DORNet: A Degradation Oriented and Regularized Network for Blind Depth Super-Resolution

Supplementary Material

Methods	Params. (M)	w/o Noisy		w/ Noisy	
		MAE	MSE	MAE	MSE
DJFR [9]	**0.08**	2.10	45.78	3.84	60.66
DKN [7]	1.16	2.07	42.34	2.54	47.64
FDKN [7]	0.69	2.19	46.07	3.13	60.26
FDSR [5]	0.60	1.67	37.05	2.27	42.73
DCTNet [24]	0.48	2.10	36.12	2.63	38.34
SUFT [12]	22.01	1.13	28.12	2.13	33.84
SFG [22]	63.53	1.72	28.20	2.89	37.09
SGNet [15]	8.97	1.33	27.01	2.20	33.68
DORNet-T	0.46	1.36	35.47	1.81	37.54
DORNet	3.05	**1.02**	**26.10**	**1.41**	**28.03**

Table 1. Quantitative comparison of additional metrics on real-world TOFDSR. MAE and MSE are calculated in centimeters.

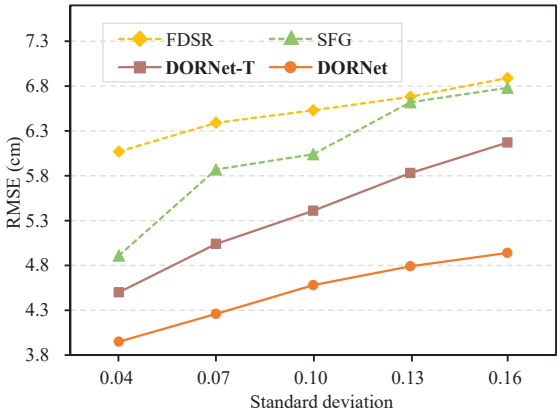

Figure 1. Robustness of adding noise at varying levels before pre-upsampling the LR depth on real-world RGB-D-D.

Methods	DKN	FDSR	DCTNet	SFG	**DORNet**
RMSE	8.01	8.06	8.10	7.83	**7.58**

Table 2. Comparison of complex lighting on RGB-D-D-Lights.

Methods	GT Depth	**HR Depth (Ours)**
RGB-D-D	3.51	**3.42**
TOFDSR	4.26	**4.21**

Table 3. Ablation study on generating degradation representations using GT depth instead of the predicted HR depth.

1. Metrics

The root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) are defined as:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{\mathbb{N}}\sum |D_{gt} - D_{hr}|^2}, \\ \text{MSE} &= \frac{1}{\mathbb{N}}\sum |D_{gt} - D_{hr}|^2, \\ \text{MAE} &= \frac{1}{\mathbb{N}}\sum |D_{gt} - D_{hr}|, \end{aligned} \tag{1}$$

where D_{hr} and D_{gt} represent the predicted HR depth and ground-truth depth, respectively. $\mathbb{N}$ is the pixel set of D_{gt}.

2. Additional Experiments

More Evaluation Metrics. Apart from the RMSE, we also employ MSE and MAE as additional metrics to evaluate our method on the real-world TOFDSR [18] dataset. Tab. 1 lists the quantitative comparison between our DORNet and state-of-the-art approaches, including DJFR [9], DKN [7], FDKN [7], FDSR [5], DCTNet [24], SUFT [12], SFG [22], and SGNet [15]. It is evident that our DORNet achieves the lowest MSE and MAE. For example, our method outperforms SGNet [15] by $0.31cm$ in MAE and $0.91cm$ in MSE (w/o Noisy), while also reducing the parameters by $5.92M$. Overall, these experimental results further demonstrate that our method can effectively recover accurate HR depth.

