3D Appearance Super-Resolution with Deep Learning: Supplementary Material

Yawei Li[1], Vagia Tsiminaki[2], Radu Timofte[1], Marc Pollefeys[2,3], and Luc van Gool[1]
[1]Computer Vision Lab, ETH Zurich, Switzerland
{yawei.li, radu.timofte, vangool}@vision.ee.ethz.ch
[2]Computer Vision and Geometry Group, ETH Zurich, Switzerland [3]Microsoft, USA
{vagia.tsiminaki, marc.pollefeys}@inf.ethz.ch

1. Details of the Provided Dataset

1.1. Mesh, images, and projection matrices

The provided dataset has 24 different scenes in total including 13 from ETH3D, 6 from Collection, 3 from SyB3R, and 2 from Middlebury. Each scene contains a 3D mesh, multi-view images, and the corresponding projection matrices. The details of those scenes are provided in Table 1 including the mesh size, the number of vertices and faces in the mesh, the resolution of the HR images, and the number of views in the scene. It is shown in Table 1 that the scenes have different complexities,

Table 1: Details of the each of the scenes in the provided dataset including the size (MB) of the mesh, number of vertices and faces (k) in the mesh, the resolution of HR images, and the number of views in each scene.

Dataset	Scene	Mesh			Image	
		Mesh size	No. vertices	No. Faces	Resolution	No. Views
ETH3D	*courtyard*	136.0	646	1,168	3096 × 2064	38
	delivery_area	100.1	511	911	3096 × 2064	44
	electro	32.7	164	293	3084 × 2052	45
	facade	121.5	583	1,026	3096 × 2052	46
	kicker	116.1	571	1,004	3096 × 2064	31
	meadow	32.0	156	282	3096 × 2064	15
	office	194.1	882	1,663	3108 × 2064	26
	pipes	171.3	850	1,502	3108 × 2064	14
	playground	44.8	233	367	3096 × 2064	38
	relief	36.7	186	324	3096 × 2064	31
	relief_2	58.5	299	519	3096 × 2064	31
	terrace	52.0	327	421	3096 × 2064	23
	terrains	55.5	265	495	3096 × 2064	42
Collection	*Beethoven*	16.0	74	144	1024 × 768	33
	Bird	16.7	78	150	1024 × 768	20
	Buddha	16.5	75	150	1404 × 936	91
	Bunny	12.3	57	111	1024 × 768	36
	Fountain	40.4	200	399	1280 × 1024	55
	Relief	47.6	233	464	1280 × 1024	40
Middlebury	*DinoRing*	143.2	619	1,237	640 × 480	48
	TempleRing	85.2	369	737	640 × 480	47
SyB3R	*GeologicalSample*	14.7	98	197	3888 × 2592	14
	Skull	2.8	19	38	3888 × 2592	14
	Toad	77	481	962	3888 × 2592	14

i.e., different mesh size and number of vertices and faces.

1.2. Texture maps

Twelve of the 24 texture maps for different resolutions (HR, $\times 2$, $\times 3$, $\times 4$ down-sampling) are shown in Fig. 1, Fig. 2, and Fig. 3, respectively. By comparing the texure maps with differnt resolutions, we find that the ground truth texture maps contain more details than the LR ones. In addition, the HR texture maps are denser than the LR ones. Since optimal UV parameters exist for the synthetic scenes *GeologicalSample*, *Toad*, and *Skull*, their texture maps have less disconnected support regions.

2. SR Results

In Table 2 and Table 3, we show the PSNR and SSIM results of different methods. Apart from the methods in the main paper, the results of FSRCNN [2], SRResNet [3], and RCAN [4] are also provided. For FSRCNN, the pre-trained models provided by the authors are directly used. SRResNet, EDSR, and RCAN are trained on DIV2K [1]. More visual results of *relief*, *facade*, *Buddha*, and *Fountain* for different methods are shown in Fig. 4, Fig. 5, Fig. 6, and Fig. 7, respectively.

Table 2: The PSNR results of different methods for scaling factor $\times 2$, $\times 3$, and $\times 4$.

