

Free Your Hands: Lightweight Turntable-Based Object Capture Pipeline

JIAHUI FAN, Nanjing University of Science and Technology, China

FUJUN LUAN, Adobe Research, USA

JIAN YANG[†], Nanjing University of Science and Technology, China

MILOŠ HAŠAN, Adobe Research, USA

BEIBEI WANG[†], Nanjing University, China

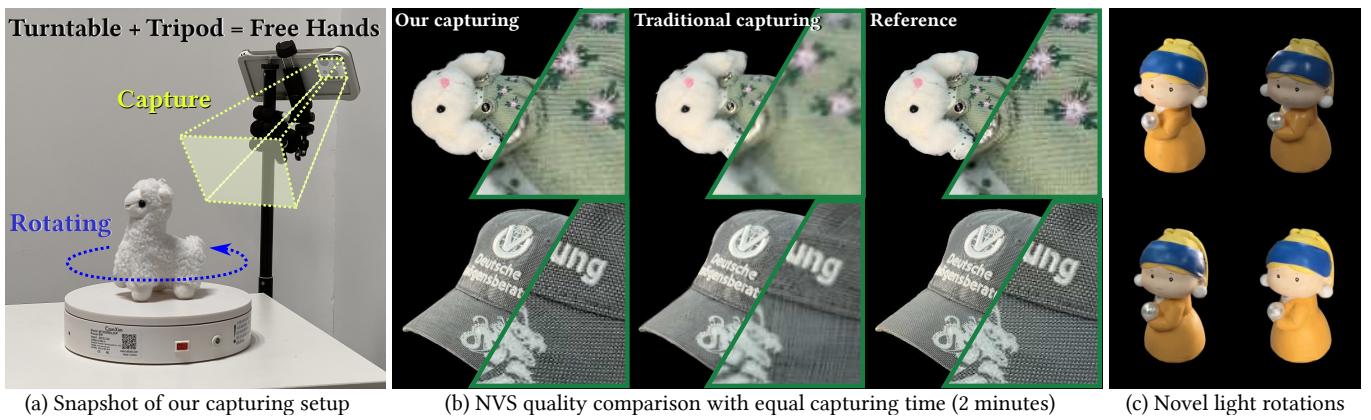


Fig. 1. We present a lightweight object capture pipeline to reduce manual effort, while increasing reconstruction quality and additionally supporting novel light rotations. (a) We use a consumer turntable to carry the object and a tripod to hold the camera, automatically capturing dense samples from various views and lighting rotations, obtaining hundreds of high-quality captures within 2 minutes. (b) We develop a rotation-conditional neural radiance representation, achieving NVS results with less blur and clearer details. (c) With our new representation, we can synthesize results under novel views and lighting rotations.

Novel view synthesis (NVS) from multiple captured photos of an object is a widely studied problem. Achieving high quality typically requires dense sampling of input views, which can lead to frustrating manual labor. Manually positioning cameras to maintain an optimal desired distribution can be difficult for humans, and if a good distribution is found, it is not easy to replicate. Additionally, the captured data can suffer from motion blur and defocus due to human error. In this paper, we use a lightweight object capture pipeline to reduce the manual workload and standardize the acquisition setup, with a consumer turntable to carry the object and a tripod to hold the camera. Of course, turntables and gantry systems have been frequently used to automatically capture dense samples under various views and lighting conditions; the key difference is that we use a turntable under *natural environment lighting*. This way, we can easily capture hundreds of valid images in several minutes without hands-on effort. However, in the object reference frame, the light conditions vary (rotate); this does not match the assumptions of standard NVS methods like 3D Gaussian splatting (3DGS). We design a neural radiance representation conditioned on light rotations, which addresses this issue and allows rendering with novel light rotations as an additional benefit. We further study the behavior of rotations and find optimal capturing strategies. We demonstrate our pipeline using 3DGS as the underlying framework, achieving higher quality and showcasing the method’s potential for novel lighting and harmonization tasks.

Additional Key Words and Phrases: Object capturing, Neural appearance, Gaussian splatting,

Authors’ addresses: Jiahui Fan, Nanjing University of Science and Technology, China, fjh@njust.edu.cn; Fujun Luan, Adobe Research, USA, fluan@adobe.com; Jian Yang[†], Nanjing University of Science and Technology, China, csjyang@njust.edu.cn; Miloš Hašan, Adobe Research, USA, milos.hasan@gmail.com; Beibei Wang[†], Nanjing University, China, beibei.wang@nju.edu.cn.

1 INTRODUCTION

Creating realistic 3D assets from real-world objects is a long-standing challenge of computer graphics, with applications to e-commerce, entertainment, digital heritage, and more. A typical pipeline captures multiple views (typically hundreds) of an object under fixed lighting and transforms the captured images into a 3D asset representation, which allows novel view synthesis (NVS). Various methods based on Neural radiance fields (NeRF) [Mildenhall et al. 2021] and 3D Gaussian splatting (3DGS) [Kerbl et al. 2023] can be used for this purpose. Still, high-quality capture requires careful manual labor and time. With hand-held cameras, one can either capture discrete images or video sequences, while in both cases, the quality of captured data can suffer from issues like motion blur and defocus, especially for capturing videos, as shown in Fig. 2. Moreover, the constraints on a reasonable sampling of viewpoints may be obvious to researchers, but manually positioning cameras to maintain the desired distribution can be difficult for average users.

Extensive research [Attal et al. 2023; Barron et al. 2021; Chen et al. 2022; Yu et al. 2024] improves upon the above NVS pipelines. For example, various lighting setups can be employed, including unknown environment lighting [Wang et al. 2021; Zhang et al. 2020], or a point light that is collocated with the camera [Bi et al. 2020a,b,c], or separate from the camera [Bi et al. 2024; Gao et al. 2020]. However, controlled lighting generally requires a dark room and more involved / expensive setups, and becomes out of reach for most non-expert users. Several works propose to utilize sparse views and cross-scene feed-forward inference techniques [Charatan

et al. 2024; Chen et al. 2024]. However, the reconstructed quality necessarily degrades without dense samples. Dense images can be captured with various specialized devices [Kang et al. 2023; Ma et al. 2021; Ye et al. 2024], which can reduce the required human labor and precision, but are not available off-the-shelf.

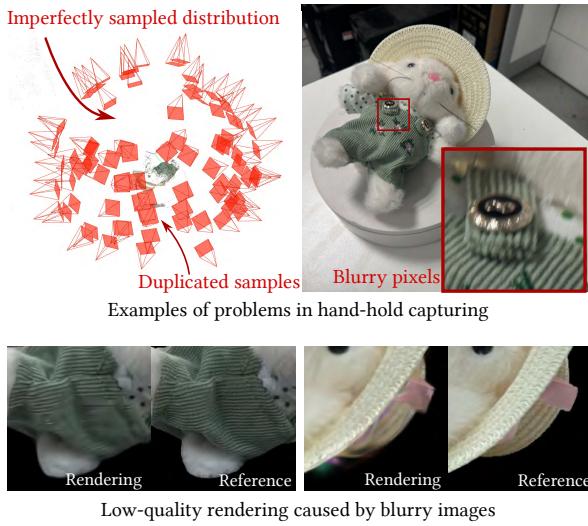


Fig. 2. Illustration of some the common problems of ordinary hand-held captured pipeline. The final NVS quality can suffer from the imperfectly sampled distribution and blurry images.

In this paper, we present *Free Your Hands*: a repeatable and robust lightweight object capture pipeline designed to reduce manual workload and standardize the acquisition process, while increasing reconstruction quality and offering novel illumination rotations as an additional benefit. We use a consumer turntable to carry the object and a tripod to support the camera. As the turntable rotates, we automatically capture dense samples from various views under rotating environment illumination (as observed from the object's frame of reference). This defines a controllable and repeatable sampling trajectory, and the tripod-mounted camera minimizes motion blur and defocus in video sequences, compared to stationary objects plus hand-held cameras. With this setup, we can easily capture hundreds of valid images in 2 minutes without much manual effort, while typically under 50 valid images with hand-held multi-view acquisition can be captured in the same time. We calibrate camera poses with a standard pipeline [Schonberger and Frahm 2016].

