DuCos: Duality Constrained Depth Super-Resolution via Foundation Model

Supplementary Material

6. Metric

Given the depth prediction $\mathbf{Y}$ and ground truth depth $\mathbf{Z}$, we use RMSE (cm), MAE (cm), and δ_1 accuracy for evaluation, which are defined as follows:

$$\begin{aligned} &\text{MAE}: \frac{1}{n}\sum|\mathbf{Z}-\mathbf{Y}|, \\ &\text{RMSE}: \sqrt{\frac{1}{n}\sum(\mathbf{Z}-\mathbf{Y})^2}, \\ &\delta_i: \frac{q}{n}\times 100\%,\ q: \max(\mathbf{Z}/\mathbf{Y}, \mathbf{Y}/\mathbf{Z}) < 1.05^i. \end{aligned} \tag{18}$$

Since δ_2 and δ_3 of different DSR approaches are very close, we focus on δ_1 for comparisons.

7. Implementation Detail

Our DuCos implementation is built in PyTorch and runs on a single NVIDIA RTX 4090 GPU. The model is trained for 200 epochs using the Adam optimizer [13], with an initial learning rate of 5×10^{-5}, which is halved at epochs 40, 80, and 120. To enhance performance, we apply data augmentation techniques, including random horizontal flipping and random 90° rotation. Given that Hypersim [31] has a high resolution (1024×768), whereas the test splits of other datasets typically have much lower resolutions, we perform random 256×256 cropping when training on Hypersim. Additionally, DuCos leverages Depth-Anything-v2-Small as the default foundation model to generate prompts.

8. Complexity Analysis

Tab. 5 gives a detailed complexity comparison on the real-world RGB-D-D benchmark. Our DuCos achieves the lowest error, highlighting its superior performance. However, this comes at the expense of significant computational complexity due to the large size of the depth foundation model. Particularly, the number of trainable parameters in DuCos remains manageable at approximately 9.6 M. The primary contributor to this complexity is the large depth foundation model. Therefore, future research could explore efficient model distillation techniques to compress these large models into more lightweight counterparts while maintaining their effectiveness in prompt-based applications.

9. Scale-cross Validation

Tab. 6 presents a cross-scale performance comparison of various DSR approaches. Specifically, we evaluate models trained for $\times 4$ super-resolution directly on the more challenging $\times 8$ and $\times 16$ super-resolution tasks without fine-tuning. Despite this substantial scale discrepancy, our DuCos method consistently outperforms other approaches, demonstrating its superior cross-scale generalization capability. These results underscore the robustness and adaptability of DuCos in handling large-scale variations, making it a promising solution for real-world scenarios where training and testing resolutions may differ significantly.

Method	Prompt	Params. (M)	Time (ms)	Speed (FPS)	RMSE (cm)
DCTNet [57]	×	**0.48**	9.15	109.29	5.43
FDKN [12]	×	0.69	5.78	173.01	5.37
DKN [12]	×	1.16	17.75	56.34	5.08
FDSR [5]	×	0.60	**5.05**	**198.02**	5.49
SUFT [33]	×	22.01	13.33	75.19	5.41
SGNet [41]	×	8.97	33.94	29.46	5.32
SFG [55]	×	63.55	21.81	45.85	3.88
DA v2-S [52]	×	24.79	29.27	34.16	87.45
DuCos†	✓	34.38	25.09	39.86	**3.68**

