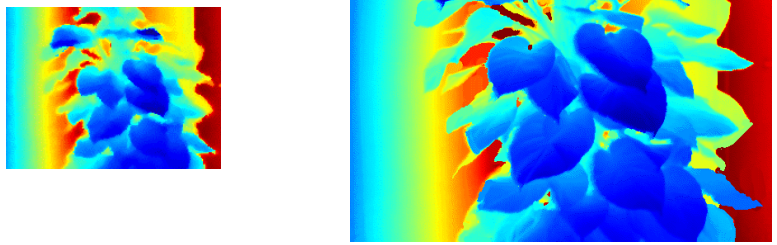
Towards Fast and Accurate Real-World Depth Super-Resolution: Benchmark Dataset and Baseline

Lingzhi He, Hongguang Zhu, Feng Li, Huihui Bai, Runmin Cong, Chunjie Zhang, Chunyu Lin, Meiqin Liu, Yao Zhao*

CVPR 2021

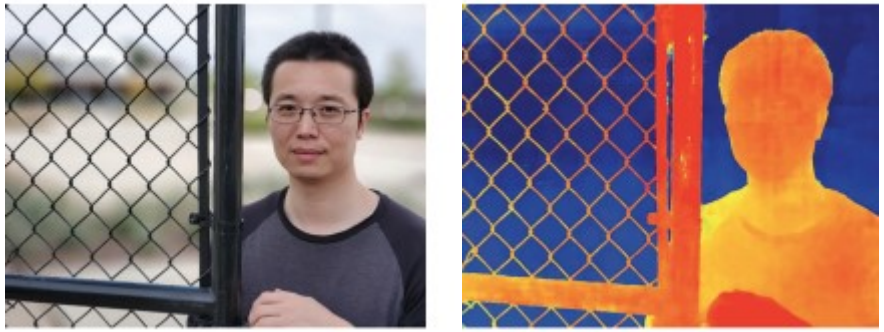
Reporter: Lingzhi He

Introduction



Depth Map Super-Resoluition

Face Recognition

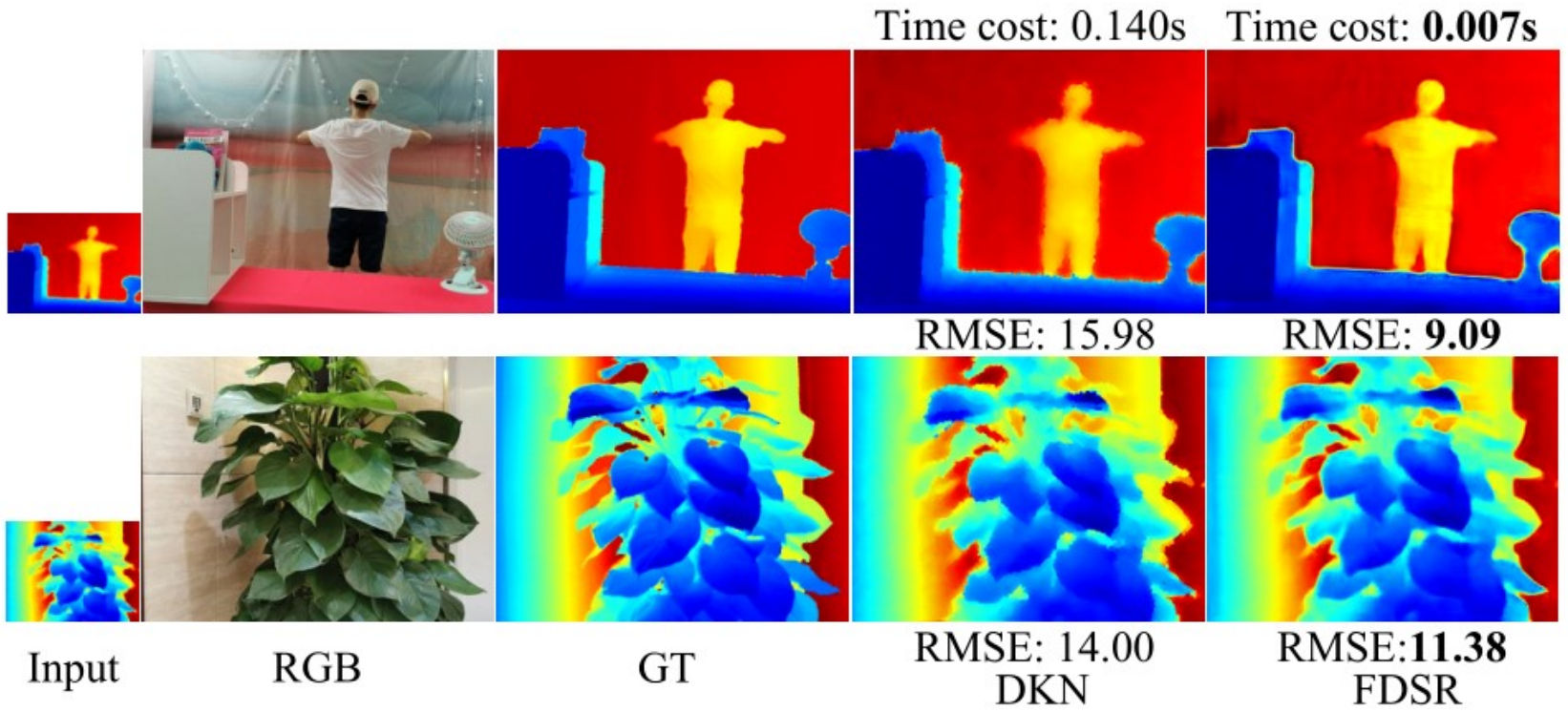


Bokeh Rendering

Gesture Recognition

- As a supplement of the RGB modality, the depth map can provide useful depth information, which has been applied in bokeh rendering, face recognition, gesture recognition, etc.
- The resolution of depth maps cannot match the resolution of RGB images, limiting practical applications to some extent. Depth map super-resolution (SR) is an effective solution.

Motivations



