3D Reconstruction and Novel View Synthesis of Indoor Environments based on a Dual Neural Radiance Field

Anonymous Authors

1 ADDITIONAL ABLATION STUDY

Effect of the depth alignment loss. We conduct experiments to verify the importance of λ_{align} to color decomposition, as shown in fig. 1. It can be seen that without λ_{align}, the generated depth image is very coarse and the wall and mirror are not smooth while the method with λ_{align} gives the accurate depth image. Besides, without λ_{align} the method seems to fail to decompose color into the view-independent component and the view-dependent component, especially in mirror region and wall regions.

Effect of self-supervised loss. We conduct experiments to verify the impact of different λ_d on reconstruction performance and rendering quality. Experimental results show that an appropriate λ_d value is crucial to balance the effects of color and depth supervision. Specifically, smaller λ_d results in enhancing the effect of the depth loss and reducing the effect of the view-independent color supervision, while larger λ_d increases the effect of color supervision and decreases the effect of depth supervision. We carry out experiments with the different values of λ_d ranging from 0.5 to 50. The results in table 1 show that $\lambda_d = 5$ gives the best performance. In addition, the performance drops whenever λ_d is larger or smaller.

Table 1: Ablation study of self-supervised loss. We evaluate the effect of different values of λ_d on the view synthesis and reconstruction. It can be seen that best performance is obtained when λ_d is 5. Worse results are achieved whenever λ_d increases or decreases.

λ_d	C-l_1 ↓	F-score ↑	PSNR ↑	LPIPS ↓
0.5	0.0132	**0.985**	36.926	0.0344
2.0	0.0134	**0.985**	37.076	0.0353
5.0 (Ours)	**0.0128**	**0.985**	**37.167**	**0.0324**
10.0	0.0132	**0.985**	37.146	0.0332
50.0	0.0131	**0.985**	36.713	0.0359

Table 2: Model storage and runtime of Du-NeRF. We list *Scene size* (m^3), *Model size*, number of parameters (*Params.*) and *Runtime*. Our method reaches convergence in at most 1 hour for all scenes.

Scene	Scene size	Model size	Params.	Runtime
Breakfast room	$4.1 \times 3.3 \times 4.7$	116 MB	28.9 M	50 min
Green room	$8.0 \times 3.1 \times 4.7$	170 MB	42.5 M	53 min
Grey-white room	$5.9 \times 3.1 \times 4.4$	127 MB	31.7 M	50 min
ICL living room	$5.3 \times 2.9 \times 5.4$	128 MB	32.1 M	51 min
Complete kitchen	$9.3 \times 3.3 \times 10.0$	298 MB	74.1 M	73 min
Kitchen	$7.0 \times 3.4 \times 8.7$	229 MB	57.0 M	61 min
Morning apartment	$3.5 \times 2.3 \times 4.0$	81 MB	20.3 M	47 min
Staircase	$6.8 \times 3.7 \times 6.5$	206 MB	51.7 M	63 min
Thin Geometry	$3.4 \times 1.2 \times 3.6$	65 MB	16.3 M	48 min
White room	$5.6 \times 3.8 \times 7.8$	201 MB	50.3 M	59 min
Office0	$4.7 \times 5.3 \times 3.3$	130 MB	32.5 M	41 min
Office1	$5.2 \times 4.5 \times 3.3$	115 MB	28.9 M	42 min
Office2	$6.8 \times 8.5 \times 3.2$	230 MB	57.4 M	52 min
Office3	$9.0 \times 9.7 \times 3.5$	298 MB	74.5 M	70 min
Office4	$6.9 \times 6.9 \times 3.2$	193 MB	48.3 M	48 min
Room0	$8.2 \times 5.1 \times 3.2$	170 MB	42.7 M	46 min
Room1	$7.1 \times 6.1 \times 3.1$	175 MB	43.7 M	45 min
Room2	$7.2 \times 5.3 \times 4.0$	184 MB	45.9 M	46 min

2 ADDITIONAL EXPERIMENTAL RESULTS

2.1 More detailed experiment results

Additional qualitative results are shown in figs. 3 and 4.

