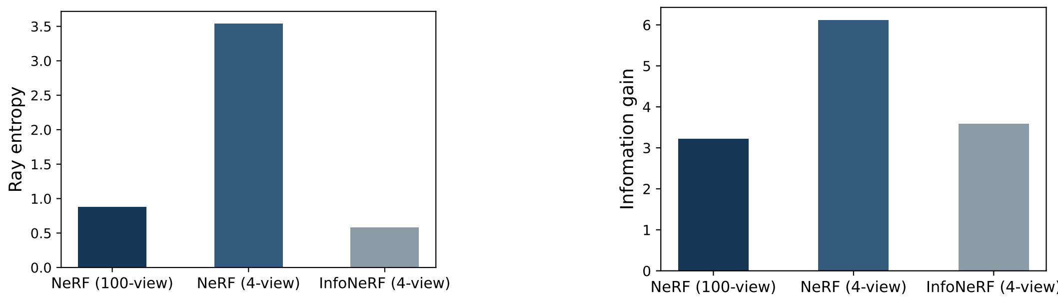
A. Validation for $\mathcal{L}_{\text{entropy}}$ and $\mathcal{L}_{\text{KL}}$

To analyze the behavior of our two regularization losses, we measure the values of ray entropy and information gain of NeRF models after training with different settings. Figure A illustrates that both ray entropy and information gain values of the few-shot NeRF model are much higher than those of the NeRF model with 100 views. This observation implies that the rendered images by the few-shot NeRF are noisier and the reduction of the two losses is desirable. Note that InfoNeRF with 4 views has almost similar values to the original NeRF model with 100 views, which presents that the proposed algorithm provides compact and consistent reconstruction results even with a small number of views.



(a) Ray entropy values after training

(b) Information gain values after training

Figure A. Quantitative comparisons of ray entropy and information gain computed in three different models. Although the few-shot NeRF model gives higher ray entropy and information gain than the NeRF with 100 views, the proposed few-shot InfoNeRF model reduces both values, which motivates the two regularization loss employed in our approach.

B. Experiments on More Complex Scenarios

To validate our algorithm in more challenging scenarios, we test InfoNeRF on a more complex real-world dataset, LLFF [4], which contains 8 natural scenes captured by a handheld cellphone. Each scene has 20 to 62 images, and we hold out 1/8 of the images as test sets following the standard protocol [4, 5]. For training, we sample 2 views in each scene. Table A presents that our algorithm still outperforms the baseline by large margins and Figure B illustrates InfoNeRF provides better visual quality with less blurring artifact.

Table A. Results on a complex dataset, LLFF [4] in the 2-view setting.

	PSNR ↑	SSIM ↑	LPIPS ↓
NeRF	12.93	0.267	0.554
InfoNeRF	**14.37**	**0.349**	**0.457**

(a) Ground-truth (b) NeRF [5] (c) InfoNeRF (d) Ground-truth (e) NeRF [5] (f) InfoNeRF

Figure B. Qualitative comparisons on the *Flower*, *T-Rex*, *Orchids*, and *Fern* scenes of the LLFF dataset in the 2-view setting.

C. Role of $\mathcal{L}_{\text{KL}}$ in Narrow-Baseline Data

InfoNeRF employ $\mathcal{L}_{\text{KL}}$ to prevent overfitting, especially for training with narrow-baseline images, and the loss term is particularly helpful for DTU. To analyze this property more thoroughly, we create two sets of views in the *Lego* scene of Realistic Synthetic 360° with narrow- and wide-baselines, where Figure C visualizes the selected images for each set.

Table B shows the full ablation results in both settings. $\mathcal{L}_{\text{KL}}$ with $\mathcal{L}_{\text{entropy}}$ is effective for the narrow-baseline setting, as the ablation results on the DTU dataset in the Table 6 of the main paper, while degrading accuracy slightly for the wide-baseline setting. Also, $\mathcal{L}_{\text{KL}}$ doesn't work well without $\mathcal{L}_{\text{entropy}}$ because enforcing the smoothness constraint between neighborhood rays is not helpful for a noisy reconstructed scene without $\mathcal{L}_{\text{entropy}}$.

Table B. Effects of the $\mathcal{L}_{\text{KL}}$ with respect to viewpoint variations on the *Lego* scene of Realistic Synthetic 360° in 4-view setting.