- Limited by the lack of real-world paired LR and HR depth maps, most existing depth map SR methods use down-sampling to obtain paired training samples. But the down-sampling manner fails to comprehensively simulate the real-world complex correspondences between the LR and HR depth maps.
- The sharp boundaries and elaborate details in the depth map SR are hard to recover especially when the scaling factor is large. How to design a fast and accurate depth map SR model to generate HR depth maps.

Contributions

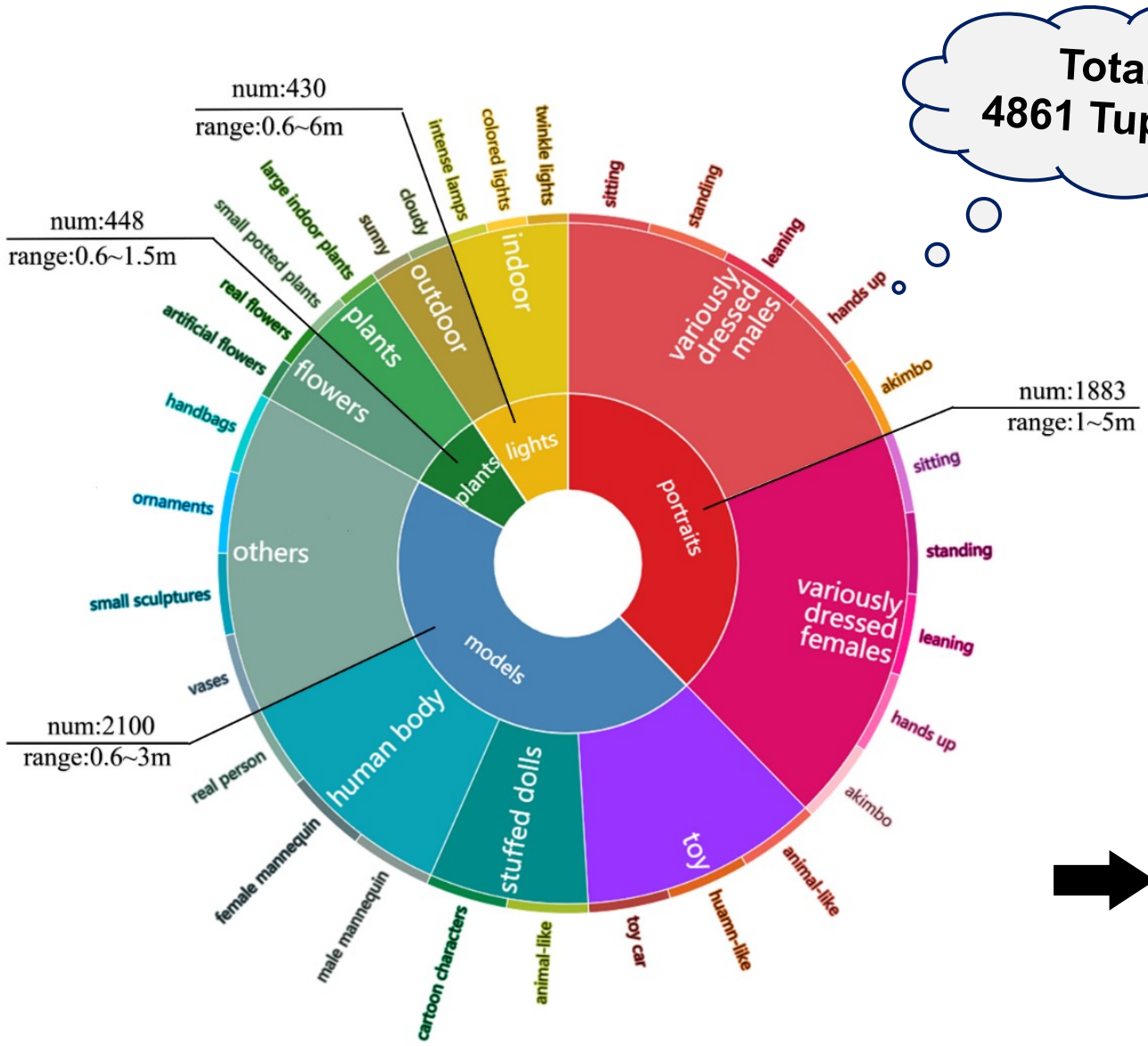
a) We build the first and large-scale depth map SR benchmark dataset named RGB-D-D dataset, towards the real scenes and real correspondences. This dataset bridges the gap between theoretical research and real-world applications, and also flourishes the depth related tasks in terms of benchmark dataset.