We compare the proposed method with *BundleFusion*, *Neus*, *VolSDF*, *NeuralRGBD* and *Go-Surf* in Scene *Grey-white room*, *Office2*, *Office3*, *Room0* for geometry reconstruction, as shown in fig. 3. Our method achieves the best visual result on all indoor scenes. It can be seen from fig. 3 that our method has richer object details and smoother planes of indoor scenes. Specifically, Our method fills in the office chair legs completely and achieves a smoother result on the background walls and floors in the second and third columns. In the last column, our method is able to reconstruct the complete table compared to *BundleFusion*, *Neus* and *NeuralRGBD*, and refine the shape of the bottle compared to *Go-Surf*.

In addition, view synthesis performance on *Whiteroom*, *Office3*, *Room0*, and *Room1* scenes of the *Neus*, *InstantNGP*, *DVGO*, *Go-Surf* and *NeuralRGBD*, are shown in fig. 4. Our approach achieves the best image rendering results, in both the texture-less areas and rich texture areas. For instance, we accurately restore the appearance of texture-less regions such as ceiling corners in the first column and walls in the second and third columns. Additionally, we reproduce the details of complex texture regions, such as the shutter in the third column and the stripes of the quilt in the last column. Methods that focus on view rendering such as *DVGO* and *InstantNGP* suffer in texture-less or sparsely observed regions, as shown in the first and third columns of fig. 4. Approaches focus on surface reconstruction, such as *NeuralRGBD* and *Go-Surf*, often struggle to get good view rendering results on areas of complex texture such as the windows in the third column and the quilt in the fourth column.

The detailed quantitative results on Replica dataset and NeuralRGBD dataset can be seen in table 5 and in tables 3 and 4, respectively. Our method obtains the SOTA rendering performance in all scenes and gives the highest performance in most of the scenes for geometry reconstruction.

2.2 More color decomposition results

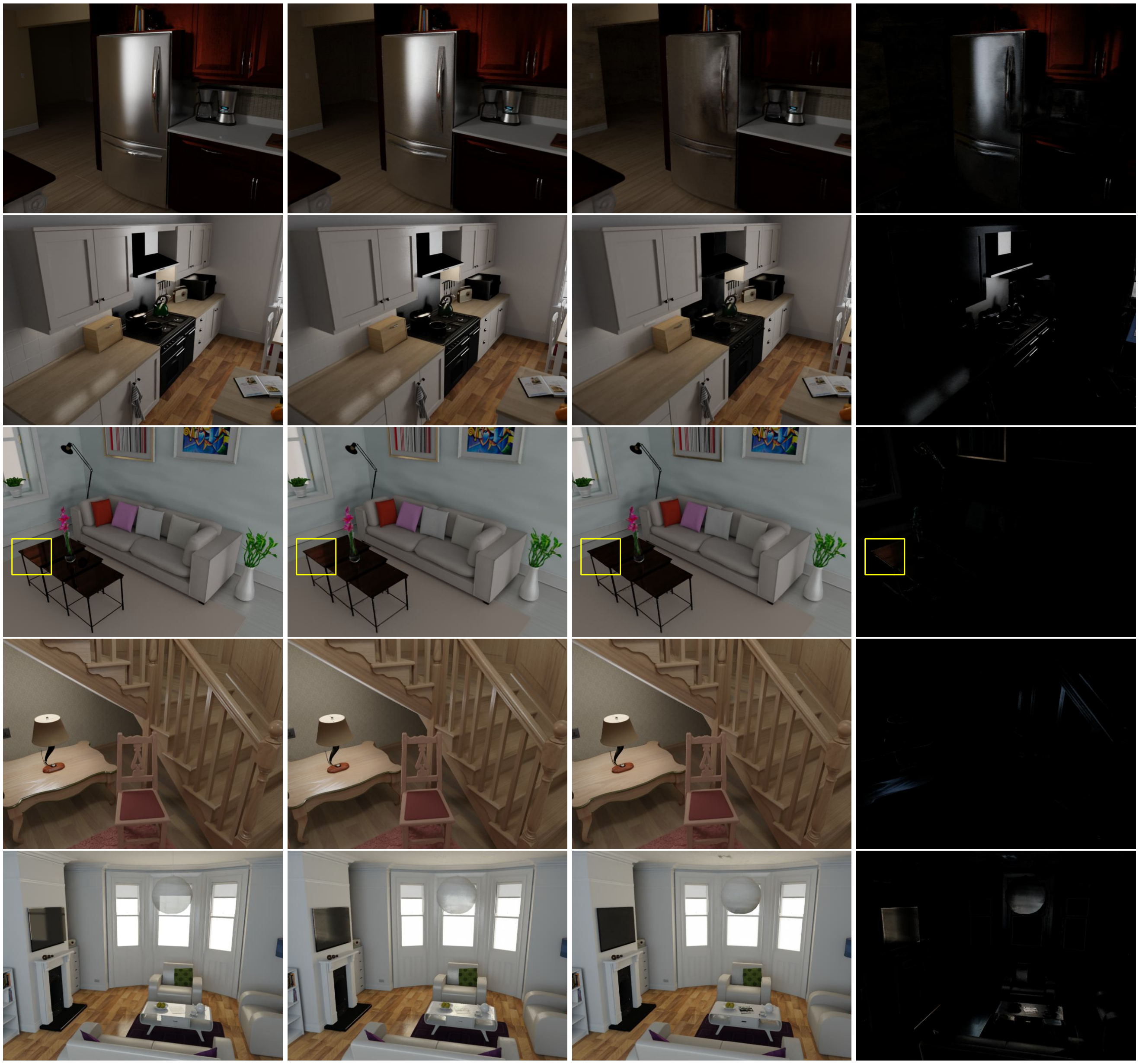
More color decomposition results are shown in fig. 2. We can see that our method successfully decomposes the full color into view-independent color and view-dependent color such as the reflective table, the book and the TV.