Comparison in Complex Lighting Scenes. Depth acquisition in real-world scenes is often affected by varying illumination conditions, leading to significant degradation. To evaluate generalization capabilities in complex lighting environments, we first select 'lights' category from the real-world RGB-D-D dataset as a new test set, RGB-D-D-Lights, which includes 430 pairs of RGB-D. Then, we directly test it using the pre-trained weights on the real-world RGB-D-D dataset that excludes the 'lights' category, without any fine-tuning. Tab. 2 shows that our method demonstrates outstanding generalization. For example, our DORNet outperforms the second-best SFG [22] by $0.25cm$ in RMSE on the RGB-D-D-Lights dataset.

Ablation Study of Degradation Representation Generation. Tab. 3 presents a comparison of degradation representation generation employing GT depth and predicted HR depth on the RGB-D-D and TOFDSR. We can observe that the predicted HR depth contributes to better performance, mainly attributed to its capability to facilitate joint optimization of degradation regularization and degradation-oriented fusion. Furthermore, the predicted HR depth is typically dense, enabling accurate estimation of degradation representations. In contrast, the GT depth captured from real-

RMSE	DJF [8]	DJFR [9]	CUNet [4]	DKN [7]	FDKN [7]	FDSR [5]	DCTNet [24]	SUFT [12]	SFG [22]	SGNet [15]	**DORNet-T**	**DORNet**
RGB-D-D	7.94	7.50	6.69	6.50	6.66	6.39	6.04	5.53	5.87	5.44	5.04	**4.26**
TOFDSR	11.45	10.92	9.76	7.42	8.13	6.31	7.52	5.08	5.46	5.11	5.07	**4.61**

Table 4. Quantitative comparison of adding noise before LR depth pre-upsampling on the real-world RGB-D-D and TOFDSR datasets.

Methods	WorldView II				GaoFen2			
	PSNR↑	SSIM↑	SAM [23]↓	ERGAS [1]↓	PSNR↑	SSIM↑	SAM [23]↓	ERGAS [1]↓
PanNet [20]	40.8176	0.9624	0.0257	1.0557	43.0659	0.9685	0.0178	0.8577
SRPPNN [2]	41.4538	0.9679	0.0233	0.9899	47.1998	0.9877	0.0106	0.5586
GPPNN [16]	41.1622	0.9684	0.0244	1.0315	44.2145	0.9815	0.0137	0.7361
MutInf [25]	41.6773	0.9705	0.0224	0.9519	47.3042	**0.9892**	**0.0102**	0.5481
PanFlow [19]	41.8584	0.9712	0.0224	0.9335	47.2533	0.9884	0.0103	0.5512
DORNet	**42.0698**	**0.9723**	**0.0215**	**0.9090**	**47.8940**	0.9890	0.0104	**0.5172**

Table 5. Quantitative comparison of Pan-Sharpening on the WorldView II and GaoFen2 datasets.

Methods	CSPN [3]	GuideNet [13]	PENet [6]	NLSPN [11]	RigNet [17]	GraphCSPN [10]	PointDC [21]	TPVD [18]	**DORNet**
Params. (M) ↓	17.4	73.5	131.5	25.8	65.2	26.4	25.1	31.2	**3.05**
RMSE ↓	0.224	0.146	0.241	0.174	0.133	0.253	0.109	0.092	**0.088**
REL ↓	0.042	0.030	0.043	0.029	0.025	0.052	0.021	**0.014**	**0.014**
δ_1 ↑	94.5	97.6	94.6	96.4	97.6	92.0	98.5	**99.1**	**99.1**
δ_2 ↑	95.3	98.9	95.3	97.9	99.1	96.9	99.2	**99.6**	**99.6**
δ_3 ↑	96.5	99.5	95.5	98.9	99.7	98.7	99.6	**99.9**	99.8

Table 6. Quantitative comparison of depth completion on the TOFDC dataset. The unit of RMSE is m.

world scenarios is often incomplete, resulting in ambiguous degradation modeling. Therefore, we leverage the predicted HR depth to model the degradation representations.

Results of Adding Noise Before Pre-Upsampling. To more accurately simulate real-world scenarios, we conduct additional noise robustness evaluations by adding Gaussian noise and Gaussian blur to the original LR depth (without pre-upsampling) as the new input.