b) We design a fast depth map super-resolution (FDSR) baseline, in which a high-frequency guided multi-scale structure is introduced to provide the frequency guidance and exploit the contextual information. Such decomposition strategy can improve the efficiency while retaining the reconstruction performance.

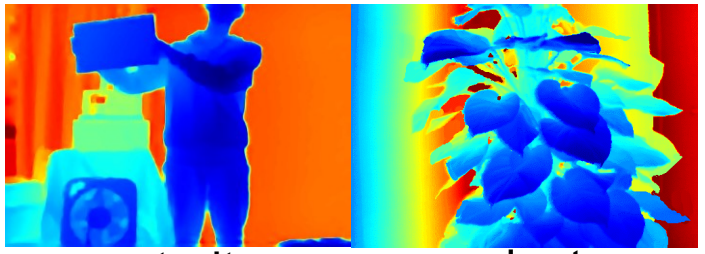
c) Our network achieves the superior performance on the public datasets and our RGB-D-D benchmark dataset in terms of the speed and accuracy. Moreover, for the real-world depth map SR task, our algorithm can generate more accurate results with clearer boundaries and to some extent correct the value errors.

http://mepro.bjtu.edu.cn/resource.html
https://github.com/lingzhi96/RGB-D-D-Dataset

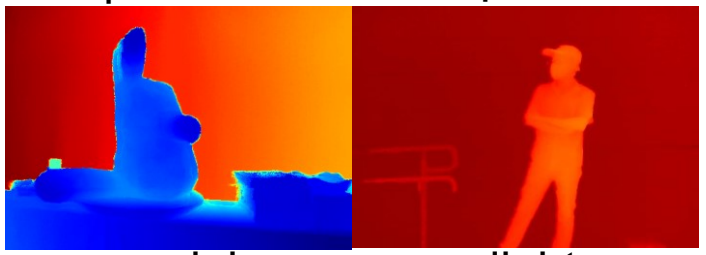
RGB-D-D Dataset



Total:
4861 Tuples



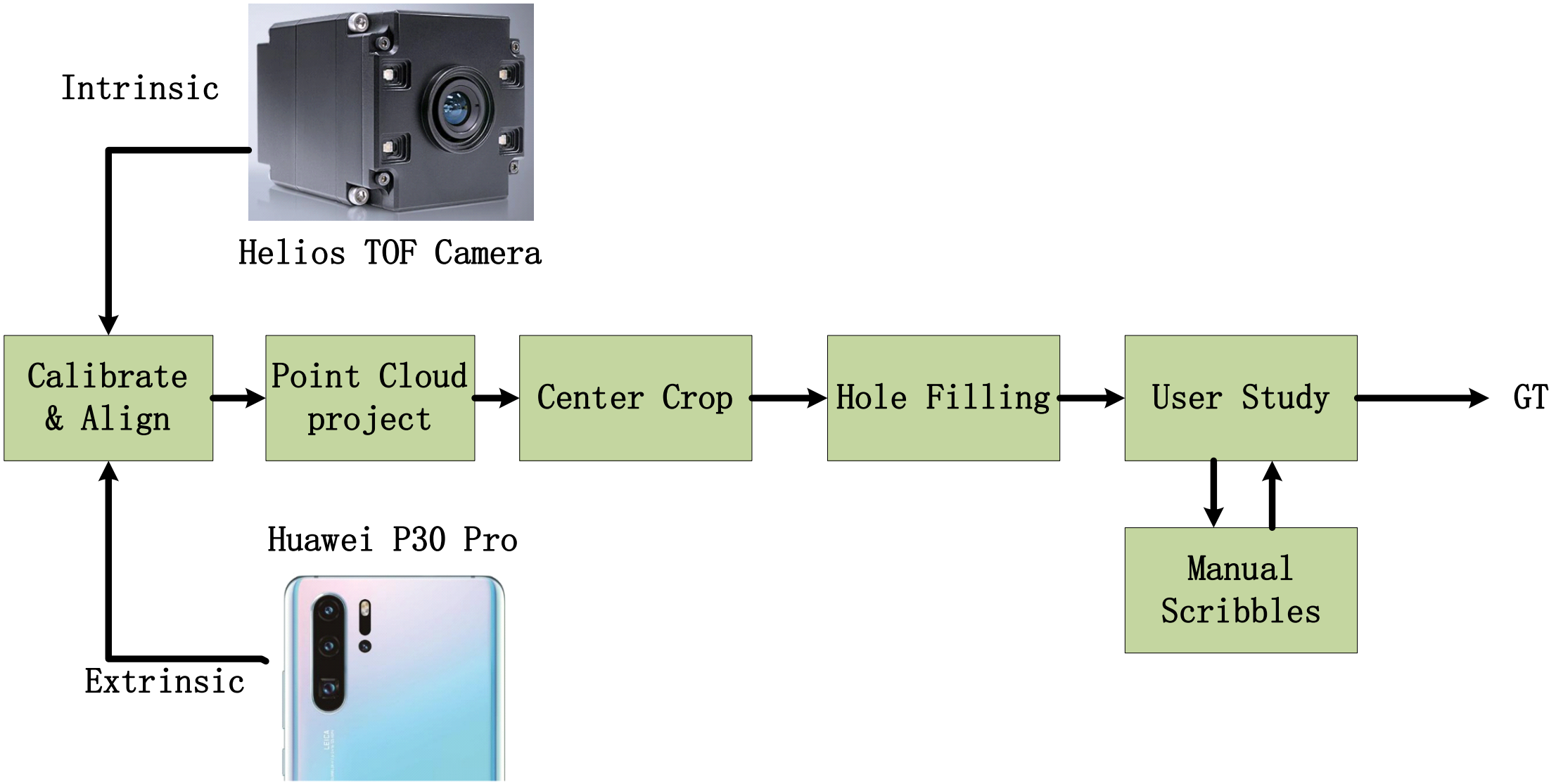
potraits plants



models lights

- Bokeh Rendering
- Optimize the edge of objects
- Low-quality depth map
- Effect of complex illumination
- ……