Method	ETH3D			Collection			Middlebury			SyB3R			Average		
	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$
Nearest	19.06	16.71	14.68	24.22	19.7	16.92	10.08	7.93	7.08	30.84	27.88	25.82	21.07	18.12	16.0
Bilinear	20.61	18.24	16.32	26.2	21.48	18.84	11.87	8.88	7.77	31.75	28.83	26.9	22.67	19.6	17.56
Bicubic	20.21	17.96	15.88	25.67	21.12	18.29	11.32	8.81	7.73	31.77	28.78	26.73	22.28	19.34	17.16
Lanczos	20.01	17.74	15.69	25.42	20.86	18.07	11.14	8.81	7.81	31.71	28.7	26.63	22.09	19.15	17.0
HRST	16.18	–	16.12	32.29	–	29.63	22.13	–	20.88	27.9	–	26.34	22.17	–	21.17
HRST+	–	–	–	32.24	–	29.9	22.76	–	21.55	–	–	–	–	–	–
FSRCNN	18.09	15.02	13.75	23.58	18.16	16.36	9.62	7.73	7.43	30.23	26.35	25.01	20.27	16.61	15.29
SRRESNET	17.61	14.89	12.83	22.99	18.1	15.22	8.92	6.98	6.5	30.04	26.88	24.52	19.79	16.53	14.36
EDSR	16.75	14.08	12.03	21.77	17.2	14.24	8.49	7.13	6.61	29.31	26.18	23.81	18.89	15.79	13.61
RCAN	16.32	13.62	11.55	21.6	16.73	13.91	8.4	7.05	6.54	29.11	25.86	23.5	18.58	15.38	13.22
EDSR+	21.13	19.75	18.44	28.25	25.53	24.19	12.73	11.21	9.9	32.78	29.9	28.31	23.66	21.75	20.4
NLR-	21.21	20.11	19.2	28.08	25.0	23.27	14.68	12.37	11.11	32.18	28.84	26.64	23.75	21.78	20.47
NLR	21.31	20.27	19.18	28.38	25.85	24.84	13.67	12.92	12.29	32.57	29.57	27.67	23.85	22.22	21.08
NHR	25.19	23.95	22.7	30.25	28.41	26.27	17.16	17.21	15.63	30.57	27.42	24.39	26.46	24.94	23.22

Table 3: The SSIM results of different methods for scaling factor $\times 2$, $\times 3$, and $\times 4$.

Method	ETH3D			Collection			Middlebury			SyB3R			Average		
	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$	$\times 2$	$\times 3$	$\times 4$
Nearest	0.81	0.74	0.68	0.88	0.77	0.7	0.54	0.43	0.39	0.91	0.84	0.79	0.82	0.74	0.67
Bilinear	0.83	0.77	0.71	0.91	0.81	0.75	0.57	0.45	0.41	0.92	0.86	0.81	0.84	0.77	0.71
Bicubic	0.83	0.76	0.7	0.9	0.8	0.73	0.56	0.45	0.41	0.92	0.86	0.81	0.84	0.76	0.7
Lanczos	0.82	0.75	0.68	0.9	0.79	0.71	0.55	0.45	0.4	0.92	0.86	0.81	0.83	0.75	0.68
HRST	0.67	–	0.66	0.96	–	0.93	0.92	–	0.9	0.88	–	0.82	0.79	–	0.77
HRST+	–	–	–	0.95	–	0.92	0.91	–	0.89	–	–	–	–	–	–
FSRCNN	0.77	0.66	0.61	0.85	0.7	0.64	0.49	0.37	0.36	0.9	0.81	0.77	0.78	0.66	0.61
SRRESNET	0.78	0.7	0.63	0.86	0.74	0.66	0.5	0.41	0.38	0.91	0.84	0.78	0.79	0.7	0.64
EDSR	0.76	0.67	0.6	0.82	0.71	0.64	0.49	0.41	0.37	0.91	0.83	0.77	0.77	0.68	0.61
RCAN	0.75	0.66	0.58	0.84	0.7	0.63	0.49	0.4	0.37	0.9	0.83	0.77	0.77	0.67	0.6
EDSR+	0.86	0.83	0.79	0.94	0.91	0.88	0.62	0.51	0.45	0.93	0.88	0.84	0.87	0.83	0.79
NLR-	0.86	0.82	0.74	0.93	0.89	0.85	0.66	0.51	0.42	0.93	0.86	0.82	0.87	0.82	0.75
NLR	0.86	0.83	0.8	0.94	0.91	0.89	0.65	0.58	0.54	0.93	0.87	0.83	0.87	0.83	0.8
NHR	0.84	0.89	0.84	0.87	0.93	0.85	0.74	0.77	0.72	0.82	0.85	0.72	0.84	0.88	0.82

References

[1] E. Agustsson and R. Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proc. CVPRW*, July 2017. 2

[2] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *ECCV*, pages 391–407. Springer, 2016. 2