Tab. 4 lists the results of adding fixed Gaussian noise (mean 0, standard deviation 0.07) and Gaussian blur (standard deviation 0.36). It can be seen that our method achieves superior noise robustness. For example, compared to the second-best method, our DORNet decreases the RMSE by $1.18cm$ on the real-world RGB-D-D dataset and by $0.47cm$ on the real-world TOFDSR dataset.

Furthermore, Fig. 1 presents a comparison across different noise levels, where the standard deviation of Gaussian noise ranges from 0.04 to 0.16, and the standard deviation of Gaussian blur is fixed at 0.36. Obviously, our method achieves excellent performance across all levels. Compared to SFG [22], DORNet reduces RMSE by 27.6% (standard deviation 0.13) and 27.1% (standard deviation 0.16).

3. Generalization on Other Restoration Tasks

To further verify the generalizability of our method in other multi-modal restoration tasks, we conduct extensive experiments on pan-sharpening and depth completion tasks, where our DORNet remains unchanged.

3.1. Pan-Sharpening

Tab. 5 demonstrates that our method achieves outstanding performance on the pan-sharpening task. Specifically, we compared DORNet with previous state-of-the-art pan-sharpening methods on the WorldView II and GaoFen2 datasets, including PanNet [20], SRPPNN [2], GPPNN [16], MutInf [25], and PanFlow [19]. Following previous approaches [2, 19, 25], due to the ground-truth is not available, we employ the Wald protocol [14] tool to generate synthetic data. Besides, PSNR, SSIM, SAM [23], and ERGAS [1] are employed as evaluation metrics.

From Tab. 5, we can clearly see that DORNet achieves the comparable performance across all four metrics on the WorldView II and GaoFen2 datasets. For example, compared to the second-best method, our DORNet increases PSNR by $0.2114dB$ on the WorldView II dataset and by $0.5898dB$ on the GaoFen2 dataset.

3.2. Depth Completion

We compare DORNet with previous state-of-the-art depth completion methods on the real-world TOFDC [18] dataset, including CSPN [3], GuideNet [13], PENet [6], NLSPN [11], RigNet [17], GraphCSPN [10], PointDC [21], and TPVD [18]. In this experiment, we maintain the same settings as TPVD [18]. REL, δ_j $(j = 1, 2, 3)$, and RMSE (in