(a) Depth (b) V_i color (c) V_d color (d) GT

Figure 1: We evaluates the effect of different λ_{align} on the generated view-independent color. The first row shows the results w/o λ_{align}, while second for w/ λ_{align}. The (a), (b), (c) are the rendering depth map, view-independent color, and view-dependent color, respectively, while (d) is the complete images. The experimental results indicate that we could not get an accurate diffuse component without λ_{align}.

3 RUNTIME AND MEMORY REQUIREMENTS

There is a detailed breakdown of the runtime and memory usage in our experiments on all of the datasets, as shown in table 2. It can be seen that our method could reach convergence in at most 1 hour on all scenes. Similar to previous methods that trade memory for time, Du-NeRF requires hundreds of MB to store the multi-resolution grid.



(a) GT color **(b) Rendering Color** **(c) $\mathrm{V_i}$ color** **(d) $\mathrm{V_d}$ color**

Figure 2: More color decoupling results. The $\mathrm{V_i}$ color represents the view-independent color, while $\mathrm{V_d}$ color is the view-dependent color. It can be seen that our method can effectively decouple the complete color into the view-independent (diffusion surfaces) and view-dependent (specular reflective surfaces) colors.

Scene	Method	Acc ↓	Com ↓	C-l_1 ↓	NC ↑	F-score ↑	PSNR ↑	SSIM ↑	LPIPS ↓
Breakfast Room	BundleFusion	0.0134	0.2000	0.1070	0.9170	0.8000	-	-	-
	Neus	0.6270	0.7720	0.7000	0.6410	0.0100	27.2880	0.8230	0.1940
	VolSDF	0.0820	0.3360	0.2090	0.6880	0.2970	28.9610	0.8730	0.1570
	NeuralRGBD	0.0145	0.0148	0.0147	0.9650	0.9900	32.5340	0.9280	0.1090
	GO-Surf	**0.0141**	0.0150	0.0145	0.9630	0.9810	29.0600	0.8830	0.1650
	InstantNGP	0.2310	0.2970	0.2640	0.5670	0.2420	31.6560	0.8970	0.0853
	DVGO	0.5310	0.9010	0.7160	0.5100	0.0184	34.2720	0.9530	0.0640
	Ours	**0.0141**	**0.0135**	**0.0138**	**0.9640**	**0.9860**	**38.2610**	**0.9830**	**0.0238**
Complete Kitchen	BundleFusion	0.0366	1.0780	0.5570	0.7470	0.4450	-	-	-
	Neus	0.2950	1.5580	0.9270	0.5610	0.0641	26.3270	0.8410	0.2540
	VolSDF	0.3260	0.9100	0.6180	0.6800	0.1470	26.0160	0.8570	0.2400
