# Neural Implicit Representation for the Image Stack: Depth, All-in-Focus, and High Dynamic Range.

## Supplementary Material

CHAO WANG, Max-Planck-Institut für Informatik, Germany
ANA SERRANO, Universidad de Zaragoza, I3A, Spain
XINGANG PAN, Max-Planck-Institut für Informatik, Germany & Nanyang Technological University, Singapore
KRZYSZTOF WOLSKI, Max-Planck-Institut für Informatik, Germany
BIN CHEN, Max-Planck-Institut für Informatik, Germany
KAROL MYSZKOWSKI, Max-Planck-Institut für Informatik, Germany
HANS-PETER SEIDEL, Max-Planck-Institut für Informatik, Germany
CHRISTIAN THEOBALT, Max-Planck-Institut für Informatik, Germany
THOMAS LEIMKÜHLER, Max-Planck-Institut für Informatik, Germany

This document supplements more details in the topics as follows:

- S1: Network architecture and implementation,
- S2: Datasets,
- S3: Depth from Defocus and All in Focus reconstructions,
- S4: Exploration of the focal distance, aperture and exposure time,
- S5: Importance of implicit tone mapper,
- S6: Proposed differntiable disk kernel,
- S7: Post-capture editing capabilities,
- S8: Sparse smapling of the TAF stack.

CCS Concepts: • **Computing methodologies → Computational photography**; **Image representations**.

Additional Key Words and Phrases: Depth from Defocus, High Dynamic Range Imaging, Neural Fields

**ACM Reference Format:**
Chao Wang, Ana Serrano, Xingang Pan, Krzysztof Wolski, Bin Chen, Karol Myszkowski, Hans-Peter Seidel, Christian Theobalt, and Thomas Leimkühler. 2023. Neural Implicit Representation for the Image Stack: Depth, All-in-Focus, and High Dynamic Range.: Supplementary Material. *ACM Trans. Graph.* 42, 6 (December 2023), 13 pages. https://doi.org/10.1145/3618367

We provide a detailed description of the network architecture and implementation in Sec. 1, followed by additional details on datasets creation in Sec. 2. In Sec. 3, we present visual comparisons for the Depth from Defocus (DfD) and All-in-Focus (Aif) tasks across three different datasets. In Sec. 4 we extend the exploration of focal distance, aperture, and exposure time. First, we introduce more details about the Time-Aperture-Focus (TAF) stack creation. Then we describe the sampling strategies employed for these parameters and present objective results for each individual scene. Furthermore, we include comparisons of the All-in-Focus High Dynamic Range (HDR) and depth quality for different sampled cross sections ($F_S$, $F_M$, $F_L$, and $F_V$). Next we compare results synthesized with and withotu use of implicit tone mapper (Sec. 5) and visualize differntiable disk kernel for different regions of the image (Sec. 6). In Sec. 7 we demonstrate post-capture editing capabilities of our pipeline. Finally, in Sec. 8, we present flexibility of our pipeline by showing the AiF images and depth maps synthesised by network that has been trained on only 2 input samples from the TAF stack.

Authors' addresses: Chao Wang, chaowang@mpi-inf.mpg.de, Max-Planck-Institut für Informatik, Saarbrücken, Germany; Ana Serrano, anase@unizar.es, Universidad de Zaragoza, I3A, Zaragoza, Spain; Xingang Pan, xingang.pan@ntu.edu.sg, Max-Planck-Institut für Informatik, Saarbrücken, Germany & and Nanyang Technological University, Singapore; Krzysztof Wolski, kwolski@mpi-inf.mpg.de, Max-Planck-Institut für Informatik, Saarbrücken, Germany; Bin Chen, binchen@mpi-inf.mpg.de, Max-Planck-Institut für Informatik, Saarbrücken, Germany; Karol Myszkowski, karol@mpi-inf.mpg.de, Max-Planck-Institut für Informatik, Saarbrücken, Germany; Hans-Peter Seidel, hpseidel@mpi-sb.mpg.de, Max-Planck-Institut für Informatik, Saarbrücken, Germany; Christian Theobalt, theobalt@mpi-inf.mpg.de, Max-Planck-Institut für Informatik, Saarbrücken, Germany; Thomas Leimkühler, thomas.leimkuehler@mpi-inf.mpg.de, Max-Planck-Institut für Informatik, Saarbrücken, Germany.


0730-0301/2023/12-ART
https://doi.org/10.1145/3618367

## 1 NETWORK ARCHITECTURE AND IMPLEMENTATION

Our pipeline consists of four components implemented by MLPs, including the flow predictor $F_f$, depth predictor $F_d$, AiF predictor $F_a$, and implicit tone mapper $F_t$. The flow predictor utilizes a two-layer SIREN [Sitzmann et al. 2020] with three input and two output channels, while both the depth predictor and all-in-focus predictor utilize a four-layer SIREN with the only difference being in the output, where the depth predictor has one channel and the all-in-focus predictor has three channels. Since the HDR image and depth map should be non-negative, we use the exponential function to constrain the output from $F_d$ and $F_a$. The tone mapper is implemented using a two-layer MLP with linear, leaky ReLU, linear, and a hyperbolic tangent activation function. We train the network with learning rate $1e^{-4}$ and Adam optimiser [Kingma and Ba 2014] based on the Pytorch framework. Training the model from scratch with a set of five focal stack images, each having a resolution of 256x384, requires approximately 22 minutes. This training process involves 2000 iterations and is performed on a GPU RTX 8000.

## 2 DATASETS

*Rendered dataset.* Since most of the published datasets do not consider the lens breathing effect, for a more reliable evaluation,

we use [1]Blender to render Time-Aperture-Focus (TAF) stacks with lens breathing for 10 production quality scenes from Blender Studio platform (https://studio.blender.org/). Selected scenes come from three unique animations: *Sprite Fright* - 3 scenes, *Spring* - 3 scenes, and *Charge* - 4 scenes. Five of the selected scenes include scenarios with high dynamic range lighting, while the other five demostrate more uniform lighting conditions. As Blender's rendering Engine *Cycles* does not include parameter specyfing working distance (distance between sensor and lens in real-life cameras) we simulated lens breathing by modifying focal length (and thus field of view) using the following equation:

$$f_{curr} = f_{ref} + \alpha * (S_{max} - S_{curr}), \tag{1}$$

where $f_{curr}$ specifies focal length at actively selected focal distance $S_{curr}$, $f_{ref}$ denotes the reference focal length selected for scene, and $S_{max}$ is the maximum focal distance selected for the scene. Both focal length and focal distance are specified in milimeters. $\alpha$ is a free parameter that allows to tune the strength of lens breathing effect. For all the scenes used in our evalution we set $\alpha$ to 0.0001.