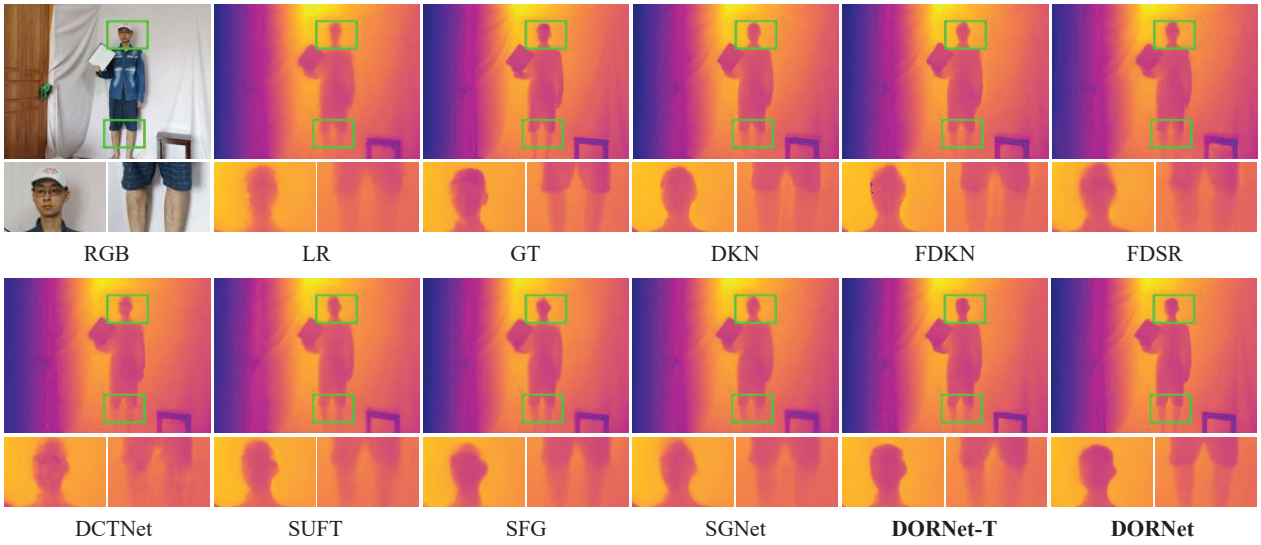

Figure 2. Visual results on the real-world RGB-D-D dataset (w/o Noise).

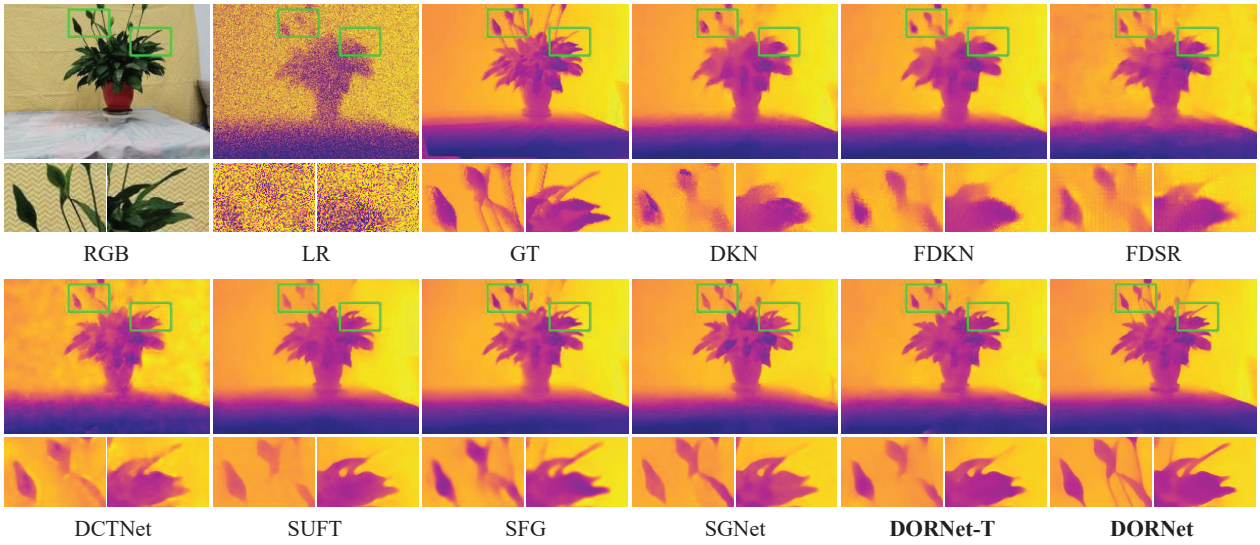

Figure 3. Visual results on the real-world RGB-D-D dataset (w/ Noise).

meters) are selected as evaluation metrics. Tab. 6 demonstrates that our method achieves satisfactory performance across all four metrics on the real-world TOFDC dataset. For example, compared to the suboptimal TPVD [18], our DORNet significantly decreases the parameters by 90% while still achieving a 4% reduction in RMSE.

4. More Visualizations

Figs. 2-5 provide more visual comparison results on both the real-world RGB-D-D [5] and TOFDSR [18] datasets. We can observe that our DORNet recovers clearer depth. For example, the person's head recovered by our method in Fig. 2 is more distinct than other approaches. Furthermore, Fig. 3 presents the visual results on the real-world RGB-D-D dataset with additional noise. It is evident that DORNet successfully removes noise and restores more accurate depth structure than other methods in noisy environments.

