19.08	75.43	1.29
VideoDA-Metric	28.00	22.14	71.49	1.54	37.96	47.73	2.29	12.23	87.57	1.20	19.76	72.82	1.39	18.61	77.84	1.27
Aether	31.00	41.45	46.34	15.66	75.90	13.61	52.34	15.91	76.80	3.89	45.73	35.55	4.55	28.27	59.39	1.88
GeometryCrafter	9.00	14.85	82.77	16.30	21.03	72.36	8.66	7.38	93.58	19.85	17.47	77.86	18.22	13.55	87.28	18.47

Table 12. Metric-scale video geometry estimation on synthetic data. We report Rel<sup>P</sup> (↓), δ<sub>1</sub><sup>P</sup> (↑), and Scale Ratio (↓), where Rel<sup>P</sup> and δ<sub>1</sub><sup>P</sup> are percentages.

Method	Rank↓	Average			Eden			IRS			MidAir			TartanAirV2		
		Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓	Rel <sup>P</sup> ↓	δ <sub>1</sub> <sup>P</sup> ↑	Scale↓
DepthAnything	19.00	45.94	52.47	623.12	77.24	48.21	298.26	29.67	57.82	247.68	31.37	50.88	1399.20	45.49	52.97	547.32
DepthAnythingV2	26.00	51.61	48.73	1305.22	85.12	43.48	582.85	36.31	53.41	671.53	35.35	46.87	2724.80	49.68	51.15	1241.70
Depth Pro	21.00	33.22	51.97	2.26	50.74	25.81	1.44	18.35	83.98	2.54	36.40	37.22	2.50	27.38	60.86	2.57
Lotus	28.00	52.73	46.48	3.05	78.71	46.47	1.53	38.88	49.61	1.57	35.03	45.30	5.93	58.31	44.53	3.19
Metric3Dv2	12.00	23.17	62.23	1.86	32.32	45.46	2.14	10.85	88.87	1.71	26.96	50.30	2.04	22.54	64.30	1.57
MoGeV1	25.00	30.09	49.22	5.78	39.86	35.81	1.42	25.28	56.71	1.88	22.53	59.02	15.99	32.70	45.36	3.85
MoGeV2	7.00	13.29	82.90	1.59	19.64	68.02	2.91	9.78	91.78	1.11	7.02	95.32	1.05	16.73	76.49	1.28
UniDepth	27.00	31.12	47.35	2.69	39.66	36.54	3.74	19.77	66.56	1.69	29.88	45.13	3.64	35.16	41.19	1.71
UniDepthV2	8.00	19.26	78.07	2.26	10.47	90.55	4.12	14.11	86.93	1.34	27.19	69.40	2.03	25.28	65.38	1.57
DUS3R-Mono	30.00	38.83	36.98	13.39	51.80	25.09	2.78	27.74	48.96	4.73	34.29	39.22	36.16	41.48	34.65	9.89
MASt3R-Mono	24.00	30.65	50.67	1.90	40.15	37.04	1.43	18.84	73.40	1.23	34.60	38.40	3.45	29.02	53.83	1.49
MapAnything-Mono	11.00	22.35	64.01	1.90	32.61	41.82	3.58	15.21	82.75	1.30	25.57	52.78	1.58	16.02	78.72	1.13
Pi3-Mono	16.00	26.63	55.58	8.24	34.42	44.78	1.91	22.59	61.26	2.22	20.49	64.71	23.80	29.02	51.56	5.02
VGGT-Mono	20.00	27.85	52.18	9.39	32.50	47.07	1.83	26.98	48.50	2.56	19.50	66.58	26.44	32.42	46.59	6.74
CREStereo	2.00	8.12	93.43	<b>1.02</b>	9.99	94.07	1.00	9.34	91.50	1.08	5.66	94.68	<b>1.00</b>	7.49	93.47	1.01
FoundationStereo	<b>1.00</b>	<b>7.16</b>	<b>95.33</b>	1.03	12.99	<b>95.67</b>	<b>1.00</b>	6.94	93.21	1.12	<b>3.77</b>	<b>96.67</b>	1.00	<b>4.95</b>	<b>95.75</b>	<b>1.00</b>
DUS3R-Stereo	29.00	34.15	43.55	15.65	46.56	30.25	3.06	26.08	52.50	4.92	27.47	50.74	43.52	36.51	40.72	11.09