*Real dataset.* For this dataset we captured TAF stacks for 25 unique scenes. Among them 13 scenes have multiple exposure information, which can be used for All-in-Focus HDR image reconstruction and depth map estimation. All of the captured scenes support the depth from defocus (DfD) task. For capturing we used a Canon RP camera with EF 50mm/F1.8 , RF 24-105mm/F4.0 and RF 85mm/F2.0 lenses and Canon 6D2 camera with EF 24-105mm F/3.5-5.6 lens. As our method primarily targets static scenes, we ensured stability of camera by using a tripod. For capturing images, we switch the camera to manual mode and manually adjust settings such as ISO, aperture, and exposure time. To facilitate the capturing process, we employ a remote application and utilize third-party libraries such as [2]*gphoto* for controlling the camera.

## 3 DEPTH FROM DEFOCUS AND ALL IN FOCUS IMAGE RECONSTRUCTION - VISUAL COMPARISON

We present additional results for depth-from-defocus (DfD) and all-in-focus (AiF) image reconstruction using various datasets, including rendered, RGBD, and real datasets. Our method consistently produces more accurate depth maps across these datasets (Fig. 1, 2, 3) and achieves sharp, clean AiF images compared to other methods (Fig. 4, 5, 6). These results further emphasize the effectiveness and benefits of our approach.

## 4 EXPLORATION OF THE FOCAL DISTANCE, APERTURE AND EXPOSURE TIME

In this section, we provide additional details about the exploration section. First we provide more detail about the dataset creation, which is important to the comprehensive exploration experiments. In details, we first render five high dynamic range scenes under different lighting conditions. Each scene contains five different f-numbers and focal distances, accounting for the lens breathing effect. Then we use the following way to project the HDR to multiple exposure LDR images. Initially, we refer to [Andersson et al. 2021] to determine the approximate exposure boundaries (top and bottom) that encompass the entire dynamic range. Subsequently, we discretize the dynamic range based on the lens f-number, ensuring compliance with the relationship defined in Eq. 2.

$$T = 2 \cdot log_2(N) + EV \tag{2}$$

The tone mapper we used is as follow:

$$\hat{I}(x,y) = \begin{cases} 12.92 \cdot I(x,y), & I(x,y) < T \\ clip((1+\phi)((I(x,y) \cdot 2^{EV})^g) - \phi), & otherwise \end{cases} \tag{3}$$

where the $I$ indicates the HDR image and $\hat{I}$ is the tone mapped LDR image. $t$ represents the threshold value, which is specifically set to $t = 0.0031308$. $\phi$, serves as an offset with $\phi$ assigned a value of $\phi = 0.055$. $EV$ indicates the exposure, and the parameter $g$ is utilized for gamma correction, with $g$ chosen as $g = \frac{1}{2.4}$.

As a result, we obtain a 3D parallelogram consisting of 125 scattered points, as Fig. 7. Each point represents a unique combination of TAF (time, aperture and focal distance);

To enhance understanding, we present visualizations of the four cross sections that used during the 2D sampling process. Fig. 7 shows the sections, which involves fixing the aperture and changing the exposure time while performing focal sweeping. Fig. 8 shows the section, which involves fixing the exposure time and changing the aperture size while performing focal sweeping. Furthermore, Fig. 9 showcases the four distinct patterns that were specifically designed for sampling purposes.

In the main paper, we present the average results of HDR-VDP3 [Mantiuk et al. 2023] and PU21-SSIM [Azimi et al. 2021] for five scenes. Here, we provide detailed results for each individual scene in Table. 2, 3, 4, 5, 6, and their previews in Fig. 13. We calculate the HDR-VDP3 using linear HDR images under the condition that the display resolution is $1024 \cdot 720$ with viewing distance 0.5m. Furthermore, we provide the visualizations of reconstructed HDR, depth map and HDR-VDP3 error map in Fig. 14 and 15. Additionally, we corroborate the findings presented in the main paper on real dataset. Specifically, we confirm that employing a moderate aperture size with variable exposure time and focal sweeping, or a fixed exposure time with changing aperture and focal sweeping, yields desirable results for both comprehensive high dynamic range (HDR) reconstruction and depth estimation. We capture three real scenes with different TAF combinations as the rendered dataset. Unlike rendered datasets, real datasets do not provide ground truth information. To assess performance in this scenario, we utilize novel view synthesis. This approach involves training the model on a sampled TAF stack, similar to the rendered dataset, which comprises five images. Once trained, we leverage the learned flow, depth, and AiF HDR to generate additional views of the dataset. By comparing these synthesized views with the real images, we can compute metrics to evaluate the model's performance. Essentially, the accuracy of the learned depth and AiF HDR is directly proportional to the proximity of the synthesized images to the real images. To maintain practical convenience while allowing for a thorough exploration of our concerns, n-rook sampling is not included in this case. We use PSNR and SSIM [Wang et al. 2004] to evaluate the quality of the synthesised images. The notations keep consistent with mainpaper $F_S$ (representing

[1]https://www.blender.org/
[2]http://www.gphoto.org/

large aperture size), $F_M$ (representing moderate aperture size), and $F_L$ (representing small aperture size) to denote the three scenarios in which the F-number is fixed while the exposure time and focal distance are varied. $F_V$ represents the case that the exposure time is fixed while the F-number and focal distance is changing. Based

Table 1. Explorations on real data.

| Methods | PSNR (↑) | SSIM (↑) |
|---|---|---|
| F_S | 31.15 ± 3.21 | 0.954 ± 0.028 |
| F_M | 31.98 ± 3.81 | 0.956 ± 0.030 |
| F_L | 30.98 ± 3.65 | 0.942 ± 0.034 |
| F_V | 31.68 ± 3.50 | 0.955 ± 0.029 |

on the results presented in Table 1, first, it is evident that after conducting a significant number of tests with various combinations, the overall quality of the results appears promising. This indicates the robustness of our method. Additionally, both $F_M$ and $F_V$ metrics outperform the others, aligning with the conclusions drawn from the rendered dataset.

As mentioned in the main paper, our method is applicable to 1D scenarios as well. We have presented numerous results for the depth from defocus (DfD) task. In addition, Fig. 12 showcases the results for HDR fusion [Debevec and Malik 1997] and Varying Aperture Photography (VAP) [Hasinoff and Kutulakos 2007]. We can see that our method performs well for the specific task. However, it should be noticed that due to the limitations of the inputs, it can be challenging to reconstruct all the information from them.

