OLATverse: A Large-scale Real-world Object Dataset with Precise Lighting Control

Xilong Zhou[1], Jianchun Chen[*,1], Pramod Rao[*,1], Timo Teufel[1], Linjie Lyu[1],
Tigran Minasian[1], Oleksandr Sotnychenko[1], Xiao-Xiao Long[2],
Marc Habermann[1], and Christian Theobalt[1]

[1]Max Planck Institute for Informatics, [2]Nanjing University

[*]Equal contribution

Abstract

*We introduce **OLATverse**, a large-scale dataset comprising around $9M$ images of 765 real-world objects, captured from multiple viewpoints under a diverse set of precisely controlled lighting conditions. While recent advances in object-centric inverse rendering, novel view synthesis and relighting have shown promising results, most techniques still heavily rely on the synthetic datasets for training and small-scale real-world datasets for benchmarking, which limits their realism and generalization. To address this gap, **OLATverse** offers two key advantages over existing datasets: large-scale coverage of real objects and high-fidelity appearance under precisely controlled illuminations. Specifically, **OLATverse** contains 765 common and uncommon real-world objects, spanning a wide range of material categories. Each object is captured using 35 DSLR cameras and 331 individually controlled light sources, enabling the simulation of diverse illumination conditions. In addition, for each object, we provide well-calibrated camera parameters, accurate object masks, photometric surface normals, and diffuse albedo as auxiliary resources. We also construct an extensive evaluation set, establishing the first comprehensive real-world object-centric benchmark for inverse rendering and normal estimation. We believe that **OLATverse** represents a pivotal step toward integrating the next generation of inverse rendering and relighting methods with real-world data. The full dataset, along with all post-processing workflows, will be publicly released at https://vcai.mpi-inf.mpg.de/projects/OLATverse/.*

1. Introduction

The appearance of real-world objects is the result of the complex interaction between geometry, material, and lighting conditions. Acquiring high-fidelity material and geometry of objects, and synthesizing photorealistic appearance under novel illuminations, remain fundamental challenges in computer vision and computer graphics, with widespread applications in film and gaming industries, autonomous driving, robotics, and VR/AR. With the advances in deep learning and generative models, recent years have witnessed rapid progress in 3D generation and reconstruction [35, 41, 46, 48, 52, 54, 56], relighting [23, 34, 60, 61, 65] and inverse rendering [3, 16, 30, 31, 62, 67]. However, due to the lack of high-fidelity large-scale real-world datasets, the majority of these techniques are trained on synthetic datasets [9, 58] or evaluated on small-scale real datasets [3, 33]. As a result, the synthesized results are typically limited in the realism, and their performance in real-world scenarios cannot be reliably assessed because of the significant domain gap between real and synthetic data.

Considerable effort has recently been devoted to constructing large-scale and high-fidelity object-centric datasets. However, due to complexity in hardware and data processing pipelines, the existing datasets remain constrained in at least one of three key aspects: quality [7, 9, 45], scale [3, 11, 18, 21, 27, 28, 33, 51, 60, 72] or precise lighting control [12, 58]. As summarized in Tab. 1, the first category of datasets [7, 9, 45] comprises synthetic and hybrid objects. While the scale of these datasets is large, they typically lack realism and exhibit significant variance in object quality. The second category of datasets [12, 58] contains thousand-scale real-world objects, offering both realism and scale. However, the material creation in these datasets heavily relies on manual annotation, and the limited illumination setups prevent accurate simulation of complex light transport. The third group of datasets [3, 11, 33] captures objects under precisely controlled illumination, but this group is typically small in both scale and diversity, restricting their applications in comprehensive benchmarking

Figure 1. **OLATverse** is a large-scale high-quality real-world OLATs dataset comprising around $9M$ images of 765 objects. We illustrate a subset of objects (top left), demonstrating the diversity and large-scale coverage of **OLATverse**, and we also show one sample object captured under multiple viewpoints and precisely controlled single light sources (top right). **OLATverse** can be utilized as a comprehensive real-world benchmark for inverse rendering, novel view synthesis and normal estimation.

and generative models training. To the best of our knowledge, no existing object dataset simultaneously provides both large-scale coverage and high-fidelity appearance.

To address these challenges, we propose **OLATverse**, the first large-scale real dataset that provides high-fidelity images of a diverse set of objects captured under precisely-controlled illumination and camera configurations. Our dataset consists of 765 real-world objects, covering a wide range of material categories (e.g., wood, stone, leather, plastic, metal, paper, plaster, fabric, ceramic) and 18.5% LVIS categories [20] from common to uncommon objects. **OLATverse** is acquired using lightstage setup [8], which enables One-Light-at-a-Time (OLAT) capture of real-world objects and provides rich information for analyzing the complex interaction between surface reflectance properties and light sources. Specifically, each object is captured using 35 well-calibrated DSLR cameras and illuminated by 331 individually controlled light sources, simulating a set of illuminations including uniform lightings, OLATs, gradient illumination and pre-defined environmental illuminations. In total, this setup yields over $9M$ high-fidelity images. In addition, we develop efficient semi-automatic mask processing pipeline to extract high-quality masks for each object captured under multiple views. Furthermore, we employ polarized gradient illumination to recover surface normal and diffuse albedo, following techniques proposed by Ma et al. [36]. These auxiliary data are particularly valuable for evaluating and supervising multi-modal tasks.

We showcase the applications of **OLATverse** across multiple tasks. The linearity of light transport allows recomposing the captured OLAT images to synthesize object appearance lit under any arbitrary novel illuminations, enabling the creation of large-scale training data for generative priors. Furthermore, we curate a subset of 42 objects with diverse material categories to construct a comprehensive evaluation benchmark. Using this benchmark, we conduct representative baseline experiments on inverse rendering, relighting, view synthesis, and normal estimation. These applications demonstrate the potential of **OLATverse** to advance future research in realistic 3D vision and relighting, facilitating the integration of data-driven methods with real-world datasets in the graphics and vision communities.

In summary, our contributions are as follows:

- We introduce **OLATverse**, the first publicly available large-scale real-world dataset, comprising around $9M$ images of 765 objects with diverse material and object categories captured under precise lighting control.
- Each object is captured under precisely controlled light-

Table 1. Comparison of object-centric datasets targeting inverse rendering and relighting tasks. We list a detailed comparison of **OLATverse** with existing datasets across several key attributes. The compared aspects include number of object (# **Objs**), whether data source is real (**Real**), lighting conditions (**IllumCond**), number of illuminations (# **Illum**), number of views (# **Views**), and capture device (**Device**). In the column of **IllumCond**, ENV denotes environment illumination, PAT represents pattern illumination. *Unspec.* indicates that the corresponding information is not specified in the dataset. () indicates that only a small portion of the dataset satisfies the criterion.