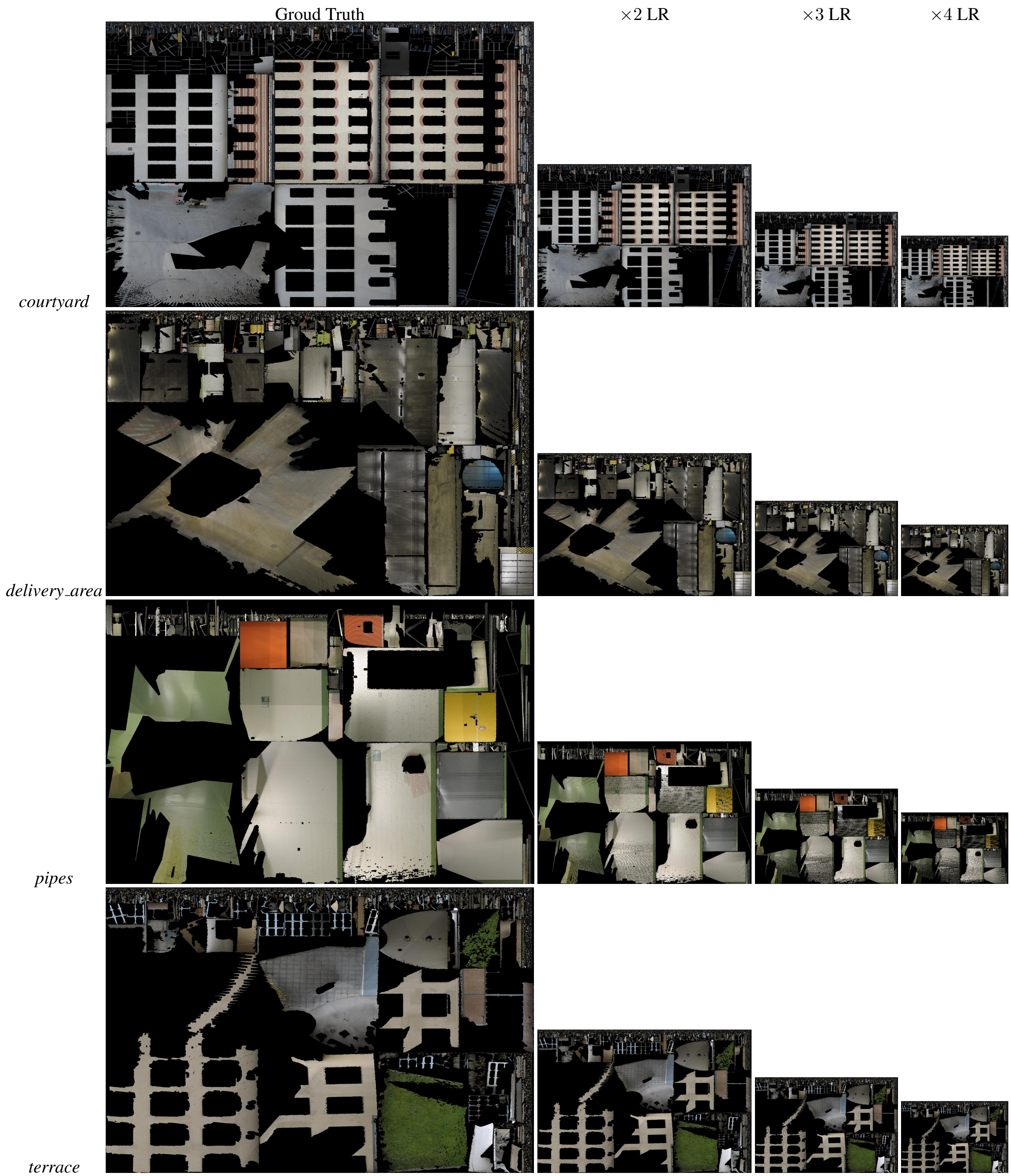

Figure 1: The texture maps of *courtyard*, *delivery_area*, *pipes*, and *terrace* for different resolutions.

[3] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proc. CVPR*, volume 2, page 4, 2017. 2

