Supplemental Document

Optics EXPRESS

View synthesis for 3D computer-generated holograms using deep neural fields: supplement

KENNETH CHEN,[1,*] ANZHOU WEN,[2,3] YUNXIANG ZHANG,[1] PRANEETH CHAKRAVARTHULA,[2,3] AND QI SUN[1]

[1] *New York University, Tandon School of Engineering, New York, NY 11201, USA*
[2] *The University of North Carolina at Chapel Hill, Department of Computer Science, Chapel Hill, NC 27599, USA*
[3] *The University of North Carolina at Chapel Hill, Department of Applied Physical Sciences, Chapel Hill, NC 27599, USA*
[*] *kennychen@nyu.edu*

Supplement DOI: https://doi.org/10.6084/m9.figshare.28776746

Parent Article DOI: https://doi.org/10.1364/OE.559364

Supplementary Materials for

Novel View Synthesis for 3D Computer-Generated Holograms Using Deep Neural Fields

1. Scene Acquisition and Reference Measurements

Scenes used throughout the manuscript were captured using an iPhone 12 Pro for training NHF. Ground-truth focal stacks (only used for qualitative comparison) were captured with a Nikon D3500 DSLR camera.

2. Training Details

The NHF neural network is trained with an Adam optimizer and cosine annealing learning rate schedule. For all results in the paper, we use a learning rate of $1e$-3, VGG loss weight of $2.5e$-2, and L2 weight of 1.0. We used a 90:10 train/test split; all results are computed on the test dataset. In a network with 5 upsampling and 5 downsampling layers, 64 dimensions, and 4 depth planes used in the focal stack loss, the training time takes about 3.5 hours to complete, or 130 training epochs. All experimental code is implemented with PyTorch. Model training and evaluation were computed on a single graphics processing unit, the NVIDIA RTX 8000, provided by the NYU Greene High-Performance Computing cluster.

3. Learning Ablation Studies

Training NHF involves two loss terms, VGG and L2. The two terms jointly enhance the inferred image quality. As visualized in Figure 1, we conducted an ablation study to quantitatively measure their effectiveness. Including both L2 and VGG improves quality by approximately 9 dB over VGG only and 2 dB over L2 only. The analysis indicates the quality enhancements achieved by jointly optimizing the NHF neural network with feature-based (VGG) and pixel-wise (L2) image loss functions.

4. Raw Smartphone Captures

Figure 2 visualized several sampled original scene captures using an iPhone 12 Pro. Such freely captured images are directly fed to NHF to generate free-viewing 3D holograms.

5. Additional Results

Figures 3, 4, 5, 6, 7 show additional simulated and experimental hardware-captured results with focal stacks. Figure 9 shows monochromatic results, for each RGB laser channel for both near and far focus. Figure 10 shows an example hologram output of the light field to hologram iSTFT transform.

6. Algorithmic Implementation

Pseudocode is defined in Algorithm 1.

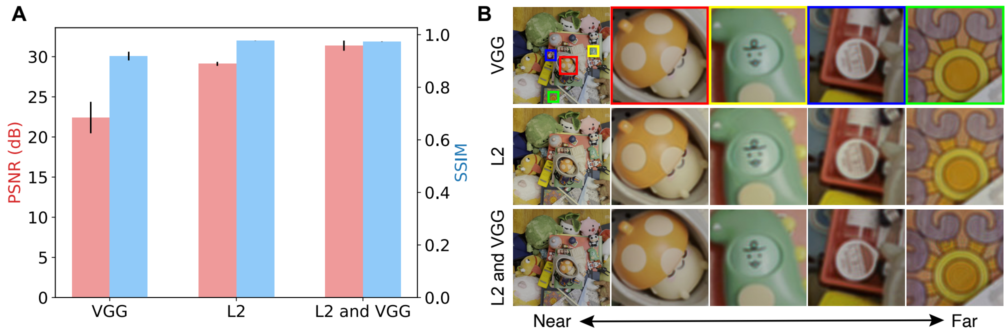

Fig. 1. *Ablation of loss functions.* **(A)** Quantitative results of ablation study for loss function terms. Error bars represent one standard deviation. **(B)** Qualitative results of ablation study, with near depth in focus.

