Supplementary Material
Sharp-NeRF: Grid-based Fast Deblurring Neural Radiance Field using Sharpness Prior

Byeonghyeon Lee[1*] Howoong Lee[2,3*] Usman Ali[2†] Eunbyung Park[1,2†]
[1]Department of Artificial Intelligence, Sungkyunkwan University
[2]Department of Electrical and Computer Engineering, Sungkyunkwan University
[3]Hanwha Vision

In this supplementary material, we provide more details on different modules of our method and describe results of experiments that we mentioned in the main paper.

1. Additional Ablation Studies

1.1. Discrete Kernel: Fixed Vs Learnable

The recently proposed Hybrid Neural Rendering Model [3] also employs several grid kernels to render clean image from the camera motion blurred images, but they are fixed. It applies convolution to clean rendered image with all blur kernels and picks the one that produces the minimal loss with the ground truth image. Fig. 1 shows rendered images from our method and our model that uses fixed discrete kernels as Hybrid Neural Rendering model does. We can see that the blur kernels cannot estimate spatially varying defocus blur accurately and fail to render clean image. We assume this is because the point spread function of defocus blur is more complex than that of camera motion blur. Defocus blur depends on various tangled optical factors, and cannot be easily estimated with fixed blur kernels.

1.2. Patch Size Analysis

Patch size is a key factor that determines the rendering quality and the training time. Larger patches mean more shared neighbors, so they can effectively decrease training time. However, rendering quality degrades as the number of patches will decrease unless we do not raise the batch size. Table 1 shows the quality metrics for rendered images, the number of required rays for each pixel, and the training times required for various patch sizes P. The number of the target pixels P'^2 for $P = 26$, $P = 22$, $P = 18$ is 16^2, 12^2, 8^2, respectively. These results show that increasing the patch size reduces the number of required neighboring pixels, which in turn reduces the training time, but results in quality degradation.

Table 1. Ablation study: Patch size evaluated for real defocus dataset. The average number of required neighboring rays per pixel is denoted as #Rays.

Patch Size	PSNR↑	SSIM↑	Brisque↓	Niqe↓	#Rays	Time
$P = 26$	23.44	0.7228	29.50	3.44	2.64	0.40
$P = 22$	23.55	0.7258	30.25	3.41	3.36	0.48
$P = 18$	23.57	0.7501	29.64	3.38	5.06	0.57

Table 2. Compassion of our method with PDRF with its different number of neighboring rays. Models are evaluated on real defocus dataset. 'n' in PDRF-n stands for the number of required neighboring rays. * means that we modified PDRF to analyze its performance for this setting.

Methods	PSNR↑	SSIM↑	Brisque↓	Niqe↓	Time
PDRF-5	23.82	0.7382	32.18	3.75	1.00
PDRF-3*	23.50	0.7217	33.35	3.92	0.88
PDRF-2*	22.34	0.6713	36.74	4.09	0.70
Ours	23.55	0.7258	30.25	3.41	0.48

1.3. Comparison with Variants of PDRF [13]

In this subsection, we compare our proposed method Sharp-NeRF with the very recently proposed method PDRF [13]. We compared them in terms of their training time and the image quality of their rendered images using various metrics. As the required number of neighboring rays is set to 5 in PDRF [13], PDRF-5 denotes this actual work. As the required number of neighboring rays is a key factor for training time, we reduced the number of neighboring rays of PDRF to decrease its training time to as low as ours. Table 2 implies that our method requires significantly lesser training time, and importantly, without any noticeable rendering quality degradation compared to PDRF.