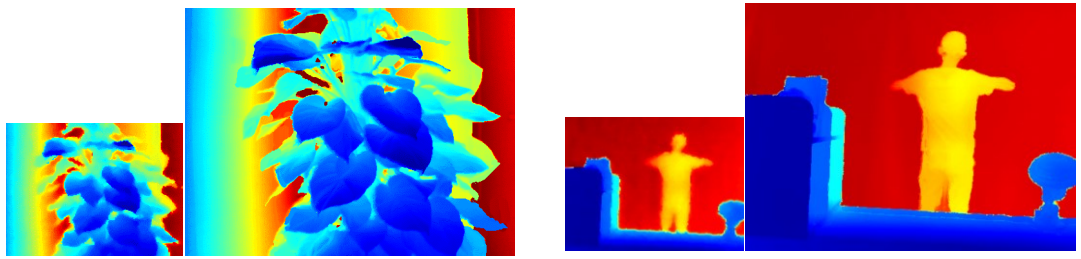
Dataset Processing



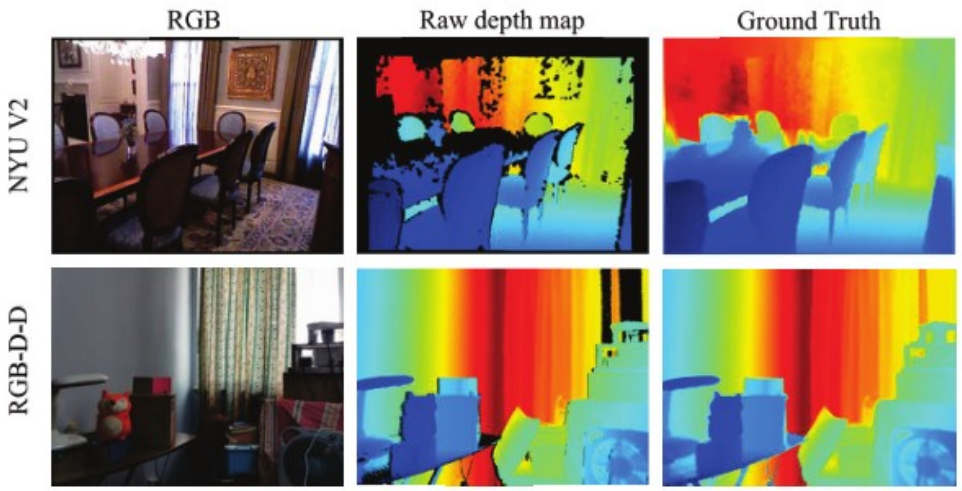
Dataset Statistic

- Real Scenes

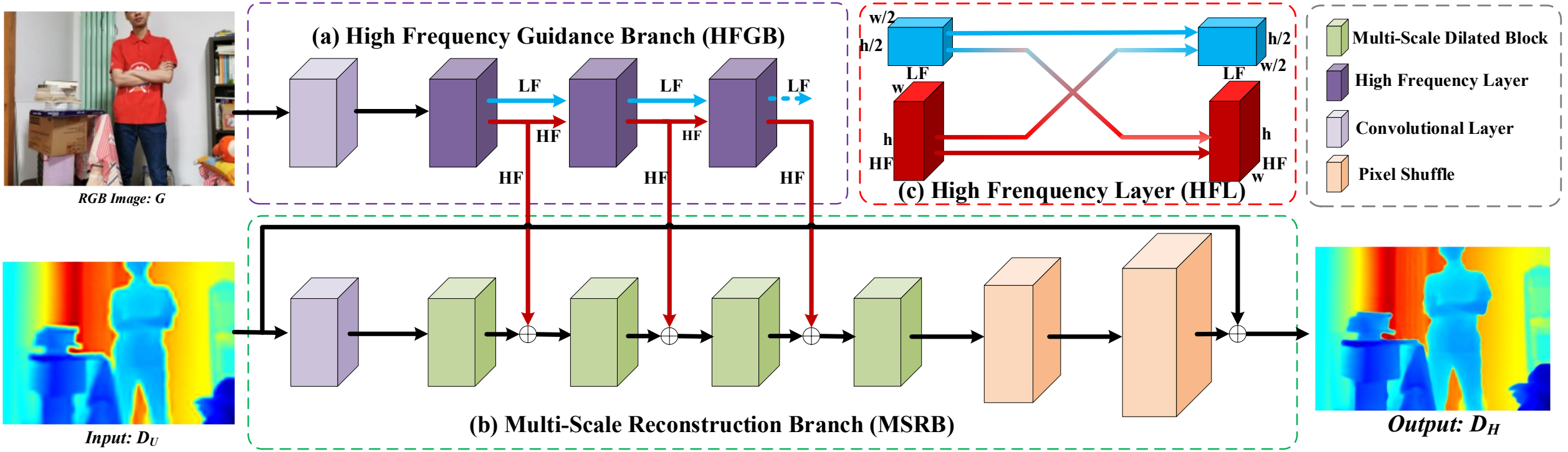
- Real Correspondences



- High Quality

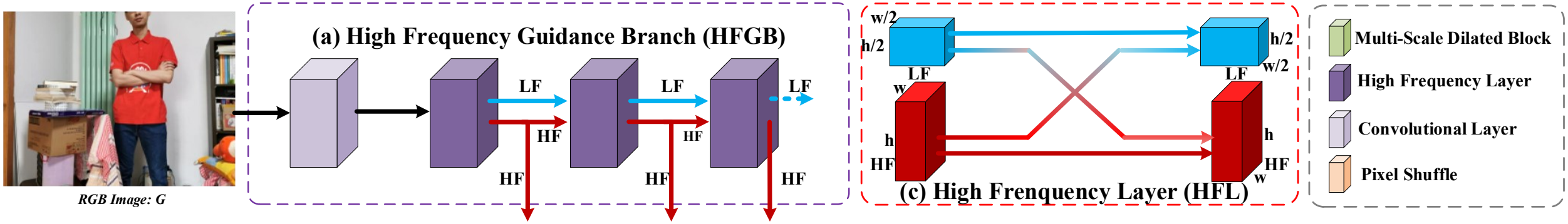


Our Method



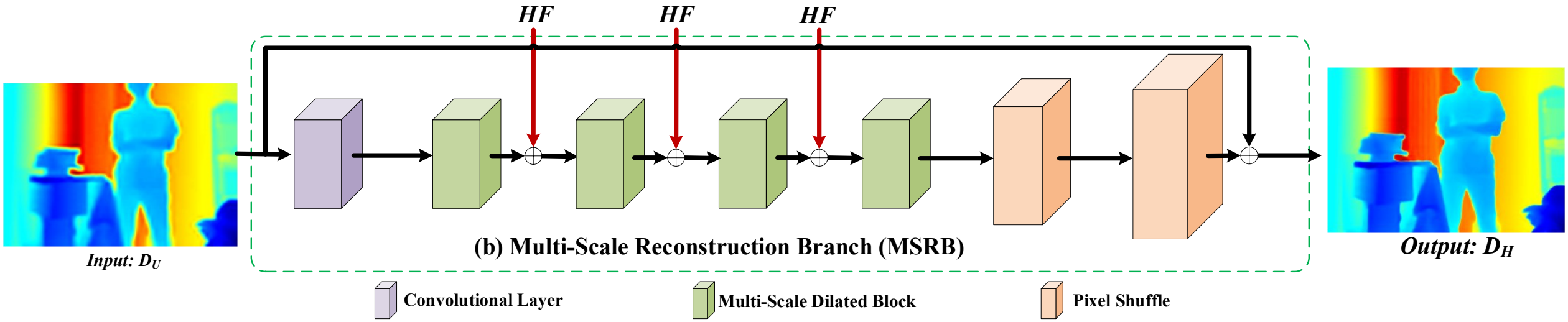
- The architecture includes a high-frequency guidance branch (HFGB) and a multiscale reconstruction branch (MSRB).
- Our framework progressively equip with four multi-scale reconstruction blocks to exploit the contextual information under different receptive fields in MSRB, meanwhile, the high-frequency guidance extracted