Fig. 2. *Example mobile phone captures.* Captures of the Shelf scene and the Pile scene.

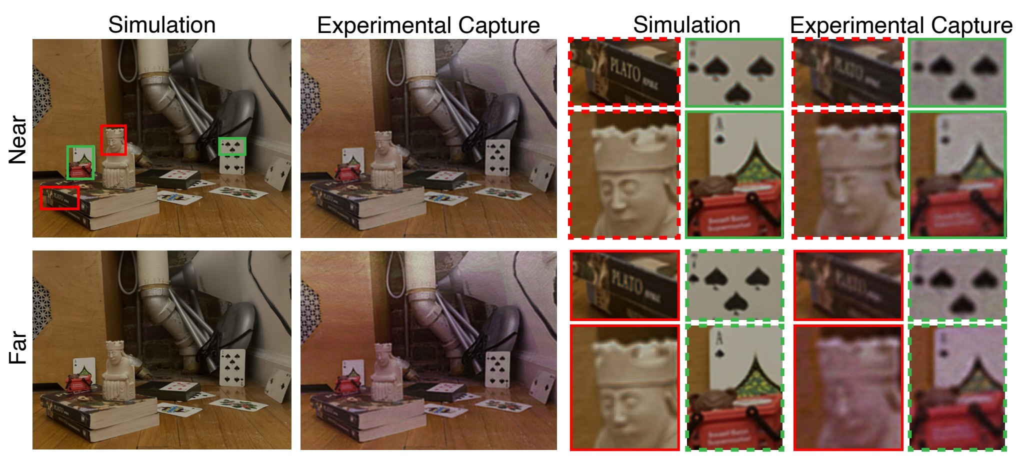

Fig. 3. *King scene.* Simulated and experimental results predicted by NHF for the King scene.

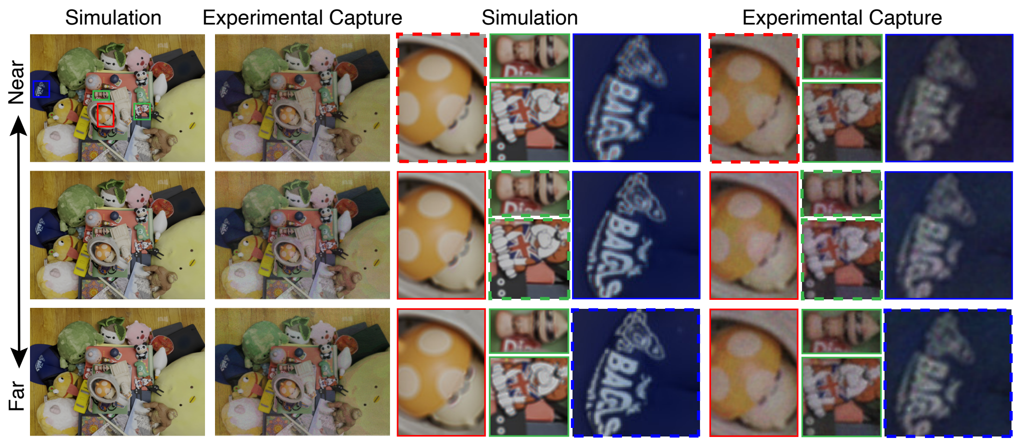

Fig. 4. *Pile scene.* Simulated and experimental results predicted by NHF for the Pile scene.