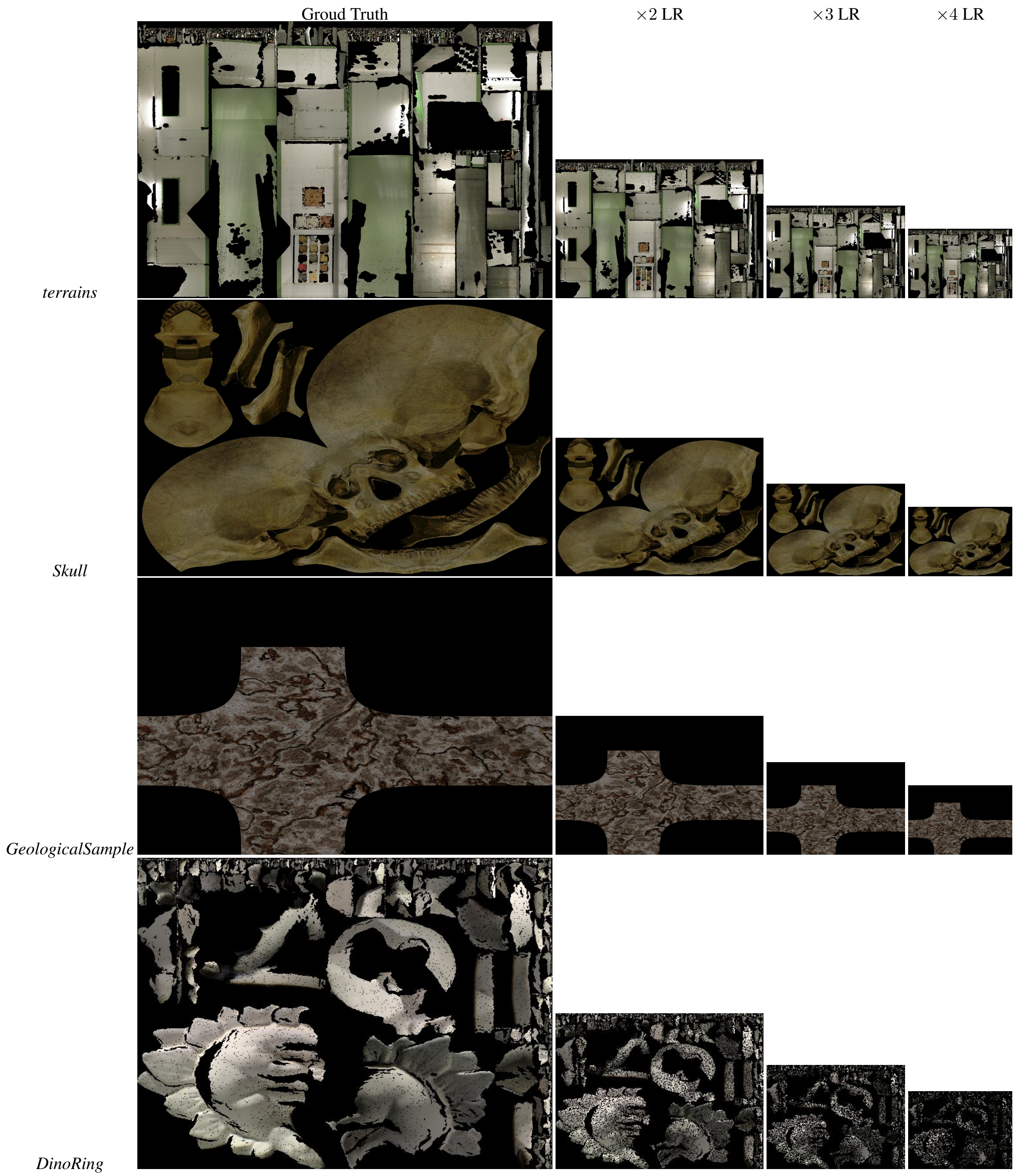

Figure 2: The texture maps of *terrains*, *Skull*, *GeologicalSample*, and *DinoRing* for different resolutions.

[4] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *Proc. ECCV*, pages 286–301, 2018. 2