Dataset	# Objs	Real	IllumCond	# Illum	# Views	Device
ABO [7]	8K	✗	ENV	3	–	–
ShapeNet-Intr [45]	31K	✗	ENV	36	–	–
TexVerse [68]	818K	(✓)	–	–	–	–
Objaverse [9]	858K	(✓)	–	–	–	–
NeROIC [27]	3	✓	ENV	4∼6	40	camera
Standford-ORB [28]	14	✓	ENV	7	70	scanner+camera
DTC [12]	2K	✓	(ENV)	2	120	scanner+camera
OmniObject3D [58]	6K	✓	–	–	–	scanner
DiLiGenT-MV [72]	5	✓	OLAT	96	20	scanner
GS^3 [3]	6	✓	OLAT	*Unspec.*	*Unspec.*	lightstage
NRHint [60]	7	✓	OLAT	*Unspec.*	*Unspec.*	smartphones
DIR [6]	16	✓	OLAT	144	2	LCD display
MIT-Intrinsic [18]	20	✓	OLAT	10	1	camera
ReNe [51]	20	✓	OLAT	40	50	robots
SSS-GS [11]	20	✓	OLAT	167	158	lightstage
OpenIllumination [33]	64	✓	PAT+OLAT	13+142	72	lightstage
OpenSubstance [40]	187	✓	PAT+OLAT	16+1620	270	lightstage
Ours	**765**	✓	**ENV+OLAT**	**11+331**	**35**	**lightstage**

ing conditions, including uniform illumination, OLATs, environmental illumination, and gradient illumination.

- We provide auxiliary data, consisting of well-calibrated camera parameters, accurate object masks, diffuse albedo, and surface normals recovered via polarized gradient illumination.
- We establish **OLATverse** as a comprehensive real-world benchmark for multiple tasks, and highlight its potential as a valuable resource for generative prior learning.

2. Related Work

2.1. Object-centric Datasets

The acquisition of high-fidelity and large-scale datasets is critical for advancing data-driven methods in inverse rendering and relighting tasks. Recent research has introduced different datasets that capture the geometry and reflectance properties of full-body human avatars [49, 50, 55, 69], human faces [42, 44, 63], planar material surfaces [10, 37, 53, 70, 71] and 3D objects [3, 6, 7, 9, 11, 12, 18, 21, 27, 28, 33, 45, 51, 58, 60, 68, 72]. In comparison, our work focuses on appearance capture of real-world objects. We categorize the existing object-centric datasets into several subgroups, as summarized in Tab. 1. The first group focuses on synthetic objects, which are highly scalable due to the procedural generation pipelines and the availability of public 3D models. ShapeNet-Intrinsics [45], for example, constructs a large-scale object intrinsic dataset from 3D models in the ShapeNet [4] collection. ABO [7] introduces synthetic household objects with complex geometry and materials. Objaverse [9] and TexVerse [68] significantly expand the diversity of objects by incorporating both synthetic and scanned real assets. While these datasets are large-scale, the quality of these datasets varies significantly across different objects, and a domain gap still remains between synthetic and real-world data. Consequently, effective curation strategies are required to filter out low-quality data, limiting their applicability in realistic 3D vision and relighting.

The second group of datasets focuses on collecting real-world objects with high-fidelity appearance. Such datasets are typically collected either in specialized studios or in-the-wild environments. Among this group, DTC [12] and OmniObject3D [58] provide the meshes of thousands of real objects via 3D scanning. Although these two datasets are large-scale, they offer limited support for appearance analysis and lack control over lighting conditions. Specifically, OmniObject3D does not support appearance capture, and DTC only contains only 50 objects captured under two environmental illuminations. In addition, the material maps provided in DTC are manually created by artists to approximate the realistic appearance, which introduces a gap to the real-world appearance.

The third group of datasets [3, 11, 33] captures OLATs in studios with a well-designed lighting setups (e.g., light-

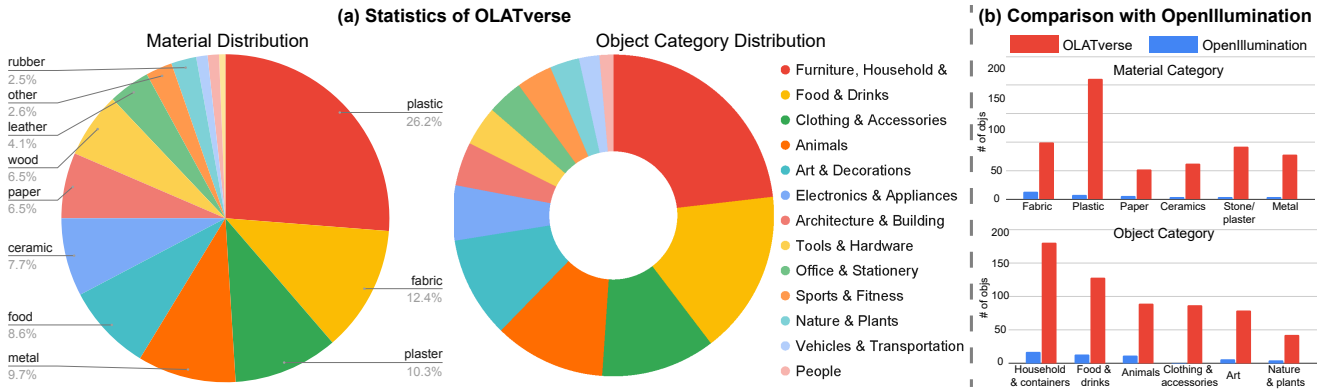

Figure 2. (a) We visualize the statistics of **OLATverse**, including the material distribution and high-level object category distribution. (b) We also show comparison against OpenIllumination [33] for the six largest material and object categories in terms of object count.

stage [8], LCD, smartphone flash, etc.), facilitating detailed analysis of light transport. However, due to the complexity of the setup and capture procedure, these datasets remain small in scale and limited in diversity. The largest OLAT dataset, OpenIllumination [33], contains only 64 objects, limiting its application for comprehensive benchmarking and generative model training. In comparison with existing object-centric datasets, **OLATverse** encompasses 765 real-world objects spanning diverse material categories and provides high-fidelity appearance under precisely controlled lighting, thereby achieving a combination of scale, diversity, and realism.