*Equal contributions
†Corresponding authors

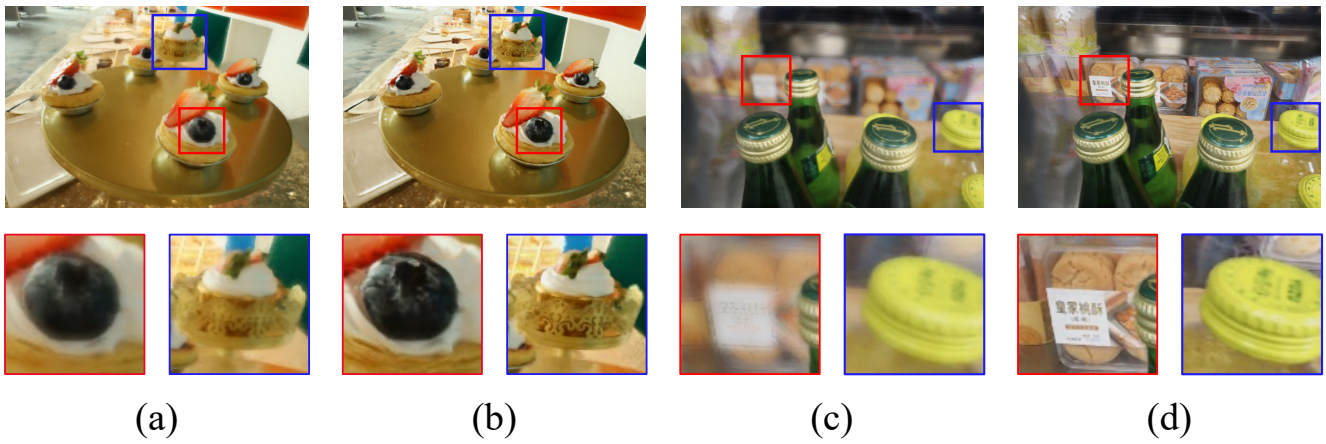

Figure 1. Comparison on learanble blur kernel and fixed blur kernel. (a), (c): Our rendering module trained with fixed discrete kernels. (b), (d): Ours (with learnable discrete kernels).