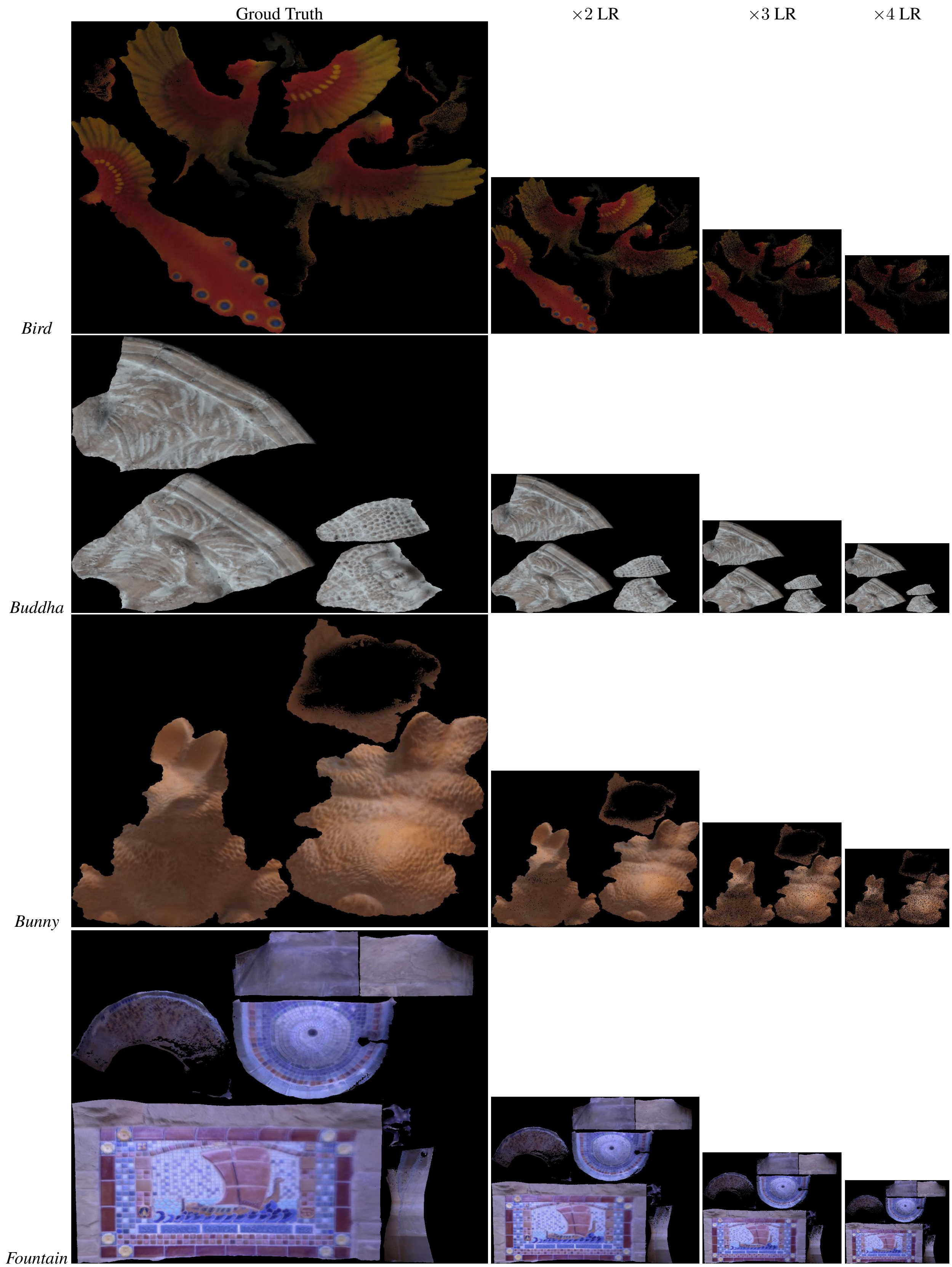

Figure 3: The texture maps of *Bird*, *Buddha*, *Bunny*, and *Fountain* for different resolutions.