2.2. Inverse Rendering & Relighting

The goal of inverse rendering is to recover the intrinsic properties of scenes or objects (e.g., albedo, normal, roughness, and metallicity) to reproduce plausible renderings under novel illuminations. Traditional approaches typically leverage advanced differentiable renderers [29, 39, 43] under reliable constraints provided by geometry to jointly optimize intrinsic properties and lighting by minimizing the discrepancy between input and rendered images. With the advances in deep learning, Neural Radiance Fields (NeRF)[38] is introduced to jointly encode geometry and appearance in a single multi-layer perceptron (MLP), producing photorealistic rendering via volume rendering. NeRF demonstrates superior performance but suffers from slow rendering speed. To address this limitation, 3D Gaussian Splatting (3DGS)[24] is proposed to represent scenes with localized Gaussian kernels and can offer real-time efficiency through an efficient differentiable splatting process. Leveraging the benefits of these two representations, many recent inverse rendering methods [13, 14, 16, 22, 31, 34, 47, 64, 67] extend NeRF or 3DGS with additional optimizable Bidirectional Reflectance Distribution Function (BRDF) modules and a differentiable renderer to recover intrinsic properties and lighting from multi-view input images. More recently, with the rapid progress of diffusion-based generative models, a new branch of research leverages the strong priors of pretrained diffusion models for inverse rendering [5, 30, 59, 62] and relighting [23, 61] tasks. These methods typically fine-tune pretrained diffusion models on datasets annotated with material maps or renderings under novel illuminations. During inference, the diffusion model extracts material reflectance properties or relit images through inverse denoising steps. In this work, we demonstrate that **OLATverse** serves as a comprehensive benchmark for object-centric inverse rendering and relighting tasks, and we evaluate several representative approaches as baselines on our selected validation dataset.

3. OLATverse

In this section, we first provide an overview of **OLATverse**, introducing the object and illumination statistics (Sec. 3.1). We then describe the hardware configuration (Sec. 3.2), and the data post-processing pipeline (Sec. 3.3), including camera and light calibration, mask segmentation, photometric surface normal estimation, and diffuse albedo extraction.

3.1. Overview

Object Composition **OLATverse** comprises 765 real-world objects with high diversity in material types, object categories, and physical sizes. As shown in (a) of Fig. 2, it covers over 13 material categories, including wood, stone, plaster, fabric, plastic, metal, food, plant, ceramics, leather, wax, rubber, and paper. Furthermore, **OLATverse** consists of a wide range of common and uncommon real-world objects, spanning over 18.5% of LVIS categories [20], substantially outperforming existing real-world object datasets, such as OmniObject3D (10.8%) [58], OpenIllumination ($4 \sim 5\%$) [33] and DTC (3%) [12]. Additionally, **OLATverse** is not limited to object physical dimension, encompassing a wide range of sizes from 5cm to 100cm, which is broader than OpenIllumination dataset ($10 \sim 20$cm). These attributes highlight the diversity and scale of **OLATverse**,

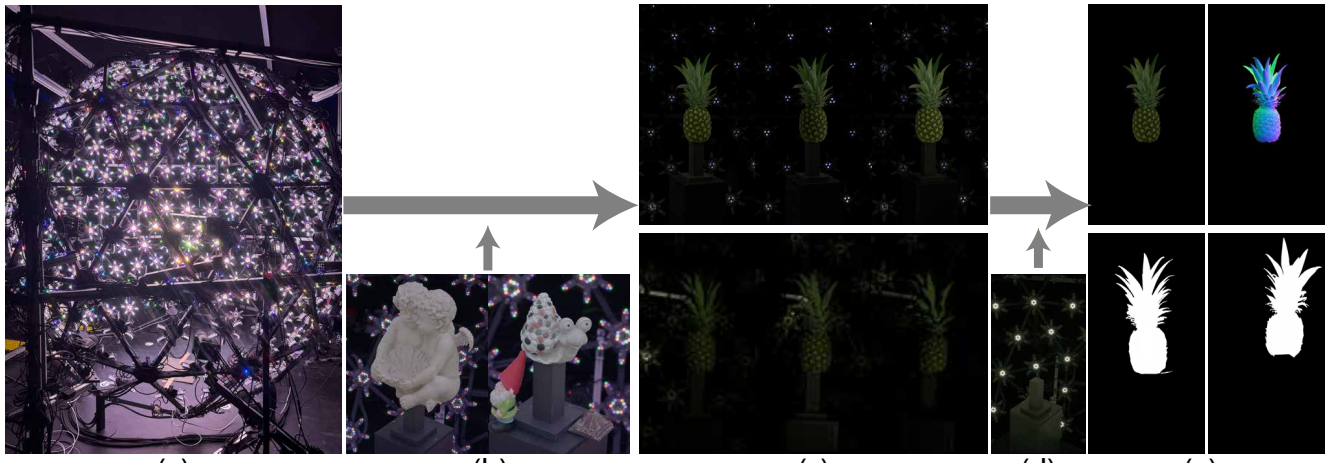

Figure 3. Illustration of the dataset capture setup and process pipeline. We utilize wooden stands with varying sizes and (a) a lightstage setup to capture raw videos of objects. During the calibration session, we record (b) reference objects to extract accurate camera parameters, which are utilized to extract (c) undistorted OLALs and relit images under varying illuminations from raw videos. Next, we capture (d) background stand image and perform (e) semi-automatic mask segmentation and normal extraction for each object.

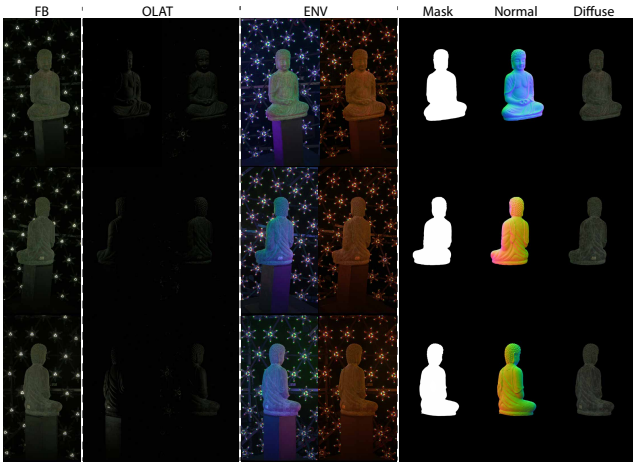

Figure 4. We visualize one sample of **OLATverse**, which includes full bright (FB), OLATs, relit images under varying pre-defined environmental illuminations (ENV), object mask, surface normals, and diffuse albedo.

establishing it as the first comprehensive OLAT dataset of real-world objects.

Illumination Composition In **OLATverse**, each object is captured using 35 DSLR cameras positioned at different viewpoints under a diverse set of lighting conditions, including 1 uniform white illumination, 12 polarized gradient illumination, 10 environment illuminations, and 331 OLATs, resulting in approximately 12K high-quality images per object. Specifically, uniform white illumination are utilized for mask segmentation and mesh reconstruction; gradient illumination are utilized in surface normals and diffuse albedo extraction; environment illuminations and OLATs capture the high-fidelity reflectance properties under precisely-controlled lighting and enable image-based rendering under novel illuminations. Some representative samples in **OLATverse** are visualized in Fig. 1 and Fig. 4.

Comparisons We compare **OLATverse** against one representative OLAT real dataset, OpenIllumination [33]. We visualize the number of objects in the six largest material and high-level object categories, as illustrated in Fig. 2 (b). While OpenIllumination covers the similar material and object categories as **OLATverse**, the scale of each category is significantly smaller than our dataset. In addition, OpenIllumination does not provide auxiliary resources such as surface normals and diffuse albedo, limiting its applications in the multi-modal tasks.