	NeuralRGBD	**0.0189**	0.1100	0.0647	0.8960	0.8790	32.0310	0.9100	0.2020
	GO-Surf	0.0254	0.0295	0.0274	0.9395	0.8930	29.6040	0.8680	0.1760
	InstantNGP	0.1950	0.8730	0.5340	0.5740	0.2350	30.4066	0.8850	0.1470
	DVGO	0.2970	1.1580	0.7280	0.5610	0.1990	31.0570	0.9010	0.2180
	Ours	0.0222	**0.0259**	**0.0240**	**0.9410**	**0.9010**	**34.5170**	**0.9520**	**0.0682**
Green Room	BundleFusion	0.0140	0.1965	0.1053	0.9090	0.8140	-	-	-
	Neus	0.2130	0.3430	0.2780	0.7400	0.1130	29.4000	0.8930	0.1450
	VolSDF	0.1360	0.4320	0.2840	0.6490	0.2250	29.8840	0.9100	0.1380
	NeuralRGBD	**0.0104**	**0.0140**	**0.0122**	**0.9340**	**0.9910**	34.0800	0.9520	0.0770
	GO-Surf	0.0124	0.0156	0.0140	0.9275	0.9825	31.0520	0.9220	0.1170
	InstantNGP	0.2440	0.9710	0.6070	0.5370	0.1180	34.9470	0.9460	0.0364
	DVGO	0.2940	0.5190	0.4070	0.5640	0.2490	34.9800	0.9550	0.0760
	Ours	0.0122	0.0150	0.0136	0.9290	0.9850	**38.5530**	**0.9780**	**0.0247**
Grey White Room	BundleFusion	0.0202	0.2743	0.1472	0.8230	0.7380	-	-	-
	Neus	0.2620	0.4820	0.3720	0.6290	0.1160	28.7820	0.8630	0.1640
	VolSDF	0.1930	0.3360	0.2650	0.7070	0.2550	30.4400	0.8950	0.1480
	NeuralRGBD	**0.0134**	**0.0151**	**0.0143**	**0.9310**	**0.9940**	35.1630	0.9470	0.0900
	GO-Surf	0.0145	0.0167	0.0156	0.9255	0.9875	30.8900	0.9115	0.1490
	InstantNGP	0.1390	0.9380	0.5390	0.5010	0.1810	32.0200	0.8790	0.1270
	DVGO	0.2410	0.4430	0.3420	0.5640	0.2990	34.7160	0.9470	0.0930
	Ours	0.0140	0.0155	0.0147	0.9260	0.9900	**37.7320**	**0.9700**	**0.0406**
Icl Living Room	BundleFusion	0.0104	0.2697	0.1400	0.9120	0.7720	-	-	-
	Neus	0.3570	0.8040	0.5810	0.6270	0.1000	31.9550	0.9010	0.1090
	VolSDF	0.2520	0.8130	0.5330	0.6370	0.1450	30.6400	0.8980	0.1400
	NeuralRGBD	**0.0089**	0.0840	0.0462	0.9070	0.9010	34.3810	0.9300	0.1960
	GO-Surf	0.0101	0.0129	0.0115	0.9670	0.9910	31.7410	0.9080	0.2425
	InstantNGP	0.7180	1.8900	1.3000	0.5140	0.0100	27.1670	0.7650	0.2760
	DVGO	0.3000	0.8670	0.5830	0.5430	0.2130	33.9930	0.9330	0.2420
	Ours	0.0112	**0.0140**	**0.0126**	**0.9690**	**0.9920**	**36.9860**	**0.9570**	**0.0638**

Table 3: Reconstruction and view synthesis results of NeuralRGBD dataset. The best performances are highlighted in bold.