from the HFGB is integrated with the multiscale contextual information to enhance the ability of detail recovery for depth map SR.

High-Frequency Guidance Branch



- A direct high-frequency decomposition method is designed, where the octave convolution is utilized to decompose the RGB features into high- and low-frequency components.
- The high-frequency components are effectively used to guide depth map SR. Such design focuses on the useful high-frequency detail information to improve the performance, while it reduces the computation complexity due to the low-frequency components are not used in the MSRB.

Multi-Scale Reconstruction Branch



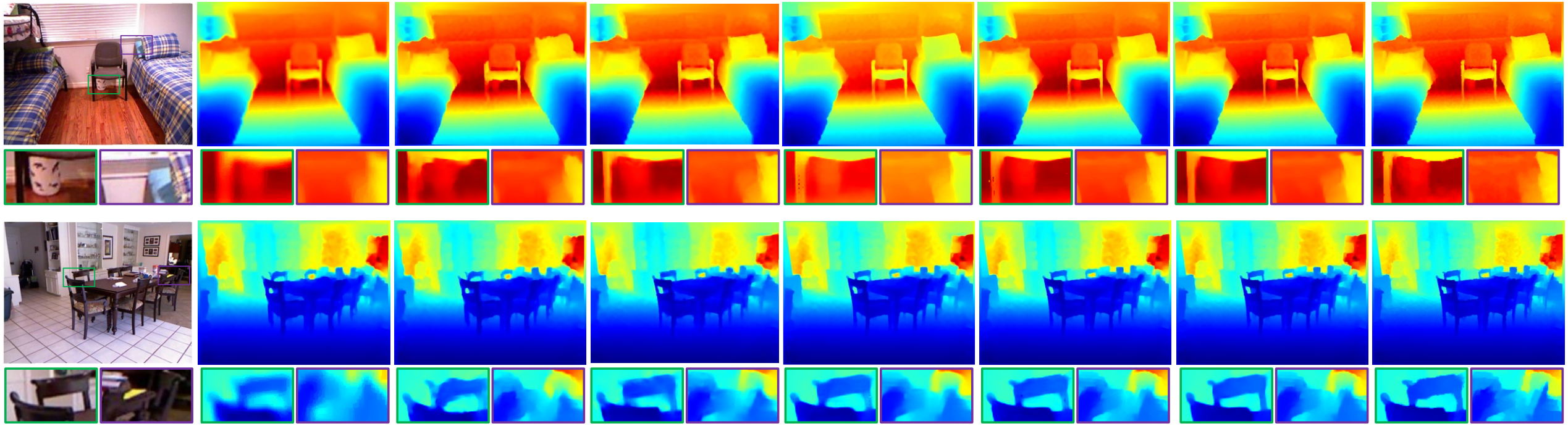
- This branch aims to progressively recover HR depth map through utilizing mulit-scale contextual information. We first use one convolution layer to initial feature extraction. Then, to exploit the contextual information under different receptive fields, we combine dilated convolutions with different dilated rates to form a multi-scale dilated block (MSDB), and one convolution layer is used to integrate the concatenated features.
- As for feature combination, three levels of high-frequency features extracted by HLFs are fused with different MSDBs respectively in the early stage of MSRB.