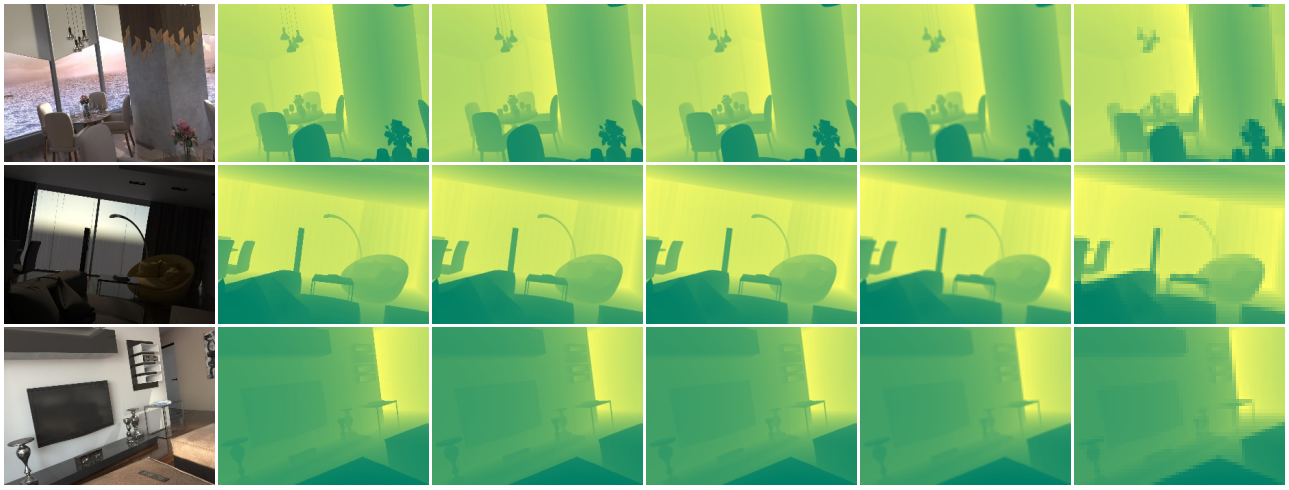
Table 5. Complexity comparisons on the real-world RGB-D-D dataset. † indicates that our DuCos employs DA v2-S to produce the prompts. All methods are measured using a single 4090 GPU.

Method	scale	RGB-D-D RMSE	RGB-D-D MAE	RGB-D-D $\delta_{1.05}$	Lu RMSE	Lu MAE	Lu $\delta_{1.05}$
CUNet [4]	×4 → ×8	3.30	1.66	97.07	4.90	2.59	93.74
FDKN [12]		3.39	1.22	96.97	4.94	1.56	94.03
DKN [12]		3.31	1.22	96.97	4.86	1.52	94.21
FDSR [5]		3.29	1.19	97.01	4.81	1.54	94.18
DCTNet [57]		3.33	1.20	97.03	4.92	1.55	94.14
DuCos		**3.19**	**1.17**	**97.11**	**4.64**	**1.46**	**94.28**
CUNet [4]	×4 → ×16	5.25	2.69	92.89	7.79	3.85	87.19
FDKN [12]		5.16	2.34	92.97	7.56	2.93	87.91
DKN [12]		5.15	2.32	93.00	7.54	2.92	87.93
FDSR [5]		5.16	2.33	93.01	7.55	2.92	87.93
DCTNet [57]		5.15	2.32	92.99	7.55	2.91	88.02
DuCos		**5.13**	**2.31**	**93.03**	**7.51**	**2.87**	**88.08**

Table 6. Cross-scale DSR on the synthetic RGB-D-D and Lu.

10. Numerical Results of Arbitrary-scale DSR

Tab. 7 provides a comprehensive numerical evaluation of arbitrary-scale DSR on the synthetic datasets, corresponding to Fig. 5 in the main text. Particularly, our DuCos consistently achieves the best performance across nearly all datasets and scales, further demonstrating its robustness and effectiveness in handling the DSR task.



RGB GT ×2 downsampling ×4 downsampling ×8 downsampling ×16 downsampling

Figure 9. Visual examples of the fully synthetic Hypersim [31] dataset.

