Cross-Guided Optimization of Radiance Fields with Multi-View Image Super-Resolution for High-Resolution Novel View Synthesis –Supplementary Material–

Youngho Yoon and Kuk-Jin Yoon
Visual Intelligence Lab., KAIST, Korea
{dudgh1732,kjyoon}@kaist.ac.kr

Abstract

Due to the lack of space in the main paper, we provide more details of the proposed methods and experimental results in the supplementary material. Specifically, in Sec.1, we provide more details of voxel-based uncertainty fields. Sec.2 explains more details about the experiments.

1. More details of voxel-based uncertainty fields

In the process of deriving voxel-based uncertainty fields, we propagate all e_{ij}^{tv} to 8 adjacent voxel grids. To simplify this process, we perform backpropagation on empty voxel grids to get the uncertainty fields in one step. The value propagate to the voxel grid v_i is as follows:

$$v_i = \frac{\sum_j w_{ij} e_{ij}^{tv}}{\sum_j w_{ij}} \tag{1}$$

At this time, w_{ij} is the trilinear interpolation weight of p_{ij}^{tv} with respect to v_i. We perform the following process to obtain this process for all voxel grids at once. First, a randomly initialized dense grid $V^{(empty)}$ is created. At this time, q_i is the i^{th} voxel grid's learnable parameter in $V^{(empty)}$. We calculate the following two functions for the sampling points for all train-view images.

$$F_h = \sum_i \sum_j q_i w_{ij} = \sum_i f_{tri}(p_i, V^{(empty)}) \tag{2}$$

$$F_g = \sum_i \sum_j q_i w_{ij} e_{ij}^{tv} = \sum_i f_{tri}(p_i, V^{(empty)}) e_{p_i}^{tv} \tag{3}$$

where f_{tri} is a trilinear interpolation fucntion. f_{tri} is implemented through the grid sample function in pytorch [5]. Now, we can derive the following equations via backpropagation:

$$\frac{\partial F_h}{\partial q_i} = \sum_j w_{ij} \ , \ \frac{\partial F_g}{\partial q_i} = \sum_j w_{ij} e_{ij}^{tv} \tag{4}$$

Finally, we derive the following equation.

$$v_i = \frac{\sum_j w_{ij} e_{ij}^{tv}}{\sum_j w_{ij}} = \frac{\partial F_g / \partial q_i}{\partial F_h / \partial q_i} \tag{5}$$

Therefore, we can obtain uncertainty fields by finding the partial derivatives of F_h and F_g for all parameters q_i.

Method	Type	Novel View Synthesis			Multi-View SR		
		PSNR(↑)	SSIM(↑)	LPIPS(↓)	PSNR(↑)	SSIM(↑)	LPIPS(↓)
Synthetic NeRF Dataset							
SwinIR-ft	Single Image	30.3931	0.9423	0.0717	32.9711	0.9598	0.0601
VRT-ft	Video	30.3856	0.9424	0.0721	32.9802	0.9602	0.0610
Ours(CROP)+SwinIR	Multi-View Image	**30.7140**	**0.9459**	**0.0671**	**33.7725**	**0.9644**	**0.0565**
BlendedMVS Dataset							
SwinIR-ft	Single Image	26.4403	0.8777	0.1505	29.2625	0.9021	0.1454
VRT-ft	Video	26.6267	0.8847	0.1451	29.6925	0.9111	0.1406
Ours(CROP)+SwinIR	Multi-View Image	**26.6914**	**0.8874**	**0.1405**	**29.7451**	**0.9126**	**0.1378**
Tanks and Temples Dataset							
SwinIR-ft	Single Image	28.5525	0.9152	0.1463	35.8514	0.9580	0.0853
VRT-ft	Video	28.5950	0.9160	0.1459	35.7594	0.9596	0.0853
Ours(CROP)+SwinIR	Multi-View Image	**28.6490**	**0.9176**	**0.1425**	**36.2430**	**0.9613**	**0.0808**

Table 1. HR novel view synthesis results and multi-view image SR results with fine-tuned models for X4 SR. SwinIR-ft means fine-tuned SwinIR and VRT-ft means fine-tuned VRT. **Bold** indicates the best results, and underline indicates the second best results.