The captured images could theoretically be used in existing NVS frameworks (NeRF or 3DGS), but the challenge lies in the varying (rotating) light conditions that break the assumptions of these techniques and are harmful to reconstruction quality. Therefore, we define a new radiance representation conditioned on light rotations. Furthermore, we investigate the relationship between the rotation of objects and the NVS problem, providing experiments and analysis to find an optimal capturing setup that balances human labor and NVS reconstruction quality.

Of course, a turntable is not novel in computer graphics and appearance capture. Previous works focus on calibration and pose estimation from such data [Cheng et al. 2023; Elms et al. 2024; Puszta and Hajder 2016], or reconstruction of complex light transport with controlled light sources [Wu et al. 2018]. Methods that use point light illumination in a dark room [Bi et al. 2020a,b,c; Zeng et al. 2023] could use a turntable with minimal changes: if camera and light poses with respect to the object are known, it does not matter whether the object is rotating. However, we are the first to investigate the relationship of a turntable under *natural environment illumination* to the NVS problem, and to practically develop such a capture pipeline. To summarize, our main contributions include:

- a novel lightweight object capture pipeline that uses a turntable under natural environment illumination, combined with a tripod to reduce manual workload and standardize the acquisition process,
- fitting a rotation-conditioned radiance field representation, allowing accurate NVS as well as novel light rotation for applications such as harmonization,
- two rotation-based sampling strategies tailored for our capture pipeline, and
- a dataset of objects with varying material properties captured and reconstructed under different rotating configurations with our pipeline for downstream research.

2 RELATED WORK

Novel view synthesis. Novel view synthesis (NVS) aims to generate new images from view directions that were not originally observed. Extensive research on 3D representations has been proposed to enable realistic novel view rendering. Notably, NeRF [Mildenhall et al. 2021], 3DGS [Kerbl et al. 2023] and their follow-ups have garnered significant attention due to their powerful representation capabilities. NeRF models radiance using integrals over a ray passing through a volume, and addresses view-dependent radiance using multilayer perceptrons (MLPs) conditioned on the view ray direction; this is also true in most follow-up methods [Chen et al. 2022; Müller et al. 2022]. Enhanced anti-aliasing techniques [Barron et al. 2021, 2022, 2023; Zhang et al. 2020] and improved reflectance modeling [Attal et al. 2023; Verbin et al. 2022] have further refined the quality and performance of NeRF-based representations. 3DGS [Kerbl et al. 2023] employs anisotropic Gaussians to represent scenes, allowing for great adaptivity to actual geometric content and enabling real-time, highly detailed renderings. The view dependence of radiance is represented using spherical harmonics, which is even more limited than MLP-based view-dependence.

Camera pose calibration. These above techniques rely on posed images to densely reconstruct 3D scenes, with the quality of the input images significantly influencing the final rendering results. Structure from Motion (SfM) methods [Hartley and Zisserman 2003; Mur-Artal et al. 2015; Schonberger and Frahm 2016; Taketomi et al. 2017], particularly COLMAP [Schonberger and Frahm 2016], are commonly employed to calibrate camera poses and provide initialization for point-based techniques. Although the quality of input data is typically assumed to be sufficient for COLMAP to succeed, this assumption is frequently invalid. Consequently, standardizing the

data acquisition process and minimizing human error is beneficial to ensure robust and reliable performance in NVS applications. Recently, VGGT [Wang et al. 2025] introduces an end-to-end transformer model to automatically predict all key 3D attributes of a scene by feeding several multi-view images, providing a powerful solution that is compatible with multiple tasks in 3D computer vision.

3 METHODS

In this section, we first briefly review the traditional multi-view NVS strategy with static objects (Sec. 3.1), and then introduce the insight of covering the "2D" sampling domain of views and light rotations with our turntable capturing pipeline (Sec. 3.2). Next, to achieve NVS in the proposed sampling domain, we present a conditional radiance representation (Sec. 3.3), and we further investigate the relationship between rotation and NVS quality to find an optimal setup (Sec. 3.4).

3.1 Static sampling

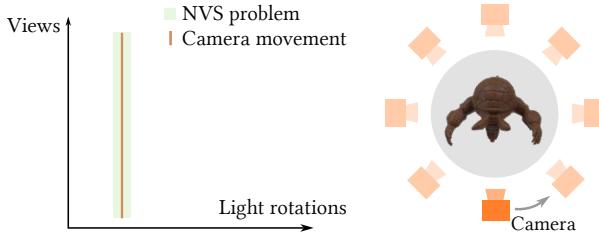


Fig. 3. The traditional capture strategy visualized in a 2D sampling domain. The object is stationary, and the camera is placed at multiple positions for a dense sampling of the NVS problem.

To create a 3D asset from a real-world object for novel view synthesis, a common approach is to capture hundreds of images by moving the camera around the object under static light conditions and then reconstruct the radiance field with NeRF or 3DGS or their variants. The entire pipeline can be formulated as follows:

$$L_{\text{static}} = \mathcal{G}(\mathbf{V}|I), \quad (1)$$

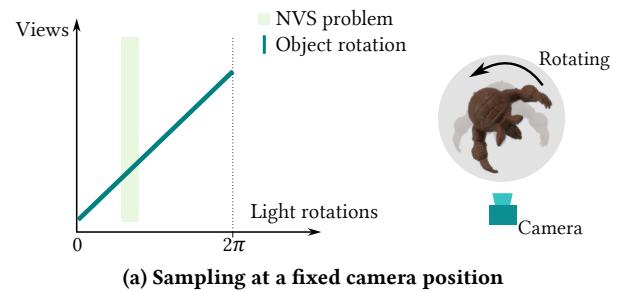
$$\{\mathbf{V}\} = \{\mathbf{V}_i | i = 1, 2, \dots, M\}, \quad (2)$$

where L_{static} represents the radiance field, \mathcal{G} denotes a reconstruction operator, which could be based on NeRF or 3DGS or their variants, \mathbf{V} is a set of camera views described by camera intrinsics and extrinsics, and I is the illumination condition under which the images are captured. The camera is placed at M different positions. The target of NVS is to predict the radiance L'_{static} at a novel viewpoint v' , while the lighting I stays the same. We can visualize this formulation in the object frame of reference as a column in a "2D" sampling domain, where the two "dimensions" represent the camera view and the light rotation angle, as shown in Fig. 3. Note that for simplicity, we visualize the camera views as one vertical dimension. Since the object (and therefore the light rotation) stays fixed throughout the capture process, we call it the *static* sampling strategy.

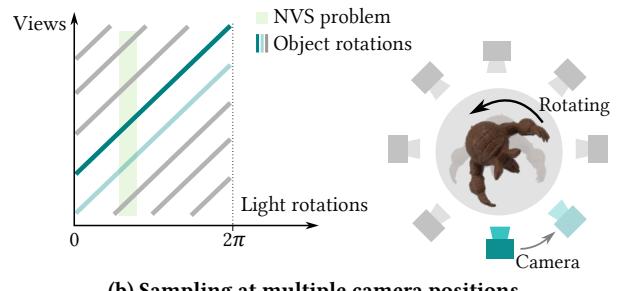
To get a faithful description of the entire "column", it is necessary to densely sample the viewpoints. The user should ensure a reasonable distribution of sampled views and be careful not to introduce motion blur and defocus issues; this is a difficult task for non-experts. The question is: can we free our hands while maintaining high-quality NVS reconstruction?

3.2 Rotating sampling

Instead of moving the camera, we can rotate the object itself. An effective approach is to use a turntable to hold the object, allowing it to rotate automatically. We can position the camera on a tripod at a fixed location to ensure video (or image burst) sequences without blur. As the turntable rotates, high-quality sequences are captured and automatically ensure a uniform sampling along the trajectory. We call this the *rotating* sampling strategy. This strategy can be illustrated in the 2D sampling domain as well, but as a diagonal line, as shown in Fig. 4 (a). Note again, we visualize the camera views as one vertical dimension; the diagonal line represents the intuition that the camera view and light rotation, in the object frame of reference, are "tied" together.