Method	Scale	Middlebury			Lu			NYU v2			RGB-D-D			TOFDSR		
		RMSE	MAE	δ_1	RMSE	MAE	δ_1	RMSE	MAE	δ_1	RMSE	MAE	δ_1	RMSE	MAE	δ_1
CUNet [4]	×1.5	0.79	0.46	99.22	0.67	0.30	99.54	1.33	0.40	99.83	0.77	0.21	99.82	2.09	0.44	99.07
FDKN [12]		1.21	0.49	98.54	1.42	0.34	99.11	1.80	0.41	99.69	0.93	0.22	99.76	1.32	0.19	99.73
DKN [12]		1.09	0.54	98.86	1.19	0.37	99.33	1.46	0.35	99.80	0.80	0.20	99.81	0.85	0.13	99.89
FDSR [5]		0.82	0.40	99.08	0.82	0.28	99.39	1.39	0.43	99.83	0.77	0.21	99.83	1.83	0.38	99.24
DCTNet [57]		0.89	0.48	99.12	0.91	0.32	99.45	1.32	0.46	99.85	0.70	0.22	99.87	0.53	0.13	99.94
DuCos		**0.57**	**0.36**	**99.65**	**0.36**	**0.17**	**99.87**	**0.76**	**0.24**	**99.96**	**0.48**	**0.16**	**99.94**	**0.26**	**0.08**	**99.98**
CUNet [4]	×2.7	1.20	0.65	98.47	1.26	0.50	98.97	2.07	0.69	99.63	1.08	0.32	99.69	2.34	0.48	99.21
FDKN [12]		1.61	0.68	98.01	1.92	0.49	98.77	2.72	0.70	99.44	1.33	0.34	99.56	2.13	0.33	99.53
DKN [12]		1.55	0.68	98.15	1.76	0.51	98.88	2.48	0.65	99.51	1.23	0.31	99.63	1.80	0.26	99.71
FDSR [5]		1.32	0.62	98.19	1.44	0.44	98.88	2.31	0.72	99.54	1.15	0.33	99.66	2.24	0.45	99.21
DCTNet [57]		1.30	0.62	98.48	1.34	**0.42**	**99.04**	2.05	0.72	99.64	**1.07**	0.32	**99.71**	1.30	0.27	99.71
DuCos		**1.12**	**0.56**	**98.83**	**1.14**	0.43	98.83	**1.98**	**0.56**	**99.67**	**1.07**	**0.29**	**99.71**	**1.10**	**0.18**	**99.84**
CUNet [4]	×3.4	1.39	0.75	98.15	1.47	0.60	98.75	2.69	0.91	99.42	1.32	0.43	99.54	2.85	0.67	98.87
FDKN [12]		1.74	0.75	97.80	2.07	0.54	98.62	3.15	0.85	99.32	1.50	0.40	99.46	2.60	0.42	99.39
DKN [12]		1.77	0.77	97.83	2.10	0.59	98.62	3.01	0.81	99.35	1.44	0.38	99.52	2.34	0.35	99.55
FDSR [5]		1.43	0.70	97.90	1.46	0.50	98.68	2.73	0.94	99.36	1.30	0.40	99.58	3.01	0.69	98.64
DCTNet [57]		1.55	0.71	98.01	1.69	0.53	98.78	2.58	0.93	99.48	1.32	0.42	99.57	1.92	0.44	99.44
DuCos		**1.33**	**0.62**	**98.52**	**1.35**	**0.40**	**99.16**	**2.37**	**0.70**	**99.55**	**1.19**	**0.33**	**99.64**	**1.69**	**0.26**	**99.71**
CUNet [4]	×5.3	1.93	0.91	97.36	**1.82**	0.68	98.33	3.69	1.25	99.07	**1.69**	0.53	99.27	3.60	0.83	98.51
FDKN [12]		2.08	0.91	97.17	2.35	0.66	98.20	4.07	1.23	99.04	1.88	0.53	99.18	3.50	0.63	98.96
DKN [12]		1.96	0.88	97.47	2.28	0.66	98.25	3.62	**1.10**	**99.20**	1.76	0.50	99.29	3.40	**0.60**	**99.03**
FDSR [5]		1.92	0.91	97.33	2.19	0.69	98.18	3.66	1.13	99.17	1.75	0.51	99.25	**3.18**	0.61	98.96
DCTNet [57]		2.06	0.91	97.08	2.27	0.67	97.96	3.72	1.31	99.05	1.90	0.58	99.11	3.27	0.72	98.58
DuCos		**1.88**	**0.81**	**97.68**	2.15	**0.56**	**98.48**	**3.60**	1.17	**99.20**	1.73	**0.49**	**99.30**	3.42	0.63	99.01
CUNet [4]	×11.6	3.65	1.73	92.72	4.14	1.47	94.79	6.70	2.72	97.08	2.94	1.09	97.93	5.82	1.76	96.11
FDKN [12]		3.25	1.50	94.28	3.86	1.25	95.36	6.04	2.36	97.74	2.85	1.00	98.14	**5.66**	1.52	96.76
DKN [12]		3.20	1.46	94.60	3.96	1.27	95.20	**5.91**	2.29	97.87	2.89	0.99	98.09	5.85	1.51	96.89
FDSR [5]		3.14	1.43	94.62	3.72	1.19	95.61	6.01	2.24	97.92	2.85	0.96	98.11	5.82	1.51	96.81
DCTNet [57]		3.88	1.82	91.86	4.68	1.57	93.58	7.14	3.09	96.33	3.45	1.25	97.48	5.95	1.61	96.35
DuCos		**2.84**	**1.22**	**95.99**	**3.66**	**1.01**	**96.56**	5.99	**2.11**	**98.13**	**2.84**	**0.90**	**98.25**	5.79	**1.29**	**97.49**

Table 7. Quantitative comparisons with arbitrary scaling factors on the synthetic DSR benchmark datasets.