Scene	**Method**	**Acc ↓**	**Com ↓**	**C-l_1 ↓**	**NC ↑**	**F-score ↑**	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**
Kitchen	BundleFusion	0.0170	0.5960	0.3065	0.8510	0.6390	-	-	-
	Neus	0.4680	0.8490	0.6580	0.5740	0.0400	25.5880	0.8240	0.2340
	VolSDF	0.2130	0.3760	0.2950	0.7000	0.2810	25.7570	0.8520	0.2300
	NeuralRGBD	**0.0198**	**0.1450**	0.0824	0.9000	0.8630	31.6270	0.9170	0.1480
	GO-Surf	0.0204	0.0265	**0.0235**	**0.9340**	**0.9430**	27.8260	0.8735	0.2000
	InstantNGP	0.1620	0.6470	0.4050	0.5520	0.2360	29.3850	0.8650	0.1380
	DVGO	0.2690	0.5120	0.3900	0.5640	0.3340	30.5230	0.9360	0.1310
	Ours	0.0207	0.0266	0.0236	0.9330	0.9400	**35.6100**	**0.9650**	**0.0458**
Morning Apartment	BundleFusion	0.0093	0.0153	0.0123	0.8880	0.9760	-	-	-
	Neus	0.2130	0.2430	0.2280	0.6660	0.2180	27.5640	0.8370	0.2090
	VolSDF	0.0804	0.1450	0.1130	0.7300	0.3700	29.2440	0.8870	0.1660
	NeuralRGBD	**0.0088**	**0.0117**	**0.0103**	**0.8920**	**0.9870**	33.1350	0.9270	0.1080
	GO-Surf	0.0106	0.0145	0.0125	0.8840	0.9750	28.3880	0.8570	0.2245
	InstantNGP	0.4260	0.4920	0.4590	0.5030	0.1000	24.2740	0.6230	0.4640
	DVGO	0.1510	0.1760	0.1630	0.5480	0.5100	33.9590	0.9480	0.0740
	Ours	0.0099	0.0136	0.0118	0.8870	0.9780	**36.6370**	**0.9660**	**0.0392**
Staircase	BundleFusion	0.0160	1.0020	0.5088	0.7960	0.4310	-	-	-
	Neus	0.3800	0.8910	0.6360	0.6190	0.1140	29.4540	0.8430	0.2710
	VolSDF	0.1070	0.7340	0.4200	0.6720	0.2440	30.9470	0.8490	0.2570
	NeuralRGBD	**0.0213**	0.0441	0.0327	0.9420	0.9010	34.8610	0.9070	0.2210
	GO-Surf	0.0233	0.0285	0.0259	0.9490	0.8855	32.1520	0.8805	0.2560
	InstantNGP	0.2880	0.3240	0.3060	0.5820	0.3130	31.1140	0.8250	0.2320
	DVGO	0.2710	0.7460	0.5080	0.5750	0.2580	34.5150	0.9160	0.2200
	Ours	0.0221	**0.0268**	**0.0244**	**0.9510**	**0.9280**	**36.7910**	**0.9550**	**0.0710**
Thin Geometry	BundleFusion	0.0227	0.0762	0.0495	0.8640	0.7160	-	-	-
	Neus	0.1550	0.5760	0.3650	0.5380	0.0860	19.2300	0.7830	0.1760
	VolSDF	0.1140	0.3960	0.2550	0.5990	0.2710	24.7540	0.8900	0.1360
	NeuralRGBD	0.0093	0.0254	0.0173	**0.9090**	0.9420	18.9310	0.6600	0.5580
	GO-Surf	0.0107	0.0186	0.0146	0.9005	0.9440	25.8900	0.8820	0.1400
	InstantNGP	0.0821	0.4290	0.2560	0.6130	0.3070	32.4840	0.9330	0.0347
	DVGO	0.1040	0.0750	0.0900	0.5720	0.4750	35.1540	0.9680	0.0330
	Ours	**0.0087**	**0.0122**	**0.0105**	0.9070	**0.9790**	**34.1260**	**0.9640**	**0.0286**
Whiteroom	BundleFusion	0.0184	0.8690	0.4440	0.8060	0.4700	-	-	-
	Neus	0.2040	0.3930	0.2980	0.6820	0.1710	29.0620	0.8770	0.1610
	VolSDF	0.1240	0.3370	0.2300	0.7450	0.3870	30.5300	0.9050	0.1230
	NeuralRGBD	**0.0202**	0.0437	0.0320	0.9200	0.9098	33.1980	0.9330	0.1240
	GO-Surf	0.0225	0.0359	0.0292	0.9285	0.9070	29.2535	0.8985	0.1640
	InstantNGP	0.1560	0.4570	0.3060	0.6060	0.3370	31.7830	0.9090	0.0956
	DVGO	0.2230	0.5060	0.3650	0.5880	0.3130	33.1580	0.9400	0.0970
	Ours	0.0214	**0.0340**	**0.0277**	**0.9270**	**0.9210**	**36.2280**	**0.9690**	**0.0409**

Table 4: Reconstruction and view synthesis results of NeuralRGBD dataset. The best performances are highlighted in bold.