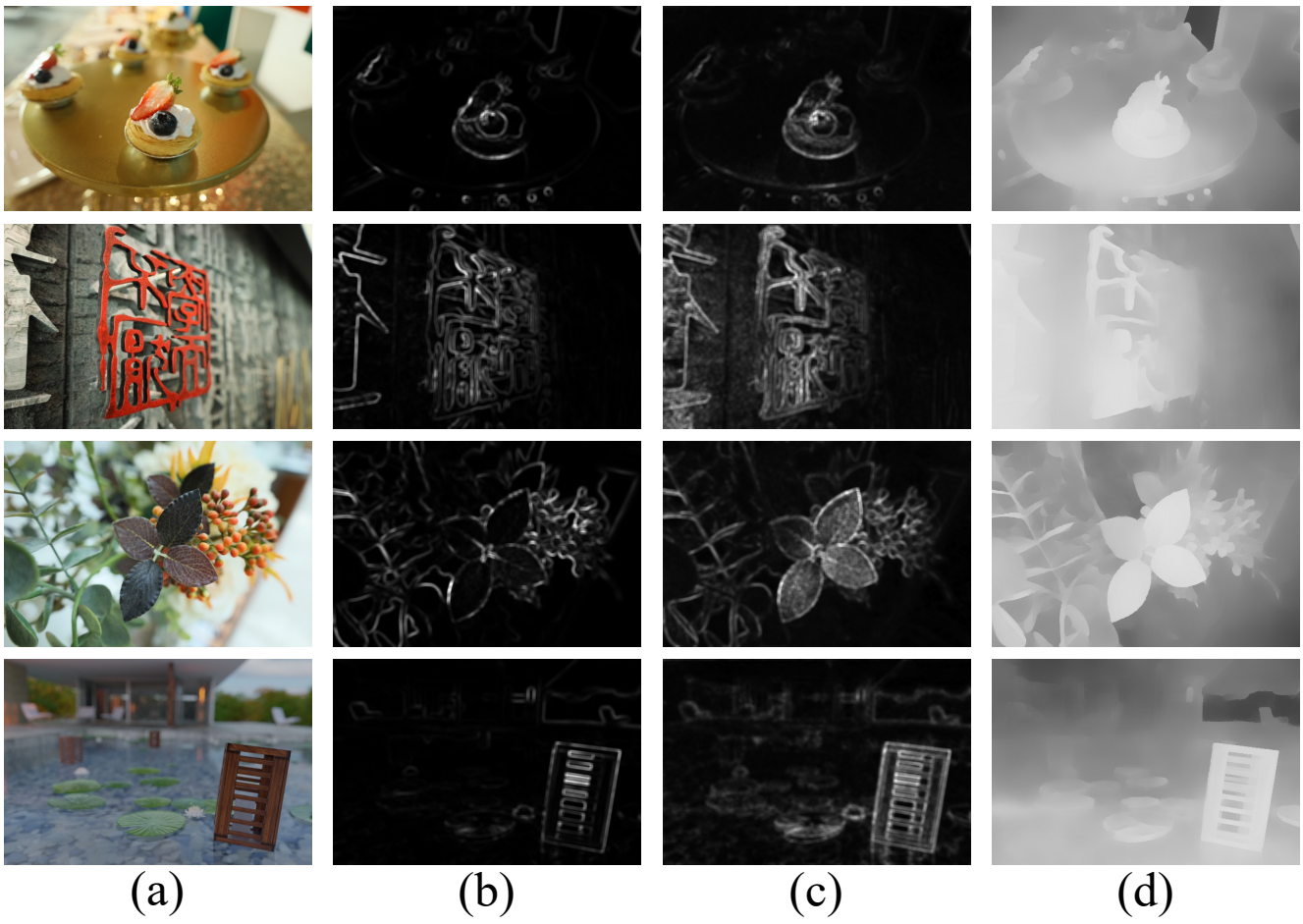

Figure 2. Comparison of sharpness maps provided by various sharpness priors. (a): Sample images, and sharpness maps from (b): Tenengrad, (c): SML, and (d): DMENet.

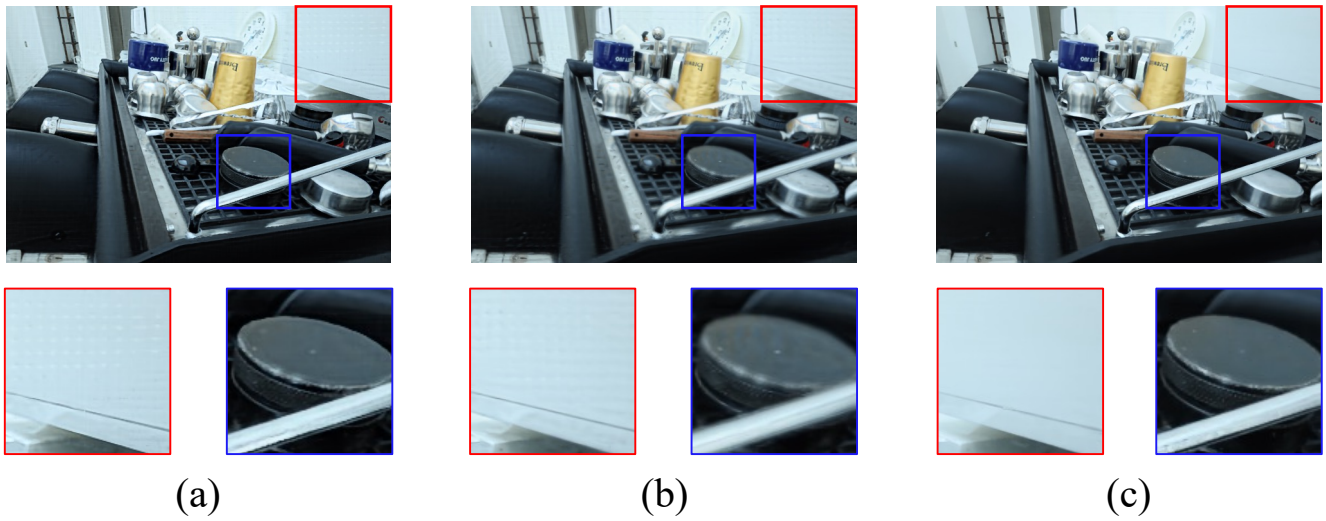

Figure 3. Comparison of various sharpness priors in terms of the quality of rendered images. (a): SML, (b): Tenengrad, (c): DMENet.

2. Selection of Sharpness Prior

In our proposed method Sharp-NeRF, sharpness prior has been used to measure the sharpness level of input image pixels. These sharpness measurements determine how much effort will be put in learning the discrete kernels for those pixels. We have tested three sharpness priors in our study. Among these priors, two are very famous focus measure operators. These operators are sum modified Laplacian (SML) [12] and Tenenbaum gradient (Tenengrad) [4], and these are hand-crafted. The third sharpness prior is DMENet [6] which is deep learning-based and it has been proposed recently.