Method	$\mathcal{L}_{\text{entropy}}$	$\mathcal{L}_{\text{KL}}$	narrow-baseline			wide-baseline		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
NeRF			13.16	0.753	0.350	18.03	0.763	0.248
InfoNeRF w/o $\mathcal{L}_{\text{entropy}}$		✓	13.33	0.750	0.349	17.96	0.753	0.254
InfoNeRF w/o $\mathcal{L}_{\text{KL}}$	✓		16.74	0.767	0.241	**19.51**	**0.792**	**0.189**
InfoNeRF	✓	✓	**18.41**	**0.775**	**0.199**	19.28	0.789	0.190

(a) Sampled view images in the narrow-baseline training set.

(b) Sampled view images in the wide-baseline training set.

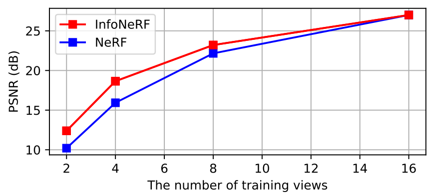
Figure C. Training set with the narrow- and wide-baselines for *Lego* scene of Realistic Synthetic 360° dataset in 4-view setting.

(a) Ground-truth (b) NeRF [5] (c) InfoNeRF w/o $\mathcal{L}_{\text{entropy}}$ (d) InfoNeRF w/o $\mathcal{L}_{\text{KL}}$ (e) InfoNeRF

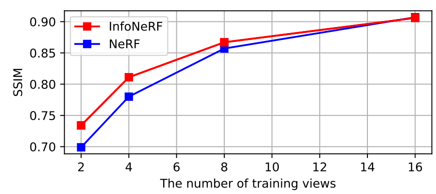
Figure D. Qualitative comparison on the *Lego* scene of the Realistic Synthetic 360° dataset in the narrow-baseline setting with 4 views.

D. Robustness to the Number of Training Views

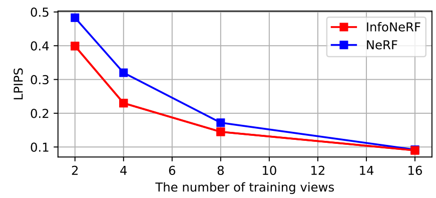
Figure E shows PSNR, SSIM, and LPIPS results of our model and baseline by varying the number of training views on the Realistic Synthetic 360° dataset, supplementing Figure 6 of the main paper, where all metrics give consistent tendencies.



(a) PSNR



(b) SSIM

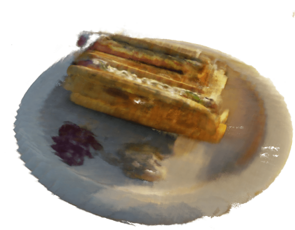


(c) LPIPS

Figure E. PSNR, SSIM and LPIPS results with respect to the number of training views on the Realistic Synthetic 360° dataset.

(a) Ground-truth

(b) NeRF [5]



(c) InfoNeRF

Figure F. Qualitative comparison on the *HotDog* scene of the Realistic Synthetic 360° dataset in the 2-view setting.

E. Integration into Other NeRF-based Models

To demonstrate the generality of our method, we incorporate the proposed regularization technique to PixelNeRF [7] and mipNeRF [1], and refer to this version of our model as InfoPixelNeRF and InfoMipNeRF, respectively. Tables C presents the qualitative results of InfoPixelNeRF and InfoMipNeRF with their baselines, where we trained all comparisons from scratch with 4 views. PixelNeRF and MipNeRF show better results than NeRF, partly thanks to their semantic prior and cone-based positional encoding, respectively, but InfoPixelNeRF and InfoMipNeRF still achieve performance gains in terms of all metrics compared to their baselines.

Table C. Experimental results of InfoPixelNeRF and InfoMipNeRF with their baselines in 4-view setting on the Realistic Synthetic 360° dataset, where all results are based on 5 runs. $*$ denotes that the model is pretrained on external training dataset with dense input views and finetuned on this dataset with a few input views.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	KID ↓
NeRF [5]	15.93±1.06	0.780±0.014	0.320±0.049	215.16±2.32	0.0740±0.0123
PixelNeRF* [7]	16.09±0.78	0.738±0.012	0.390±0.030	265.25±6.73	0.1274±0.0063
InfoPixelNeRF* (ours)	**16.30±0.80**	**0.745±0.015**	**0.372±0.030**	**264.33±6.21**	**0.1263±0.0058**
mipNeRF [1]	19.04±0.02	0.814±0.000	0.219±0.000	142.96±1.89	0.0422±0.0010
InfoMipNeRF (ours)	**19.33±0.04**	**0.818±0.000**	**0.213±0.001**	**142.95±0.54**	**0.0418±0.0004**

F. Per-Scene Breakdown on the Realistic Synthetic 360° Dataset

Supplementing Table 1 of the main paper, we show the experimental results from individual scenes in terms of PSNR, SSIM, and LPIPS in the subsequent tables (Table E, F, and G). Table D is the same as Table 1 of the main paper, added for convenience. As shown in the tables, InfoNeRF achieves consistent and meaningful improvement over its baselines, considering the standard deviations of the results.