Algorithm 1: NHF Pipeline

Input: $\mathbf{I}$: set of N input camera captures

$\mathbf{C} \leftarrow \mathbf{SfM}(\mathbf{I})$ // estimate camera parameters from input images with a SfM pipeline such as COLMAP

$f \leftarrow$ **trainGaussianSplatting(I, C)** // train scene representation

/* Generate NHF training dataset */

init: $\mathbf{I_r}$

init: $\mathbf{H}$

init: γ_x, γ_y, N

for *cam* : ***C*** **do**

  init: L

  for j : *range*($-N$, N) **do**

    for k : *range*($-N$, N) **do**

      /* Render light field elemental view from scene representation (Eq. 2) */

$$\mathbf{t}_{(j,k)} \leftarrow cam.\mathbf{t} + cam.\mathbf{R} \begin{pmatrix} j\gamma_x \\ k\gamma_y \\ 0 \end{pmatrix}$$

      $L(j,k) \leftarrow f\left(cam.\mathbf{R}, \mathbf{t}_{(j,k)}\right)$

  $\mathbf{I_r}[cam] \leftarrow L(0,0)$

  $\mathbf{H}[cam] \leftarrow \mathrm{STFT}^{-1}(L)$

/* Train NHF CNN */

init: $\mathcal{M}_\Theta$

while *not converged* **do**

  for *cam* : ***C*** **do**

    $\Theta \leftarrow \Theta - \alpha \cdot \frac{\partial}{\partial\Theta} \sum_z \mathcal{L}\left(\mathcal{P}(\mathbf{H}[cam], z), \mathcal{P}(\mathcal{M}_\Theta(\mathbf{I_r}[cam]), z)\right)$