(a) Sampling at a fixed camera position



(b) Sampling at multiple camera positions

Fig. 4. Our *rotating* sampling strategy illustrated in the 2D sampling domain. The object is rotating on a turntable, and the camera is placed at a fixed position (a) or multiple positions (b). Each camera position leads to an individual diagonal line in the sampling domain.

Suppose we place the camera at M different locations, and the turntable completes a full rotation each time. This strategy can be illustrated as multiple diagonal lines, as shown in Fig. 4 (b). Here the intuition is that the camera view and light rotation, in the object frame of reference, are "tied" together differently for each turntable sequence. From each video sequence corresponding to a

camera position, we sample N images, resulting in MN images for reconstruction. Now the reconstruction formulation becomes:

$$L_{\text{rotating}} = \mathcal{G}(\mathbf{V}, \mathbf{I}), \quad (3)$$

$$\{\mathbf{V}\} = \{\mathbf{V}_i | i = 1, 2, \dots, M\}, \quad (4)$$

$$\{\mathbf{I}\} = \{\mathbf{I}_j | j = 1, 2, \dots, N\}, \quad (5)$$

where L_{rotating} is a radiance field that is conditioned on view directions and light rotations.

With M different camera positions and a full 360-degree turntable rotation per camera position, we have M diagonal lines in this 2D sampling domain. With well distributed camera positions, we can get good coverage of the 2D domain with multiple diagonal lines, which allows high-quality novel view and light rotation synthesis. (We can further optimize this strategy by letting the turntable rotate for a shorter angle per camera view, which we will present in Sec. 3.4.)

3.3 Conditional radiance field with rotating lighting

The next question is how to reconstruct the conditional radiance field from these captured images with light rotations. The original formulation for radiance field reconstruction, as presented in Eqn. (2), essentially operates as a one-variable regression problem, predicting radiance values from different viewpoints. In contrast, our conditional radiance field reconstruction approach incorporates a rotating light environment (in the object frame), making it a regression problem involving two variables. The key insight is to fit a radiance field as a function of both camera views and light rotations, and interpret the NVS problem as a one-dimensional slice of this field.

Either 3DGS or NeRF variants can be used as the baseline 3D representation with our capturing pipeline. In the following, we choose 3DGS for demonstration, and we also provide results for a NeRF-based variant in our supplementary materials. In standard 3DGS, radiance is represented using spherical harmonics (SH) basis functions; making them conditional on varying light rotations is not convenient. Instead, we use a neural network conditioned on light rotations (in addition to view directions). Specifically, each Gaussian has an appearance latent (feature) vector, which can be interpreted by an MLP. The final radiance (color) at each Gaussian is computed as:

$$c = \text{MLP}(x|v, \theta), \quad (6)$$

where x is the latent vector stored in each Gaussian point, v is the view direction (a normalized 3D direction), and θ is the light rotation angle. See Sec. 5 for more details on the MLP architecture. We evaluate the network before splatting the Gaussians, and the final pixel colors are blended using standard 3DGS differentiable rasterization, which does not require any modifications.

3.4 Swing sampling for optimal NVS quality

With the *rotating* sampling strategy ($s = 2\pi$), many valid image samples can be obtained. However, when the object rotates, the illumination also rotates, introducing more variation of the radiance at the same time. As a result, the rotation makes the reconstruction of the radiance field more difficult, and not all images from a rotation are equally helpful to the final NVS, since some samples are far from the problem in the 2D sampling domain. Therefore, we need to find

an optimal rotating strategy to balance the sample number and the difficulty in reconstruction. In this section, we further investigate the behavior of rotating capturing and find an optimal setting (rotating angle s) for our capturing pipeline.

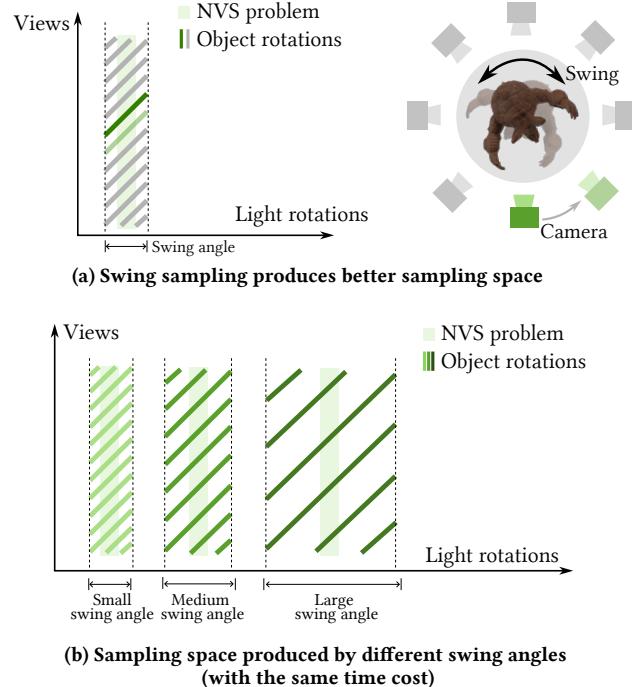


Fig. 5. Our *swing* sampling strategy illustrated in the 2D sampling domain. (a) The object is swinging around the original position on a turntable, and the camera is placed at multiple positions. (b) Different swing angles can lead to different shapes and densities of the sampling region, given the same amount of capturing time.

To balance between the sample number and NVS quality, the key is to find the most helpful samples, which means to sample near the NVS problem (i.e., the original light condition). Instead of rotating for full circles, we can let the object *swing* around its original position back and forth. In this case, the turntable first rotates for s radians (clockwise) and stops for m seconds. Then, it starts rotating backward (counter-clockwise) for the same angle s , stops for another m seconds, and repeats. During each rotation, we still set the camera on the tripod to capture video sequences, and during the m seconds between each rotation, we manually relocate the tripod and camera to get ready for the next round. We also illustrate the *swing* sampling strategy in Fig. 5 (a).

We can compute the total time cost t of such a *swing* capture as:

$$T = M \frac{s}{v} + (M - 1)m, s \in [0, 2\pi], \quad (7)$$

where v is the angular speed of the turntable. The total number of valid samples can also be represented as:

$$P = Msn, \quad (8)$$

where n is the number of samples per radian. With the formulation above, given the same time cost t , we can derive and conclude that:

- A larger swing angle generates more samples.
- A larger swing angle involves less human labor, since we need to relocate the camera for $M - 1$ times.
- A larger swing angle makes a wider sampling region in the 2D sampling domain, leading to images that are less helpful to the final NVS quality. We also illustrate the effect of different swing angles in Fig. 5 (b).
- When $s = 0$, this *swing* strategy degrades to *static* sampling (Sec. 3.1).
- When $s = 2\pi$, this *swing* strategy degrades to *rotating* sampling (Sec. 3.2).

Unfortunately, it is nontrivial to find an analytical solution for the optimal swing angle \hat{s} . However, with experiments, we show the relationship between swing angle s and the final NVS quality in Fig. 6 and the peak signal-to-noise ratio (PSNR) values in Table 1. As a conclusion, we find that $s^* = 0.2\pi$ achieves a good balance between manual labor and NVS quality, and we also provide some rendering comparisons in our supplementary.

Table 1. The comparison of NVS qualities with different swing angles in the same capturing time (2 minutes). The qualities are evaluated in PSNR (\uparrow), and we take the Clown scene as an example. Overall, $s^* = 0.2\pi$ achieves a good balance between the manual labor, sample number and final NVS quality.