3.2. Capture Setup

We capture **OLATverse** using lightstage, as shown in (a) of Fig. 3, a spherical dome equipped with 35 RED Komodo 6K cameras and 331 controllable LEDs capable of emitting red, green, blue, amber, and white light (RGBAW). The cameras and lights are distributed 360° around the center of the dome, enabling synchronized multi-illumination capture at 30 FPS. Consistently capturing large real-world objects of varying sizes requires a carefully designed setup. At the center of the lightstage, we place a table and wooden solid stands with different sizes to support the objects of varying physical sizes. We select the stand with a surface area slightly smaller than the objects to minimize occlusion from the lighting located on the lower lightstage dome. Additionally, the table and stands are all covered with a black blanket or dark matte paper, to avoid unnecessary artifacts introduced by color bleeding and specular reflections from the supports. Furthermore, we manually adjust the focus of camera lenses to maintain consistent image quality, ensuring that objects of varying sizes appear properly scaled

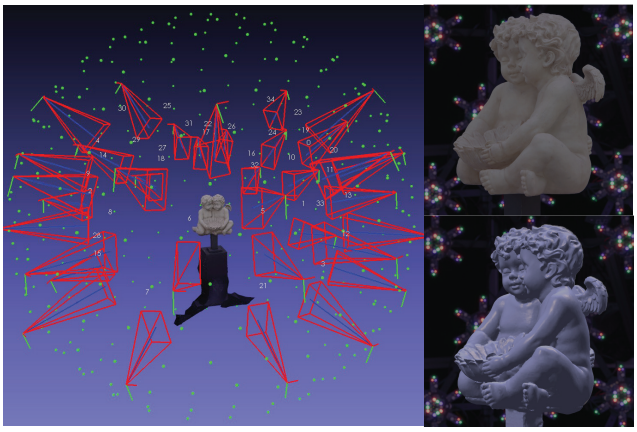

Figure 5. Visualization of calibrated cameras (marked as red), light sources (marked as green) and object mesh in our capture setup. In the left part of the figure, the reconstructed mesh decently matches the original image, demonstrating the correctness of the calibration process.

within the capture frame.

3.3. Data Processing

Camera and Lighting Calibration Acquiring accurate camera parameters consistently across large-scale 765 objects in a studio setup is a non-trivial task. Directly applying feature-based calibration algorithms to different objects leads to unstable and ambiguous calibration quality since the real-world objects vary widely in physical dimension, texture richness, and material properties, which limit the accuracy and consistency of the camera calibration process. To address this challenge, we leverage the fixed camera configuration of the lightstage, and perform a calibration session every $20 \sim 30$ regular capture sessions, demonstrated in (b) of Fig. 3. During calibration session, we lock down several fixed objects with rich texture and Lambertian surfaces as calibration references, and employ the feature-based algorithms implemented in Metashape [1] to recover both intrinsic and extrinsic camera parameters. For subsequent capture sessions, the previously estimated camera parameters are reused. All calibration captures are conducted under uniform white illumination to ensure robust and consistent feature detection. We quantitatively evaluate the calibration accuracy by computing the mean reprojection error of triangulated keypoints across the reference objects, yielding an average error of 0.86 pixels. The positions of individual lights are also provided within the same canonical system by measuring their physical location. We visualize the calibrated cameras, light sources, and the reconstructed mesh of one calibration reference object in Fig. 5. As is shown, the reconstructed mesh qualitatively aligns with real captured frames, demonstrating the accuracy of calibration process.

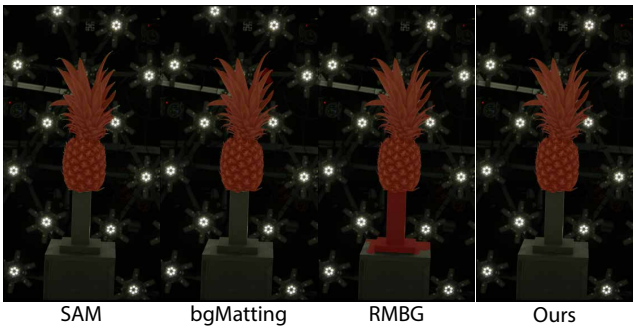

Figure 6. Semi-automatic mask processing. We show the object masks generated by SAM, bgMatting, RMGB-2.0 as well as our final mask produced by our proposed mask segmentation strategy.

Mask Segmentation Unlike human-centric datasets [50], where Sapiens [25] is tailored specifically for human body segmentation, efficiently producing high-quality masks for large-scale objects in our capture setup remains a challenging task. A straightforward solution is to employ Segment-Anything (SAM) [26] with multiple bounding-box prompts for instance segmentation, similar to OpenIllumination [33]. However, this strategy is inefficient and difficult to scale for a large-scale dataset. To efficiently tackle this challenge, we develop a simple yet effective semi-automatic segmentation pipeline that combines background matting (bgMatting) [32], SAM [26] and RMBG-2.0 [2]. Specifically, we capture a foreground image $\mathbf{I}_{\text{fg}}$, containing the object and stand, and a background image $\mathbf{I}_{\text{bg}}$ only for the stand, as is shown in (d) of Fig. 3. We empirically observe that bgMatting is capable of roughly separating the object from the stand, but fails to preserve fine details around the contour of objects. SAM, when guided by bounding boxes generated from bgMatting, can achieve a clean separation between the stand and the object but at the cost of mask quality. In comparison to the other two methods, RMBG-2.0 captures the most detailed contours but consistently treats the stand as part of the foreground, yielding a wrong object mask. Therefore, we leverage the strengths of all three segmentation strategies, and compute the final object mask $\mathbf{M}_{\text{obj}}$ as follows:

$$\mathbf{M}_{\text{stup}} = \begin{cases} \mathbf{M}_2(\mathbf{I}_{\text{bg}}) & \text{(a)}, \\ \mathbf{M}_2(\mathbf{I}_{\text{bg}})[1 - \mathbf{M}_3(\mathbf{M}_1(\mathbf{I}_{\text{bg}}, \mathbf{I}_{\text{fg}}))] & \text{(b)}, \end{cases} \tag{1}$$

$$\mathbf{M}^*_{\text{obj}} = \mathbf{M}_2(\mathbf{I}_{\text{fg}})(1 - \mathbf{M}_{\text{stup}}), \tag{2}$$

where $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{M}_3$ represent bgMatting [32], SAM [26] and RMBG-2.0 [2], respectively. $\mathbf{M}_{\text{stup}}$ represents the mask of the stand. In Eq. 1, *case a* is applied to lower camera views, where part of the objects may be occluded by the stand, while *case b* is used for all other views. Then we refine the intermediate results $\mathbf{M}^*_{\text{obj}}$ using morphological transformations and remove the disconnected components to obtain the final clean mask $\mathbf{M}_{\text{obj}}$. This pipeline

