Supplementary Material for "Spatial Hierarchy Aware Residual Pyramid Network for Time-of-Flight Depth Denoising"

Guanting Dong, Yueyi Zhang, and Zhiwei Xiong

University of Science and Technology of China
gtdong@mail.ustc.edu.cn, {zhyuey, zwxiong}@ustc.edu.cn

1 Introduction

In this supplementary material, we first provide the detailed network architectures of our proposed SHARP-Net in Section 2. We then show more visualizations of experimental results on qualitative comparison in Section 3.

2 Detailed Network Architectures

Our proposed Spatial Hierarchy Aware Residual Pyramid Network (SHARP-Net) consists of three parts: a Residual Regression Module as the backbone for multi-scale feature extraction, a Residual Fusion Module and a Depth Refinement Module to optimize the performance. The details of SHARP-Net are shown in Table 1. The ($\times$) represents the upsample operation based on bicubic interpolation. For example, ($\times 2$) means that interpolating the input image to twice over its original size. The 'all ©' represents concating the upsample of output residuals of all the residual regression blocks. The $\bigoplus$ and the © respectively represent the addition operation and the concatenation operation.

3 Additional Experimental Results on Synthetic Dataset

In this section, we provide more visualizations of experiment results. We first show more error maps in Fig. 1 to compare our SHARP-Net with DeepToF [2] and ToF-KPN [4] on synthetic datasets and realistic datasets, including the ToF-FlyingThings3D (TFT3D) dataset [4], FLAT dataset [1] and True Box dataset [5]. It can be seen that the error of our SHARP-Net is smaller compared with other methods. To further demonstrate the performance of our proposed SHARP-Net in depth denoising, we visualize the depth values along a scan-line for more scenes on the TFT3D dataset in Figure 2, following the experimental settings in Section 5.5. It is obviously observed that our proposed SHARP-Net achieves the best performance on eliminating the MPI noise and the shot noise.

Table 1. The detailed network architecture of our proposed SHARP-Net.

Module Name	Layer Name	Kernel Size	Stride	Input Channels	Output Channels	Input Layer
Residual Regression Block (RRB)	conv1	3×3	1	hold	128	hold
	conv2	3×3	1	128	96	conv1
	conv3	3×3	1	96	64	conv2
	conv4	3×3	1	64	32	conv3
	conv5	3×3	1	32	16	conv4
	conv6	3×3	1	16	1	conv5
Residual Regression Module (RRM)	conv1_1	3×3	1	2	16	depth©amplitude
	conv1_2	3×3	1	16	16	conv1_1
	conv2_1	3×3	2	16	32	conv1_2
	conv2_2	3×3	1	32	32	conv2_1
	conv3_1	3×3	2	32	64	conv2_2
	conv3_2	3×3	1	64	64	conv3_1
	conv4_1	3×3	2	64	96	conv3_2
	conv4_2	3×3	1	96	96	conv4_1
	conv5_1	3×3	2	96	128	conv4_2
	conv5_2	3×3	1	128	128	conv5_1
	conv6_1	3×3	2	128	192	conv5_2
	conv6_2	3×3	1	192	192	conv6_1
	RRB6	3×3	1	192	1	conv6_2
	RRB5	3×3	1	129	1	conv5_2©(RRB6×2)
	RRB4	3×3	1	97	1	conv4_2©(RRB5×2)
	RRB3	3×3	1	65	1	conv3_2©(RRB4×2)
	RRB2	3×3	1	33	1	conv2_2©(RRB3×2)
	RRB1	3×3	1	17	1	conv1_2©(RRB2×2)
Residual Fusion Module (RFM)	RFM	1×1	1	6	1	all © (RRB$i \times 2^{i-1}$) $i \in [1, 6]$
Depth Refinement Module (DRM)	conv1_1	3×3	1	1	16	RFM$\bigoplus$depth
	conv1_2	3×3	1	16	16	conv1_1
	conv2_1	3×3	2	16	32	conv1_2
	conv2_2	3×3	1	32	32	conv2_1
	conv3_1	3×3	2	32	64	conv2_2
	conv3_2	3×3	1	64	64	conv3_1
	conv4_1	3×3	2	64	128	conv3_2
	conv4_2	3×3	1	128	128	conv4_1
	upconv1_1	3×3	2	128	64	conv4_2
	upconv1_2	3×3	1	128	64	conv3_2©upconv1_1
	upconv2_1	3×3	2	64	32	upconv1_2
	upconv2_2	3×3	1	64	32	conv2_2©upconv2_1
	upconv3_1	3×3	2	32	16	upconv2_2
	upconv3_2	3×3	1	32	16	conv1_2©upconv3_1
	w	3×3	1	16	9	upconv3_2