Swing angle s	Total images P	NVS quality (PSNR)
0 (<i>static</i>)	50	37.73
0.05π	54	36.34
0.1π	96	38.47
0.2π	140	38.80
0.25π	162	38.08
0.5π	234	36.79
π	324	36.61
2π	360	37.17

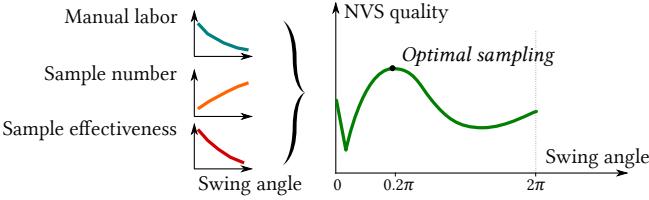


Fig. 6. The trends of the manual labor, sample number and sample effectiveness, and their combined effects on the final NVS quality when increasing the swing angle. We observe that $s^* = 0.2\pi$ produces the highest NVS quality, achieving a good balance among all these factors.

4 APPLICATIONS

With our proposed capturing pipeline, we can capture images and predict not only the NVS results at the original light condition,

but also any light rotations at any views, allowing for extensive applications in different scenarios. Here, we outline several of these applications, and we also demonstrate them in Sec. 6.3.

NVS with light rotations. One obvious application of our pipeline is for NVS, by predicting the radiance field at any given light rotation. Additionally, our pipeline also naturally supports adjusting the light condition for objects at given views, which can be used in tasks like harmonization. Note that our prediction only works for the rotation angles within the range of captured data. Therefore, in terms of adjusting light conditions, we suggest that users increase the swing angle to have more flexibility.

NVS with distilled SH representation. Besides using the neural representation directly, it can be distilled into other representations (e.g., SHs) by fixing a light rotation angle. The distilled representation allows novel view synthesis, while being compatible with existing 3DGS-based applications.

Novel lights from combination of rotations. Since our neural representation encodes the radiance field with multiple light conditions, we can relight the object by combining different light rotations (e.g., front-lit and back-lit ones), achieving extensive light conditions. Furthermore, if the original environment lighting is known, it is possible to approximately fit a target lighting using the known rotations, in order to achieve even more complex relighting tasks.

5 IMPLEMENTATION DETAILS

Capturing and dataset setups. We use the 48 mm camera from an iPhone 15 Pro Max for capturing, and the snapshot of our capturing environment is shown in Fig. 1 (a). We provide both the *rotating* dataset (i.e., $s = 2\pi$) and *swing* dataset (i.e., $s = 0.2\pi$) for extensive validation and downstream research. Both dataset consists of 35 real-world objects with diffuse, furry, or glossy appearances. For *rotating* capturing, we capture $M = 8$ turntable videos and select $N = 60$ frames from each video, resulting in a total of 480 images. For *swing* capturing, we set $M = 20$ and $N = 7$, resulting in 140 images. During capturing, we position the camera approximately every $\frac{\pi}{4}$ radians around the object, while varying elevation angles. To obtain sufficient samples, a capturing process takes us from 2 minutes ($s = 0.2\pi$) to 3 minutes ($s = 2\pi$).

Dataset preparation. After capturing, we crop the video into 1440×1440 rotation segments and randomly select frames from each rotation. Since the turntable rotates at a fixed angular speed, we divide the rotation angle by the number of frames to determine the rotation angle at each selected timestamp. We then conduct calibration using COLMAP [Schonberger and Frahm 2016], and finally remove the background and mask them using SAM2 [Ravi et al. 2024]. Compared to the traditional *static* capturing, our pipeline does not introduce any extra procedures in data preparation. Preparing a dataset of an object usually takes us 30 minutes.

MLP architecture and optimization. In our neural radiance representation, the MLP is 128-channel with 2 hidden layers, using a level-1 frequency encoding for the input view directions and light rotations. The latent vectors are 8-dimensional. During optimization, the network weights and the latent vectors are jointly updated and

optimized to fit the appearance under the rotating light conditions. We use the same loss functions from the standard 3DGS.

6 RESULTS

6.1 Experiment setup

We validate our pipeline on our captured and synthetic datasets. We run all experiments on an Ubuntu 22.04 LTS distribution powered by Windows Subsystem Linux 2. The optimization typically takes about 5-10 minutes on an RTX4090 GPU. We implement our neural radiance representation based on 3DGS [Kerbl et al. 2023] for comparison. There are other advanced approaches [Liu et al. 2024; Meng et al. 2024; Yu et al. 2024] that can enhance the NVS quality, while they are orthogonal to our pipeline. We did not perform comparisons with them, since it is straightforward to implement our method based on any 3DGS-based NVS approaches.

6.2 Quality validation

Table 2. The NVS quality of our neural radiance representation and 3DGS on our captured datasets. We choose $s = 0.2\pi$ for the best NVS quality. All testing sets are with *static* capturing strategy. The results are evaluated in PSNR (\uparrow), and the best results are colored in red.

Scene	Ours	3DGS	Scene	Ours	3DGS
Alpaca	34.48	33.11	Jar	35.40	34.24
Bear	35.68	35.03	JoyCon	37.43	36.76
BirdGreen	34.84	33.71	Kirby	38.20	37.54
BirdPink	35.25	34.80	Kodak	34.42	33.06
Burger	37.92	36.22	Monza	38.36	37.46
Cap	28.55	26.31	Mug	31.08	30.22
Capybara	36.09	35.21	Octopus	40.35	38.70
CapySmall	38.71	38.41	Panda	35.19	33.06
Cat	34.22	33.73	Pearl	38.47	37.02
China	37.71	36.96	Penguin	34.73	33.20
Clown	38.42	35.92	Pine	31.63	29.69
Controller	34.83	34.00	Rabbit	30.69	29.83
Dinosaur	41.80	39.66	Rider	37.04	36.33
Dolphin	39.35	38.16	RiderSmall	38.98	37.49
Eggplant	36.09	34.30	Sparrow	35.13	33.83
Ferrari	36.52	35.45	Sunglasses	33.91	33.71
Fries	39.38	38.69	Truck	37.42	35.31
HanSolo	37.45	36.29	Tuan	36.01	35.00
Average	36.15	34.95			

The validation of our method includes two aspects: the NVS and the light rotation. Since we can choose different swing angles or different purposes, we provide both the *rotating* dataset ($s = 2\pi$) and *swing* dataset ($s = 0.2\pi$) for validation.

For NVS quality, we compare with traditional NVS methods on our *swing* datasets in Table 2 and Fig. 8. The training set is captured with swing angle $s = 0.2\pi$ and the testing set is *static*. Our method can produce higher quality compared to existing NVS methods on all scenes.

Table 3. The rendering quality at novel light rotations of our neural radiance representation and 3DGS on our captured datasets. We choose $s = 2\pi$ to best support light rotations. All testing sets are with *rotating* capturing strategy, where the rotating angles vary from 0 to 2π . The results are evaluated in PSNR (\uparrow), and the best results are colored in red.

Scene	Ours	3DGS	Scene	Ours	3DGS
Alpaca	31.83	29.83	Jar	27.82	23.48
Bear	33.50	25.67	JoyCon	37.74	30.54
BirdGreen	32.94	23.96	Kirby	36.44	29.55
BirdPink	32.90	27.05	Kodak	30.43	27.51
Burger	37.20	29.80	Monza	36.11	29.54
Cap	27.43	24.97	Mug	29.71	26.56
Capybara	33.16	22.79	Octopus	36.61	27.54
CapySmall	34.70	25.04	Panda	28.58	25.80
Cat	30.35	26.01	Pearl	37.26	29.09
China	33.95	30.07	Penguin	28.80	24.00
Clown	39.95	35.96	Pine	24.16	21.86
Controller	33.83	29.73	Rabbit	30.05	22.99
Dinosaur	35.25	32.52	Rider	28.12	25.41
Dolphin	38.03	33.51	RiderSmall	35.56	31.04
Eggplant	37.82	27.96	Sparrow	29.72	25.53
Ferrari	32.37	30.63	Sunglasses	32.20	28.75
Fries	34.96	28.77	Truck	35.44	25.62
HanSolo	36.06	31.88	Tuan	30.79	23.86
Average	33.10	27.63			

Table 4. The comparison of NVS qualities on synthetic datasets. All results are evaluated in PSNR (\uparrow), and the top-3 results are marked in red / orange / yellow. Note that in this comparison, the testing set is *static*. Our method produces higher-quality results from training images with rotating lights than 3DGS does with common dataset setups (100 static images), and our *swing* sampling strategy produces the overall highest quality, even compared to 3DGS with fairly sufficient ideally sampled training images.