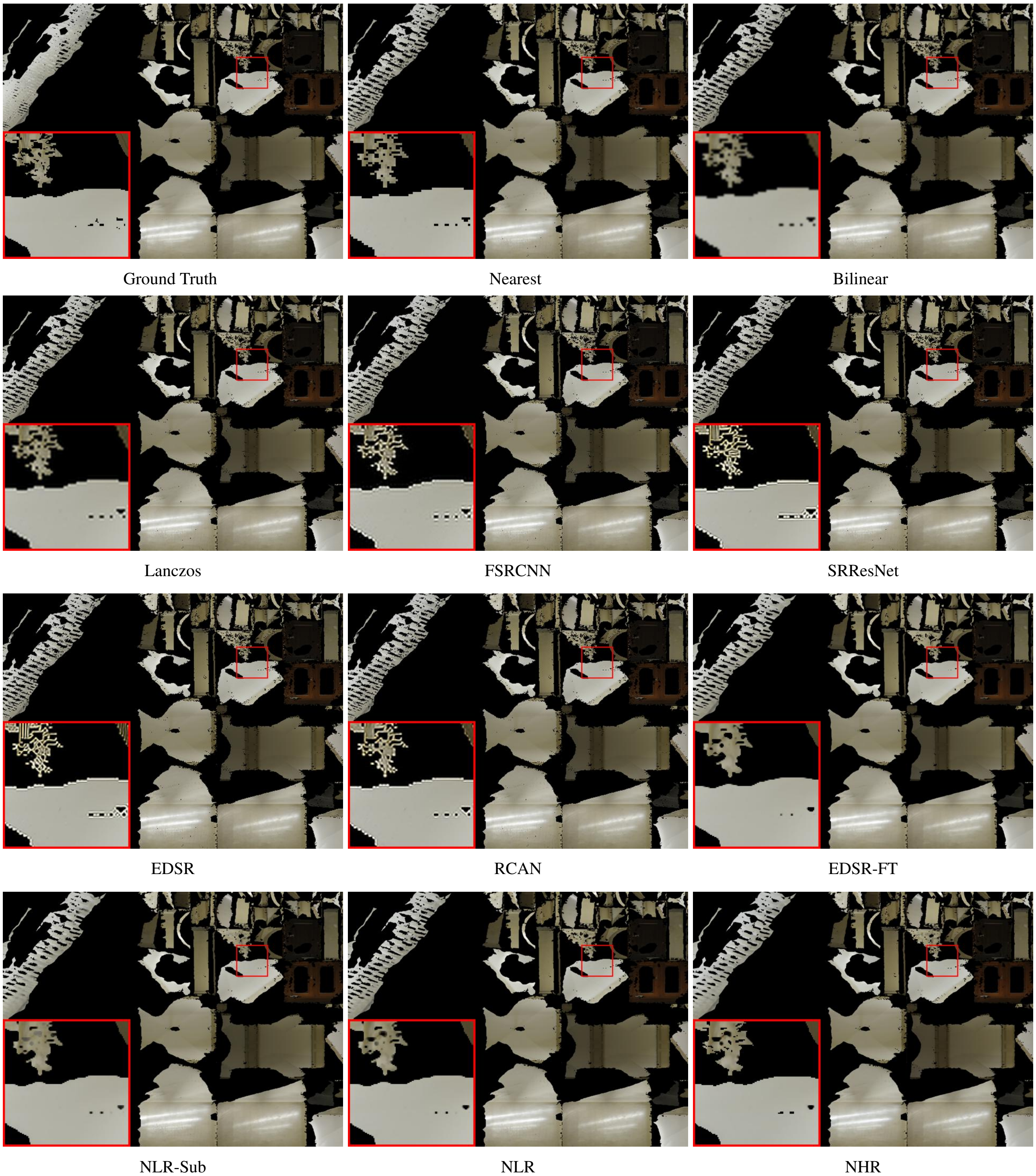

Figure 4: Texture map SR results of *relief* by different methods.