MASt3R-Stereo	14.00	26.26	58.13	1.96	31.63	49.42	1.26	17.48	77.16	1.23	30.61	45.47	3.84	25.33	60.45	1.50
UFM-Base	6.00	15.55	86.20	1.02	13.72	89.53	1.01	12.87	88.79	<b>1.06</b>	16.97	83.90	1.01	18.65	82.56	1.01
UniMatchFlow	5.00	12.23	88.65	1.24	9.17	94.13	1.00	22.15	77.95	1.88	8.33	90.91	1.00	9.29	91.60	1.08
MapAnything-Stereo	10.00	21.44	65.46	1.90	31.25	43.60	3.59	14.74	83.46	1.31	24.46	54.79	1.58	15.32	79.99	1.13
Pi3-Stereo	13.00	24.99	58.65	8.27	30.65	51.51	1.92	22.30	61.97	2.21	19.71	66.87	23.87	27.30	54.24	5.08
VGGT-Stereo	18.00	27.10	53.48	9.46	31.86	48.22	1.85	26.45	49.22	2.58	19.42	67.08	26.48	30.69	49.40	6.94
MapAnything-Video	9.00	18.06	72.55	1.85	28.50	49.55	3.38	10.31	91.27	1.28	20.46	64.43	1.62	12.97	84.95	1.13
Pi3-Video	4.00	8.38	91.57	19.66	12.68	83.93	4.56	<b>6.42</b>	<b>94.30</b>	3.33	6.75	94.99	60.55	7.66	93.08	10.21
VGGT-Video	3.00	7.81	92.51	35.51	<b>7.59</b>	93.83	5.73	7.18	93.55	5.41	4.94	96.20	117.40	11.52	86.44	13.51
MegaSAM	23.00	33.94	51.48	8.09	50.27	27.10	11.07	14.91	80.73	1.48	44.06	35.93	17.85	26.53	62.17	1.98
ViPE	22.00	30.41	51.89	2.61	43.05	34.07	5.83	19.64	73.91	1.41	27.79	48.71	1.33	31.18	50.86	1.86
VideoDA-Metric	17.00	32.06	54.80	2.69	64.05	23.06	6.08	10.10	91.12	1.34	28.77	46.30	1.64	25.34	58.71	1.71
Aether	31.00	58.30	26.22	550.00	98.70	0.22	2163.90	38.34	39.73	7.01	30.53	46.07	3.03	65.61	18.85	26.04
GeometryCrafter	15.00	33.36	56.25	5.33	29.26	48.86	2.17	20.42	78.92	2.36	57.00	36.75	10.43	26.77	60.47	6.37

Table 13. Metric-scale video geometry estimation on real data. We report  $\text{Rel}^P$  ( $\downarrow$ ),  $\delta_1^P$  ( $\uparrow$ ), and Scale Ratio ( $\downarrow$ ), where  $\text{Rel}^P$  and  $\delta_1^P$  are percentages.

Method	Rank $\downarrow$	Average			BotanicGarden			KITTI			MS2			VB-Rome		
		$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$	$\text{Rel}^P \downarrow$	$\delta_1^P \uparrow$	Scale $\downarrow$
DepthAnything	15.00	25.66	66.12	1230.76	38.51	52.36	553.62	13.40	84.68	1446.90	24.93	61.60	1468.80	25.79	65.82	1453.70
DepthAnythingV2	19.00	28.52	62.40	2648.05	45.01	43.82	1193.70	15.24	80.46	3213.60	26.67	59.45	3332.50	27.16	65.87	2852.40
Depth Pro	20.00	23.80	62.26	1.28	36.77	35.69	1.53	11.42	87.23	1.11	20.95	68.09	1.18	26.08	58.03	1.31
Lotus	16.00	24.08	65.56	5.00	29.57	57.90	2.14	17.40	75.41	6.14	25.06	64.11	6.04	24.28	64.81	5.68
Metric3Dv2	2.00	<b>14.17</b>	83.91	1.22	<b>19.39</b>	74.70	1.52	7.17	93.60	1.09	<b>14.96</b>	<b>81.86</b>	1.11	15.17	85.48	1.14
MoGeV1	14.00	21.49	67.09	7.57	27.81	52.98	3.94	10.61	89.91	10.13	23.12	63.50	8.29	24.43	61.98	7.91
MoGeV2	4.00	16.37	79.75	1.19	22.95	65.56	1.15	8.13	93.44	1.28	18.40	75.97	1.17	16.00	84.05	1.16