Scene	Ours (140 swing)	Ours (480 rotating)	3DGS (100 static)	3DGS (480 static)
Armadillo	36.73	35.63	35.94	36.52
Ficus	35.45	34.15	34.00	35.12
Flowers	32.05	31.89	30.23	31.24
Lego	33.89	33.32	32.13	32.47
Average	34.53	33.74	33.07	33.83

For the rendering quality of light rotations, we compare with traditional NVS methods on our *rotating* datasets in Table 3 and Fig. 9. The training set is captured with rotating angle $s = 2\pi$, and the testing set is also *rotating*. By comparison, we find that our method has a much higher quality consistently. In particular, 3DGS exhibits an overly dark appearance at grazing angles and obvious blurriness on the furry surface, as it cannot handle rotating light conditions. Although they can reconstruct good geometries, the radiance (especially the shadow effects) is incorrectly predicted.

In contrast, our method produces results that closely match the reference, thanks to the conditional radiance representation.

In Fig. 1 (b), we compare the NVS quality from our *swing* capturing with that from the traditional *static* capturing in equal capturing time (2 minutes). We apply our neural radiance representation on both datasets to show the difference in the qualities of captured data. We set the rotation angle to be zero for the traditional capturing. Our capturing pipeline is not only easier to perform, but also producing higher-quality reconstruction with more detailed appearances.

In Table 4 and Fig. 7, we compare the NVS quality with 3DGS, assuming that we provide it with ideally captured datasets. Our neural radiance representation achieves higher-quality renderings than those from 3DGS with a common setup (100 static images) and is still competitive with results of 3DGS from abundantly sampled static views (480 static images). Note that in practice, even for 480 images with rotating lights, we only need less than 4 minutes to capture them, while capturing 100 high-quality static images already costs more than 4 minutes of difficult manual labor.

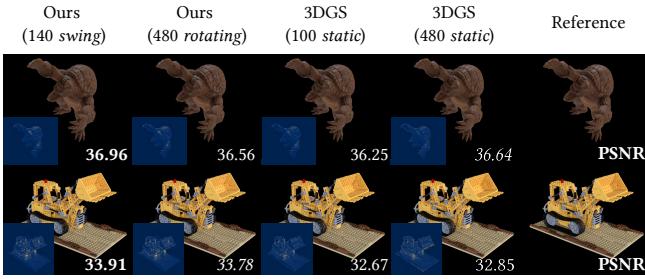


Fig. 7. Comparison of NVS results by our captured datasets and 3DGS with ideally sampled *static* datasets. The best/second-best results are marked as **bold/italic**. Our neural radiance representation achieves high-quality renderings, and is competitive even with results of 3DGS from abundantly sufficient training images.

In our supplementary material, we provide ablation studies on two crucial factors (the number of camera positions M and the swing angle s). In addition, we also show the optimal swing angles under different frequencies of the environment lights. Please refer to them for more details.

6.3 Validation on applications

NVS with light rotations. In Figs. 1 (c) and 10, we showcase the rendered result under different light rotations. With varying lighting conditions, our rendered results exhibit overall reasonable lighting outcomes. We also show the harmonization results in Fig. 11 and demonstrate this application in our supplementary video.

NVS with distilled SH representation. After selecting a lighting angle, our neural representation can be easily distilled into an SH representation. This way, it can fit existing 3DGS-based applications. In Fig. 13, we show the NVS results rendered with the distilled SHs under three different lighting rotations after a 3-minute distillation process. The distilled SHs achieve close qualities to the reference at the specified rotation angles. Note that all the results are obtained from only one training pass.

Novel light from combination of rotations. Our neural representation encodes the radiance field with multiple light conditions, which allows us to combine different light rotations (e.g., front-lit and back-lit ones), achieving a fused light condition. In Fig. 12, we demonstrate the rendering results by combining the front-lit and the back-lit scenarios, with different lighting colors. This way, our method enables the creation of rich and diverse assets.

6.4 Discussion and limitations

More flexible relighting. Our capture pipeline enables NVS with different light rotations or a combination of various lighting conditions. However, our method still cannot fully or flexibly support relighting in entirely new environments. We will address this in future work.

Data processing. Our capture pipeline relies on COLMAP [Schonberger and Frahm 2016] to calibrate our camera poses. However, the changing light condition during the capture slightly raises the difficulty of calibration. While using a checkerboard to carry the object can partially alleviate this issue, a COLMAP-free calibration approach will further improve our capture quality. Besides, like traditional capturing methods, the final quality of our datasets might be occasionally hurt due to the inconsistent segmentation from the large model [Ravi et al. 2024]. Therefore, a more accurate background removal approach will also help our pipeline.

Alternative NVS frameworks. We chose the basic 3DGS framework to validate the effectiveness of our capturing pipeline. However, there are still many advanced approaches that provide even higher NVS qualities. Their work is orthogonal to ours, and it is straightforward to adapt them into our pipeline to achieve advanced NVS qualities.

7 CONCLUSION

In this paper, we have presented *Free Your Hands*, a lightweight object-capturing pipeline to reduce manual workload, standardize the acquisition process, and ensure repeatability. The proposed capture pipeline consists of a simple setup: a consumer turntable to hold the target object and a tripod to hold the camera. As the turntable rotates, we can easily capture hundreds of valid images in several minutes without hands-on effort, minimizing human errors. Then, we design a neural radiance representation conditioned on light rotations tailored for the captured images, as well as an optimal rotation configuration for our pipeline in terms of the final NVS qualities. Our capture pipeline can be integrated into both NeRF- and 3DGS-based frameworks. We have demonstrated the effectiveness of our pipeline in the 3DGS-based framework across various applications, including NVS under different light rotations or combined lighting conditions and harmonization in novel light conditions, showing higher quality in NVS.

There are still many potential future research directions. One promising avenue is to expand the neural radiance representation into a fully relightable representation. Additionally, improving the NVS and relighting quality on some difficult types of objects, such as reflective or transparent ones, is also an interesting and challenging direction.