Methods	Evaluation	office0	office1	office2	office3	office4	room0	room1	room2	Mean
BundleFusion	C-l_1 ↓	0.0109	0.0110	0.0310	0.0660	0.0165	0.0587	0.0110	0.0339	0.0299
	F-score ↑	0.9880	0.9910	0.9190	0.8410	0.9690	0.8820	0.9900	0.9100	0.9360
	PSNR ↑	-	-	-	-	-	-	-	-	-
	LPIPS ↓	-	-	-	-	-	-	-	-	-
Neus	C-l_1 ↓	0.0154	0.2280	0.2390	0.2000	0.3440	0.2140	0.1570	0.2940	0.2288
	F-score ↑	0.2340	0.1800	0.1730	0.3050	0.0850	0.2300	0.2210	0.1240	0.1940
	PSNR ↑	33.274	33.874	26.372	26.893	28.546	25.913	28.068	28.570	28.939
	LPIPS ↓	0.1570	0.1240	0.1790	0.1640	0.1660	0.2660	0.2120	0.1790	0.1810
VolSDF	C-l_1 ↓	0.1270	0.2050	0.1820	0.2830	0.2980	0.2750	0.1580	0.2160	0.2180
	F-score ↑	0.4200	0.2440	0.4150	0.3250	0.2980	0.3280	0.3810	0.3010	0.3390
	PSNR ↑	34.327	35.733	28.996	27.943	30.595	26.374	28.807	30.227	30.375
	LPIPS ↓	0.1570	0.1120	0.1650	0.1730	0.1590	0.2660	0.2140	0.1560	0.1750
NeuralRGBD	C-l_1 ↓	0.0635	**0.0088**	0.0544	0.0177	0.3340	0.0892	0.1010	0.3480	0.1271
	F-score ↑	0.8550	0.9890	0.9380	0.9790	0.7040	0.8740	0.8420	0.5960	0.8470
	PSNR ↑	36.563	38.148	31.559	30.509	33.934	28.311	30.866	31.452	32.668
	LPIPS ↓	0.1640	0.1590	0.2000	0.1840	0.1600	0.2750	0.2410	0.2000	0.1980
GO-Surf	C-l_1 ↓	0.0099	0.0104	0.0125	0.0156	0.0133	0.0131	0.0103	0.0117	0.0121
	F-score ↑	0.9880	0.9925	0.9875	0.9825	0.9890	0.9935	0.9960	**0.9880**	0.9896
	PSNR ↑	35.105	35.655	29.059	29.138	31.784	27.378	29.522	30.096	30.967
	LPIPS ↓	0.1620	0.1740	0.2355	0.1940	0.1835	0.2975	0.2650	0.2270	0.2170
InstantNGP	C-l_1 ↓	0.2560	1.2320	0.4940	0.4940	1.1500	0.7870	0.6270	0.7900	0.7288
	F-score ↑	0.2010	0.1020	0.2660	0.2660	0.1030	0.1050	0.1510	0.0720	0.1583
	PSNR ↑	36.667	37.679	30.793	30.793	32.772	26.912	30.786	32.411	32.352
	LPIPS ↓	0.1710	0.0754	0.1240	0.1240	0.2140	0.2620	0.1350	0.0976	0.1500
DVGO	C-l_1 ↓	0.2160	0.2790	0.2010	0.3270	0.3660	0.4160	0.2470	0.3120	0.2955
	F-score ↑	0.3170	0.2040	0.3010	0.2470	0.1330	0.2360	0.2630	0.2110	0.2390
	PSNR ↑	36.859	37.878	31.596	28.349	32.583	26.985	30.250	31.192	31.962
	LPIPS ↓	0.1670	0.1640	0.2100	0.2260	0.2100	0.3500	0.2510	0.2090	0.2230
Du-NeRF	C-l_1 ↓	**0.0096**	0.0102	**0.0116**	**0.0141**	**0.0118**	**0.0119**	**0.0093**	**0.0108**	**0.0112**
	F-score ↑	**0.9890**	**0.9930**	**0.9900**	**0.9860**	**0.9910**	**0.9950**	**0.9970**	**0.9880**	**0.9911**
	PSNR ↑	**41.365**	**41.922**	**34.913**	**34.622**	**37.776**	**34.184**	**36.123**	**35.928**	**37.104**
	LPIPS ↓	**0.0546**	**0.0630**	**0.0933**	**0.0790**	**0.0775**	**0.0824**	**0.0673**	**0.0790**	**0.0740**

Table 5: Reconstruction and view synthesis results of Replica dataset. The best performances are highlighted in bold.