Fig. 6 further shows the visual results ($\times 16$) on the synthetic NYU-v2 dataset, demonstrating that our DORNet achieves superior performance in reconstructing precise geometric structure. For instance, the edges of the window predicted by our method in Fig. 6 exhibit closer alignment with the GT depth, showing fewer error.

In summary, these visual comparisons demonstrate that our method effectively reconstructs sharp and accurate structural details, achieving satisfactory DSR performance.

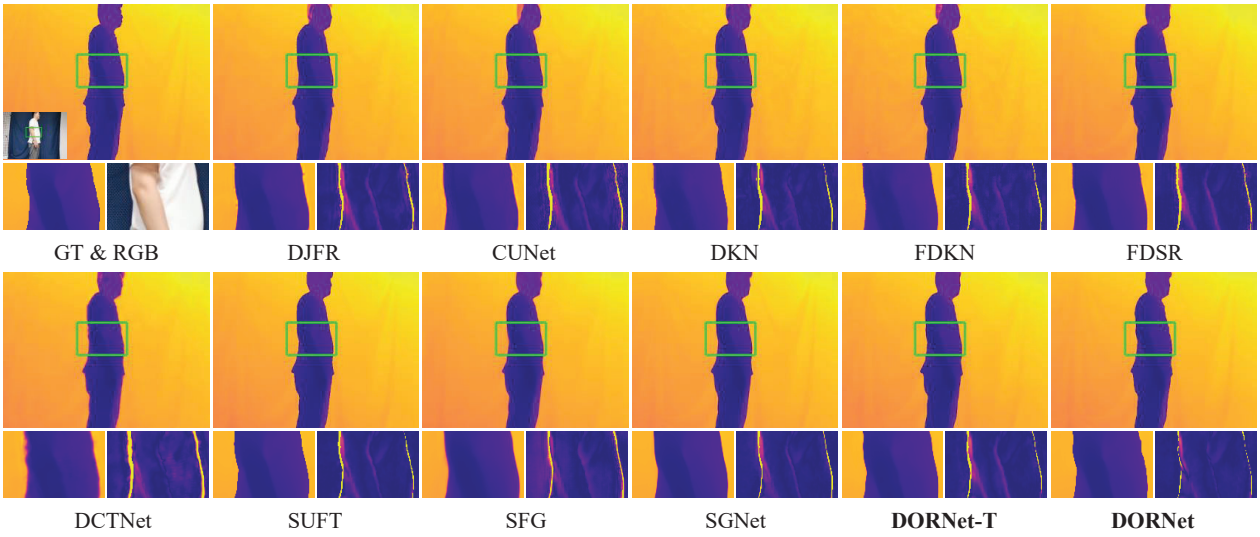

Figure 4. Visual results (left) and error maps (right) on the real-world TOFDSR dataset (w/o Noise).