Table 2. Explorations of different combinations for scene 1.

| Methods | HDR-VDP3 (↑) | PU21-SSIM (↑) | Abs_rel (↓) |
|---|---|---|---|
| F_S | 9.257 ± 0.153 | 0.914 ± 0.010 | 0.770 ± 0.350 |
| F_M | 9.780 ± 0.047 | 0.953 ± 0.007 | 0.430 ± 0.244 |
| F_L | 9.837 ± 0.001 | 0.951 ± 0.001 | 0.462 ± 0.001 |
| F_V | 9.770 ± 0.019 | 0.951 ± 0.047 | 0.268 ± 0.041 |
| n-rooks | 9.741 ± 0.004 | 0.943 ± 0.018 | 0.262 ± 0.017 |

Table 3. Explorations of different combinations for scene 2.

| Methods | HDR-VDP3 (↑) | PU21-SSIM (↑) | Abs_rel (↓) |
|---|---|---|---|
| F_S | 9.721 ± 0.031 | 0.959 ± 0.003 | 0.496 ± 0.127 |
| F_M | 9.852 ± 0.008 | 0.967 ± 0.001 | 0.467 ± 0.074 |
| F_L | 9.810 ± 0.008 | 0.958 ± 0.002 | 0.768 ± 0.002 |
| F_V | 9.834 ± 0.010 | 0.965 ± 0.001 | 0.432 ± 0.084 |
| n-rooks | 9.829 ± 0.021 | 0.963 ± 0.003 | 0.464 ± 0.029 |

## 5 IMPORTANCE OF IMPLICIT TONE MAPPER

In Fig. 10, we show an additional visualization that highlights the superiority of our implicit tone mapper over the explicit one in terms of the quality of reconstructed HDR (High Dynamic Range) images. We can see the explicit tone mapper introduces more artifacts to the recovered HDR, resulting in poor overall quality.

Table 4. Explorations of different combinations for scene 3.

| Methods | HDR-VDP3 (↑) | PU21-SSIM (↑) | Abs_rel (↓) |
|---|---|---|---|
| F_S | 9.791 ± 0.035 | 0.962 ± 0.006 | 0.146 ± 0.019 |
| F_M | 9.869 ± 0.013 | 0.971 ± 0.003 | 0.208 ± 0.027 |
| F_L | 9.835 ± 0.021 | 0.964 ± 0.003 | 0.455 ± 0.043 |
| F_V | 9.847 ± 0.016 | 0.968 ± 0.004 | 0.164 ± 0.022 |
| n-rooks | 9.846 ± 0.010 | 0.966 ± 0.003 | 0.168 ± 0.015 |

Table 5. Explorations of different combinations for scene 4.

| Methods | HDR-VDP3 (↑) | PU21-SSIM (↑) | Abs_rel (↓) |
|---|---|---|---|
| F_S | 9.514 ± 0.044 | 0.944 ± 0.003 | 0.136 ± 0.046 |
| F_M | 9.828 ± 0.014 | 0.970 ± 0.003 | 0.083 ± 0.006 |
| F_L | 9.917 ± 0.015 | 0.982 ± 0.003 | 0.402 ± 0.029 |
| F_V | 9.810 ± 0.024 | 0.970 ± 0.004 | 0.153 ±0.005 |
| n-rooks | 9.831 ± 0.099 | 0.967 ± 0.012 | 0.146 ± 0.051 |

Table 6. Explorations of different combinations for scene 5.

| Methods | HDR-VDP3 (↑) | PU21-SSIM (↑) | Abs_rel (↓) |
|---|---|---|---|
| F_S | 9.238 ± 0.148 | 0.907 ± 0.019 | 0.473 ± 0.017 |
| F_M | 9.817 ± 0.046 | 0.962 ± 0.007 | 0.482 ± 0.051 |
| F_L | 9.910 ± 0.009 | 0.971 ± 0.002 | 0.686 ± 0.003 |
| F_V | 9.774 ± 0.003 | 0.957 ± 0.001 | 0.454 ± 0.010 |
| n-rooks | 9.830 ± 0.012 | 0.962 ± 0.002 | 0.478 ± 0.018 |

## 6 DISK KERNEL - VISUALIZATION

In Fig. 11, we provide visualizations of the defocus map and the various sizes of our disk kernel, corresponding to the specific location in the image.

## 7 EDITING CAPABILITES

We demonstrate the editing capabilities our method enables. With only 5 inputs of the TAF stack, the proposed pipeline is able to synthesise an image with any exposure, aperture and focal distance selected from continuous 3D space (Fig. 16 and 17).

## 8 SPARSE SAMPLING OF THE TAF STACK

In the paper all our evaluations were conducted on the models trained with the use of 5 input samples from the TAF stack. However, even for lower number of samples, our method produces plausible all-in-focus HDR image nad Depth map. In Fig. 18, we show results synthesized by our method trained with only two input samples from the TAF stack.

### 8.1 The Effect of Lens Breathing

While our method can handle the lens breathing in the input data, a subset of SOTA methods (DFFV and DEReD) do not account for this effect, which might lead to a drop in their performance. In this section, we investigate how our method compares to the existing ones when the lens breathing is removed from the data. The results of the ablation run on the modified NYUv2 dataset are presented in Table 7. As expected, removing the lens breathing improves the

Table 7. Evaluation of DfD on the NYU dataset without lens breathing.

| Method | MAE (↓) | MSE (↓) | Abs_rel (↓) |
|---|---|---|---|
| DFFMobile | 0.20 ± 0.09 | 0.07 ± 0.06 | 0.95 ± 0.01 |
| DFFWild | **0.11 ± 0.03** | **0.02 ± 0.01** | 0.16 ± 0.12 |
| DEReD | 0.19 ± 0.11 | 0.07 ± 0.07 | 0.82 ± 0.06 |
| DFFV | 0.32 ± 0.15 | 0.16 ± 0.13 | 0.65 ± 0.66 |
| **Ours** | **0.11± 0.04** | **0.02 ± 0.01** | **0.07 ± 0.03** |

Table 8. Evaluation of AiF image reconstruction on the NYU dataset without lens breathing.

| Method | PSNR (↑) | SSIM (↑) |
|---|---|---|
| DFFMobile | 33.2 ± 5.2 | 0.96 ± 0.04 |
| DEReD | 25.9 ± 2.4 | 0.90± 0.05 |
| **Ours** | **37.4 ± 5.6** | **0.98 ± 0.02** |

performance of all SOTA methods (refer to Table 2 in the main manuscript), yet they do not surpass the proposed solution in the DfD task.

In addition, we conduct experiments on All-in-Focus reconstruction to illustrate the substantial influence of the lens breathing effect on other methods. As indicated in Table 8, the absence of lens breathing notably improves the quality of the SOTA methods. Nevertheless, our method still demonstrates superior performance.