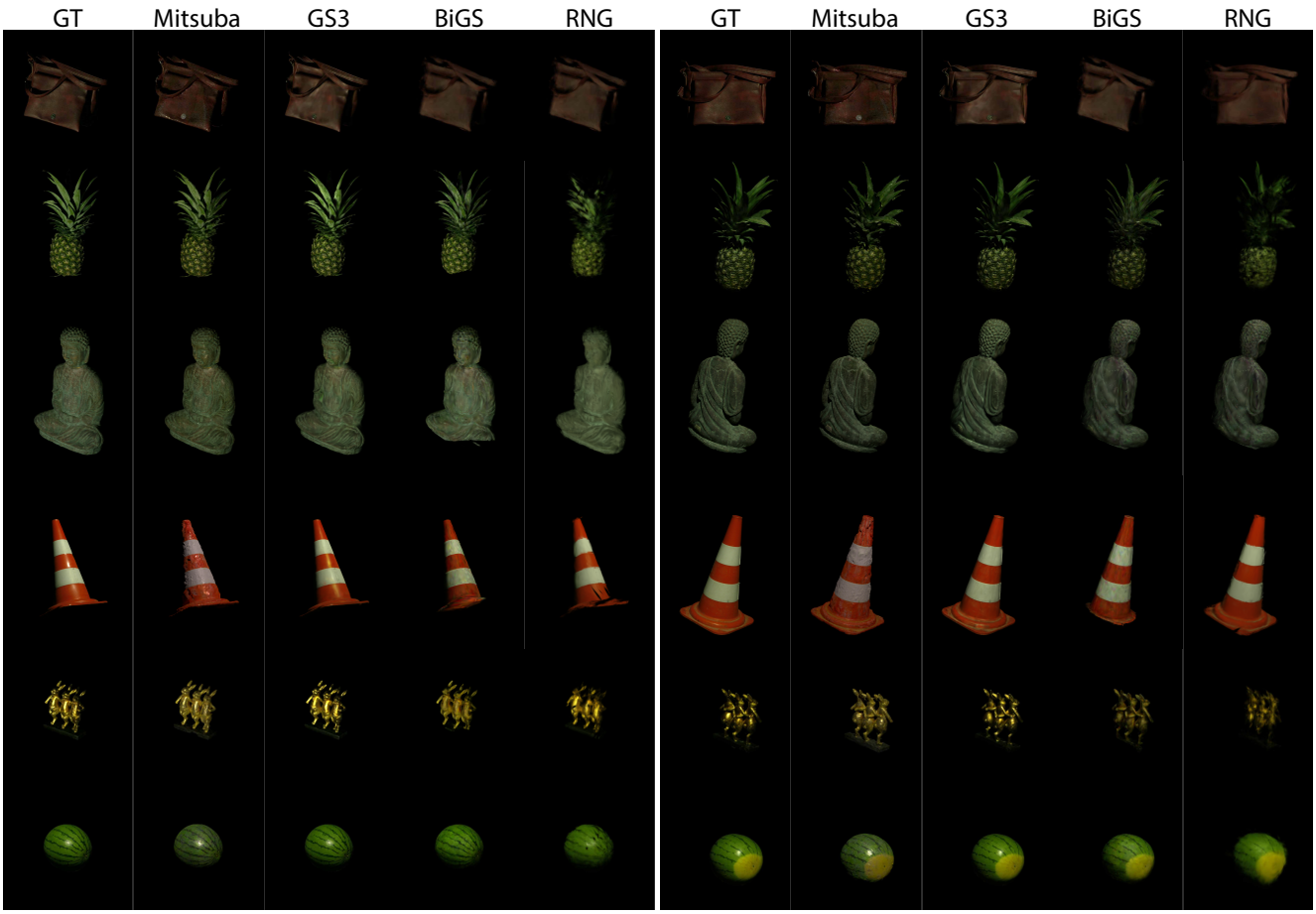

Figure 7. We visualize the inverse rendering and novel view synthesis results of several baseline methods (Mitsuba [39], GS^3 [3],BiGS [34] and RNG [14]) evaluated on our validation dataset. In this figure, we show relit objects from inference views and light directions.

achieves a success rate of 95% across all objects from all views in our setup. For the remaining failure cases, we create a lightweight user interface for manual correction. In Fig. 6, we illustrate masks generated by three individual segmentation strategies and our method.

Normal and Albedo Extraction In addition to diverse illuminations, **OLATverse** provides pseudo ground truth surface normals and diffuse albedo as auxiliary data to support multimodal tasks. We recover surface normals through photometric stereo solutions [19] by analyzing radiance variations of images captured under gradient illuminations. For all cameras, we captured images $\mathbf{I}_{\text{cg}}$ under color gradient illumination. To further cancel out the view-dependent specular reflection on the non-Lambertian surfaces [17, 36], we apply linear polarized filters to five fixed cameras, and capture images under polarized gradient full bright illuminations, denoted as $\mathbf{I}_{\perp}$. The diffuse albedo $\mathbf{D}$ and surface normals $\mathbf{N}$ are computed as follows:

$$\mathbf{D} = 0.5(\mathbf{I}_{\perp}^{+} + \mathbf{I}_{\perp}^{-}), \tag{3}$$

$$\mathbf{N}^{*} = \frac{(\mathbf{I}^{+} - \mathbf{I}^{-})}{(\mathbf{I}^{+} + \mathbf{I}^{-})}, \quad \mathbf{N} = \frac{\mathbf{N}^{*}}{|\mathbf{N}^{*}|}, \tag{4}$$

where $\mathbf{I}^{+}$ and $\mathbf{I}^{-}$ denote image captured under opposite gradient directions in either polarized or non-polarized settings. For polarized normals, we utilize polarized captures $\mathbf{I}_{\perp}^{+}$ and $\mathbf{I}_{\perp}^{-}$, whereas for non-polarized normals, the images under color gradient illuminations $\mathbf{I}_{\text{cg}}^{+}$ and $\mathbf{I}_{\text{cg}}^{-}$ are used. Both polarized and non-polarized normals are incorporated in the final dataset, and we empirically observe that polarized normals yield higher accuracy for most objects.

4. Application

In this section, we demonstrate several applications of **OLATverse**. First, we exploit the linearity of light transport to synthesize relit images under arbitrary novel illuminations using OLATs (Sec. 4.1). Then we set up a validation dataset and conduct baseline experiments on inverse rendering and view synthesis (Sec. 4.2), as well as surface normal estimation (Sec. 4.3).