REFERENCES

- Benjamin Attal, Jia-Bin Huang, Christian Richardt, Michael Zollhöfer, Johannes Kopf, Matthew O'Toole, and Changil Kim. 2023. HyperReel: High-Fidelity 6-DoF Video With Ray-Conditioned Sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 16610–16620.
- Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5855–5864.
- Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5470–5479.
- Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2023. Zip-NeRF: Anti-aliased grid-based neural radiance fields. *arXiv preprint arXiv:2304.06706* (2023).
- Sai Bi, Zexiang Xu, Pratul P Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020a. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824* (2020).
- Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020b. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 294–311.
- Sai Bi, Zexiang Xu, Kalyan Sunkavalli, David Kriegman, and Ravi Ramamoorthi. 2020c. Deep 3d capture: Geometry and reflectance from sparse multi-view images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5960–5969.
- Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. 2024. GS³: Efficient Relighting with Triple Gaussian Splatting. In *SIGGRAPH Asia 2024 Conference Papers*.
- David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, and Vincent Sitzmann. 2024. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19457–19467.
- Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. TensorRF: Tensorial radiance fields. In *European conference on computer vision*. Springer, 333–350.
- Yuedong Chen, Haofei Xu, Chuanyia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. 2024. MVSpplat: Efficient 3D Gaussian Splatting from Sparse Multi-view Images. In *Computer Vision – ECCV 2024*, Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Günter Wiedenfeld (Eds.). Springer Nature Switzerland, Cham, 370–386.
- Zezhou Cheng, Matheus Gadelha, and Subhransu Maji. 2023. Accidental turntables: Learning 3d pose by watching objects turn. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2113–2122.
- Ethan Elms, Yasir Latif, Tae Ha Park, and Tat-Jun Chin. 2024. Event-based structure-from-orbit. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19541–19550.
- Duan Gao, Guojun Chen, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. 2020. Deferred neural lighting: free-viewpoint relighting from unstructured photographs. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 1–15.
- Richard Hartley and Andrew Zisserman. 2003. *Multiple view geometry in computer vision*.
- Kaizhang Kang, Zoubin Bi, Xiang Feng, Yican Dong, Kun Zhou, and Hongzhi Wu. 2023. Differentiable Dynamic Visible-Light Tomography. In *SIGGRAPH Asia 2023 Conference Papers*. 1–12.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)* 42, 4 (2023), 1–14.
- Xi Liu, Chaoyi Zhou, and Siyu Huang. 2024. 3dgs-enhancer: Enhancing unbounded 3d gaussian splatting with view-consistent 2d diffusion priors. *arXiv preprint arXiv:2410.16266* (2024).
- Xiaohu Ma, Kaizhang Kang, Ruisheng Zhu, Hongzhi Wu, and Kun Zhou. 2021. Free-Form Scanning of Non-Planar Appearance with Neural Trace Photography. *ACM Trans. Graph.* 40, 4, Article 124 (July 2021), 13 pages. <https://doi.org/10.1145/3450626.3459679>
- Jiarui Meng, Haijie Li, Yanmin Wu, Qiankun Gao, Shuzhou Yang, Jian Zhang, and Siwei Ma. 2024. Mirror-3dgs: Incorporating mirror reflections into 3d gaussian splatting. *arXiv preprint arXiv:2404.01168* (2024).
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* (2021).
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* 41, 4, Article 102 (July 2022), 15 pages. <https://doi.org/10.1145/3528223.3530127>
- Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics* (2015).
- Zoltán Pusztai and Levente Hajder. 2016. A turntable-based approach for ground truth tracking data generation. (2016).
- Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khader, Roman Rädele, Chloé Rolland, Laura Gustafson, et al. 2024. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024).
- Johannes L Schonberger and Jan-Michael Frahm. 2016. Structure-from-motion revisited. In *CVPR*.
- Takafumi Taketomi, Hideaki Uchiyama, and Sei Ikeda. 2017. Visual SLAM algorithms: A survey from 2010 to 2016. *IPSJ Transactions on Computer Vision and Applications* 9, 1 (2017), 1–11.
- Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. 2022. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 5481–5490.
- Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. 2025. Vggt: Visual geometry grounded transformer. *arXiv preprint arXiv:2503.11651* (2025).
- Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. 2021. NeRF+: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064* (2021).
- Bojian Wu, Yang Zhou, Yiming Qian, Minglun Gong, and Hui Huang. 2018. Full 3D reconstruction of transparent objects. *arXiv preprint arXiv:1805.03482* (2018).
- Sheng Ye, Yuze He, Matthieu Lin, Jenny Sheng, Ruoyu Fan, Yiheng Han, Yubin Hu, Ran Yi, Yu-Hui Wen, Yong-Jin Liu, et al. 2024. PVP-Recon: Progressive View Planning via Warping Consistency for Sparse-View Surface Reconstruction. *ACM Transactions on Graphics (TOG)* 43, 6 (2024), 1–13.
- Mulin Yu, Tao Lu, Linning Xu, Lihan Jiang, Yuanbo Xiangli, and Bo Dai. 2024. GSDF: 3DGS meets SDF for improved rendering and reconstruction. *arXiv preprint arXiv:2403.16964* (2024).
- Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. 2023. Relighting Neural Radiance Fields with Shadow and Highlight Hints. In *ACM SIGGRAPH 2023 Conference Proceedings*.
- Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. 2020. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492* (2020).

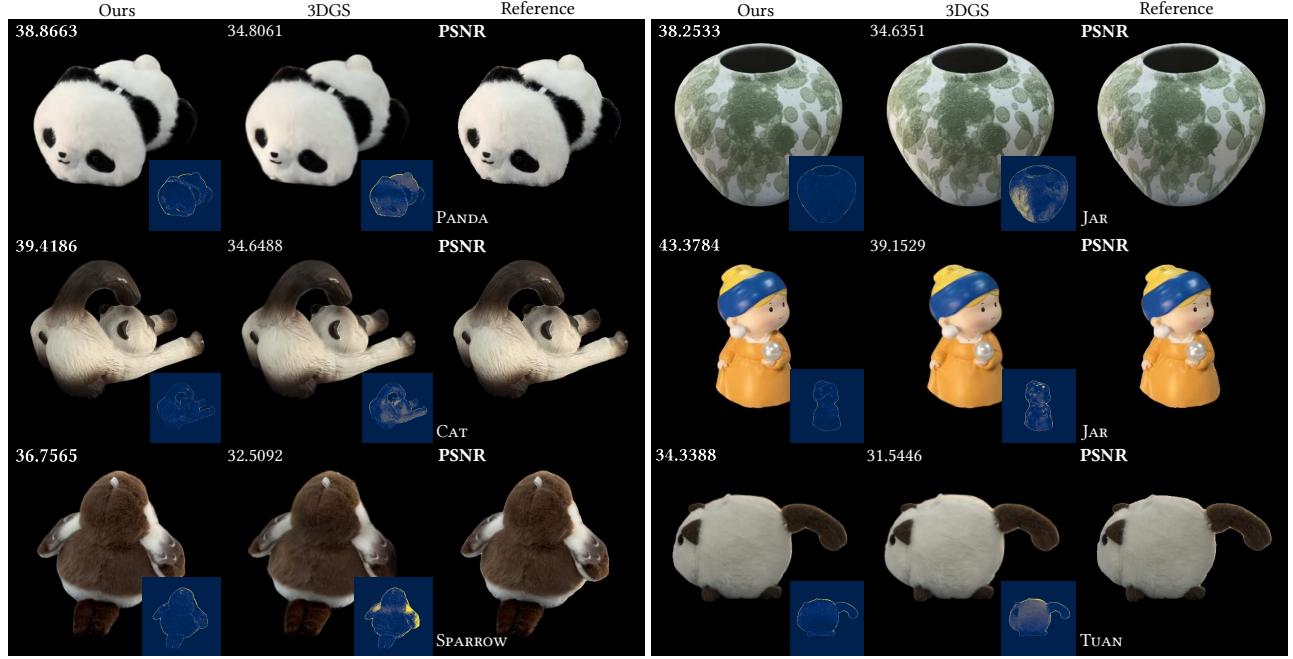


Fig. 8. Comparison of NVS results by our neural radiance representation and 3DGS on our *swing* datasets. The best results are marked as **bold**. Our capturing pipeline can produce higher-quality NVS results compared to 3DGS, and the whole capturing process is without much manual efforts.