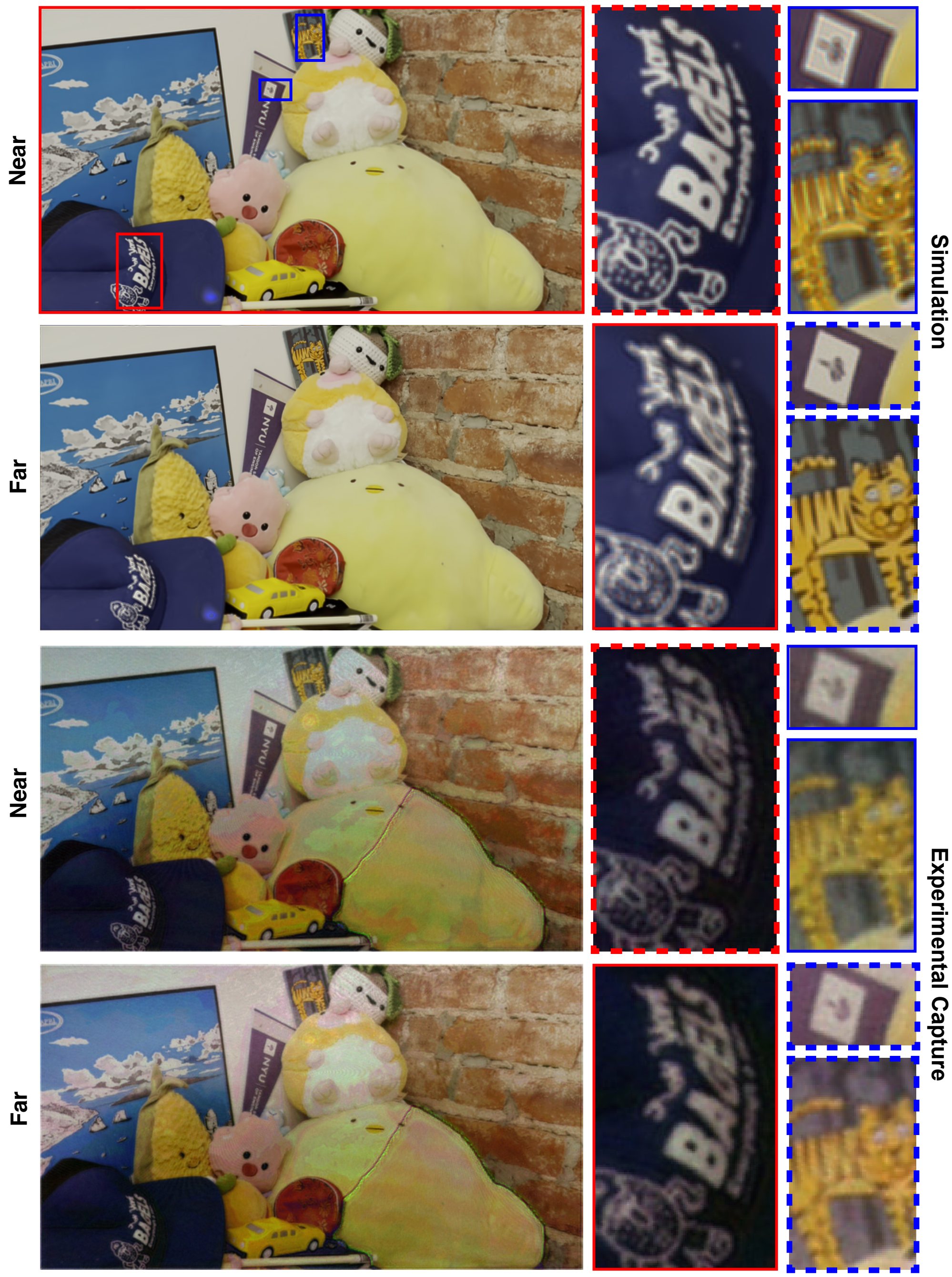

Fig. 5. *Pile 2 scene.* Simulated and experimental results predicted by NHF for the Pile 2 scene.