UniDepth	3.00	15.52	80.28	1.27	21.68	66.43	1.90	7.24	92.37	1.02	16.32	78.14	1.05	16.84	84.17	<b>1.09</b>
UniDepthV2	6.00	18.90	77.82	1.30	27.22	62.90	1.64	8.11	93.85	1.05	15.66	81.54	1.09	24.60	72.99	1.44
DUST3R-Mono	26.00	26.18	58.94	23.90	35.89	40.94	12.12	14.14	81.49	27.11	25.12	58.33	28.50	29.57	54.99	27.88
MASt3R-Mono	30.00	31.60	47.10	2.00	32.14	47.26	1.69	25.88	53.31	2.23	33.54	43.30	1.95	34.86	44.52	2.11
MapAnything-Mono	12.00	21.30	72.61	1.24	29.98	57.83	1.59	11.03	89.32	1.07	20.09	73.49	1.08	24.10	69.79	1.23
Pi3-Mono	22.00	24.11	61.53	11.37	28.96	49.14	4.57	12.30	86.96	14.10	25.14	57.66	14.46	30.06	52.35	12.35
VGGT-Mono	27.00	24.61	58.32	13.34	32.43	39.66	5.80	12.80	84.36	19.09	25.93	55.01	15.88	27.26	54.23	12.57
CREStereo	7.00	20.07	77.32	<b>1.10</b>	31.82	61.10	1.18	6.49	94.46	1.01	18.78	78.33	<b>1.03</b>	23.18	75.38	1.17
FoundationStereo	13.00	22.06	70.93	1.18	37.95	43.96	1.46	5.89	<b>95.02</b>	1.01	19.36	77.61	1.03	25.06	67.14	1.22
DUST3R-Stereo	21.00	24.75	61.98	24.54	32.64	46.59	12.66	12.68	85.26	27.73	23.85	60.79	29.34	29.85	55.28	28.44
MASt3R-Stereo	28.00	30.00	53.02	2.07	29.46	57.42	1.83	23.42	60.57	2.49	32.95	45.90	1.93	34.17	48.20	2.03
UFM-Base	23.00	35.19	60.32	2.23	47.75	52.82	<b>1.04</b>	14.39	86.80	1.01	23.76	71.09	1.08	54.88	30.57	5.79
UniMatchFlow	9.00	22.03	74.27	1.85	21.07	<b>80.54</b>	1.05	<b>5.58</b>	94.51	<b>1.01</b>	18.85	76.73	1.04	42.64	45.31	4.32
MapAnything-Stereo	10.00	20.58	74.12	1.23	28.70	60.57	1.57	10.51	90.22	1.07	19.50	74.91	1.07	23.61	70.79	1.22
Pi3-Stereo	18.00	23.25	63.52	11.67	27.03	53.49	4.62	11.40	88.73	14.01	23.84	60.18	14.61	30.73	51.69	13.45
VGGT-Stereo	25.00	24.18	59.37	13.35	31.02	41.81	5.82	12.39	85.74	19.10	25.42	55.98	15.93	27.87	53.96	12.54
MapAnything-Video	8.00	19.39	77.03	1.24	27.43	63.44	1.50	9.72	91.78	1.06	19.62	74.75	1.11	20.78	78.15	1.27
Pi3-Video	<b>1.00</b>	14.25	<b>84.73</b>	40.03	19.46	75.35	21.85	6.70	94.02	28.58	17.18	78.91	69.58	<b>13.65</b>	<b>90.64</b>	40.12
VGGT-Video	5.00	15.51	78.76	92.84	20.19	68.54	41.19	8.33	91.99	55.66	17.49	73.62	208.72	16.01	80.91	65.80
MegaSAM	29.00	32.32	48.10	5.73	38.62	39.94	2.92	16.76	74.29	8.97	39.20	34.05	4.37	34.72	44.13	6.64
ViPE	24.00	25.75	59.91	1.30	31.78	49.47	1.40	15.74	75.97	1.17	28.57	49.70	1.17	26.91	64.50	1.47
VideoDA-Metric	17.00	24.88	65.05	1.52	41.13	37.53	2.20	12.70	86.91	1.20	20.59	71.07	1.38	25.11	64.71	1.29
Aether	31.00	56.90	29.44	161.62	98.99	0.24	620.51	18.70	70.80	4.56	66.29	16.52	16.54	43.61	30.20	4.86
GeometryCrafter	11.00	19.22	73.27	15.62	27.87	55.73	6.54	8.59	93.11	19.86	20.07	72.41	18.07	20.36	71.86	18.03