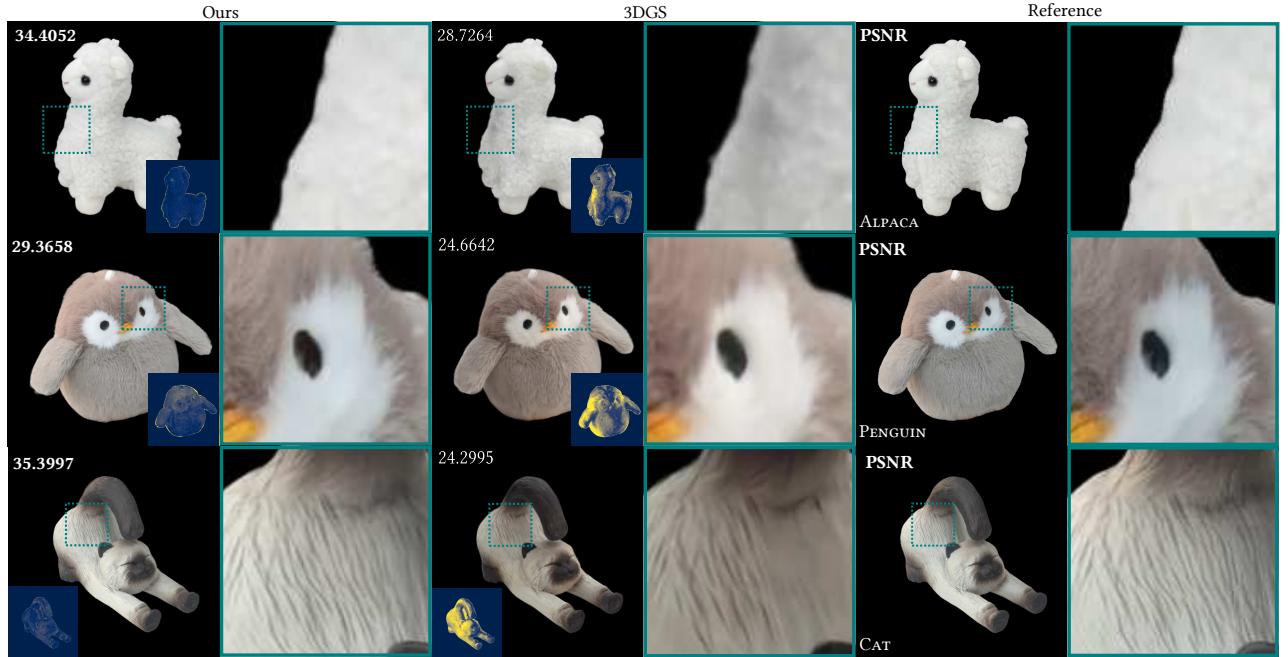


Fig. 9. Comparison of rendering results with novel light rotations by our neural radiance representation and 3DGS on our *rotating* datasets. The best results are marked as **bold**. 3DGS cannot handle the rotating light conditions, resulting in wrongly predicted shadow/light effects. In contrast, our pipeline provides closer results to the reference.

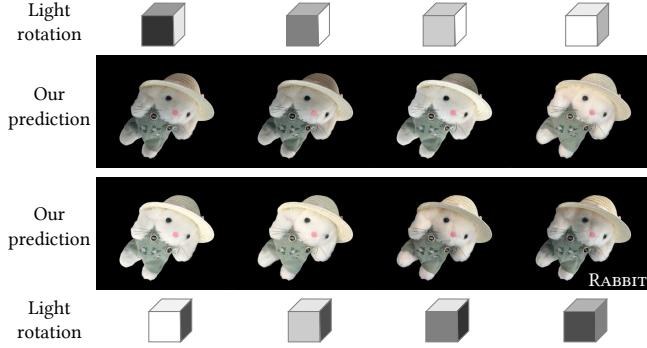


Fig. 10. The rendering results of our model from a novel view and novel light rotations. Our method can effectively predict reasonable light transition when light rotates, as our light-conditioned neural representation is learned based on samples from multiple light conditions.

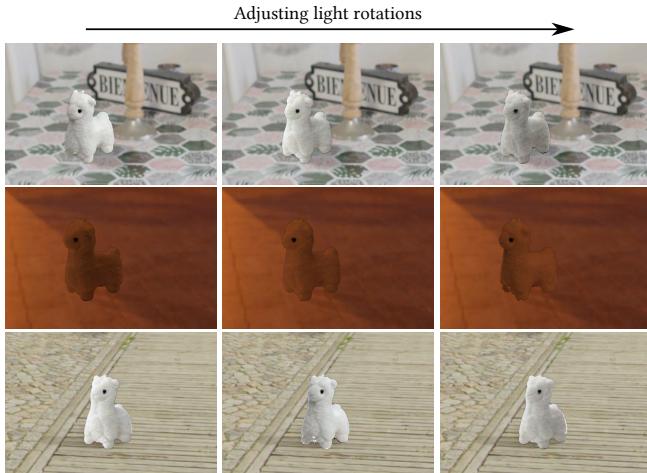


Fig. 11. Since our conditional neural representation encodes light rotations, we can achieve harmonization by simply adjusting the rotation angle to find the best-suitable appearance of lights and shadows.

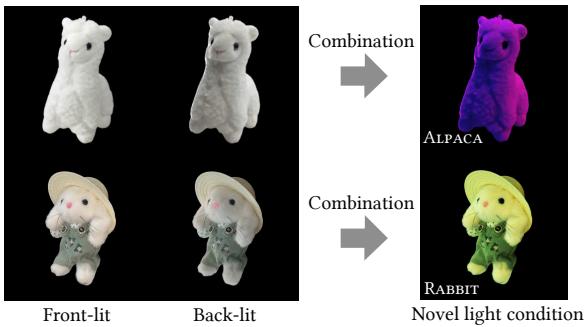


Fig. 12. Relighting the objects by linear combinations of light rotations with RGB weights. Top: blue and purple light combined. Bottom: yellow and green light combined.

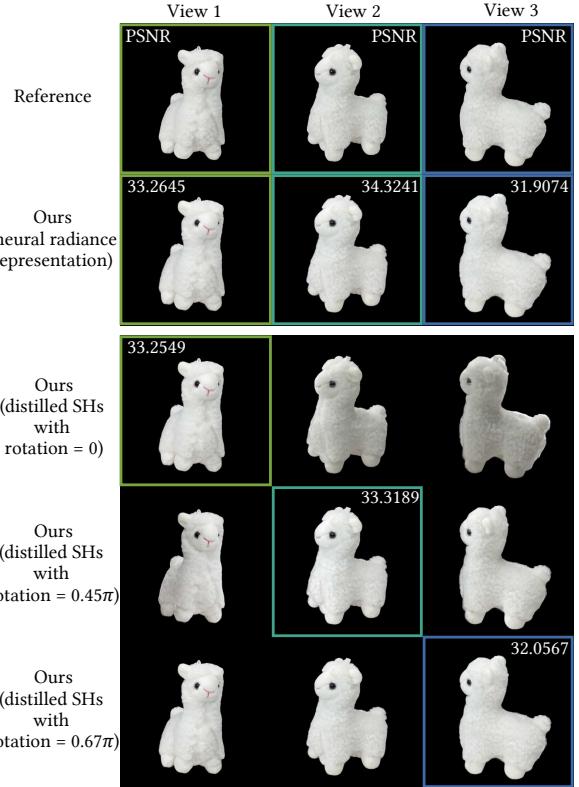


Fig. 13. The NVS result with SHs distilled from our trained model. By a simple distillation process, our conditional radiance representation can be exported into a static radiance field at a specific light rotation angle for simple NVS and easy cooperation with other applications.

Supplementary Materials:

Free Your Hands: Lightweight Turntable-Based Object Capture Pipeline

JIAHUI FAN, Nanjing University of Science and Technology, China

FUJUN LUAN, Adobe Research, USA

JIAN YANG[†], Nanjing University of Science and Technology, China

MILOŠ HAŠAN, Adobe Research, USA

BEIBEI WANG[†], Nanjing University, China

In the main paper, we have proposed *Free Your Hands*, a lightweight turntable-based object capturing pipeline. This pipeline allows for rendering with novel views and light rotations. We have also introduced a novel formulation of NVS problems in a 2D sampling space, and managed to find an optimal sampling strategy to build our pipeline. In this supplementary, we provide:

- results with a NeRF-based underlying representation [Chen et al. 2022],
- ablation studies on two crucial factors (number of camera positions M and swing angle s) regards the final NVS quality,
- and the trends of optimal swing angles with the environment lighting of different frequencies.

1 NERF-BASED IMPLEMENTATION

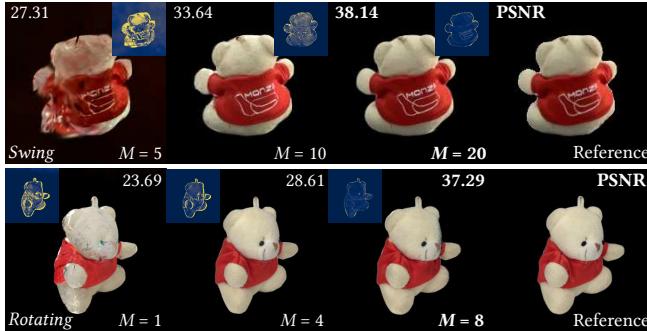


Fig. 1. The effect of the number of camera positions (M) in our pipeline. Given only few camera position, we observe one “diagonal line” in the 2D sampling space, resulting in severe over-fitting. In practice, we choose $M = 8$ for our *rotating* datasets and $M = 20$ for *swing* datasets.