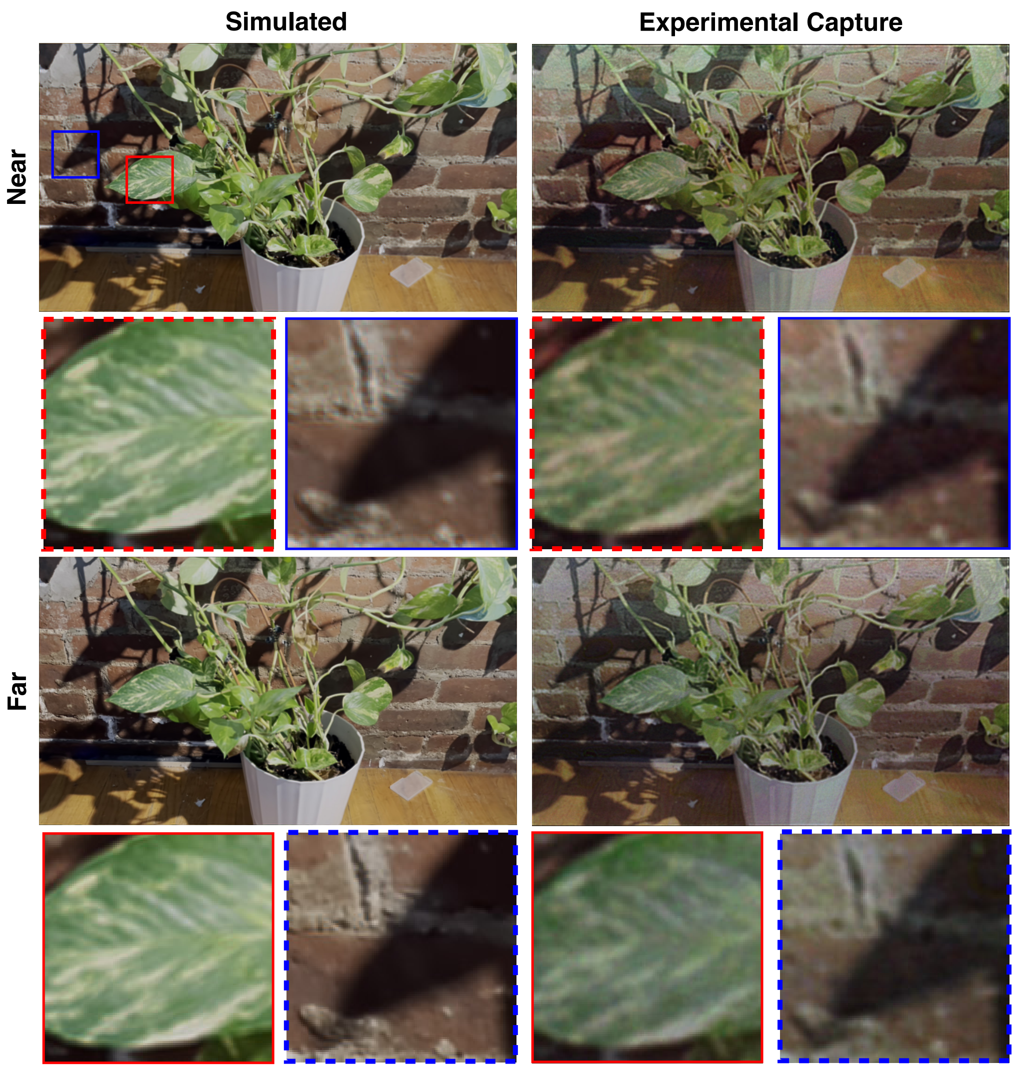

Fig. 6. *Plant scene.* Simulated and experimental results predicted by NHF for the Plant scene.