Table D. Experimental results of few-shot novel view synthesis on the Realistic Synthetic 360° dataset in 4-view setting. $*$ denotes that the model is pretrained on external training dataset with dense input views and finetuned on this dataset with a few input views.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	KID ↓
NeRF, 100 views	31.01	0.947	0.081	42.83	0.002
PixelNeRF* [7]	16.09±0.78	0.738±0.012	0.390±0.030	265.25±6.73	0.127±0.006
NeRF [5]	15.93±1.06	0.780±0.014	0.320±0.049	215.16±2.32	0.074±0.012
DietNeRF [2]	16.06±1.13	0.793±0.019	0.306±0.050	197.02±12.87	0.065±0.004
InfoNeRF (ours)	**18.65**±**0.18**	**0.811**±**0.008**	**0.230**±**0.008**	**181.47**±**4.97**	**0.062**±**0.004**

Table E. Average PSNRs and standard deviations of individual scenes on the Realistic Synthetic 360° dataset in 4-view setting.

Method	Lego	Chair	Drums	Ficus	Hotdog	Materials	Mic	Ship	Avg.
NeRF, 100 views	32.54	33.00	25.01	30.13	36.18	29.62	32.91	28.65	31.01
PixelNeRF* [7]	15.14±0.75	18.87±1.38	15.10±0.63	16.60±0.70	19.37±1.78	12.31±1.02	16.35±0.97	14.96±0.75	16.09±0.78
NeRF [5]	15.61±4.53	18.57±1.64	12.50±0.98	16.37±2.24	19.64±2.26	15.65±4.16	14.78±2.37	14.30±4.04	15.93±1.06
DietNeRF [2]	17.13±4.77	19.37±3.12	13.74±1.55	15.76±3.56	18.24±5.28	15.00±5.18	17.71±1.55	11.51±4.27	16.06±1.13
InfoNeRF (ours)	**18.92**±**0.51**	**20.06**±**1.11**	**14.33**±**0.62**	**19.41**±**0.07**	**21.30**±**2.31**	**18.34**±**0.88**	**18.55**±**1.71**	**18.27**±**0.71**	**18.65**±**0.18**

Table F. Average SSIMs and standard deviations of individual scenes on the Realistic Synthetic 360° dataset in the 4-view setting.