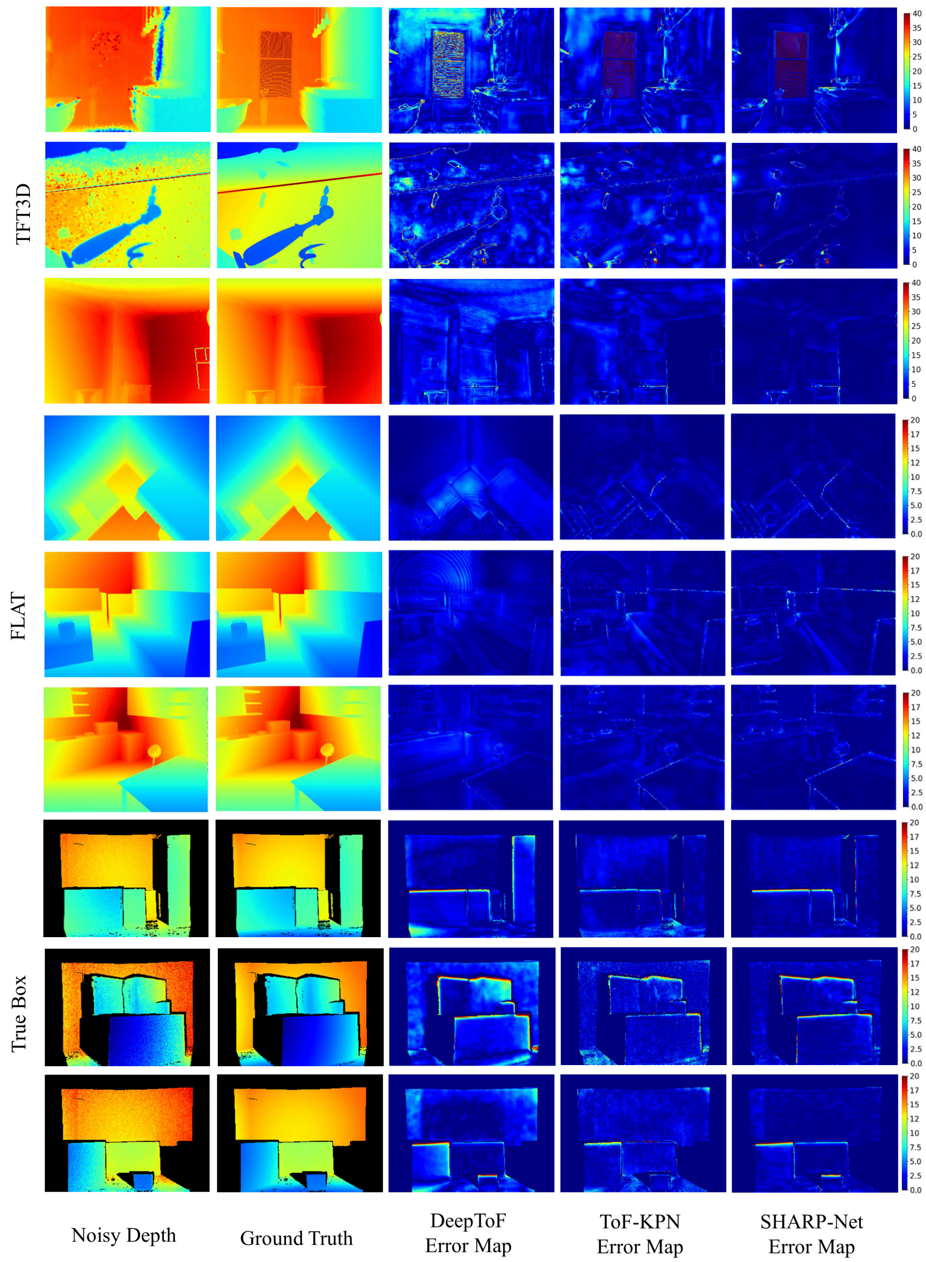
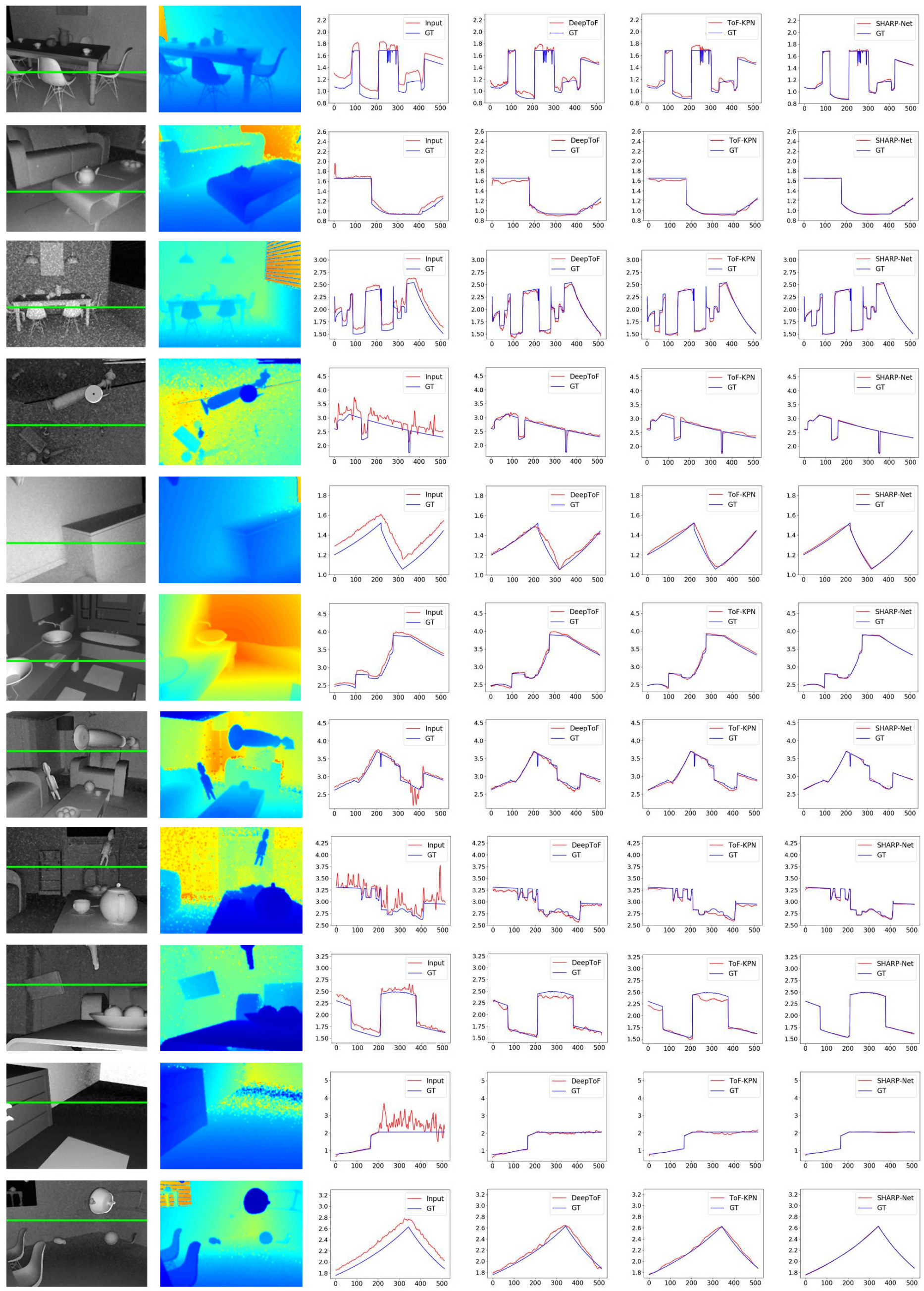