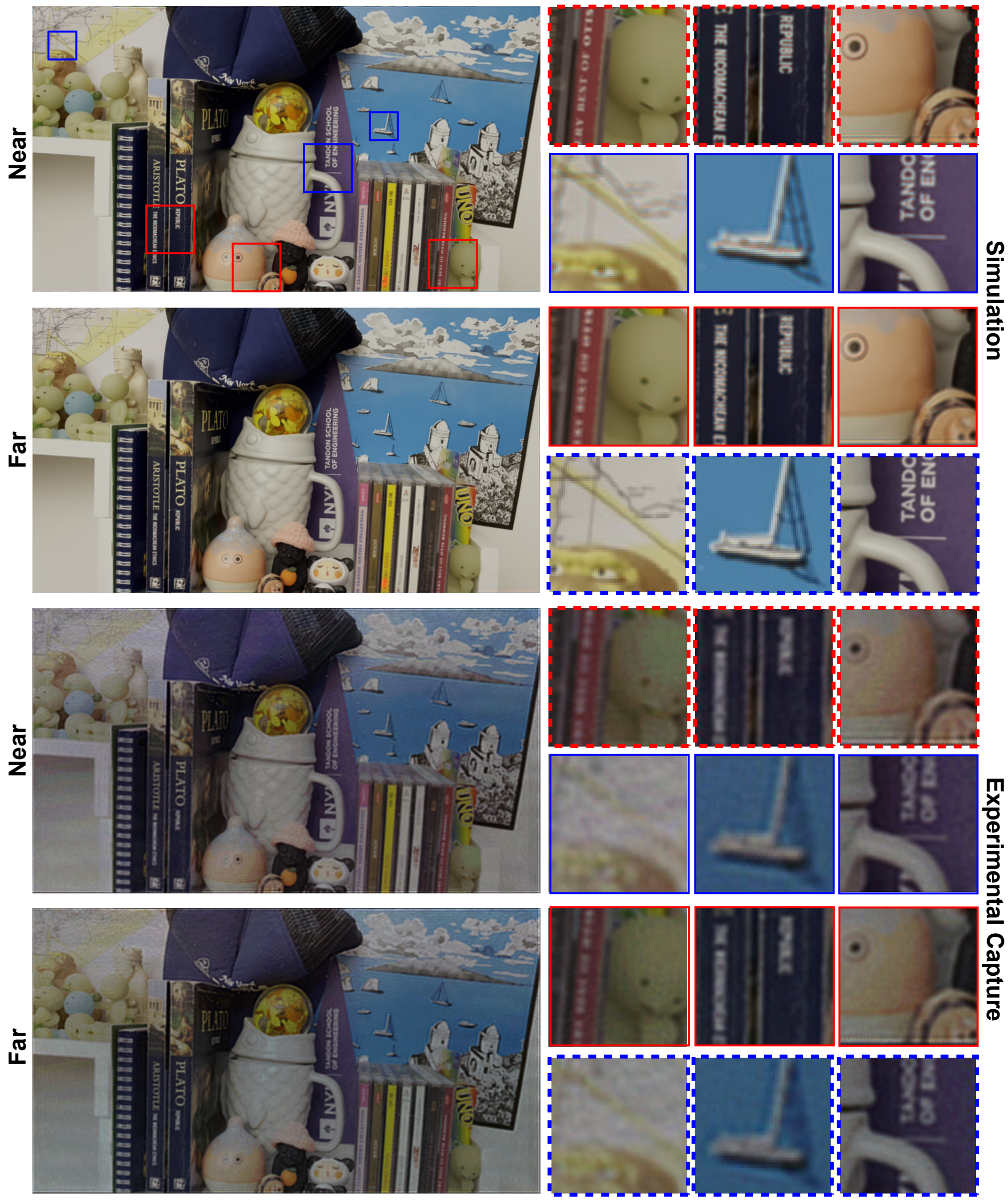

Fig. 7. *Shelf scene.* Simulated and experimental results predicted by NHF for the Shelf scene.