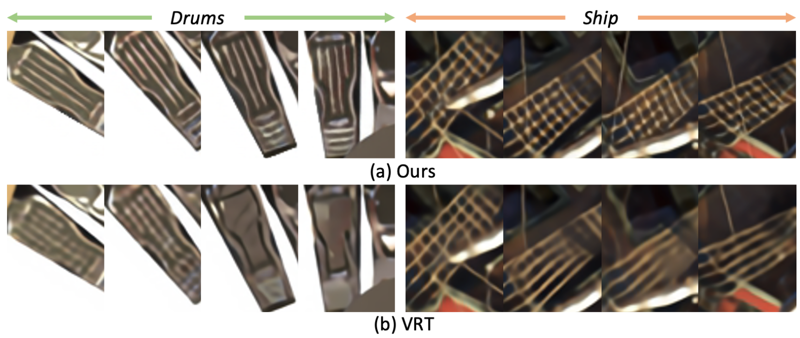

Figure 1. SR results of adjacent images on the Synthetic-NeRF dataset obtained from our model (a) and VRT (b).

2. More details of Experiments

2.1. Comparison with fine-tuned models

As in Table 1 in the main paper, we conducted experiments on three test datasets. Among them, the 'Synthetic NeRF' dataset and the 'Tanks and Temple' dataset are not used for training. The 'BlendedMVS' dataset is split into the training set and the test set, and the training split of 'BlendedMVS' is used for the training of SUM along with the 'RTMV' dataset. The training dataset used for SUM has a smaller domain gap with the test datasets,therefore, for a fair evaluation, we additionally fine-tuned SwinIR and VRT models using the same training dataset used for SUM and conducted a quantitative comparison. As shown in Table 1, our model demonstrates better performances in HRNVS and MVSR compared to the fine-tuned SwinIR (SwinIR-ft) and fine-tuned VRT (VRT-ft). Through this experiment, we also demonstrated that our SR models can utilize multi-view images as a training dataset more effectively than existing SR models.

2.2. Discussion about view consistency

The HRNVS performance of our model serves as indirect proof of improved view consistency. For example, poor multi-view consistency in train-view images leads to inaccurate scene geometry in the radiance fields model, negatively impacting HRNVS results. As such, our HRNVS performance can be inferred as evidence of improved view consistency. Additionally, as shown in Fig. 1, our sequential results for the same scene demonstrate visually high multi-view consistency.

2.3. Datasets for training SUM

We additionally visualize the datasets to train the SR update module (SUM). We use 40 scenes from the BlendedMVS dataset and 60 scenes from the RTMV dataset. As shown in Fig. 2, we visualize the results of 4 scenes for BlendedMVS dataset (a,b,c,d) and RTMV dataset (e,f,g,h), respectively. The first to fourth columns are, in order, low-resolution image, high-resolution image, rendered RGB output, and uncertainty map.

2.4. Quantitative Results

We additionally show the results of high-resolution novel view synthesis (HRNVS) and multi-view image super-resolution (MVSR) for each scene in the three datasets. Table. 2, Table. 4, and Table. 6 are results for HRNVS, and Table. 3,Table. 5, and Table. 7 are results for MVSR.

2.5. Qualitative Comparison

We additionally visualize the qualitative comparison of HRNVS and MVSR. Figure. 3 shows the results for HRNVS, and Fig. 4 and Fig. 5 show the results for MVSR.

References

[1] Ri Cheng, Yuqi Sun, Bo Yan, Weimin Tan, and Chenxi Ma. Geometry-aware reference synthesis for multi-view image super-resolution. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6083–6093, 2022. 4, 5, 6

[2] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *arXiv preprint arXiv:2201.12288*, 2022. 4, 5, 6

[3] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 4, 5, 6

[4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 4, 5, 6

[5] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 1

[6] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. Nerf-sr: High quality neural radiance fields using supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6445–6454, 2022. 4, 5, 6

Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg.
				PSNR(↑)					
LR	29.7755	23.8981	27.9507	33.3681	29.1700	27.1356	30.3919	26.9668	28.5821
NeRF-SR [6]	30.2272	24.2528	29.4971	33.4723	29.5622	26.9168	30.3976	26.8422	28.8960
EDSR [4]	30.7208	24.6691	29.9158	34.8040	30.4855	28.8484	31.2529	27.6478	29.7931
SwinIR [3]	31.2154	24.9985	31.0068	35.4589	32.2369	29.2156	31.6982	28.0719	30.4878
MVSRnet [1]	30.9722	24.7673	30.1083	35.1703	30.9997	28.8978	31.4135	27.7022	30.0039
VRT [2]	31.3786	**25.0779**	30.8626	35.5274	32.1902	**29.2180**	31.7118	28.0914	30.5072
Ours(CROP)+EDSR	30.8831	24.6148	30.7071	35.1610	30.7669	28.8738	31.4810	27.7933	30.0351
Ours(CROP)+SwinIR	**31.5335**	24.9992	**31.5063**	**35.6218**	**32.8868**	29.1645	**31.7679**	**28.2320**	**30.7140**
				SSIM(↑)					
LR	0.9263	0.8976	0.9436	0.9591	0.9240	0.9138	0.9645	0.8397	0.9211
NeRF-SR [6]	0.9372	0.9062	0.9552	0.9608	0.9318	0.9173	0.9637	0.8403	0.9266
EDSR [4]	0.9398	0.9155	0.9591	0.9679	0.9412	0.9432	0.9697	0.8526	0.9361
SwinIR [3]	0.9459	0.9226	0.9677	0.9708	0.9588	0.9464	0.9728	0.8599	0.9431
MVSRnet [1]	0.9440	0.9178	0.9612	0.9702	0.9477	0.9438	0.9709	0.8545	0.9388
VRT [2]	0.9481	0.9230	0.9668	0.9714	0.9588	0.9463	0.9728	0.8606	0.9435
Ours(CROP)+EDSR	0.9458	0.9183	0.9652	0.9701	0.9486	0.9443	0.9719	0.8581	0.9403
Ours(CROP)+SwinIR	**0.9513**	**0.9236**	**0.9709**	**0.9725**	**0.9641**	**0.9468**	**0.9740**	**0.8637**	**0.9459**
				LPIPS(↓)					
LR	0.0841	0.1217	0.0715	0.0754	0.1075	0.1094	0.0456	0.1997	0.1019
NeRF-SR [6]	0.0786	0.1154	0.0510	0.0747	0.1019	0.1105	0.0486	0.2130	0.0992
EDSR [4]	0.0708	0.1007	0.0461	0.0555	0.0870	0.0709	0.0336	0.1912	0.0820
SwinIR [3]	0.0634	0.0890	0.0352	0.0505	0.0569	0.0642	0.0268	0.1805	0.0708
MVSRnet [1]	0.0642	0.0974	0.0449	0.0514	0.0757	0.0678	0.0299	0.1871	0.0773
VRT [2]	0.0615	0.0887	0.0363	0.0495	0.0585	0.0649	0.0261	0.1805	0.0708
Ours(CROP)+EDSR	0.0604	0.0927	0.0375	0.0528	0.0658	0.0661	0.0280	0.1801	0.0729
Ours(CROP)+SwinIR	**0.0567**	**0.0856**	**0.0317**	**0.0481**	**0.0496**	**0.0622**	**0.0251**	**0.1776**	**0.0671**

Table 2. HR novel view synthesis results on the Synthetic NeRF dataset for X4 SR. **Bold** indicates the best results, and underline indicates the second best results.

Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg.
				PSNR(↑)					
EDSR [4]	31.0280	29.0701	33.1699	36.0150	31.2824	33.4061	31.0502	29.2705	31.7865
SwinIR [3]	31.5610	30.3991	34.9668	37.0812	33.8057	35.3058	31.7382	30.1294	33.1234
MVSRnet [1]	31.1493	29.2505	32.9244	36.4229	31.7389	33.5732	31.3718	29.2801	31.9639
VRT [2]	31.8339	30.5930	34.8324	37.1716	33.9033	**35.3397**	31.9428	30.2596	33.2345
Ours(CROP)+EDSR	31.0952	28.9368	34.5458	36.2754	31.4471	33.7515	31.5975	29.5630	32.1515
Ours(CROP)+SwinIR	**32.1797**	**30.6643**	**36.5222**	**37.6013**	**35.3525**	35.1576	**32.0609**	**30.6411**	**33.7725**
				SSIM(↑)					
EDSR [4]	0.9390	0.9570	0.9752	0.9686	0.9430	0.9740	0.9664	0.8765	0.9500
SwinIR [3]	0.9465	0.9700	0.9843	0.9741	0.9658	**0.9825**	0.9732	0.8922	0.9611
MVSRnet [1]	0.9438	0.9623	0.9766	0.9725	0.9507	0.9776	0.9704	0.8811	0.9544
VRT [2]	0.9488	0.9700	0.9841	0.9740	0.9650	0.9820	0.9733	0.8912	0.9610
Ours(CROP)+EDSR	0.9462	0.9624	0.9817	0.9717	0.9518	0.9782	0.9721	0.8860	0.9563
Ours(CROP)+SwinIR	**0.9535**	**0.9706**	**0.9875**	**0.9763**	**0.9731**	0.9824	**0.9748**	**0.8967**	**0.9644**
				LPIPS(↓)					
EDSR [4]	0.0742	0.0744	0.0354	0.0562	0.0954	0.0568	0.0457	0.1830	0.0776
SwinIR [3]	0.0671	0.0533	0.0268	0.0464	0.0582	**0.0373**	0.0265	0.1642	0.0600
MVSRnet [1]	0.0675	0.0631	0.0366	0.0496	0.0796	0.0446	0.0313	0.1770	0.0687
VRT [2]	0.0641	0.0531	0.0265	0.0475	0.0599	0.0386	0.0253	0.1643	0.0599
Ours(CROP)+EDSR	0.0645	0.0595	0.0297	0.0513	0.0693	0.0407	0.0267	0.1670	0.0636
Ours(CROP)+SwinIR	**0.0599**	**0.0528**	**0.0250**	**0.0442**	**0.0480**	0.0377	**0.0252**	**0.1596**	**0.0565**