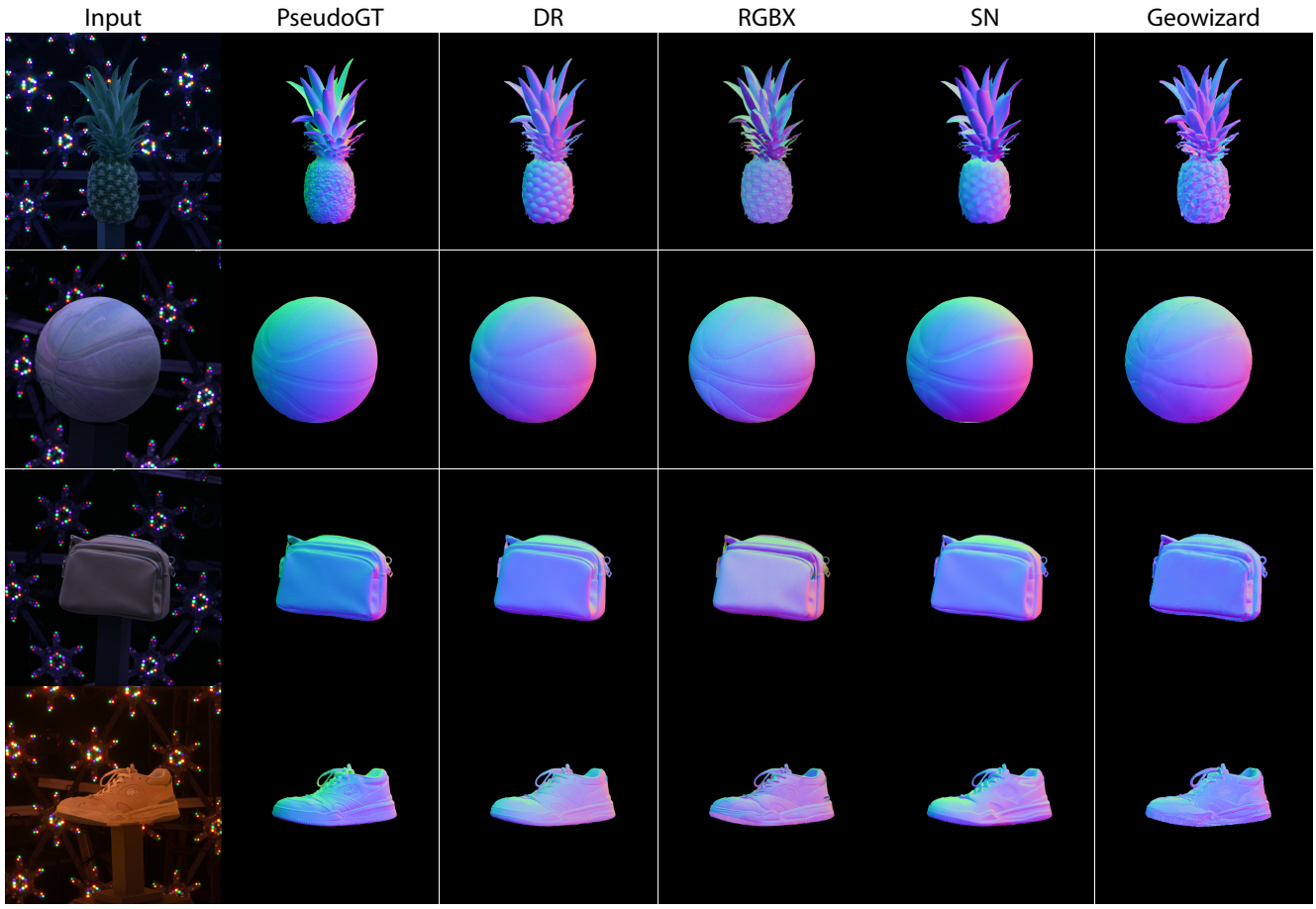

Figure 8. Visual comparison of pseudo ground truth normals with normals estimated by DR [30], RGBX [62], SN [59] and GW [15]. To ensure a robust and generalized comparison, we provide input images of each validation object under four different illuminations.

4.1. Relighting

With OLATs, we can produce image-based renderings under any arbitrary illuminations, leveraging the linearity of light transport. Specifically, given a target equirectangular environment map $\mathbf{E}$, the relit image $\mathbf{I}_{\text{relit}}$ of a certain object can be obtained as:

$$\mathbf{I}_{\text{relit}} = \sum_{i=1}^{N_{olat}} \left(\mathcal{F}(\mathbf{E} \odot \mathbf{M}_{\text{i}}) \cdot \mathbf{I}_{\text{i}}\right), \tag{5}$$

where $\odot$ denotes pixel-wise multiplication, $\mathcal{F}$ represents per-channel averaging operator, N_{olat} denotes the number of OLATs. $\mathbf{I}_{\text{i}}$ and $\mathbf{M}_{\text{i}}$ denote the ith OLAT image and the corresponding environmental mask. Examples of some relighting results are shown in Fig. 1. Leveraging this property, **OLATverse** can be efficiently scaled to a relighting dataset of real objects under diverse illuminations.

4.2. Inverse Rendering and View Synthesis

To construct a comprehensive validation dataset, we carefully select 42 objects with rich textures spanning 14 material categories. For each object, the raw images are downsampled to 750×1.4k resolution, and every fifth camera and light are used as inference sets, while the remaining data are utilized for training. Object meshes are reconstructed using the feature-based algorithm provided in MetaShape [1]. We evaluate Mitsuba [39] (using MetaShape-reconstructed meshes) as well as several 3DGS-based inverse rendering methods [3, 14, 34] on our dataset. All baseline methods are executed using the source code released by the authors and initialized with the same reconstructed meshes. We visualize relighting results under inference views and lighting conditions in Fig. 7, and report quantitative metrics (SSIM [57], PSNR, and LPIPS [66]) in Tab. 2. As shown, GS^3 consistently outperforms other inverse rendering methods both visually and numerically. Compared to the other methods, GS^3 accurately captures the specular reflections on glossy surfaces such as the watermelon, metallic rabbit, and plastic roadblock.

4.3. Normal Estimation

Next, we benchmark several diffusion-based normal estimation methods [15, 30, 59, 62] on our validation dataset.

Table 2. Numerical comparison of inverse rendering baselines on our validation dataset using PSNR, LPIPS, and SSIM metrics.

Method	PSNR ↑	LPIPS ↓	SSIM ↑
Mitsuba+Mshape [39]	35.906	0.0260	0.976
GS^3 [3]	38.538	0.0263	0.982
RNG [14]	32.065	0.051	0.962
BiGS [34]	32.98	0.0451	0.940

Table 3. Numerical comparison of normal estimation methods on our validation dataset using normal angular metrics.

Method	Mean↓	Med↓	11.25° ↑	22.5° ↑	30° ↑
SN [59]	**31.85**	**30.25**	8.93	34.00	**55.40**
RGBX [62]	51.95	49.70	6.40	22.80	35.85
DR [30]	34.88	33.28	8.13	31.00	50.15
GW [15]	34.42	32.03	**10.98**	**34.10**	50.05

In **OLATverse**, the extracted surface normals from 5 polarized views and 35 non-polarized views are in the world coordinate. During preprocessing, we select normal maps extracted from 5 polarized views for evaluation and reproject the surface normals from world to camera coordinates using camera parameters, ensuring compatibility with the monocular normal estimation methods. To enable lighting-invariant evaluation, each object is evaluated under four different environmental illuminations. For all experiments, we use official source code, and present both qualitative and quantitative results in Fig. 8 and Tab. 3. We follow the same evaluation protocol as StableNormal (SN) [59], reporting the mean and median angular errors, where the lower values indicates better accuracy. We also measure the percentage of pixels with angular error below thresholds of 11.25°, 22.5°, and 30°. Quantitatively, SN and GeoWizard (GW), which are specifically designed for normal estimation tasks, outperform RGBX and DiffusionRender(DR), which are originally developed for image/video relighting. Visually RGBX and DR produce the surface normals with finer high-frequency details than SN and GW. However, none of these methods recover accurate surface normals for real-world objects with complex geometric structures in the evaluation dataset, highlighting the importance of **OLATverse** for advancing normal estimation research.