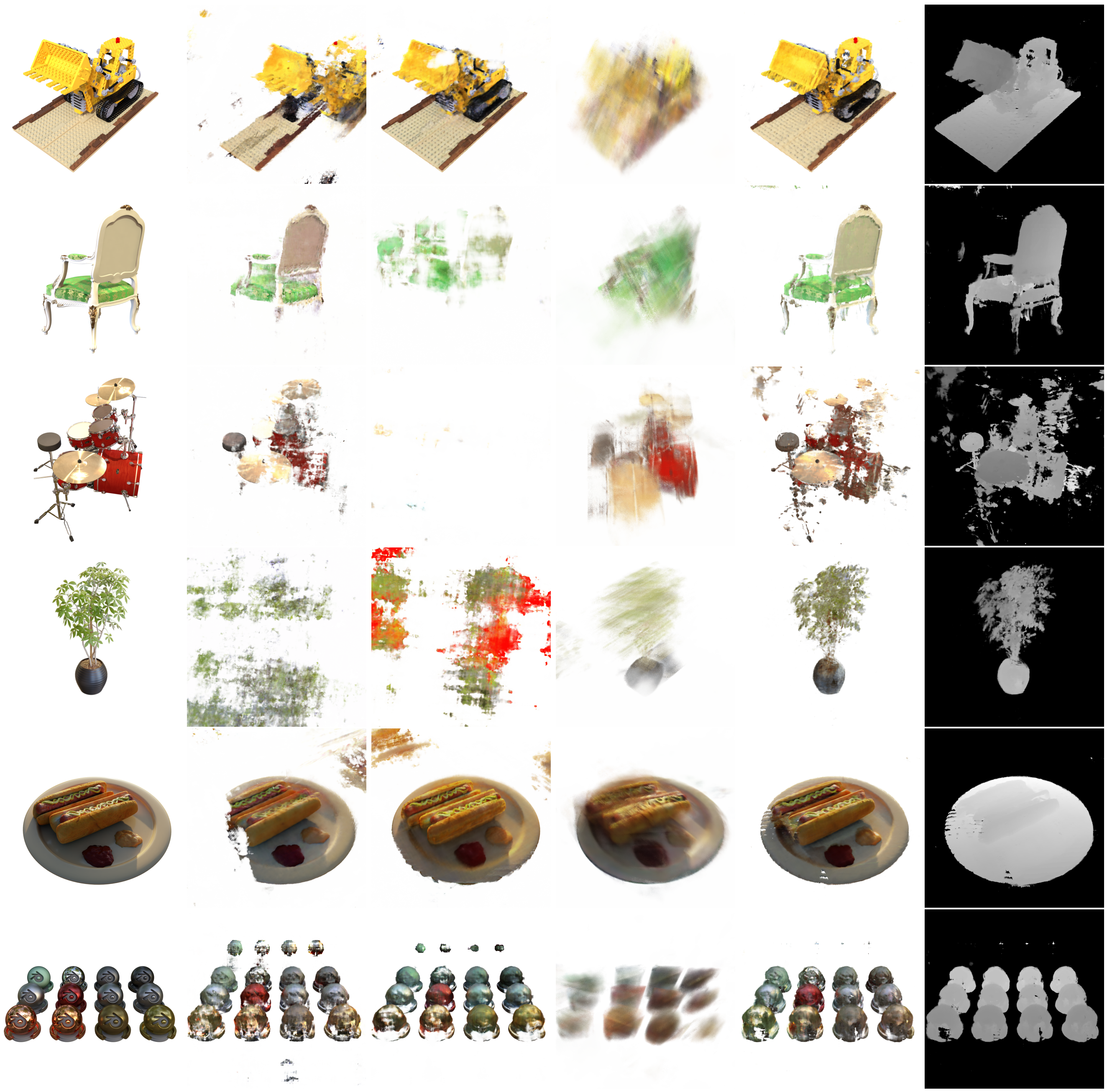
Method	Lego	Chair	Drums	Ficus	Hotdog	Materials	Mic	Ship	Avg.
NeRF, 100 views	0.961	0.967	0.925	0.964	0.974	0.949	0.980	0.856	0.947
PixelNeRF* [7]	0.703±0.014	0.802±0.026	0.699±0.016	0.802±0.017	0.836±0.023	0.644±0.027	0.767±0.021	0.655±0.014	0.738±0.012
NeRF [5]	0.739±0.065	0.818±0.019	0.721±0.022	0.833±0.030	0.863±0.024	0.768±0.070	0.826±0.043	0.675±0.047	0.780±0.014
DietNeRF [2]	0.766±0.079	**0.846**±**0.022**	**0.750**±**0.021**	0.812±0.046	0.851±0.070	0.789±0.050	0.879±0.028	0.649±0.057	0.793±0.019
InfoNeRF (ours)	**0.788**±**0.008**	0.840±0.011	0.730±0.015	**0.851**±**0.001**	**0.871**±**0.027**	**0.799**±**0.052**	**0.883**±**0.012**	**0.723**±**0.012**	**0.811**±**0.008**

Table G. Average LPIPS's and standard deviations of individual scenes on the Realistic Synthetic 360° dataset in the 4-view setting.

Method	Lego	Chair	Drums	Ficus	Hotdog	Materials	Mic	Ship	Avg.
NeRF (100 views)	0.050	0.046	0.091	0.044	0.121	0.063	0.028	0.206	0.081
PixelNeRF* [7]	0.410±0.031	0.325±0.065	0.462±0.026	0.326±0.039	0.266±0.050	0.490±0.034	0.395±0.033	0.446±0.037	0.390±0.030
NeRF [5]	0.318±0.149	0.284±0.047	0.452±0.057	0.224±0.089	0.232±0.049	0.294±0.178	0.351±0.094	0.412±0.095	0.320±0.049
DietNeRF [2]	0.285±0.178	0.314±0.152	**0.304**±**0.094**	0.315±0.173	0.229±0.113	0.324±0.200	0.210±0.059	0.464±0.137	0.306±0.050
InfoNeRF (ours)	**0.182**±**0.020**	**0.196**±**0.016**	0.374±0.026	**0.148**±**0.011**	**0.188**±**0.044**	**0.218**±**0.073**	**0.207**±**0.038**	**0.324**±**0.015**	**0.230**±**0.008**