Table 3. Multi-view image SR results on the Synthetic NeRF dataset for X4 SR. **Bold** indicates the best results, and underline indicates the second best results.

Method	Character	Fountain	Jade	Statues	Avg.
			PSNR(↑)		
LR	25.6300	25.4200	26.4100	23.7200	25.2950
NeRF-SR [6]	26.7233	26.0939	**27.6352**	**25.2847**	26.4342
EDSR [4]	26.8400	26.2013	26.9429	24.5000	26.1210
SwinIR [3]	27.3800	26.5087	27.2382	24.9400	26.5167
MVSRnet [1]	27.2900	26.5225	27.2478	24.8300	26.4726
VRT [2]	**27.8000**	**26.7089**	27.1691	24.6400	<u>26.5795</u>
Ours(CROP)+EDSR	27.1285	26.1634	27.1293	24.6163	26.2594
Ours(CROP)+SwinIR	<u>27.7476</u>	<u>26.5683</u>	<u>27.4310</u>	<u>25.0186</u>	**26.6914**
			SSIM(↑)		
LR	0.9000	0.8400	0.8600	0.7900	0.8475
NeRF-SR [6]	0.9149	0.8575	0.8857	<u>0.8405</u>	0.8747
EDSR [4]	0.9200	0.8581	0.8811	0.8300	0.8723
SwinIR [3]	0.9200	0.8672	0.8874	0.8400	0.8787
MVSRnet [1]	0.9200	0.8701	0.8889	0.8400	0.8798
VRT [2]	<u>0.9300</u>	**0.8797**	<u>0.8895</u>	0.8400	<u>0.8848</u>
Ours(CROP)+EDSR	0.9250	0.8677	0.8860	0.8386	0.8793
Ours(CROP)+SwinIR	**0.9316**	<u>0.8777</u>	**0.8939**	**0.8462**	**0.8874**
			LPIPS(↓)		
LR	0.1100	0.2100	0.1800	0.2100	0.1775
NeRF-SR [6]	0.0951	0.2006	0.1615	0.1969	0.1635
EDSR [4]	0.0900	0.1951	0.1589	0.1900	0.1585
SwinIR [3]	<u>0.0800</u>	0.1856	0.1513	0.1800	0.1492
MVSRnet [1]	<u>0.0800</u>	0.1811	0.1486	0.1800	0.1474
VRT [2]	<u>0.0800</u>	<u>0.1730</u>	0.1502	0.1800	0.1458
Ours(CROP)+EDSR	0.0805	0.1798	<u>0.1472</u>	<u>0.1751</u>	<u>0.1456</u>
Ours(CROP)+SwinIR	**0.0754**	**0.1723**	**0.1447**	**0.1696**	**0.1405**

Table 4. HR novel view synthesis results on the BlendedMVS dataset for X4 SR. **Bold** indicates the best results, and <u>underline</u> indicates the second best results.

Method	Character	Fountain	Jade	Statues	Avg.
			PSNR(↑)		
EDSR [4]	27.5832	27.5499	31.8448	28.0176	28.7489
SwinIR [3]	28.2957	27.9407	32.3207	28.5672	29.2811
MVSRnet [1]	28.0870	27.6973	32.1936	28.2734	29.0628
VRT [2]	**28.9996**	**28.5479**	<u>32.4471</u>	<u>28.7717</u>	<u>29.6916</u>
Ours(CROP)+EDSR	27.9300	27.3423	31.7107	28.2041	28.7968
Ours(CROP)+SwinIR	<u>28.9957</u>	<u>28.4602</u>	**32.6519**	**28.8726**	**29.7451**
			SSIM(↑)		
EDSR [4]	0.9233	0.8697	0.9170	0.8665	0.8941
SwinIR [3]	0.9350	0.8796	0.9236	0.8753	0.9034
MVSRnet [1]	0.9344	0.8808	0.9250	0.8732	0.9034
VRT [2]	<u>0.9422</u>	**0.8988**	<u>0.9253</u>	<u>0.8784</u>	<u>0.9112</u>
Ours(CROP)+EDSR	0.9343	0.8780	0.9197	0.8738	0.9015
Ours(CROP)+SwinIR	**0.9435**	<u>0.8949</u>	**0.9304**	**0.8818**	**0.9126**
			LPIPS(↓)		
EDSR [4]	0.0969	0.1927	0.1478	0.1904	0.1570
SwinIR [3]	0.0847	0.1826	0.1480	0.1752	0.1476
MVSRnet [1]	0.0849	0.1803	0.1480	0.1768	0.1475
VRT [2]	**0.0786**	**0.1654**	<u>0.1450</u>	<u>0.1730</u>	<u>0.1405</u>
Ours(CROP)+EDSR	0.0843	0.1806	0.1556	0.1748	0.1488
Ours(CROP)+SwinIR	<u>0.0788</u>	<u>0.1675</u>	**0.1397**	**0.1653**	**0.1378**