5. Limitation

Although we apply a linear polarized filter to cancel out the specular reflection during normal extraction, artifacts remain noticeable for objects with glossy materials or low-reflectance texture, consistent with the observation in RNHA [69]. These artifacts arise from the weak signal-to-noise ratio of low-reflectance objects and the ambiguity between view-dependent reflection and surface normals. While the extracted surface normals and diffuse albedo are not exact ground truth, they can still provide valuable supervision signals for multi-modal training tasks. In addition, the ground truth meshes are not incorporated in our dataset due to hardware limitations. In the future, it would be interesting to integrate an advanced scanning system to jointly capture both appearance and geometry of real objects.

6. Conclusion and Discussion

In this work, we introduce **OLATverse**, the first large-scale real-world dataset, comprising around $9M$ images of 765 real objects with diverse physical sizes and materials under precise lighting control. Compared to the existing object-centric datasets, **OLATverse** offers two substantial advantages: large-scale and high-fidelity appearance under precise lighting control. In addition, our dataset provides auxiliary resources such as surface normals, diffuse albedo, and accurate object masks. We demonstrate that **OLATverse** can be used to construct a relighting resource under arbitrary illuminations and serves as a comprehensive real-world benchmark for inverse rendering, view synthesis, relighting, and normal estimation tasks. A promising avenue for future work is to leverage this dataset to train data-driven generative priors for realistic relighting and appearance modeling. We believe that **OLATverse** will advance research toward bridging the gap between synthetic and real-world data in graphics and vision communities.

References

[1] Agisoft LLC. Metashape. https://www.agisoft.com/downloads/installer/, 2025. Version retrieved July 2025. 6, 8

[2] BRIA AI. BRIA-RMBG-2.0: High-quality real-time background removal. https://huggingface.co/spaces/briaai/BRIA-RMBG-2.0, 2023. Accessed: 2025-07-15. 6

[3] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. Gs3: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–12, 2024. 1, 3, 7, 8, 9

[4] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3

[5] Xi Chen, Sida Peng, Dongchen Yang, Yuan Liu, Bowen Pan, Chengfei Lv, and Xiaowei Zhou. Intrinsicanything: Learning diffusion priors for inverse rendering under unknown illumination. In *European Conference on Computer Vision*, pages 450–467. Springer, 2024. 4

[6] Seokjun Choi, Hoon-Gyu Chung, Yujin Jeon, Giljoo Nam, and Seung-Hawn Baek. A real-world display inverse rendering dataset. *ICCV*, 2025. 3

[7] Jasmine Collins, Shubham Goel, Kenan Deng, Achleshwar Luthra, Leon Xu, Erhan Gundogdu, Xi Zhang, Tomas F Yago Vicente, Thomas Dideriksen, Himanshu Arora, et al.

Abo: Dataset and benchmarks for real-world 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21126–21136, 2022. 1, 3

[8] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, 2000. 2, 4

[9] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13142–13153, 2023. 1, 3

[10] Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. Single-image svbrdf capture with a rendering-aware deep network. *ACM Transactions on Graphics (ToG)*, 37(4):1–15, 2018. 3

[11] Jan-Niklas Dihlmann, Arjun Majumdar, Andreas Engelhardt, Raphael Braun, and Hendrik Lensch. Subsurface scattering for 3d gaussian splatting. *arXiv preprint arXiv:2408.12282*, 2024. 1, 3

[12] Zhao Dong, Ka Chen, Zhaoyang Lv, Hong-Xing Yu, Yunzhi Zhang, Cheng Zhang, Yufeng Zhu, Stephen Tian, Zhengqin Li, Geordie Moffatt, et al. Digital twin catalog: A large-scale photorealistic 3d object digital twin dataset. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 753–763, 2025. 1, 3, 4

[13] Kang Du, Zhihao Liang, and Zeyu Wang. Gs-id: Illumination decomposition on gaussian splatting via diffusion prior and parametric light source optimization. *arXiv preprint arXiv:2408.08524*, 2024. 4

[14] Jiahui Fan, Fujun Luan, Jian Yang, Milos Hasan, and Beibei Wang. Rng: Relightable neural gaussians. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26525–26534, 2025. 4, 7, 8, 9

[15] Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. Geowizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In *European Conference on Computer Vision*, pages 241–258. Springer, 2024. 8, 9

[16] Jian Gao, Chun Gu, Youtian Lin, Zhihao Li, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *European Conference on Computer Vision*, pages 73–89. Springer, 2024. 1, 4

[17] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A Wilson, and Paul Debevec. Estimating specular roughness and anisotropy from second order spherical gradient illumination. In *Computer Graphics Forum*, pages 1161–1170. Wiley Online Library, 2009. 7

[18] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2335–2342. IEEE, 2009. 1, 3

[19] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (ToG)*, 38(6):1–19, 2019. 7

[20] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5356–5364, 2019. 2, 4

[21] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 406–413, 2014. 1, 3

[22] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensoir: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 165–174, 2023. 4

[23] Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. Neural gaffer: Relighting any object via diffusion. *Advances in Neural Information Processing Systems*, 37:141129–141152, 2024. 1, 4

[24] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 4

[25] Minchul Kim, Dingqiang Ye, Yiyang Su, Feng Liu, and Xiaoming Liu. Sapiensid: Foundation for human recognition. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 13937–13947, 2025. 6

[26] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 6

[27] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. Neroic: Neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 1, 3

[28] Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Elliott Wu, Jiajun Wu, et al. Stanford-orb: a real-world 3d object inverse rendering benchmark. *Advances in Neural Information Processing Systems*, 36:46938–46957, 2023. 1, 3

[29] Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. Modular primitives for high-performance differentiable rendering. *ACM Transactions on Graphics*, 39(6), 2020. 4

[30] Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Chih-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, et al. Diffusion renderer: Neural inverse and forward rendering with video diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26069–26080, 2025. 1, 4, 8, 9

[31] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21644–21653, 2024. 1, 4

[32] Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian L Curless, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Real-time high-resolution background matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8762–8771, 2021. 6

[33] Isabella Liu, Linghao Chen, Ziyang Fu, Liwen Wu, Haian Jin, Zhong Li, Chin Ming Ryan Wong, Yi Xu, Ravi Ramamoorthi, Zexiang Xu, et al. Openillumination: A multi-illumination dataset for inverse rendering evaluation on real objects. *Advances in Neural Information Processing Systems*, 36:36951–36962, 2023. 1, 3, 4, 5, 6