11. More Visualizations

Fig. 9 presents some RGB-D examples from the fully synthetic Hypersim [31] dataset, demonstrating its high-quality and realistic scenes. Fig. 11 shows visual comparisons at ×2, ×4, ×8, and ×16 scales on the synthetic NYU v2, while Fig. 12 illustrates results for arbitrary-scale DSR. Additionally, Fig. 10 showcases visual results on the real-world RGB-D-D and TOFDSR benchmark datasets. These visualizations further confirm that our DuCos effectively enhances depth predictions, yielding more precise shapes, sharper edges, and improved structural consistency.

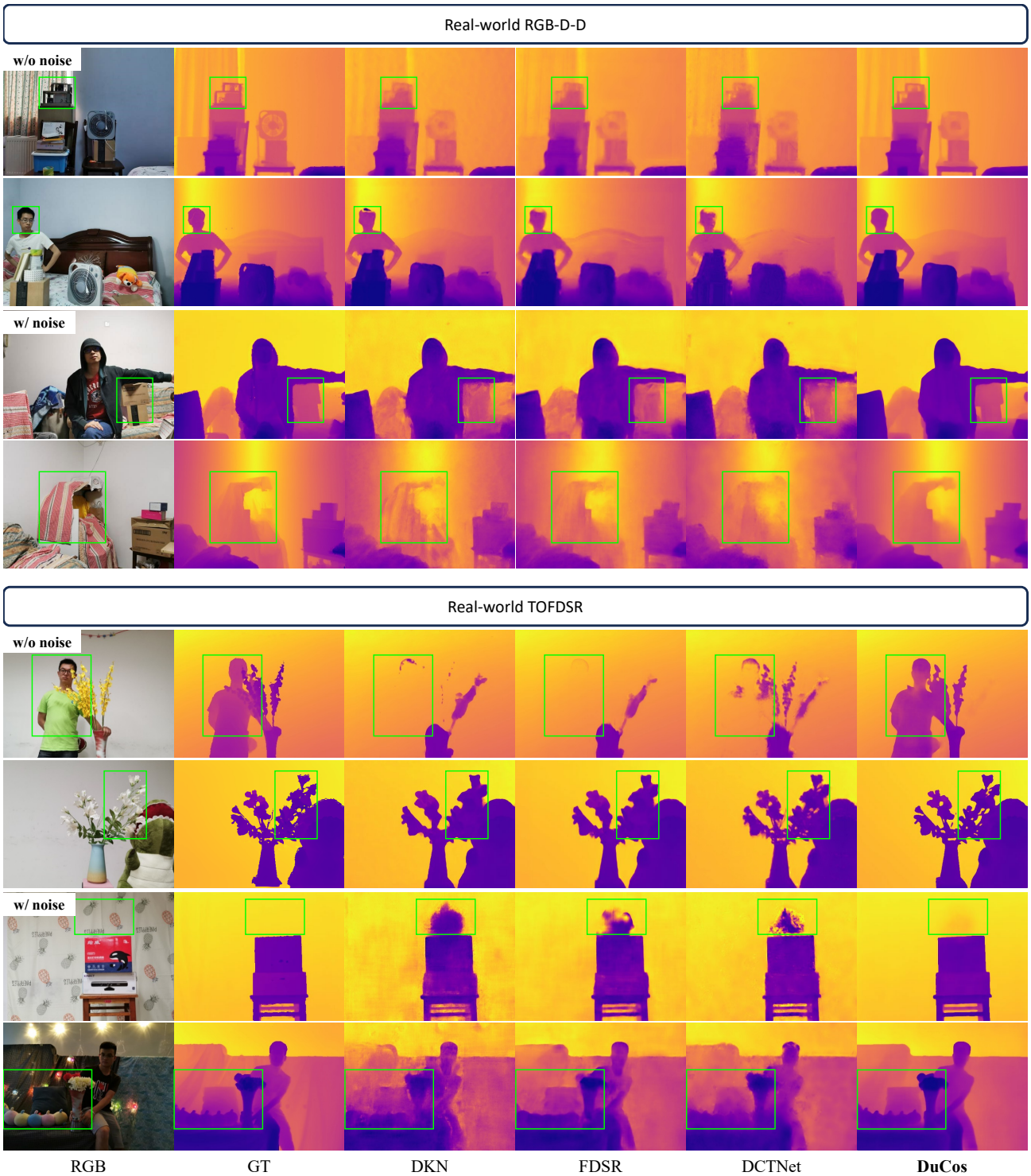


Figure 10. Visual comparisons of different DSR methods with and without noise on the real-world RGB-D-D and TOFDSR datasets.