Table 5. Multi-view image SR results on the BlendedMVS dataset for X4 SR. **Bold** indicates the best results, and <u>underline</u> indicates the second best results.

Method	Barn	Caterpillar	Family	Ignatius	Truck	Avg.
			PSNR(↑)			
LR	27.0067	25.7908	32.8656	23.9114	26.9718	27.3093
NeRF-SR [6]	25.2235	24.2541	31.0649	27.5498	25.7183	26.7621
EDSR [4]	27.2225	26.0106	33.6156	28.2580	27.3131	28.4840
SwinIR [3]	27.3489	26.0865	33.8695	<u>28.2542</u>	27.3796	28.5877
MVSRnet [1]	27.2672	26.0559	33.6632	28.2535	27.3558	28.5191
VRT [2]	27.3022	26.0871	<u>33.9469</u>	**28.2591**	<u>27.3912</u>	<u>28.5973</u>
Ours(CROP)+EDSR	<u>27.3676</u>	<u>26.0909</u>	33.8230	28.1396	27.3484	28.5539
Ours(CROP)+SwinIR	**27.4261**	**26.1932**	**33.9849**	28.2154	**27.4255**	**28.6490**
			SSIM(↑)			
LR	0.8363	0.9064	0.9558	0.9148	0.9030	0.9033
NeRF-SR [6]	0.8059	0.8793	0.9392	0.9298	0.8801	0.8869
EDSR [4]	0.8459	0.9104	0.9612	0.9448	0.9100	0.9144
SwinIR [3]	0.8483	0.9112	0.9625	0.9447	0.9118	0.9157
MVSRnet [1]	0.8469	0.9107	0.9613	0.9446	0.9104	0.9148
VRT [2]	0.8488	0.9113	0.9632	<u>0.9448</u>	0.9116	0.9159
Ours(CROP)+EDSR	<u>0.8498</u>	<u>0.9117</u>	<u>0.9635</u>	0.9446	<u>0.9125</u>	<u>0.9164</u>
Ours(CROP)+SwinIR	**0.8528**	**0.9126**	**0.9642**	**0.9449**	**0.9136**	**0.9176**
			LPIPS(↓)			
LR	0.2879	0.1638	0.0753	0.1036	0.1583	0.1578
NeRF-SR [6]	0.3505	0.2048	0.1024	0.1041	0.1982	0.1920
EDSR [4]	0.2786	0.1597	0.0700	0.0820	0.1518	0.1484
SwinIR [3]	0.2757	0.1576	0.0674	0.0817	0.1484	0.1462
MVSRnet [1]	0.2770	0.1578	0.0687	<u>0.0814</u>	0.1500	0.1470
VRT [2]	0.2750	0.1578	<u>0.0663</u>	0.0818	0.1483	0.1459
Ours(CROP)+EDSR	<u>0.2721</u>	<u>0.1564</u>	<u>0.0663</u>	0.0820	<u>0.1451</u>	<u>0.1444</u>
Ours(CROP)+SwinIR	**0.2681**	**0.1546**	**0.0645**	**0.0810**	**0.1441**	**0.1425**

Table 6. HR novel view synthesis results on the Tanks and Temples dataset for X4 SR. **Bold** indicates the best results, and <u>underline</u> indicates the second best results.