Figure 3: We show additional reconstruction results on the scene *Grey-white room*, *Office2*, *Office3* and *Room0*. Our approach allows for rich details and smoother planes highlighted in the yellow box.

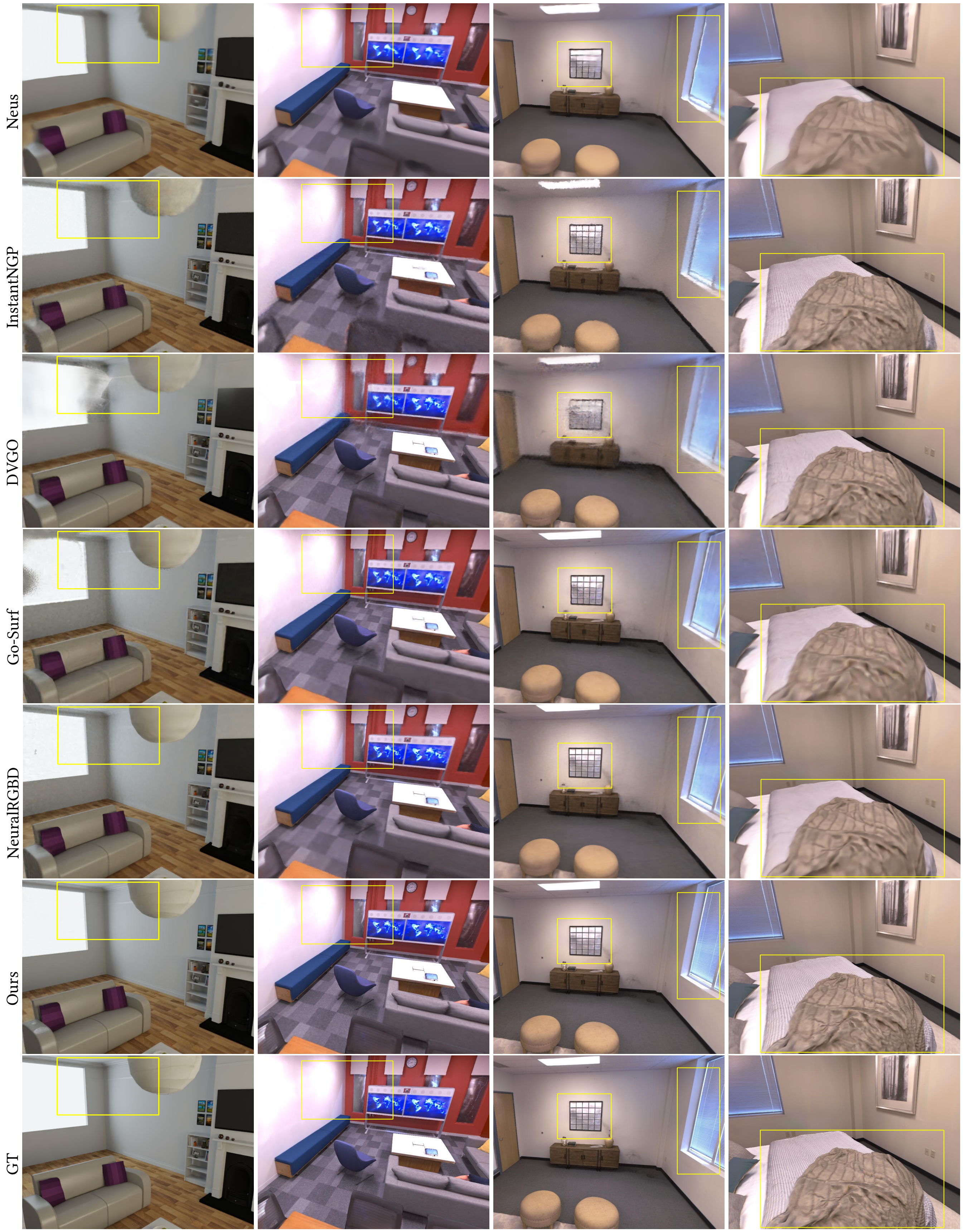


Figure 4: Additional results of view synthesis on scenes *Whiteroom, Office3, Room0, Room1.* The methods for novel view synthesis, such as *InstantNGP, DVGO*, fail to render clear results at texture-less regions, and methods focusing on geometry reconstruction, such as *Neural-RGBD, Go-Surf*, fail to restore the appearance of the regions with complex texture.