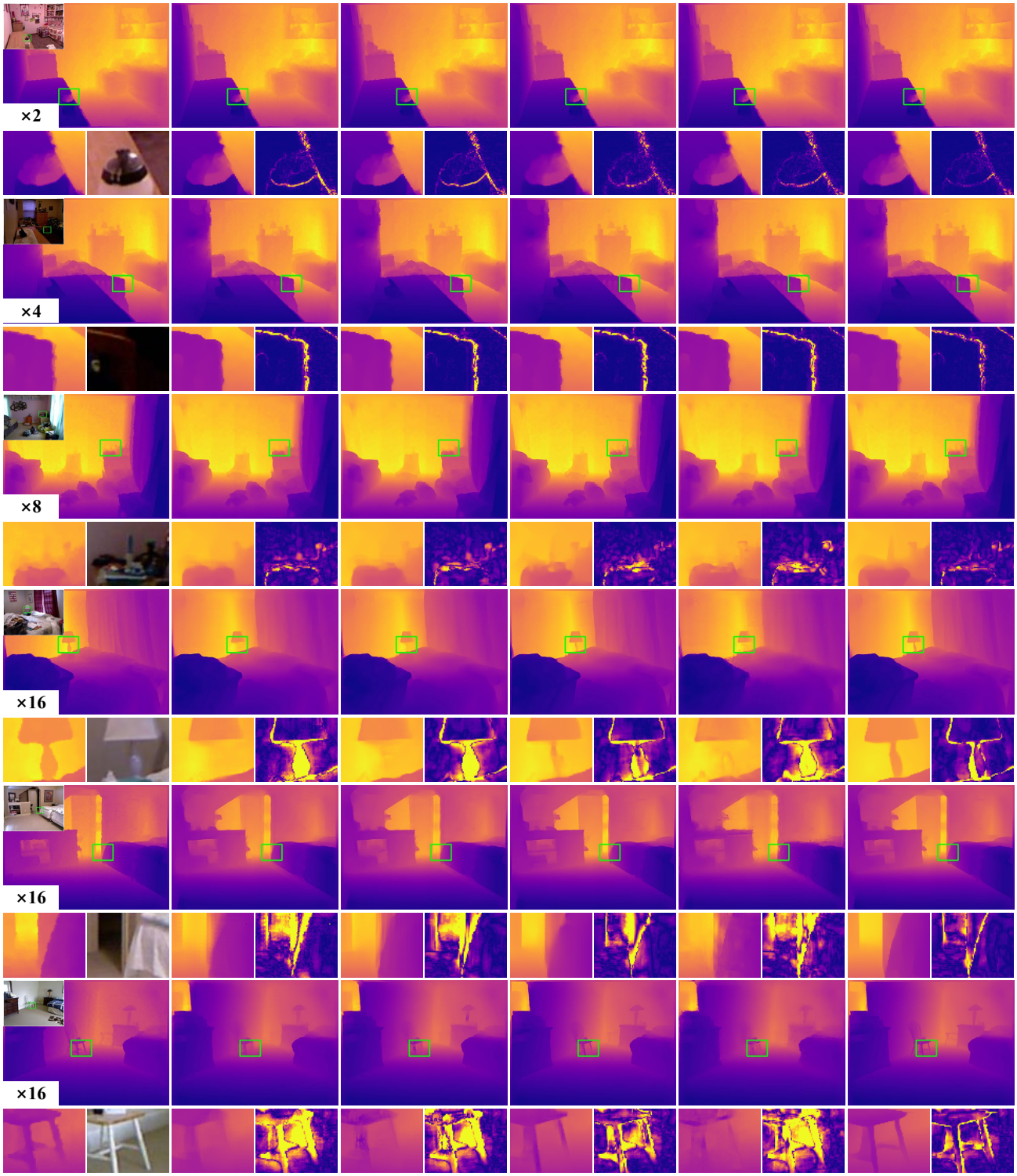


Figure 11. Visual comparisons of different DSR methods with $\times 2$, $\times 4$, $\times 8$, and $\times 16$ scaling factors on the synthetic NYU v2 dataset.