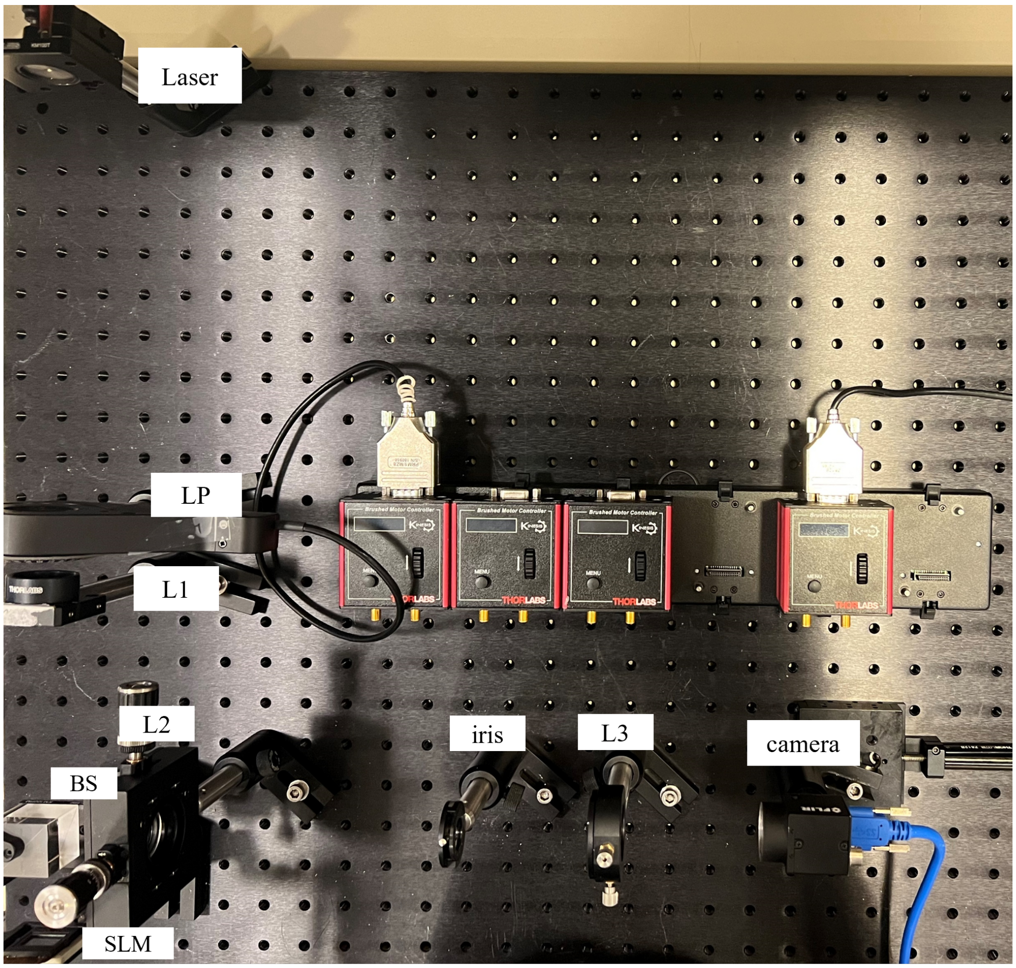

Fig. 8. *Prototype holographic display.*

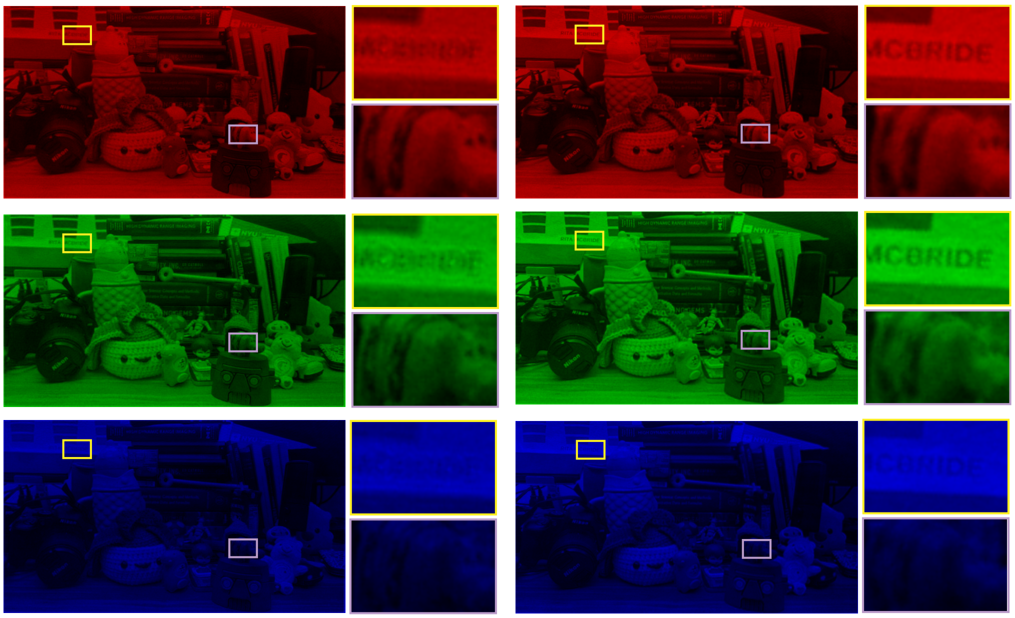

Fig. 9. *Monochromatic results are shown here for each laser. Left column is near focus, and right is far focus.*

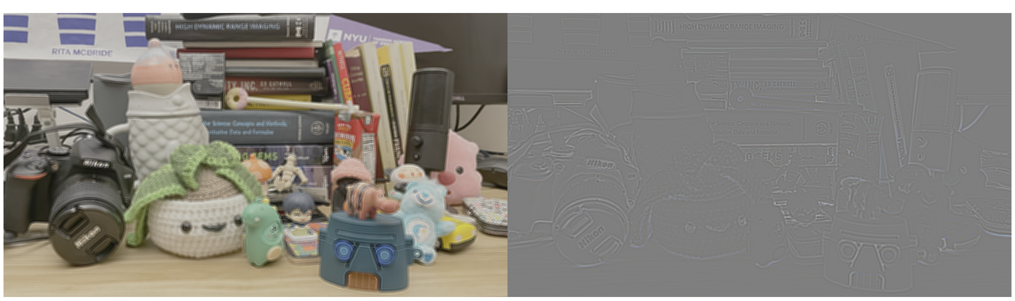

Fig. 10. *Holographic Stereogram.* Here, we show the output of the holographic stereogram method (inverse STFT), which converts light field to hologram. In this figure, the amplitude is shown on the left and phase to the right.