Experiments

- Benchmark Datasets: NYU v2 (1449 RGB-D images), RGB-D-D (4861 RGB-D images).
- We sample 1000 RGB-D image pairs from the NYU v2 dataset for training and the rest 449 image pairs for testing. As for RGB-D-D dataset, we randomly split 1586 portraits, 380 plants, 249 models for training and 297 portraits, 68 plants, 40 models for testing.
- Evaluation Metrics:

RMSE:	pixel wise depth map SR accuracy
Depth Value Errors:	global confidence depth map SR accuracy
Depth Edge Errors:	edge area depth map SR accuracy

Experiments on NYU v2



(a) RGB	(b) SDF	(c) SVLRM	(d) DJFR	(e) FDKN	(f) DKN	(g) FDSR	(h) GT
Running Time (s)	25 (CPU)	0.10 (GPU)	0.01 (GPU)	0.01 (GPU)	0.17 (GPU)	0.01 (GPU)	

Experiments on NYU v2

RMSE	Bicubic	MRF [7]	GF [12]	JBU [18]	TGV [8]	Park [31]	SDF [22]	FBS [4]	DMSG [14]	PAC [38]	DJF [22]	DJFR [23]	DKN [16]	FDKN [16]	FDSR
×4	8.16	7.84	7.32	4.07	4.98	5.21	5.27	4.29	3.02	2.39	3.54	3.38	1.62	1.86	**1.61**
×8	14.22	13.98	13.62	8.29	11.23	9.56	12.31	8.94	5.38	4.59	6.2	5.86	3.26	3.58	**3.18**
×16	22.32	22.2	22.03	13.35	28.13	18.1	19.24	14.59	9.17	8.09	10.21	10.11	6.51	6.96	**5.86**

Table 1. Comparisons with the state-of-the-art methods in terms of RMSE on NYU v2 [28]. The depth values are measured in centimeter.

Percentage	Value Errors (in 10 m)			Edge Errors		
	×4	×8	×16	×4	×8	×16
SDF [22]	0.42	1.28	3.52	4.20	10.19	25.06
SVLRM [30]	1.08	2.56	5.76	6.04	24.28	49.26
DJF [22]	1.05	2.74	6.25	9.87	30.38	55.35
DJFR [23]	1.04	2.72	6.25	6.78	25.01	53.98
FDKN [16]	0.04	0.24	1.00	0.83	3.27	13.03
DKN [16]	0.05	0.20	1.10	0.95	2.95	13.78
FDSR	**0.04**	**0.18**	**0.69**	**0.78**	**2.60**	**9.44**

Table 2. Value errors and edge errors on NYU v2 [28].

Experiments

RMSE	SDF [22]	SVLRM [30]	DJF [22]	DJFR [23]	FDKN [16]	DKN [16]	FDSR	FDSR$^+$
×4	2.00	3.39	3.41	3.35	1.18	1.30	1.16	**1.11**
×8	3.23	5.59	5.57	5.57	1.91	1.96	1.82	**1.71**
×16	5.16	8.28	8.15	7.99	3.41	3.42	3.06	**3.01**

Table 3. Quantitative depth map SR results on RGB-D-D. FDSR$^+$ is trained in downsampling manner on RGB-D-D)

Percentage	Value Errors (in 3 m)			Edge Errors		
	×4	×8	×16	×4	×8	×16
SDF [22]	0.33	0.90	2.37	3.22	8.74	20.71
SVLRM [30]	0.80	2.11	4.58	5.08	15.18	34.30
DJF [22]	0.82	2.19	4.89	5.65	17.07	35.32
DJFR [23]	0.79	2.15	4.78	5.26	15.66	34.54
FDKN [16]	0.11	0.28	0.94	1.39	3.41	11.73
DKN [16]	0.14	0.33	1.54	2.11	3.55	12.93
FDSR	0.10	0.26	0.76	1.38	3.09	12.47
FDSR$^+$	**0.09**	**0.21**	**0.67**	**1.15**	**2.79**	**11.68**

Table 5. Value errors and edge errors of depth SR results on RGB-D-D. FDSR$^+$ is trained in downsampling training manner.