G. Additional Qualitative Results

G.1. Realistic Synthetic 360°



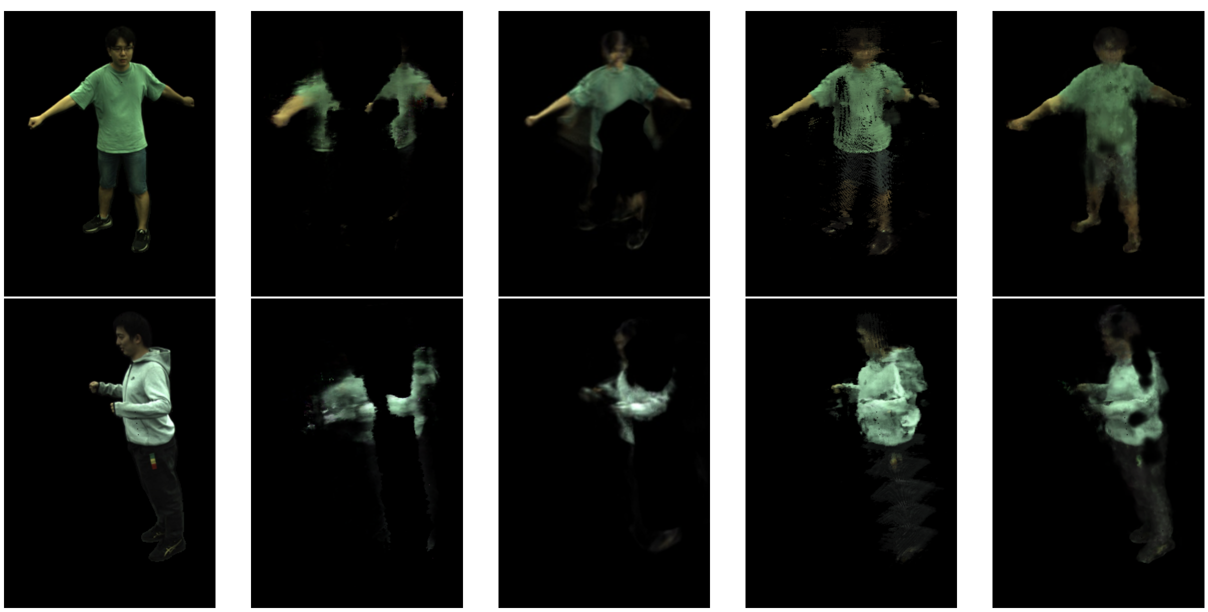
(a) Ground-truth (b) NeRF [5] (c) DietNeRF [2] (d) PixelNeRF [7] (e) InfoNeRF (f) InfoNeRF (depth)

Figure G. Qualitative comparison of our method with other NeRF-based models on the *Lego*, *Chair*, *Drums*, *Ficus*, *HotDog* and *Materials* scenes of the Realistic Synthetic 360° dataset in 4-view setting. Existing works often suffer from noise (b), color distortion (c), or blur effect (d), while InfoNeRF provides distinguished rendering quality. Figure (f) visualizes depth maps estimated by InfoNeRF, which provide clear boundaries and fine details.

(a) Ground-truth (b) NeRF [5] (c) DietNeRF [2] (d) PixelNeRF [7] (e) InfoNeRF (f) InfoNeRF (depth)

Figure H. Additional qualitative comparison of our method with other NeRF-based models on the *Mic* and *Ship* scenes of the Realistic Synthetic 360° dataset in 4-view setting. As same with Figure G, InfoNeRF provides clear boundaries and fine details compared to existing works.

G.2. ZJU-MoCap



(a) Ground-truth (b) NeRF [5] (c) NV [3] (d) InfoNeRF (e) NB [6]

Figure I. Qualitative comparison on the ZJU-MoCap dataset in 4-view setting. We visualized the rendering results of prior-free algorithms (b-d), including ours, and prior-based algorithm (e). While existing prior-free algorithms (b-c) often suffer from inconsistent reconstruction and missing parts of the human body, InfoNeRF can render most of the human body comparable to prior-based algorithm (e).

H. Potential Negative Societal Impact & Limitations

The proposed few-shot view synthesis algorithm works well with only a small number of views, so it is rather more vulnerable to adversarial attacks, which can be problematic when it is applied to real-world scenarios, *e.g.*, AR/VR systems. InfoNeRF shows outstanding results on the few-shot volume rendering, but is still struggling from the need for calibrated cameras albeit few in number, which we leave as future work.

References

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *ICCV*, 2021. 3

[2] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *ICCV*, 2021. 4, 5, 6

[3] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *ACM TOG*. 6

[4] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM TOG*, 2019. 1

[5] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 2, 3, 4, 5, 6

[6] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *CVPR*, 2021. 6

[7] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *CVPR*, 2021. 3, 4, 5, 6