Method	Barn	Caterpillar	Family	Ignatius	Truck	Avg.
			PSNR(↑)			
EDSR [4]	32.1484	33.9276	36.3531	36.6149	33.4264	34.4941
SwinIR [3]	33.4641	<u>35.6484</u>	37.3917	37.2503	<u>34.7209</u>	35.6951
MVSRnet [1]	32.3777	34.4622	36.4573	36.8019	33.6488	34.7496
VRT [2]	<u>33.4813</u>	35.5446	<u>37.8286</u>	<u>37.3337</u>	34.6037	<u>35.7584</u>
Ours(CROP)+EDSR	32.5761	34.9083	37.2558	37.1567	34.1593	35.2112
Ours(CROP)+SwinIR	**33.6049**	**36.2304**	**38.2025**	**37.9578**	**35.2193**	**36.2430**
			SSIM(↑)			
EDSR [4]	0.9047	0.9670	0.9704	0.9659	0.9536	0.9523
SwinIR [3]	**0.9219**	<u>0.9751</u>	0.9742	<u>0.9683</u>	<u>0.9623</u>	<u>0.9604</u>
MVSRnet [1]	0.9063	0.9675	0.9702	0.9652	0.9536	0.9526
VRT [2]	0.9205	0.9734	<u>0.9755</u>	0.9680	0.9602	0.9595
Ours(CROP)+EDSR	0.9109	0.9711	0.9746	0.9680	0.9592	0.9568
Ours(CROP)+SwinIR	<u>0.9215</u>	**0.9754**	**0.9766**	**0.9692**	**0.9637**	**0.9613**
			LPIPS(↓)			
EDSR [4]	0.1714	0.0797	0.0617	0.0621	0.0942	0.0938
SwinIR [3]	<u>0.1520</u>	<u>0.0691</u>	0.0570	<u>0.0597</u>	<u>0.0829</u>	<u>0.0841</u>
MVSRnet [1]	0.1727	0.0829	0.0660	0.0676	0.0984	0.0975
VRT [2]	0.1545	0.0717	<u>0.0548</u>	0.0606	0.0855	0.0854
Ours(CROP)+EDSR	0.1669	0.0739	0.0585	0.0606	0.0886	0.0897
Ours(CROP)+SwinIR	**0.1500**	**0.0661**	**0.0526**	**0.0566**	**0.0789**	**0.0808**

Table 7. Multi-view image SR results on the Tanks and Temples dataset for X4 SR. **Bold** indicates the best results, and <u>underline</u> indicates the second best results.