Our pipeline is flexible and compatible to both 3DGS- and NeRF-based NVS approaches. We use neural features to represent the radiance, together with an MLP that is able to decode the feature vectors into RGB colors, shared by all neural features. In NeRF, the radiance is already represented by neural features. Therefore, we simply add the light rotation as a conditional input to the MLP decoder. For efficiency, we choose TensoRF [Chen et al. 2022] as the

Authors’ addresses: Jiahui Fan, Nanjing University of Science and Technology, China, fjh@njust.edu.cn; Fujun Luan, Adobe Research, USA, fluan@adobe.com; Jian Yang[†], Nanjing University of Science and Technology, China, csjyang@njust.edu.cn; Miloš Hašan, Adobe Research, USA, milos.hasan@gmail.com; Beibei Wang[†], Nanjing University, China, beibei.wang@nju.edu.cn.

Table 1. The NVS quality of our neural radiance representation and TensoRF on our captured datasets. We choose $s = 0.2\pi$ for the best NVS quality. All testing sets are with *static* capturing strategy. The results are evaluated in PSNR (\uparrow), and the best results are colored in red.

Scene	Ours	TensoRF	Scene	Ours	TensoRF
Alpaca	39.05	38.58	Jar	40.14	39.54
Bear	40.83	40.37	JoyCon	37.49	36.94
BirdGreen	36.94	35.52	Kirby	44.32	43.88
BirdPink	41.27	40.69	Kodak	33.02	32.60
Burger	38.75	38.33	Monza	43.23	42.78
Cap	38.91	38.45	Mug	30.96	30.43
Capybara	39.77	39.17	Octopus	41.67	41.08
CapySmall	30.83	30.35	Panda	37.36	37.03
Cat	36.05	35.54	Pearl	40.74	40.77
China	42.80	42.32	Penguin	36.17	35.71
Clown	43.28	43.01	Pine	35.25	34.66
Controller	38.96	38.38	Rabbit	34.56	34.05
Dinosaur	44.22	43.76	Rider	37.98	37.48
Dolphin	44.37	43.93	RiderSmall	39.12	38.78
Eggplant	38.20	37.98	Sparrow	36.76	36.49
Ferrari	35.82	35.27	Sunglasses	32.46	32.08
Fries	41.64	41.26	Truck	39.07	38.45
HanSolo	35.14	34.54	Tuan	39.51	38.94
Average	38.51	38.03			

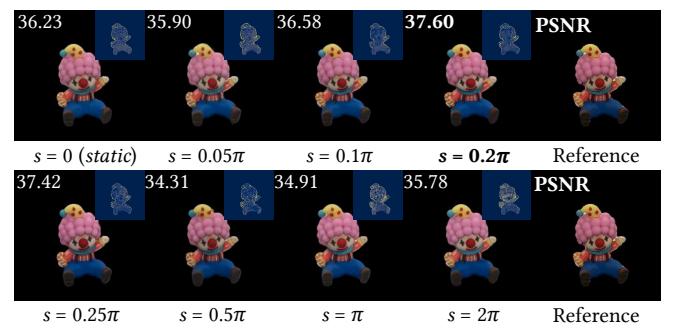


Fig. 2. The NVS results of different swing angle with the same capturing time (2 minutes). By comparison, $s^* = 0.2\pi$ achieves a good balance between the manual labor and final quality.

baseline method to validate our pipeline, since its training time is much shorter than the original NeRF [Mildenhall et al. 2021].

Table 2. The rendering quality at novel light rotations of our neural radiance representation and TensoRF on our captured datasets. We choose $s = 2\pi$ to best support light rotations. All testing sets are with *rotating* capturing strategy, where the rotating angles vary from 0 to 2π . The results are evaluated in PSNR (\uparrow), and the best results are colored in red.

Scene	Ours	TensoRF	Scene	Ours	TensoRF
Alpaca	36.24	26.10	Jar	30.40	25.05
Bear	36.70	26.89	JoyCon	36.34	30.99
BirdGreen	35.77	28.52	Kirby	41.29	32.77
BirdPink	37.05	26.44	Kodak	29.40	23.73
Burger	39.45	33.06	Monza	38.44	27.94
Cap	32.38	16.57	Mug	29.42	25.85
Capybara	34.90	24.39	Octopus	39.88	30.37
CapySmall	28.66	25.32	Panda	28.02	25.82
Cat	31.54	23.88	Pearl	38.17	27.25
China	34.91	28.57	Penguin	28.35	22.36
Clown	40.59	34.47	Pine	27.65	20.11
Controller	36.91	20.69	Rabbit	31.72	22.50
Dinosaur	34.19	30.89	Rider	28.58	24.97
Dolphin	37.81	33.92	RiderSmall	37.09	26.71
Eggplant	39.00	28.92	Sparrow	29.74	21.32
Ferrari	31.22	30.11	Sunglasses	32.18	28.78
Fries	37.28	29.06	Truck	35.89	26.15
HanSolo	34.28	31.32	Tuan	30.84	26.61
Average	34.23	26.90			

We provide the quantitative results in Tables 1 (for NVS qualities) and 2 (for light rotations). Overall, our method outperforms TensoRF on both datasets.

2 ABLATION STUDY

There are two key factors that decide the final NVS quality reconstructed from our captured dataset, and we provide ablation studies on them.

In Fig. 1, we show the rendering results of different numbers of camera positions (M) with our *swing* and *rotating* captures. For fairness, we provide the same total number of samples by increasing the frame numbers per rotation (N). By comparison, increasing the number of camera locations improves the rendering quality. The main reason is that sparse locations lead to uneven sampling on the 2D sampling space, which can lead to overfitting during optimization and hurt the rendering quality.

In Fig. 2, we compare the rendering quality of different swing angles (s) within the same capturing time (2 minutes). By comparison, $s^* = 0.2\pi$ provides the best rendering quality, as we described in the main text of this paper.

3 ENVIRONMENT LIGHT FREQUENCIES

By introducing the swing angles, we balance between the NVS difficulty and the number of samples. Generally, with a larger swing angle, the radiance variation when rotating objects is more obvious. We experiment and validate our choice of the optimal swing angle

with a common indoor environment lighting. However, if the environment light is of low frequency, which means the illumination varies little with rotation, it might leads to different optimal swing angles.

Theoretically, with lower-frequent (i.e., softer) lights, the swing angle should be increased. We validate this by gradually applying a Gaussian blur on an environment map and find their corresponding optimal swing angles. The blurred environment maps and their quantitative results are shown in Fig. 3. In conclusion, for common environments, $s^* = 0.2\pi$ is a suggested choice for balancing the human labor and NVS quality.

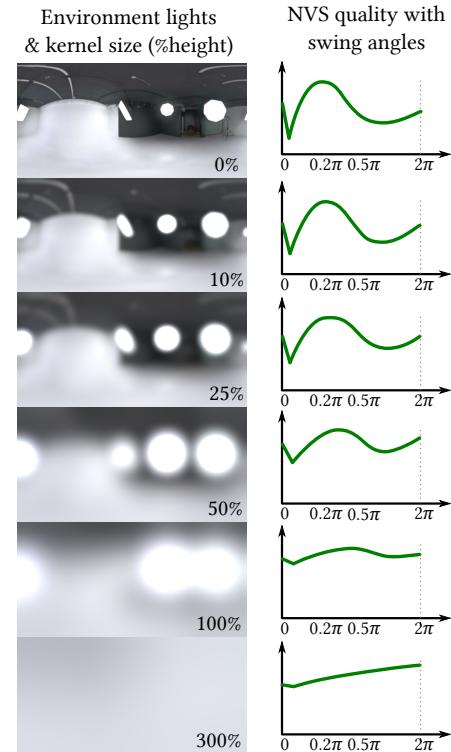


Fig. 3. Relationship between the NVS quality and swing angle in capturing with environment lights of different frequency. When the light becomes softer, the optimal swing angle is also relatively larger (from 0.2π to 0.5π). Overall, for common environments, $s^* = 0.2\pi$ can achieve the best quality.

REFERENCES

- Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. TensoRF: Tensorial radiance fields. In *European conference on computer vision*. Springer, 333–350.
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* (2021).