Fig. 2 displays sharpness maps of different sharpness priors. Depending on their working principle, SML and Tenengrad measure sharpness using derivatives of pixels which becomes hard to compute on textureless regions. This is indicated by the blue and red boxes, respectively. It's evident that SML and Tenengrad tend to focus on sharp regions primarily along edges. In contrast, DMENet does a better work in distinguishing between sharp and blurry regions even there is no or little texture. On the basis of this observation, we decided to use DMENet [6] as our sharpness prior. Fig. 3 shows a qualitative result for the performance of these three priors.

3. Results on Synthetic Data

Table 4 and 5 show the quantitative results of experiments conducted on synthetic defocus dataset. We evaluated both Full-reference (PSNR, SSIM) and No-reference metrics (Brisque, Niqe). Our rendering quality under Full-reference metrics is not as good as compared to the result on real deblur dataset. However, our method achieved sound performance on par with state-of-the-art methods under no-refernce metrics. On top of that, this all good performance has been achieved while keeping the training time to less than half an hour, which is the shortest time among all deblurring neural fields. In short, our method has set a new state-of-the-art time for the deblurring neural fields.

4. No-Reference Metrics for Deblurring NeRF

The commercially available optical imaging devices (cameras) have limited depth of field. Using these cameras, usually, when an image is captured of a scene that has a larger depth of field, some parts of the scene look blurred in the image. Unless, the camera is tuned for the appropriate larger depth of field, or some possessing like image stitching or defocus deconvolution is performed, those regions of the scene that fall out of the range of depth of field of the camera, appear blurred in the images. In contrast, an image in which all the regions of the captured scene are sharp is called an all-in-focus (AIF) image. Although, it is possible to acquire defocused and all-in-focus pair by DSLR camera in two sequential shots, such as [1] which collects 500 defocused and all-in-focus pairs by dual-pixel (DP) camera via adjusting aperture size and exposure time in two separate shots. However, datasets collected by such methods suffer from several artifacts such as inconsistent brightness and mismatched contents. In other words, obtaining a reliable AIF and defocus image pair is very challenging. This is the reason why, until recently, there has been a lack of real image datasets that contain defocused and AIF image pairs [14]. The same difficulty arises for NeRF frameworks. The dataset that has been used in our study is the Deblur-NeRF dataset [8]. In this dataset, the defocus blurred images were also captured using the large aperture, and hence

Table 3. Full-reference and No-reference image quality metrics.

PSNR ↑	Peak signal-to-noise ratio. It is derived from MSE and indicates the ratio of the maximum pixel intensity to the power of the distortion.
SSIM [15] ↑	Structural similarity index. It combines the local image structure, brightness, and contrast into a single local quality score
Brisque [10] ↓	Blind/referenceless image spatial quality evaluator. This model is trained on a database of images with known distortions and as a result it can only assess the quality of images with the same kind of distortion.
Niqe [11] ↓	Natural image quality evaluator. Despite being trained on a database of reference images, this model can assess the quality of images that have arbitrary distortion.

Table 4. Quantitative results on synthetic defocus dataset under Full-reference metrics.

	Cozy2room		Factory		Pools		Tanabata		Trolley		Average		Time
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	Hours
NeRF [9]	30.03	0.8926	25.36	0.7847	27.77	0.7266	23.90	0.7811	22.67	0.7103	25.93	0.7791	2.30
TensoRF [2]	30.04	0.8949	25.27	0.7966	26.81	0.6658	22.73	0.7701	22.18	0.7143	25.40	0.7683	0.42
Deblur-NeRF [8]	31.85	0.9175	28.03	0.8628	30.52	0.8246	26.26	0.8517	25.18	0.8067	28.37	0.8527	10.35
DP-NeRF [5]	32.11	0.9215	29.26	0.8793	31.44	0.8529	27.05	0.8635	26.79	0.8395	29.33	0.8713	20.20
PDRF-5 [13]	32.01	0.9310	25.60	0.7786	31.53	0.8686	27.70	0.8851	27.90	0.8841	28.95	0.8695	1.00
PDRF-10 [13]	31.90	0.9321	26.56	0.8102	31.29	0.8657	28.21	0.8952	28.48	0.8956	29.29	0.8798	1.67
Ours	31.32	0.9133	28.67	0.8979	30.51	0.8264	24.95	0.8536	26.03	0.8498	28.30	0.8682	0.43

Table 5. Quantitative results on synthetic defocus dataset under No-reference metrics.