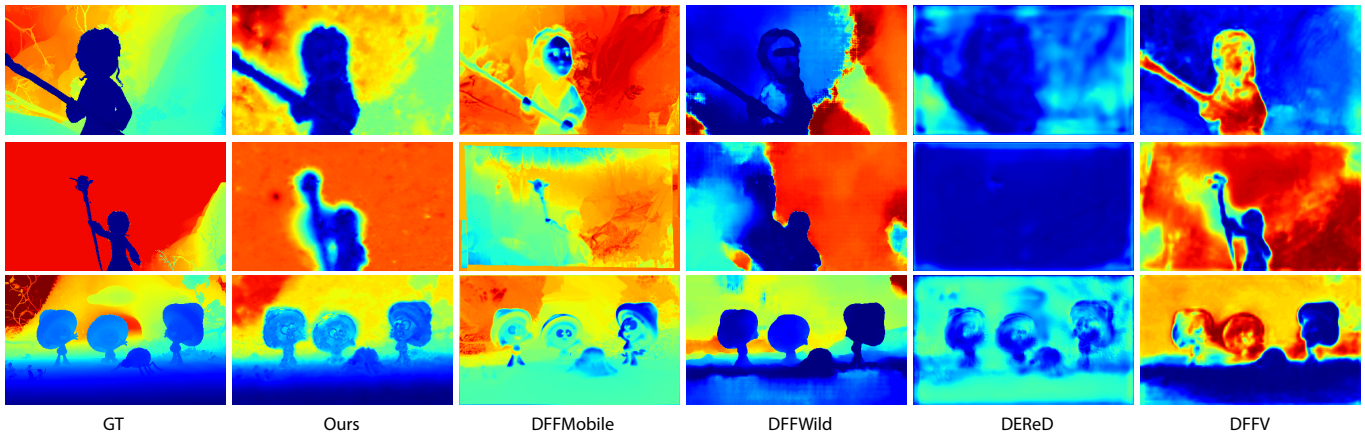


Fig. 1. More depth visualization for render datasets used in evaluation.

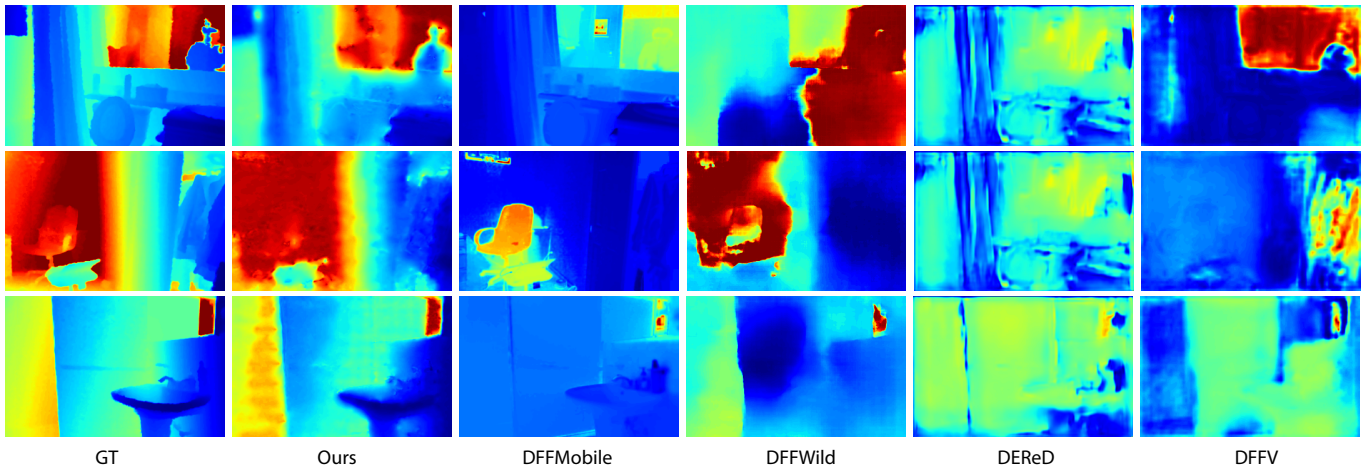


Fig. 2. More depth visualization for RGBD datasets used in evaluation.

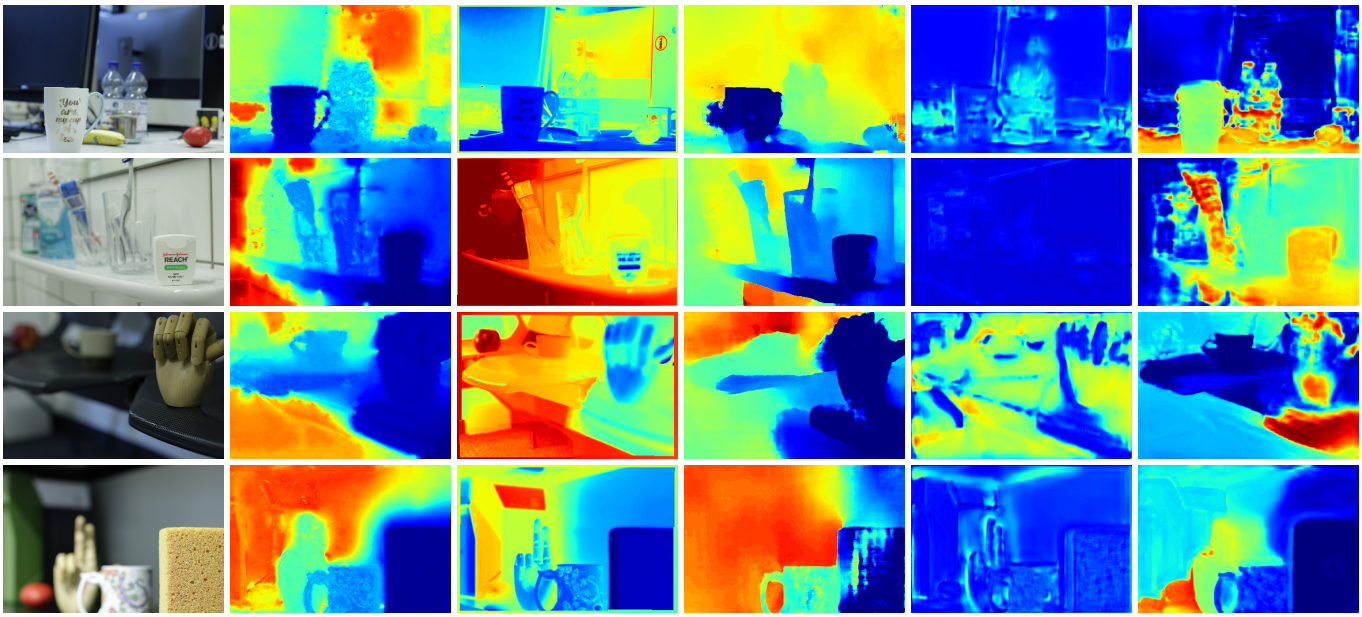


Fig. 3. More depth visualization for real datasets used in evaluation. As for *Real Dataset* there is no Ground Truth available, we provide one of the images from the focal stack for more intuitive comparison.


Fig. 4. More all-in-focus image reconstruction for render datasets used in evaluation.