Fig. 1. Qualitative comparison on the TFT3D dataset, the FLAT dataset and the True Box dataset for ToF image denoising. For each dataset, three scenes are selected for comparison. The colorbars in the right show the color scale for error maps with the unit in cm.



ToF Amplitude | Depth Input | Depth Input | DeepToF | ToF-KPN | SHARP-Net

Fig. 2. Quantitative comparison with previous works along a green scan line in a depth image from the TFT3D dataset. 'GT' means the ground truth depth. Our proposed SHARP-Net demonstrates the best performance on depth denoising.

4 Additional Experimental Results on Realistic Dataset

To further evaluate our proposed SHARP-Net, we conduct experiments on another real-world ToF dataset named CoRBS [6], which is captured by Kinect One [3] and contains the ground truth depth for training and testing. We use the same training configuration as the TFT3D dataset. The results are shown in the following Table 2.

Table 2. Quantitative comparison with DeepToF and ToF-KPN on CoRBS dataset.

Model Name	MAE(cm)	Relative Error
DeepToF	1.94	34.3%
ToF-KPN	1.82	32.2%
SHARP-Net	1.18	20.8%

References

1. Guo, Q., Frosio, I., Gallo, O., Zickler, T., Kautz, J.: Tackling 3d tof artifacts through learning and the flat dataset. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 368–383 (2018)
2. Marco, J., Hernandez, Q., Munoz, A., Dong, Y., Jarabo, A., Kim, M.H., Tong, X., Gutierrez, D.: Deeptof: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. ACM Transactions on Graphics (ToG) **36**(6), 219 (2017)
3. Payne, A., Daniel, A., Mehta, A., Thompson, B., Bamji, C.S., Snow, D., Oshima, H., Prather, L., Fenton, M., Kordus, L., et al.: 7.6 a 512× 424 cmos 3d time-of-flight image sensor with multi-frequency photo-demodulation up to 130mhz and 2gs/s adc. In: 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC). pp. 134–135. IEEE (2014)
4. Qiu, D., Pang, J., Sun, W., Yang, C.: Deep end-to-end alignment and refinement for time-of-flight rgb-d module. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 9994–10003 (2019)
5. Su, S., Heide, F., Wetzstein, G., Heidrich, W.: Deep end-to-end time-of-flight imaging. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6383–6392 (2018)
6. Wasenmüller, O., Meyer, M., Stricker, D.: Corbs: Comprehensive rgb-d benchmark for slam using kinect v2. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–7. IEEE (2016)