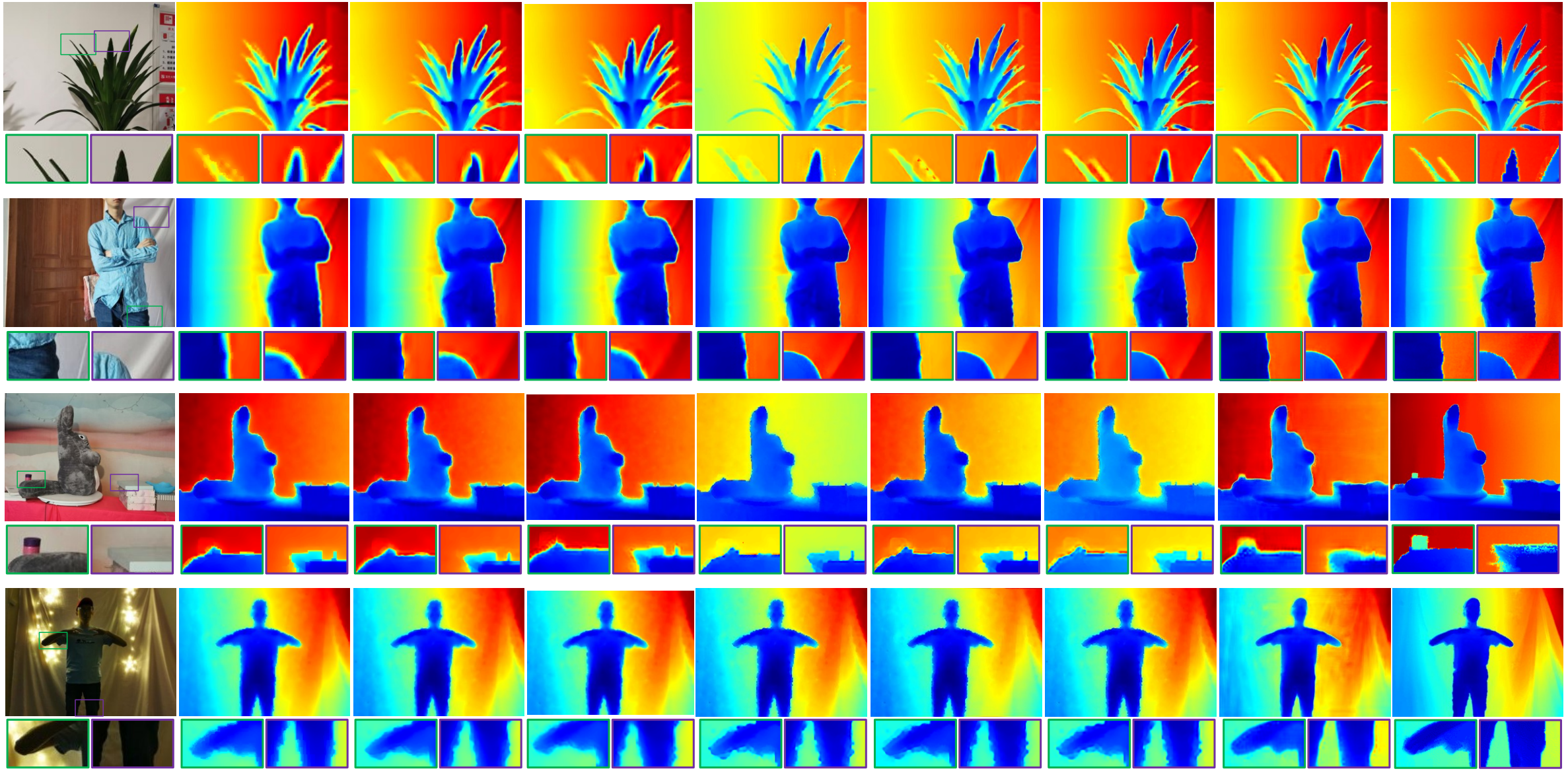
Methods	NYU v2 [28]			RGB-D-D		
	×4	×8	×16	×4	×8	×16
w/o HFGB	2.02	3.90	7.58	1.16	1.88	3.47
w/o HFL	1.68	3.21	5.89	1.13	1.85	3.20
FDSR	**1.61**	**3.18**	**5.86**	**1.11**	**1.71**	**3.01**

Table 6. RMSE evaluation of HFL and HFGB.

	SDF [22]	SVLRM [30]	DJF [22]	DJFR [23]	FDKN [16]	DKN [16]	FDSR	FDSR^{++}
RMSE	7.16	8.05	7.90	8.01	7.50	7.38	7.50	**5.49**
Value Errors	2.86	3.62	3.62	3.67	2.85	2.83	2.90	**1.71**
Edge Errors	52.78	51.87	50.56	52.28	51.73	51.90	51.89	**42.89**

Table 4. RMSE, value errors and edge errors of depth SR results. FDSR^{++} is trained on RGB-D-D in real-world training manner.

Experiments on RGB-D-D



(a) RGB (b) SDF (c) SVLRM (d) DJFR (e) FDKN (f) DKN (g) FDSR (h) FDSR^{+} / FDSR^{++} (i) GT

Conclusion

- We build the first benchmark dataset which satisfy both real scene and real correspondence. The dataset contains paired LR and HR depth maps in multiple scenarios, and contributes the completely new benchmark dataset for real-world depth map SR research.
- Furthermore, the “RGB-D-D” triples not only can complete the traditional depth-related tasks, such as depth estimation, depth completion, etc. but also have significant potential to promote the application of depth maps on portable intelligent electronics.
- We also provide a fast and accurate depth map SR baseline adaptively focusing on the high-frequency components of the guidance and suppress the low-frequency components.
- Our algorithm achieves the competitive performance on public datasets and our proposed dataset, what’s more, it has an ability to cope with the task of real-world depth map SR.

Thanks

Lingzhi He (何凌志)
Beijing Jiaotong University

E-mail: 19112002@bjtu.edu.cn