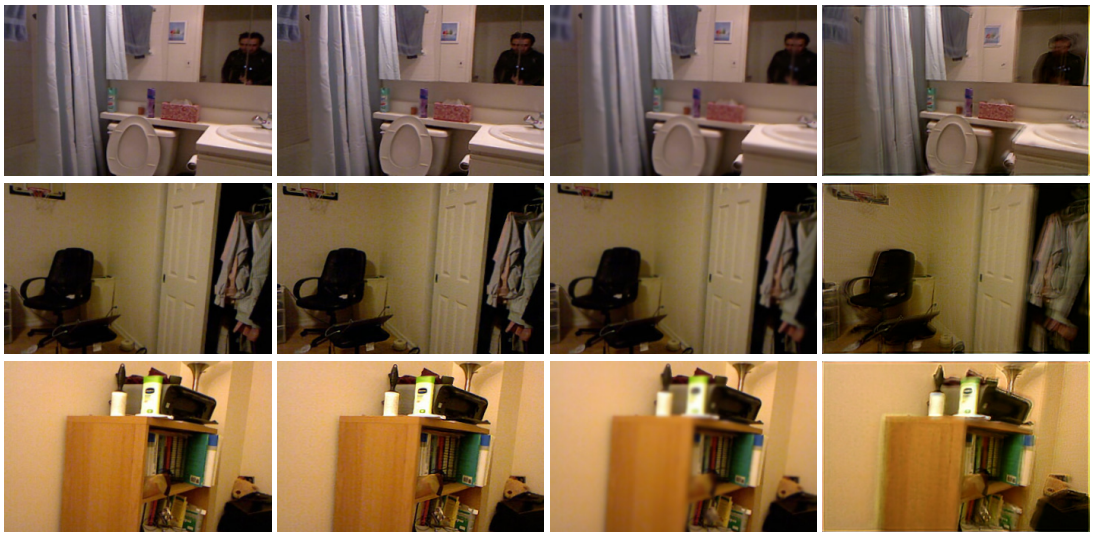


Fig. 5. More all-in-focus image reconstruction for RGBD datasets used in evaluation.

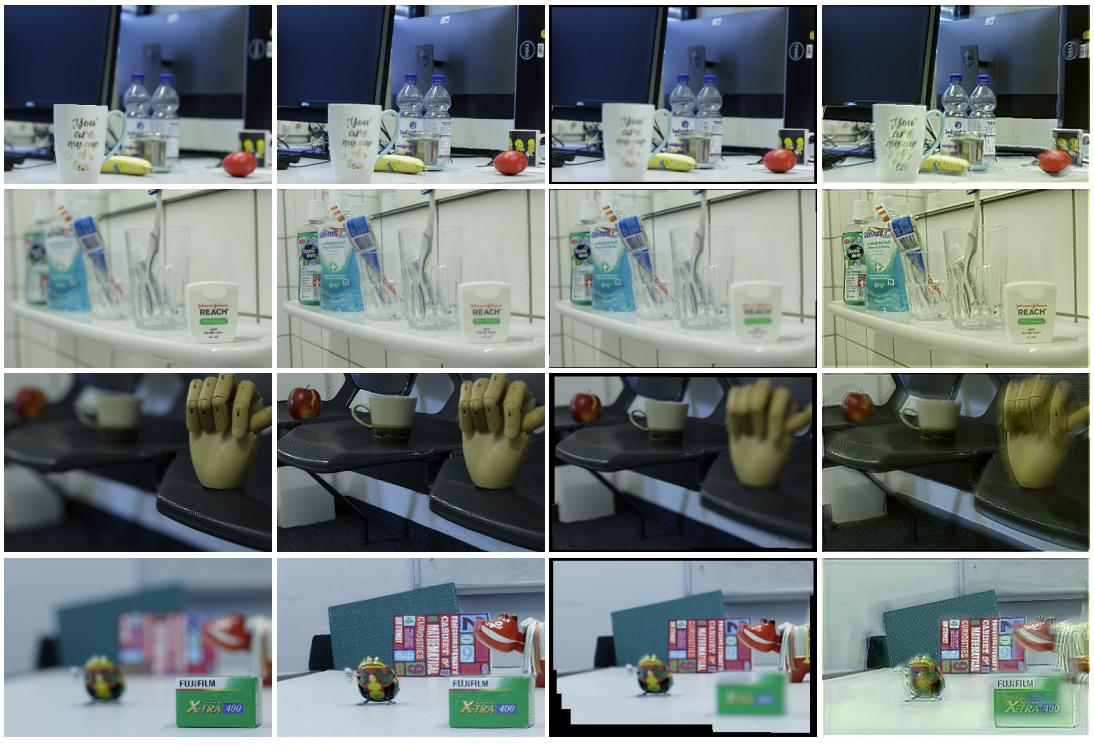


Fig. 6. More all-in-focus image reconstruction for real datasets used in evaluation. As for *Real Dataset* there is no Ground Truth available, we provide one of the images from the focal stack for more intuitive comparison.

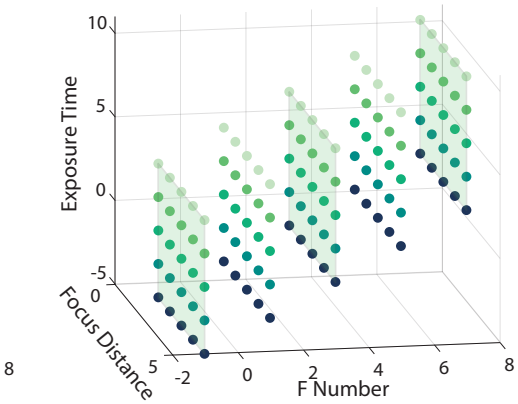


Fig. 7. Sampled cross sections of fixed aperture with changing exposure time and focal sweeping. In these cases, the exposure conditions vary in conjunction with changes in the exposure time. The left side cross section represents images captured using a small F-number (large aperture size), while systematically varying the exposure time and focal distance in different combinations.

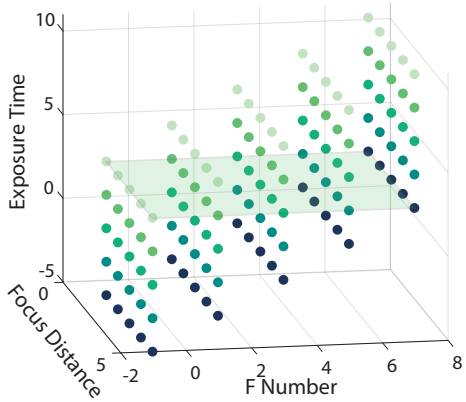


Fig. 8. Sampled cross sections of fixed exposure time with changing aperture size and focal sweeping. In this case, the exposure conditions vary in conjunction with changes in the aperture size. A small F-number (large aperture) corresponds to higher exposure, while a large F-number (small aperture) corresponds to lower exposure.