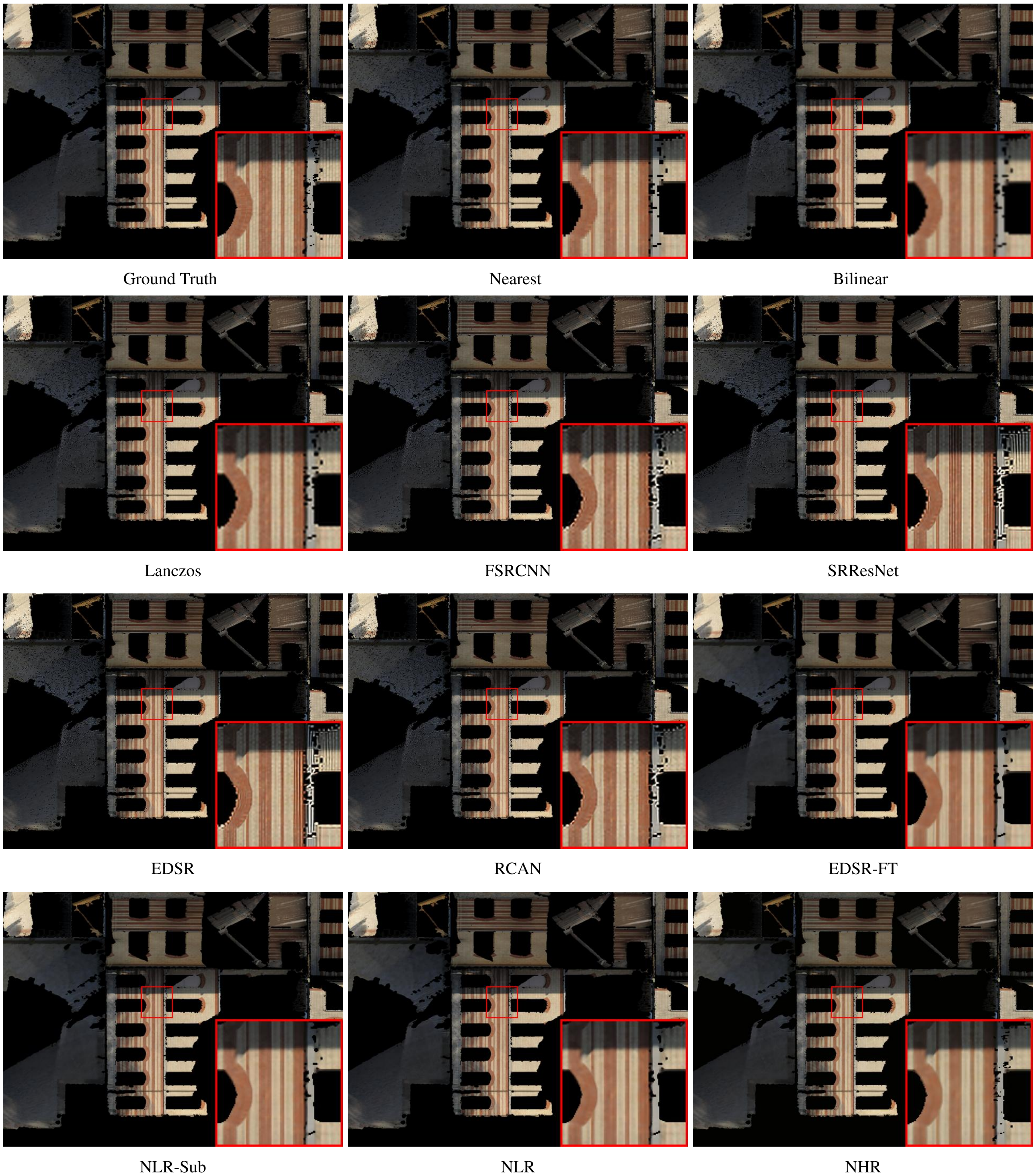

Figure 5: Texture map SR results of *facade* by different methods.

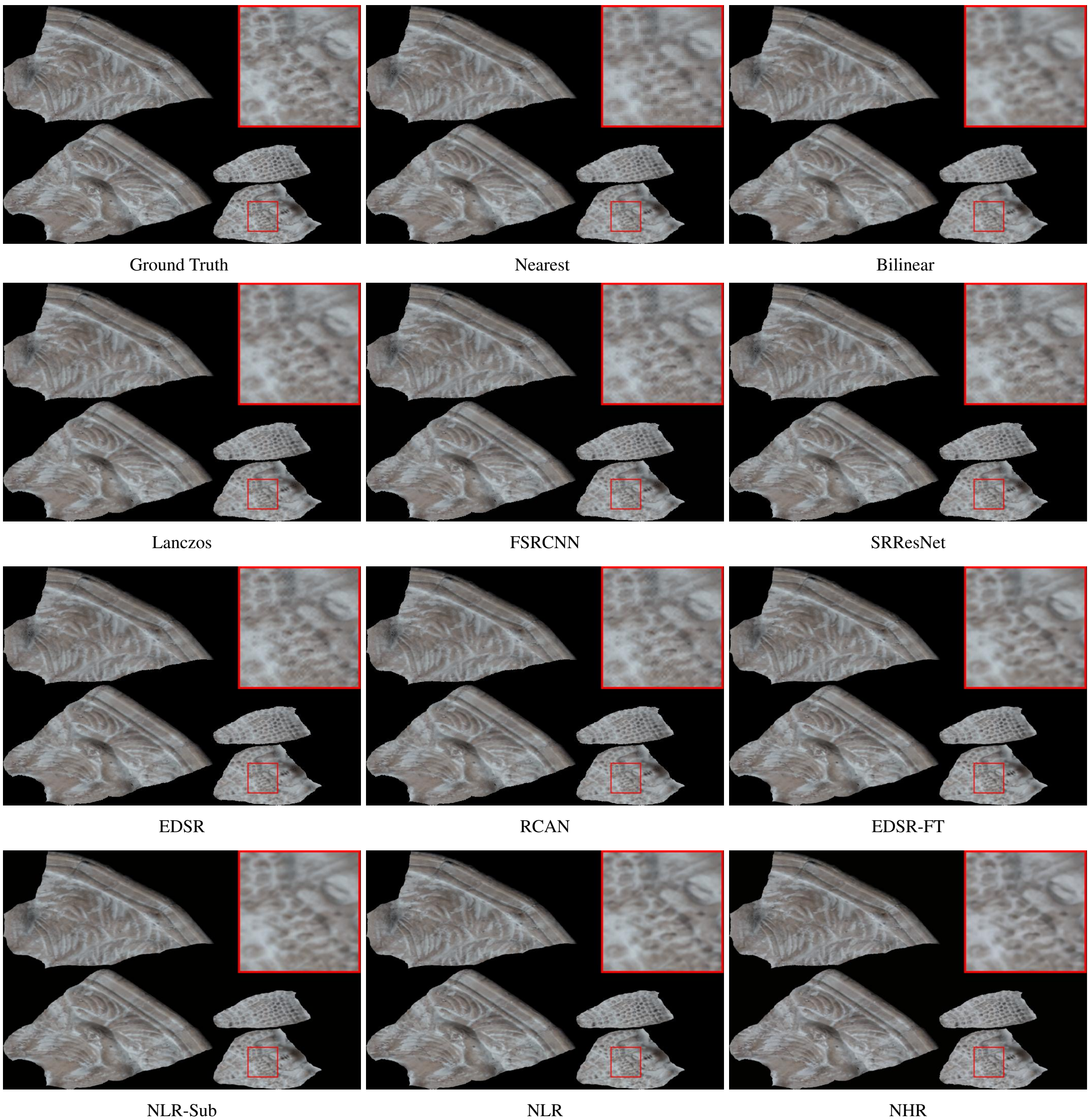

Figure 6: Texture map SR results of *Buddha* by different methods.