[34] Zhenyuan Liu, Yu Guo, Xinyuan Li, Bernd Bickel, and Ran Zhang. Bigs: Bidirectional gaussian primitives for relightable 3d gaussian splatting. *arXiv preprint arXiv:2408.13370*, 2024. 1, 4, 7, 8, 9

[35] Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, et al. Wonder3d: Single image to 3d using cross-domain diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9970–9980, 2024. 1

[36] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, Paul E Debevec, et al. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques*, 9(10):2, 2007. 2, 7

[37] Xiaohe Ma, Xianmin Xu, Leyao Zhang, Kun Zhou, and Hongzhi Wu. Opensvbrdf: A database of measured spatially-varying reflectance. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023. 3

[38] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 4

[39] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: A retargetable forward and inverse renderer. *ACM Transactions on Graphics (ToG)*, 38(6):1–17, 2019. 4, 7, 8, 9

[40] Fan Pei, Jinchen Bai, Xiang Feng, Zoubin Bi, Kun Zhou, and Hongzhi Wu. Opensubstance: A high-quality measured dataset of multi-view and-lighting images and shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5221–5231, 2025. 3

[41] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*, 2022. 1

[42] Pramod Rao, Xilong Zhou, Abhimitra Meka, Gereon Fox, Mallikarjun B R, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, and Christian Theobalt. 3dpr: Single image 3d portrait relighting with generative priors. 2025. 3

[43] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv preprint arXiv:2007.08501*, 2020. 4

[44] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 130–141, 2024. 3

[45] Jian Shi, Yue Dong, Hao Su, and Stella X Yu. Learning non-lambertian object intrinsics across shapenet categories. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1685–1694, 2017. 1, 3

[46] Yichun Shi, Peng Wang, Jianglong Ye, Long Mai, Kejie Li, and Xiao Yang. Mvdream: Multi-view diffusion for 3d generation. In *The Twelfth International Conference on Learning Representations*. 1

[47] Yahao Shi, Yanmin Wu, Chenming Wu, Xing Liu, Chen Zhao, Haocheng Feng, Jian Zhang, Bin Zhou, Errui Ding, and Jingdong Wang. Gir: 3d gaussian inverse rendering for relightable scene factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 4

[48] Yawar Siddiqui, Tom Monnier, Filippos Kokkinos, Mahendra Kariya, Yanir Kleiman, Emilien Garreau, Oran Gafni, Natalia Neverova, Andrea Vedaldi, Roman Shapovalov, et al. Meta 3d assetgen: Text-to-mesh generation with high-quality geometry, texture, and pbr materials. *Advances in Neural Information Processing Systems*, 37:9532–9564, 2024. 1

[49] Giota Stratou, Abhijeet Ghosh, Paul Debevec, and Louis-Philippe Morency. Effect of illumination on automatic expression recognition: a novel 3d relightable facial database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 611–618. IEEE, 2011. 3

[50] Timo Teufel, Pulkit Gera, Xilong Zhou, Umar Iqbal, Pramod Rao, Jan Kautz, Vladislav Golyanik, and Christian Theobalt. Humanolat: A large-scale dataset for full-body human relighting and novel-view synthesis. *arXiv preprint arXiv:2508.09137*, 2025. 3, 6

[51] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20762–20772, 2023. 1, 3

[52] Briac Toussaint, Diego Thomas, and Jean-Sébastien Franco. Probesdf: Light field probes for neural surface reconstruction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 11026–11035, 2025. 1

[53] Giuseppe Vecchio and Valentin Deschaintre. Matsynth: A modern pbr materials dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22109–22118, 2024. 3

[54] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 1

[55] Shaofei Wang, Tomas Simon, Igor Santesteban, Timur Bagautdinov, Junxuan Li, Vasu Agrawal, Fabian Prada,

Shoou-I Yu, Pace Nalbone, Matt Gramlich, et al. Relightable full-body gaussian codec avatars. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pages 1–12, 2025. 3

[56] Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3295–3306, 2023. 1

[57] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 8

[58] Tong Wu, Jiarui Zhang, Xiao Fu, Yuxin Wang, Jiawei Ren, Liang Pan, Wayne Wu, Lei Yang, Jiaqi Wang, Chen Qian, et al. Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 803–814, 2023. 1, 3, 4

[59] Chongjie Ye, Lingteng Qiu, Xiaodong Gu, Qi Zuo, Yushuang Wu, Zilong Dong, Liefeng Bo, Yuliang Xiu, and Xiaoguang Han. Stablenormal: Reducing diffusion variance for stable and sharp normal. *ACM Transactions on Graphics (TOG)*, 43(6):1–18, 2024. 4, 8, 9

[60] Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. Relighting neural radiance fields with shadow and highlight hints. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 1, 3

[61] Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. Dilightnet: Fine-grained lighting control for diffusion-based image generation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. 1, 4

[62] Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. Rgb↔x: Image decomposition and synthesis using material- and lighting-aware diffusion models. In *ACM SIGGRAPH 2024 Conference Papers*, New York, NY, USA, 2024. Association for Computing Machinery. 1, 4, 8, 9

[63] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Neural video portrait relighting in real-time via consistency modeling. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 802–812, 2021. 3

[64] Libo Zhang, Yuxuan Han, Wenbin Lin, Jingwang Ling, and Feng Xu. Prtgaussian: Efficient relighting using 3d gaussians with precomputed radiance transfer. In *2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1–6. IEEE, 2024. 4

[65] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In *The Thirteenth International Conference on Learning Representations*, 2025. 1

[66] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 8

[67] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 1, 4

[68] Yibo Zhang, Li Zhang, Rui Ma, and Nan Cao. Texverse: A universe of 3d objects with high-resolution textures. *arXiv preprint arXiv:2508.10868*, 2025. 3

[69] Taotao Zhou, Kai He, Di Wu, Teng Xu, Qixuan Zhang, Kuixiang Shao, Wenzheng Chen, Lan Xu, and Jingyi Yu. Relightable neural human assets from multi-view gradient illuminations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4315–4327, 2023. 3, 9

[70] Xilong Zhou and Nima Khademi Kalantari. Adversarial single-image svbrdf estimation with hybrid training. In *Computer Graphics Forum*, pages 315–325. Wiley Online Library, 2021. 3

[71] Xilong Zhou, Milos Hasan, Valentin Deschaintre, Paul Guerrero, Yannick Hold-Geoffroy, Kalyan Sunkavalli, and Nima Khademi Kalantari. Photomat: A material generator learned from single flash photos. In *ACM SIGGRAPH 2023 conference proceedings*, pages 1–11, 2023. 3

[72] Zhenglong Zhou, Zhe Wu, and Ping Tan. Multi-view photometric stereo with spatially varying isotropic materials. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1482–1489, 2013. 1, 3