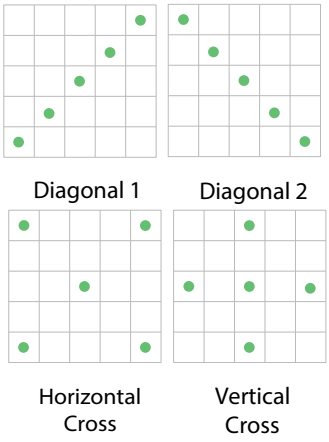


Fig. 9. Sampling patterns we use for the 2D cross section sampling.

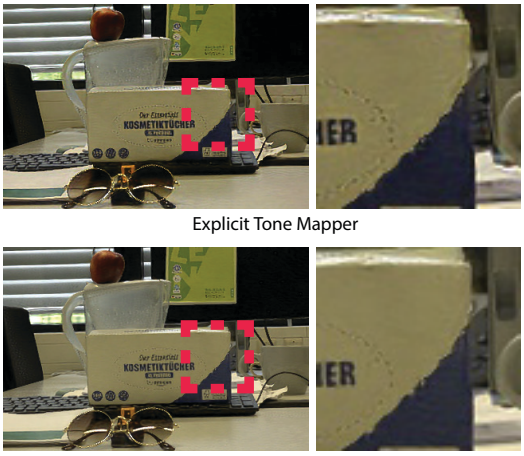


Fig. 10. Comparisons of the reconstructed HDR between the explicit tone mapper and ours.Figure showcases the presence of artifacts in the reconstructed HDR images from the explicit tone mapper.

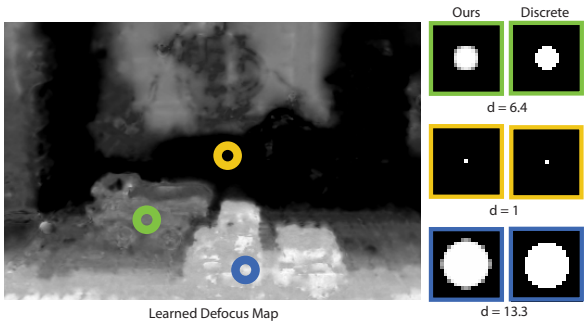


Fig. 11. Visualization of the learned defocus map and our disk kernels. We choose three representative points in defocus map to display the corresponding kernel, *d* denotes the value at the defocus point, which also serves as the diameter of the Circle of Confusion (CoC). For comparisons, we provide the discrete disk kernel in the rights side as well.

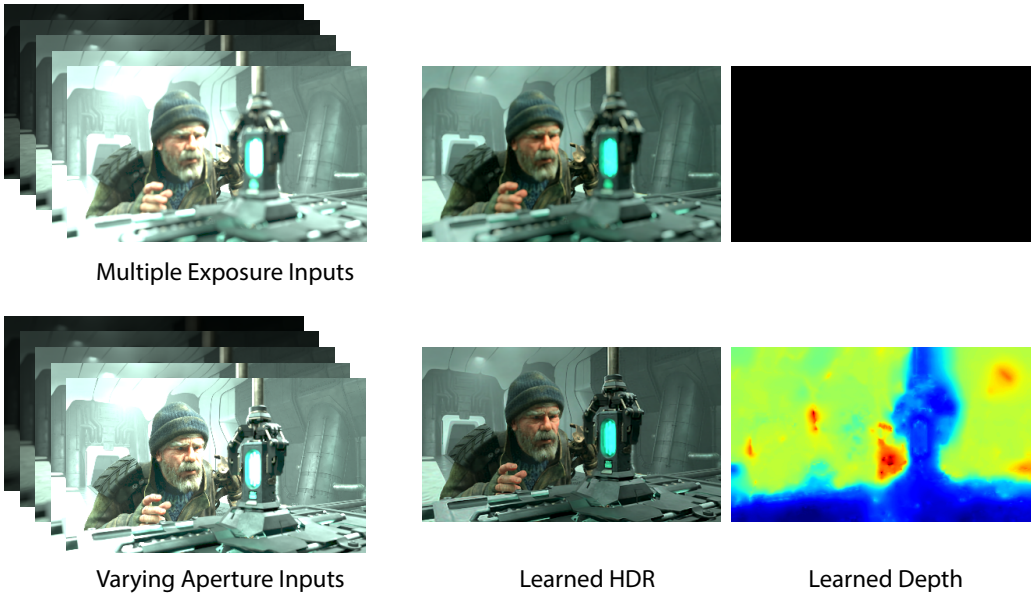


Fig. 12. HDR and VAP results visualizations. In the first row, the HDR fusion task [Debevec and Malik 1997] is demonstrated, which successfully reconstructs the HDR but fails to estimate the depth map due to the absence of defocus blur clues. In the second row, the VAP (Varying Aperture Photography) [Hasinoff and Kutulakos 2007] results is showcased. By changing the aperture size, VAP incorporates multiple exposure information and defocus blur cues, enabling the reconstruction of both HDR and depth. However, the supervision based solely on aperture is insufficient to produce accurate results.


Fig. 13. Visualization of five scenes we use in the exploration.

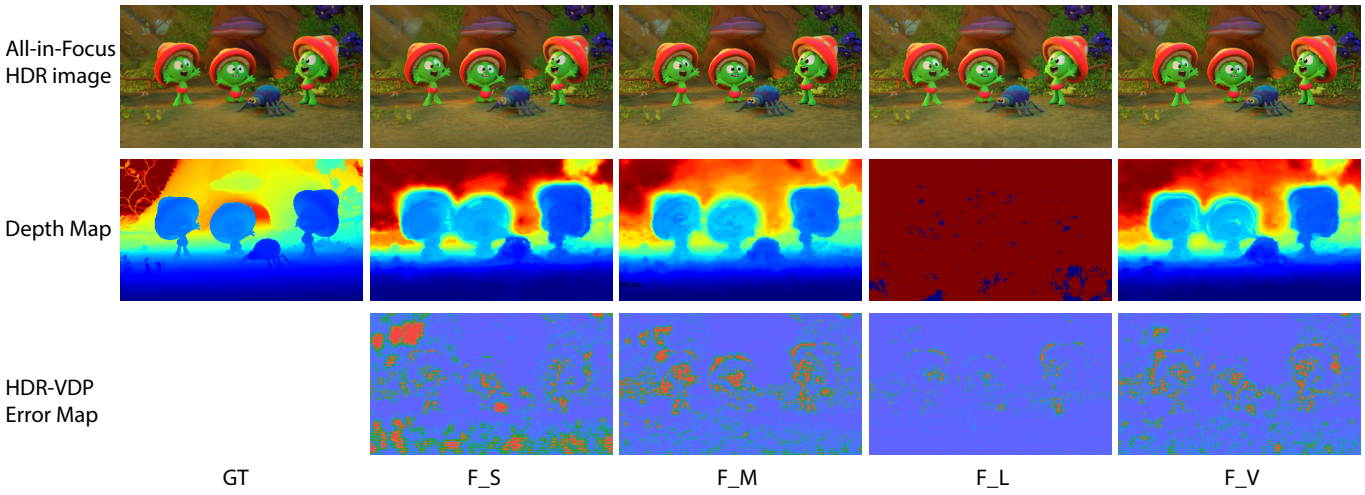


Fig. 14. Visualizations of the explorations of the focal distance, aperture, and exposure. The results demonstrate that a small F-number ($F_S$), which corresponds to a large aperture size, leads to a decrease in the quality of all-in-focus HDR reconstruction. Conversely, a large F-number ($F_L$), indicating a small aperture size, results in poorer depth map estimation. However, utilizing a middle aperture ($F_M$) proves advantageous for both aspects. Furthermore, fixing the exposure time while varying the aperture size ($F_V$) produces promising outcomes in terms of both all-in-focus HDR reconstruction and depth map estimation, similar to the performance achieved with the $F_M$ configuration.