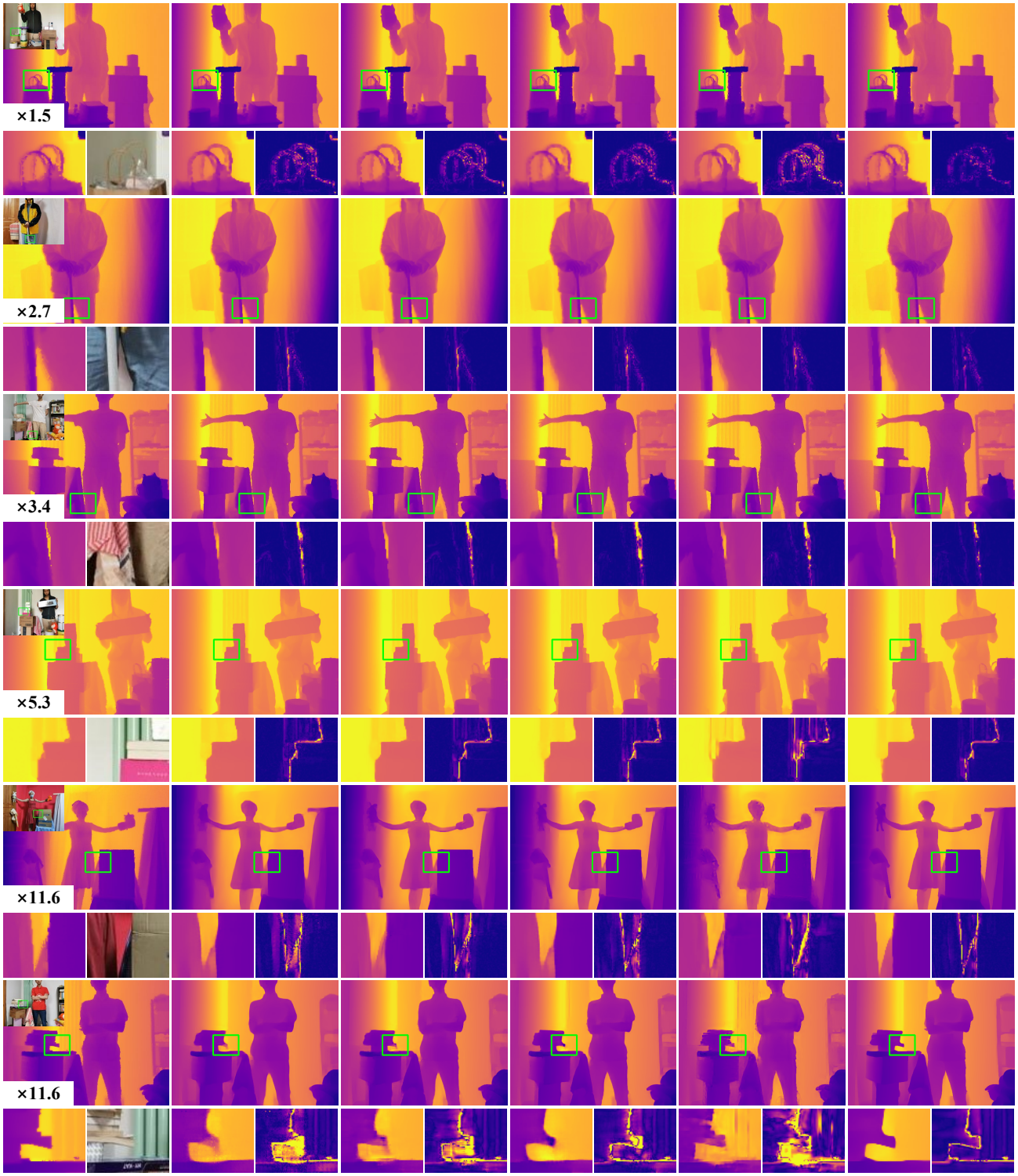


Figure 12. Visual comparisons of different methods with $\times 1.5$, $\times 2.7$, $\times 3.4$, $\times 5.3$, and $\times 11.6$ scales on the synthetic RGB-D-D dataset.