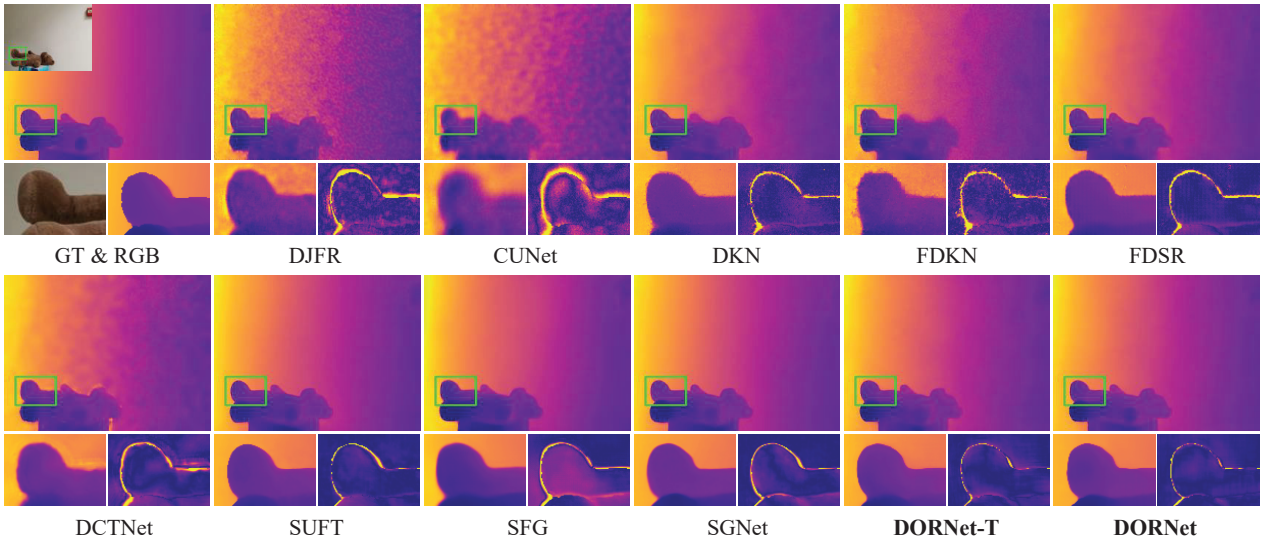

Figure 5. Visual results (left) and error maps (right) on the real-world TOFDSR dataset (w/ Noise).

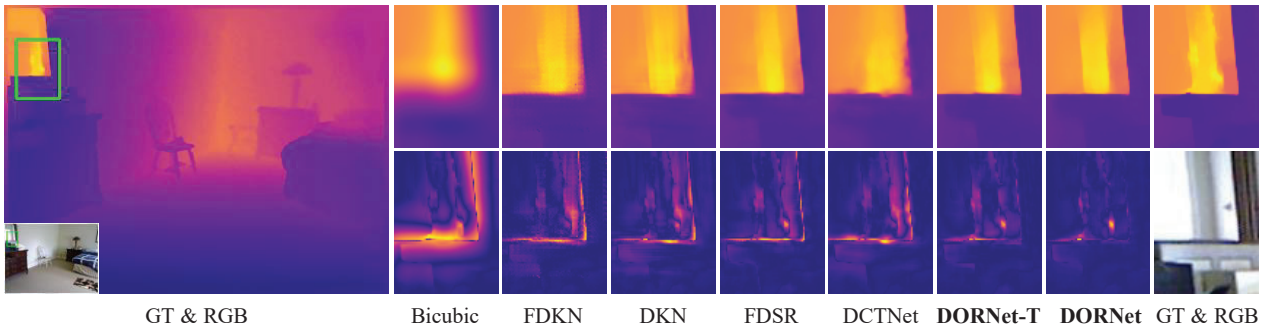

Figure 6. Visual results (top) and error maps (bottom) on the synthetic NYU-v2 dataset ($\times 16$).

References

[1] Luciano Alparone, Lucien Wald, Jocelyn Chanussot, Claire Thomas, Paolo Gamba, and Lori Mann Bruce. Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, 2007. 2

[2] Jiajun Cai and Bo Huang. Super-resolution-guided progressive pansharpening based on a deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):5206–5220, 2020. 2

[3] Xinjing Cheng, Peng Wang, and Ruigang Yang. Learning depth with convolutional spatial propagation network. In *ECCV*, pages 103–119, 2018. 2

[4] Xin Deng and Pier Luigi Dragotti. Deep convolutional neural network for multi-modal image restoration and fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3333–3348, 2020. 2

[5] Lingzhi He, Hongguang Zhu, Feng Li, Huihui Bai, Runmin Cong, Chunjie Zhang, Chunyu Lin, Meiqin Liu, and Yao Zhao. Towards fast and accurate real-world depth super-resolution: Benchmark dataset and baseline. In *CVPR*, pages 9229–9238, 2021. 1, 2, 3

[6] Mu Hu, Shuling Wang, Bin Li, Shiyu Ning, Li Fan, and Xiaojin Gong. Penet: Towards precise and efficient image guided depth completion. In *ICRA*, 2021. 2