.

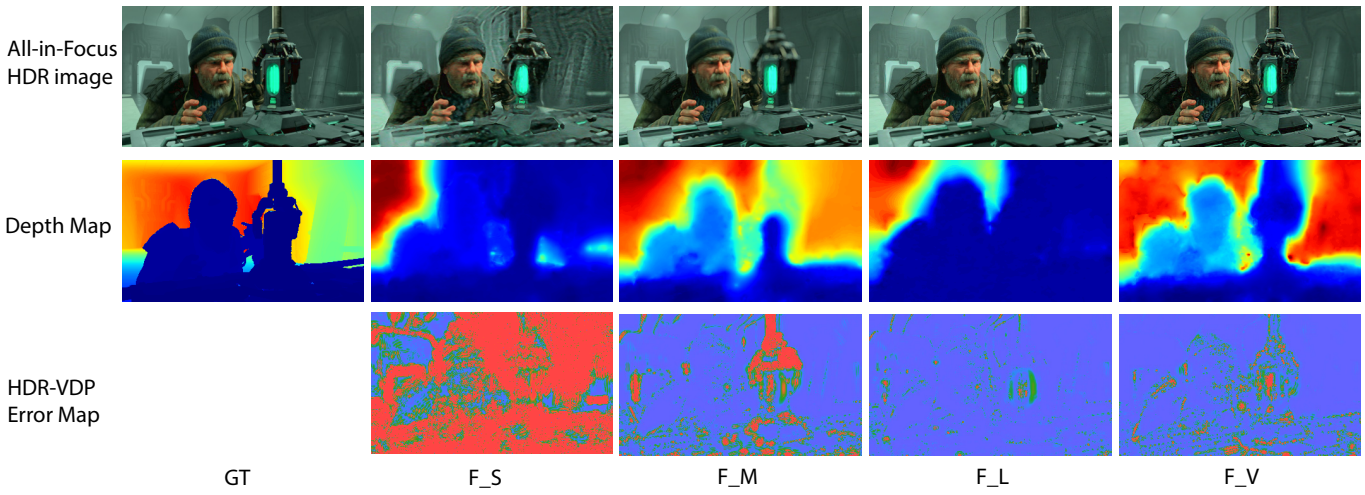


Fig. 15. Visualizations of the explorations of the focal distance, aperture, and exposure. The results demonstrate that a small F-number ($F_S$), which corresponds to a large aperture size, leads to a decrease in the quality of all-in-focus HDR reconstruction. Conversely, a large F-number ($F_L$), indicating a small aperture size, results in poorer depth map estimation. However, utilizing a middle aperture ($F_M$) proves advantageous for both aspects. Furthermore, fixing the exposure time while varying the aperture size ($F_V$) produces promising outcomes in terms of both all-in-focus HDR reconstruction and depth map estimation, similar to the performance achieved with the $F_M$ configuration.

.

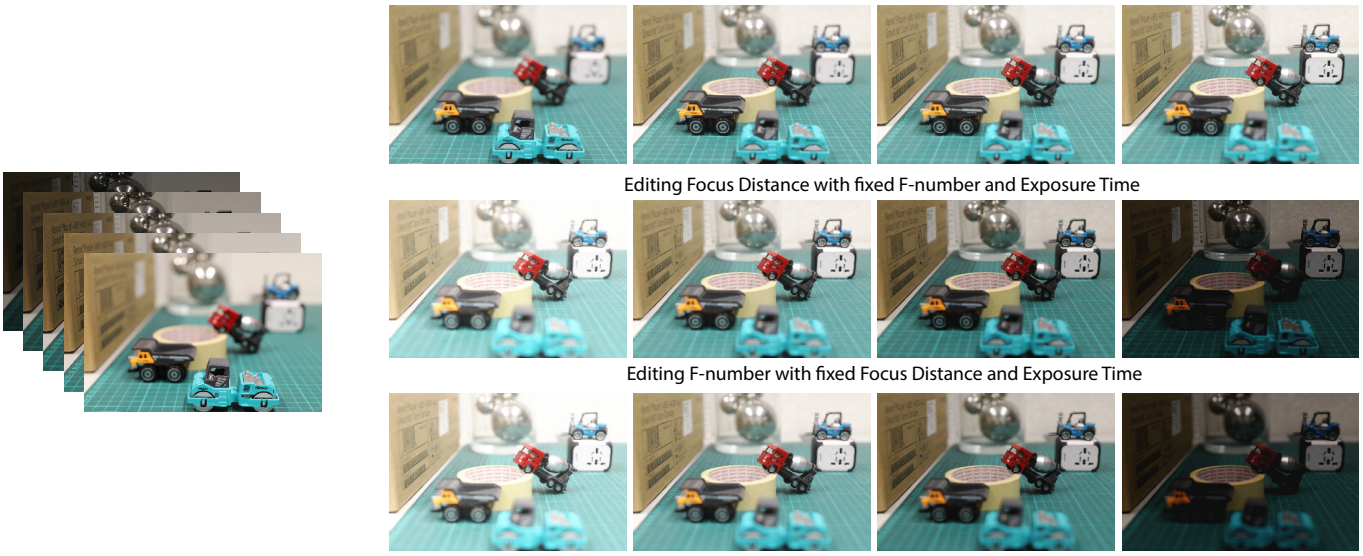


Fig. 16. Visualizations of post editing. After training our model on the left side TAF stack, we gain the capacity to perform various post editing. As shown in the right side, in the upper row, adjusting the focal distance with constant F-number and exposure time yields promising refocus results. In the second row, changing the F-number while fixing focal distance and exposure time influences blur and exposure levels. Lastly, varying exposure time with fixed focal distance and F-number impacts only exposure. From the last two rows, it is clear that exposure can be influenced by both the F-number and exposure time. However, altering the F-number uniquely impacts the defocus blur too.