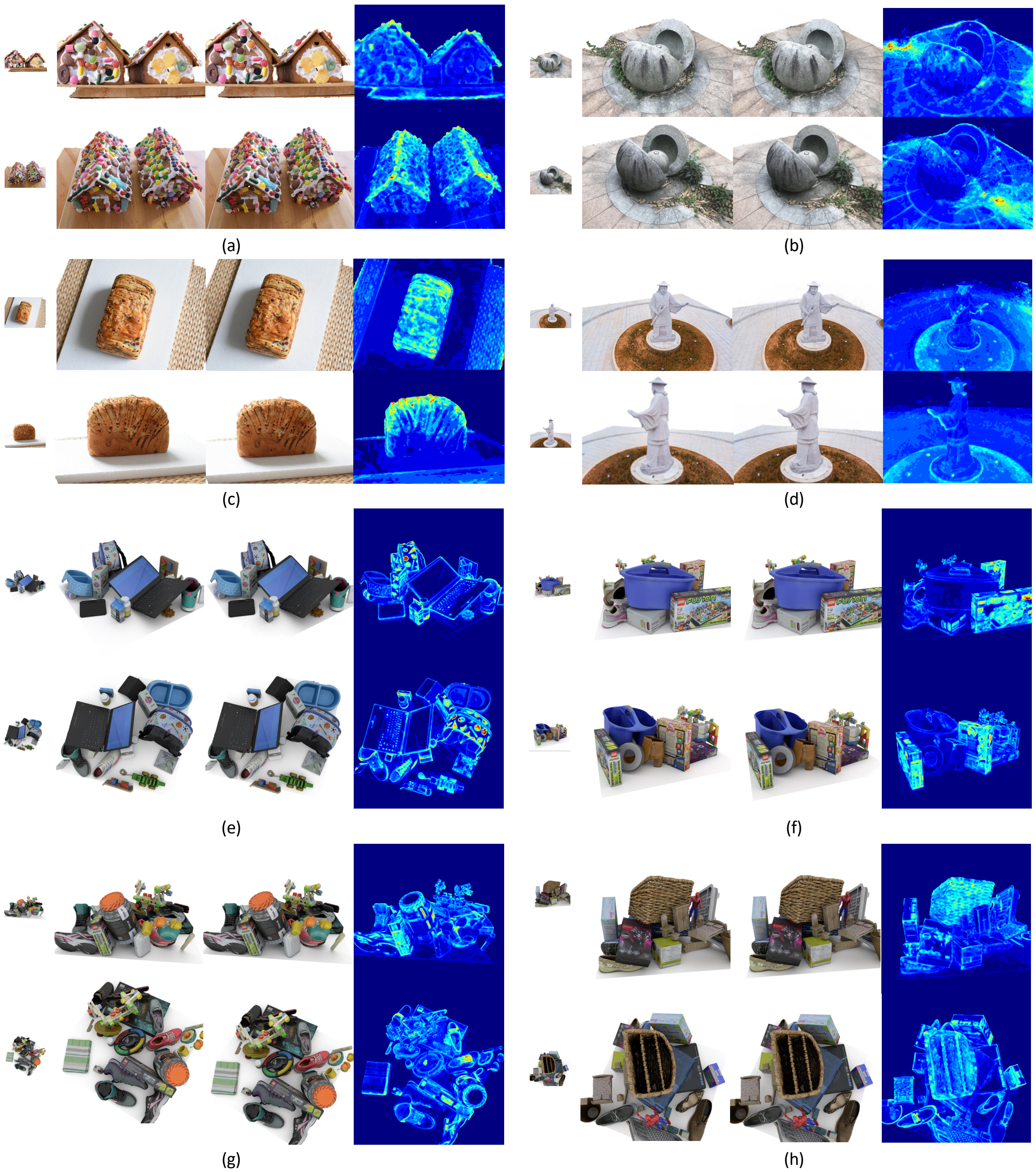

Figure 2. BlendedMVS(a,b,c,d) and RTMV(e,f,g,h) datasets for training SR update module (SUM).

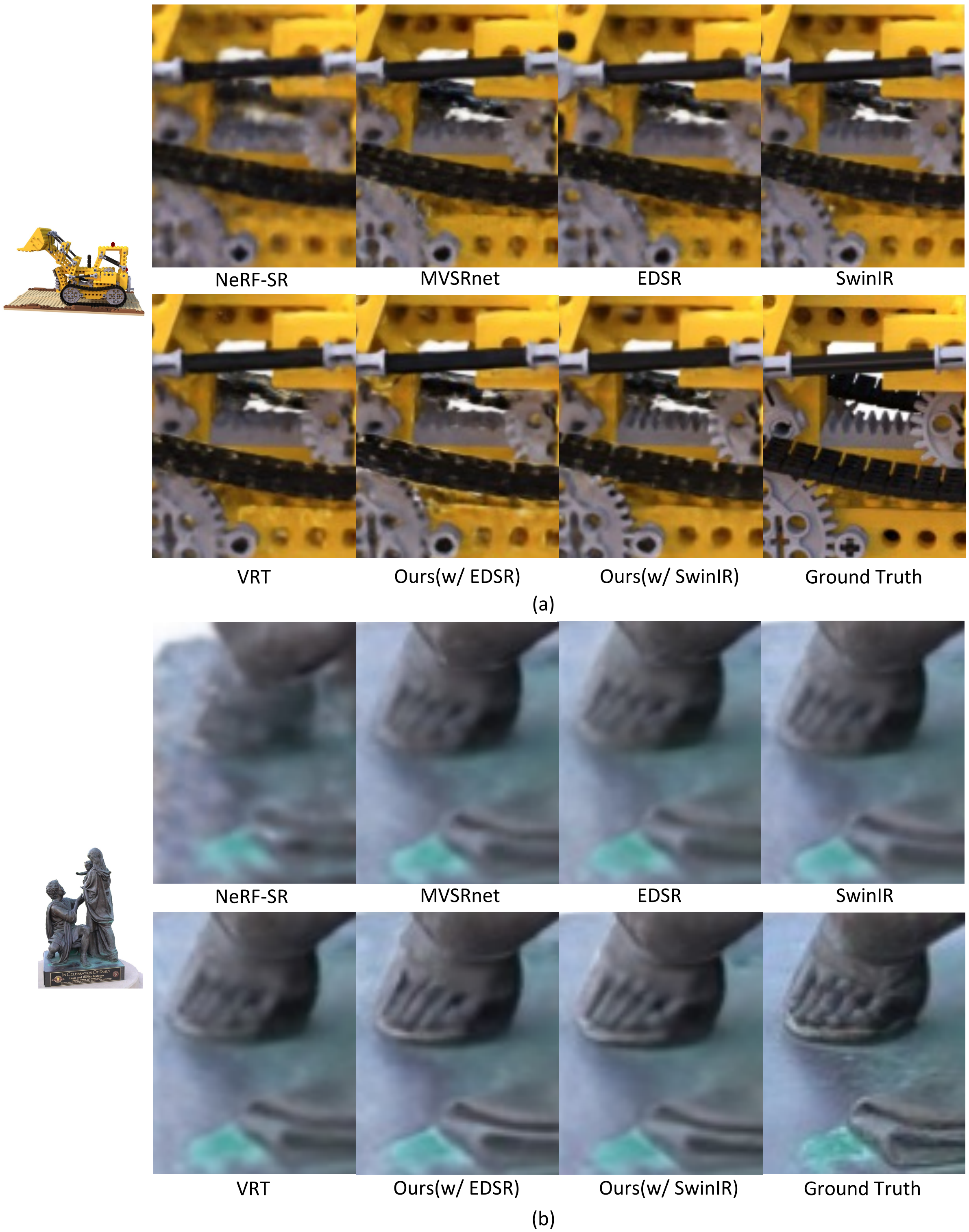

Figure 3. Qualitative comparisons of HR novel view synthesis of different methods.