[7] Beomjun Kim, Jean Ponce, and Bumsub Ham. Deformable kernel networks for joint image filtering. *International Journal of Computer Vision*, 129(2):579–600, 2021. 1, 2

[8] Yijun Li, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep joint image filtering. In *ECCV*, pages 154–169, 2016. 2

[9] Yijun Li, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Joint image filtering with deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1909–1923, 2019. 1, 2

[10] Xin Liu, Xiaofei Shao, Bo Wang, Yali Li, and Shengjin Wang. Graphcspn: Geometry-aware depth completion via dynamic gcns. In *ECCV*, pages 90–107, 2022. 2

[11] Jinsun Park, Kyungdon Joo, Zhe Hu, Chi-Kuei Liu, and In So Kweon. Non-local spatial propagation network for depth completion. In *ECCV*, 2020. 2

[12] Wuxuan Shi, Mang Ye, and Bo Du. Symmetric uncertainty-aware feature transmission for depth super-resolution. In *ACMMM*, pages 3867–3876, 2022. 1, 2

[13] Jie Tang, Fei-Peng Tian, Wei Feng, Jian Li, and Ping Tan. Learning guided convolutional network for depth completion. *IEEE Transactions on Image Processing*, 30:1116–1129, 2020. 2

[14] Lucien Wald, Thierry Ranchin, and Marc Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 63(6):691–699, 1997. 2

[15] Zhengxue Wang, Zhiqiang Yan, and Jian Yang. Sgnet: Structure guided network via gradient-frequency awareness for depth map super-resolution. In *AAAI*, pages 5823–5831, 2024. 1, 2

[16] Shuang Xu, Jiangshe Zhang, Zixiang Zhao, Kai Sun, Junmin Liu, and Chunxia Zhang. Deep gradient projection networks for pan-sharpening. In *CVPR*, pages 1366–1375, 2021. 2

[17] Zhiqiang Yan, Kun Wang, Xiang Li, Zhenyu Zhang, Jun Li, and Jian Yang. Rignet: Repetitive image guided network for depth completion. In *ECCV*, pages 214–230, 2022. 2

[18] Zhiqiang Yan, Yuankai Lin, Kun Wang, Yupeng Zheng, Yufei Wang, Zhenyu Zhang, Jun Li, and Jian Yang. Tri-perspective view decomposition for geometry-aware depth completion. In *CVPR*, pages 4874–4884, 2024. 1, 2, 3

[19] Gang Yang, Xiangyong Cao, Wenzhe Xiao, Man Zhou, Aiping Liu, Xun Chen, and Deyu Meng. Panflownet: A flow-based deep network for pan-sharpening. In *ICCV*, pages 16857–16867, 2023. 2

[20] Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. Pannet: A deep network architecture for pan-sharpening. In *ICCV*, pages 5449–5457, 2017. 2

[21] Zhu Yu, Zehua Sheng, Zili Zhou, Lun Luo, Si-Yuan Cao, Hong Gu, Huaqi Zhang, and Hui-Liang Shen. Aggregating feature point cloud for depth completion. In *ICCV*, pages 8732–8743, 2023. 2

[22] Jiayi Yuan, Haobo Jiang, Xiang Li, Jianjun Qian, Jun Li, and Jian Yang. Structure flow-guided network for real depth super-resolution. In *AAAI*, pages 3340–3348, 2023. 1, 2

[23] Roberta H Yuhas, Alexander FH Goetz, and Joe W Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*, 1992. 2

[24] Zixiang Zhao, Jiangshe Zhang, Shuang Xu, Zudi Lin, and Hanspeter Pfister. Discrete cosine transform network for guided depth map super-resolution. In *CVPR*, pages 5697–5707, 2022. 1, 2

[25] Man Zhou, Keyu Yan, Jie Huang, Zihe Yang, Xueyang Fu, and Feng Zhao. Mutual information-driven pan-sharpening. In *CVPR*, pages 1798–1808, 2022. 2