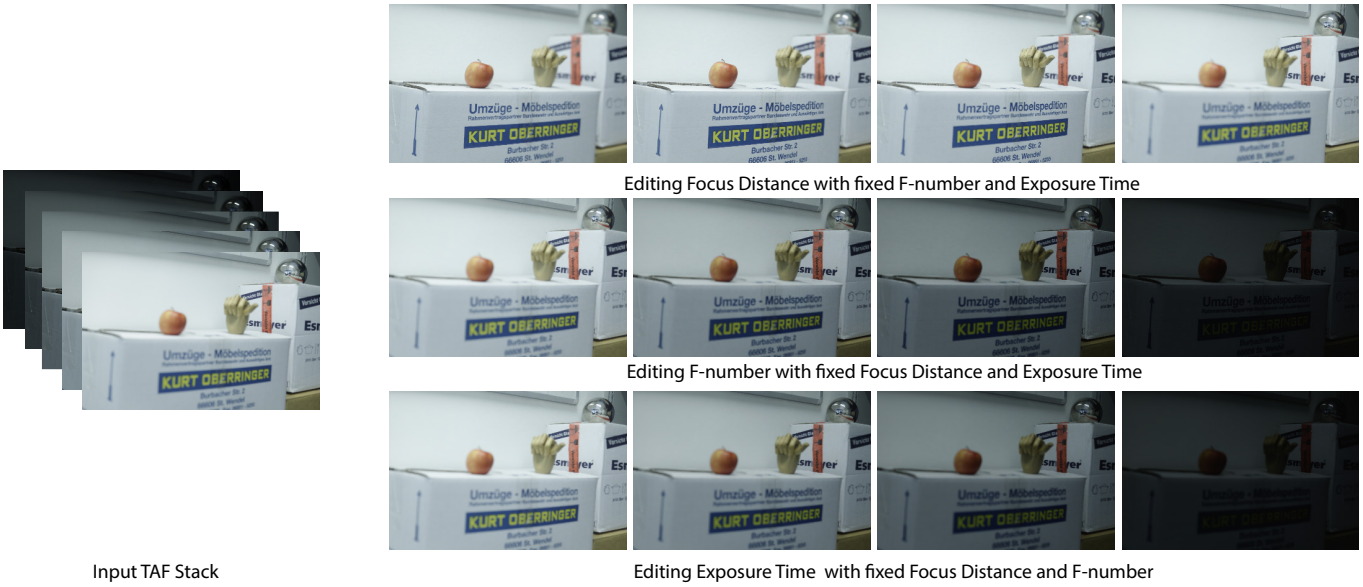


Fig. 17. Visualizations of post editing. After training our model on the left TAF stack, we gain the capacity to perform various post editing. As shown in the right side, in the upper row, adjusting the focal distance with constant F-number and exposure time yields promising refocus results. In the second row, changing the F-number while fixing focal distance and exposure time influences blur and exposure levels. Lastly, varying exposure time with fixed focal distance and F-number impacts only exposure. From the last two rows, it's clear that exposure can be influenced by both the F-number and exposure time. However, altering the F-number uniquely impacts the defocus blur too.

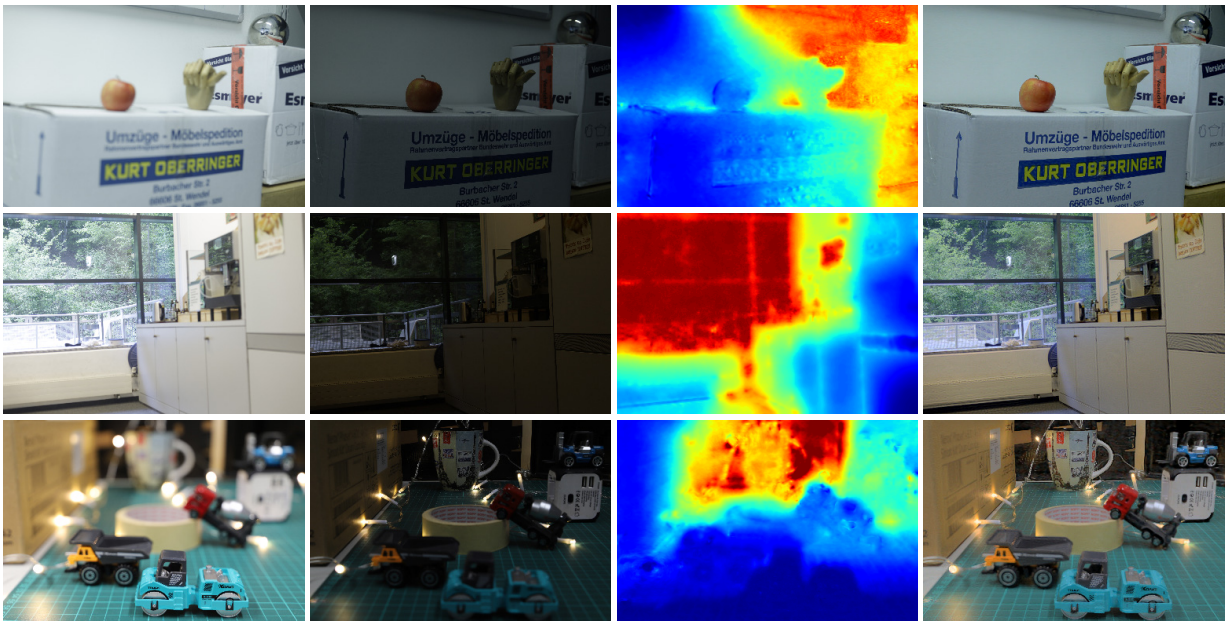


Fig. 18. Results on two inputs. Our methods offer flexibility in terms of inputs, as it can be applied to arbitrary inputs. This means that even if two different TAF stack images are provided, our approach can still generate a reasonable All-in-Focus HDR image and depth map. For better visualization, we use [Reinhard et al. 2002]to tone map the learned HDR to LDR. This versatility allows for a wider range of applications and accommodates various input scenarios.

## REFERENCES


Pontus Andersson, Jim Nilsson, Peter Shirley, and Tomas Akenine-Möller. 2021. Visualizing the Error in Rendered High Dynamic Range Images. In *Eurographics Short Papers*. https://doi.org/10.2312/egs.20211015

Maryam Azimi et al. 2021. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In *2021 Picture Coding Symposium (PCS)*. IEEE, 1–5.

Paul E. Debevec and Jitendra Malik. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., USA, 369–378. https://doi.org/10.1145/258734.258884

Samuel W Hasinoff and Kiriakos N Kutulakos. 2007. A layer-based restoration framework for variable-aperture photography. In *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 1–8.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Rafal K Mantiuk, Dounia Hammou, and Param Hanji. 2023. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625* (2023).

Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. 2002. Photographic tone reproduction for digital images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 267–276.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. 2020. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems* 33 (2020), 7462–7473.

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