	Cozy2room		Factory		Pools		Tanabata		Trolley		Average	
	Brisque↓	Niqe↓	Brisque↓	Niqe↓	Brisque↓	Niqe↓	Brisque↓	Niqe↓	Brisque↓	Niqe↓	Brisque↓	Niqe↓
TensoRF [2]	27.52	3.13	42.26	3.77	44.27	4.29	42.29	3.65	36.56	3.65	38.58	3.70
Deblur-NeRF [8]	17.85	3.09	35.60	3.57	39.93	4.08	37.20	3.63	35.35	3.56	33.19	3.59
DP-NeRF [5]	17.83	3.20	35.21	3.54	38.33	3.89	36.37	3.41	33.95	3.39	32.34	3.49
PDRF-5 [13]	21.19	3.11	27.96	3.74	44.27	4.62	36.75	4.00	34.94	3.46	33.02	3.79
PDRF-10 [13]	19.65	3.17	28.24	3.58	40.42	4.37	35.97	3.81	34.40	3.37	31.74	3.66
Ours	21.65	3.22	35.19	3.44	36.91	3.87	36.10	3.58	34.77	3.44	32.92	3.51

they are not free from errors. Owing to the mismatch between AIF and defocus blurred images, deblurring techniques may suffer from inconsistent results when employing the full-reference metrics. The inconsistency among full-reference metrics (especially PSNR) has been observed and highlighted in the recent deblurring NeRFs [3, 7]. For example, it has been mentioned in [3] that if the reference images are blurry, the PSNR (and SSIM) values may be worse. It has been observed in [7] that few blurry images can still achieve higher PSNR than their sharp (deblurred) counterparts. Due to these reasons, we have opted to include the No-reference metrics as well. When compared to Full-reference metrics, generally, all No-reference quality metrics typically perform better in terms of agreement with a subjective human quality score.

References

[1] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pages 111–126. Springer, 2020. 3

[2] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pages 333–350. Springer, 2022. 4

[3] Peng Dai, Yinda Zhang, Xin Yu, Xiaoyang Lyu, and Xiaojuan Qi. Hybrid neural rendering for large-scale scenes with motion blur. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1, 4

[4] Eric Krotkov. Focusing. *International Journal of Computer Vision*, 1(3):223–237, 1988. 3

[5] Dogyoon Lee, Minhyeok Lee, Chajin Shin, and Sangyoun Lee. Deblurred neural radiance field with physical scene priors. *arXiv preprint arXiv:2211.12046*, 2022. 4

[6] Junyong Lee, Sungkil Lee, Sunghyun Cho, and Seungyong Lee. Deep defocus map estimation using domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[7] Kai-En Lin, Yen-Chen Lin, Wei-Sheng Lai, Tsung-Yi Lin, Yi-Chang Shih, and Ravi Ramamoorthi. Vision transformer

for nerf-based view synthesis from a single input image. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 806–815, 2023. 4

[8] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. Deblur-nerf: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12861–12870, 2022. 3, 4

[9] 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. 4

[10] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012. 4

[11] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 4

[12] Shree K Nayar and Yasuo Nakagawa. Shape from focus. *IEEE Transactions on Pattern analysis and machine intelligence*, 16(8):824–831, 1994. 3

[13] Cheng Peng and Rama Chellappa. Pdrf: Progressively deblurring radiance field for fast and robust scene reconstruction from blurry images, 2022. 1, 4

[14] Lingyan Ruan, Bin Chen, Jizhou Li, and Miu-Ling Lam. Aifnet: All-in-focus image restoration network using a light field-based dataset. *IEEE Transactions on Computational Imaging*, 7:675–688, 2021. 3

[15] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4