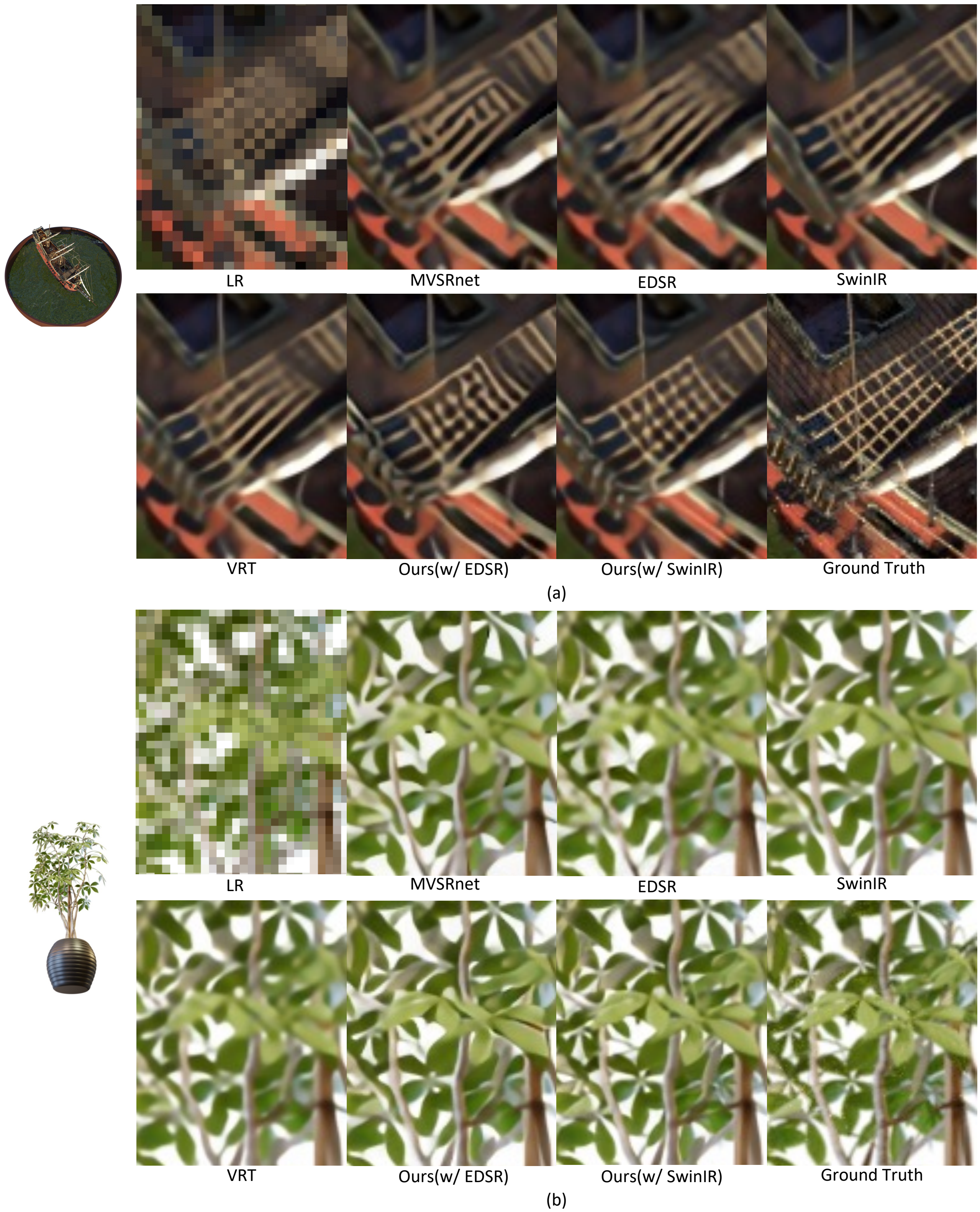

Figure 4. Qualitative comparisons of multi-view image SR of different methods.

Figure 5. Qualitative comparisons of multi-view image SR of different methods.