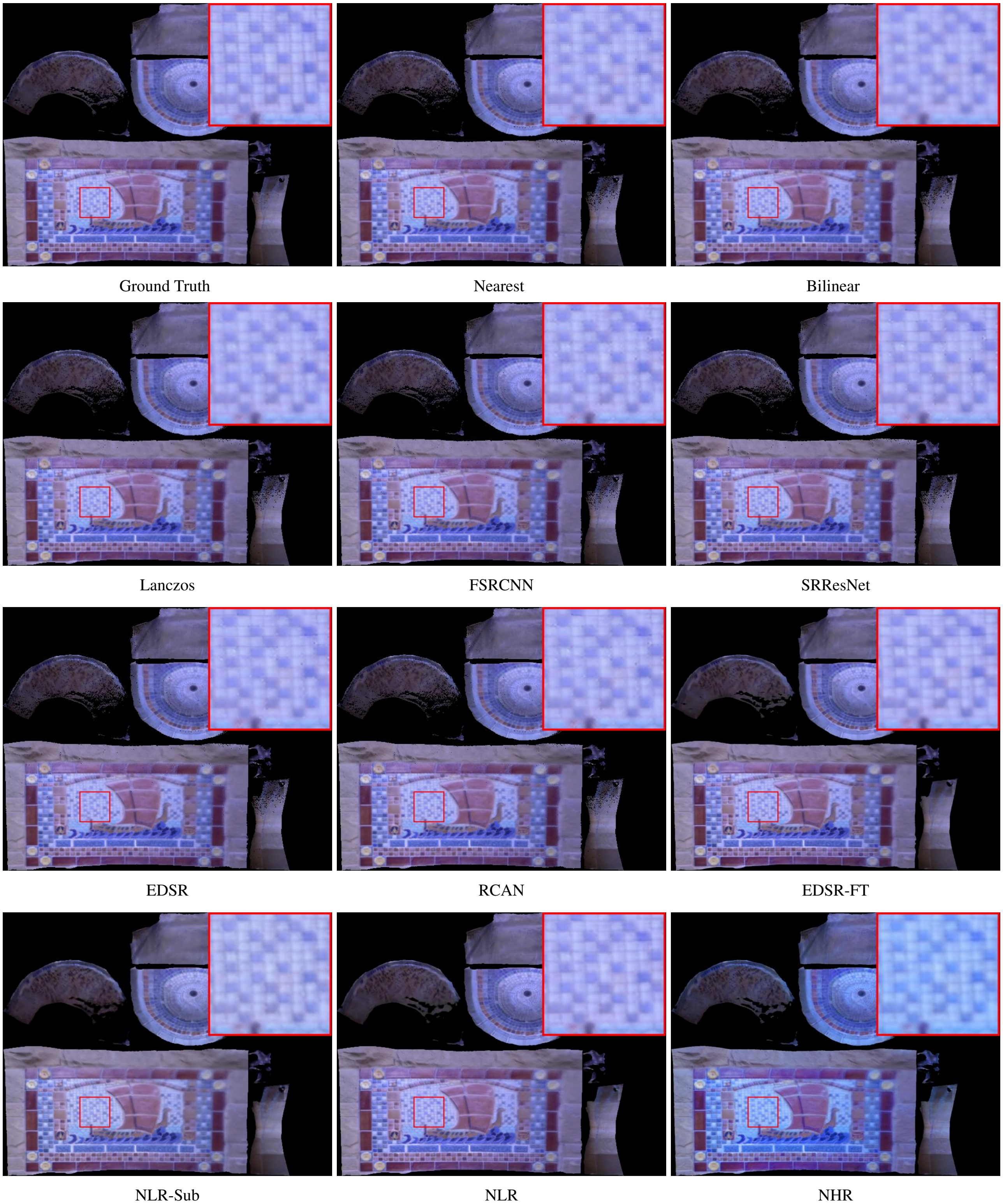

Figure 7: Texture map SR results of *Fountain* by different methods.