
MetaGS: A Meta-Learned Gaussian-Phong Model for Out-of-Distribution 3D Scene Relighting

Yumeng He Yunbo Wang*

MoE Key Lab of Artificial Intelligence, AI Institute, School of Computer Science
Shanghai Jiao Tong University
{ymhe, yunbow}@sjtu.edu.cn

Abstract

Out-of-distribution (OOD) 3D relighting requires novel view synthesis under unseen lighting conditions that differ significantly from the observed images. Existing relighting methods, which assume consistent light source distributions between training and testing, often degrade in OOD scenarios. We introduce **MetaGS** to tackle this challenge from two perspectives. First, we propose a meta-learning approach to train 3D Gaussian splatting, which explicitly promotes learning generalizable Gaussian geometries and appearance attributes across diverse lighting conditions, even with biased training data. Second, we embed fundamental physical priors from the *Blinn-Phong* reflection model into Gaussian splatting, which enhances the decoupling of shading components and leads to more accurate 3D scene reconstruction. Results on both synthetic and real-world datasets demonstrate the effectiveness of MetaGS in challenging OOD relighting tasks, supporting efficient point-light relighting and generalizing well to unseen environment lighting maps.

1 Introduction

3D scene relighting generates novel lighting effects that interact with the observed 3D environment, with recent advances in learning-based volume rendering offering effective solutions [21, 36, 23, 40, 24, 22, 3, 30]. A typical approach captures a substantial number of multi-view images of a scene under individual lighting conditions, then trains a NeRF or 3D Gaussian splatting (3DGS) model to generalize to new point light positions [28, 13, 9, 41, 8]. Notably, such idealized data requirements are often impractical in real-world scenarios. To simulate a more realistic capturing process, we build on prior work using a **One Light At a Time (OLAT)** setup [34, 14, 1], where training data is collected with a moving point light and a moving camera. Recent studies have further explored low-cost OLAT capture using a smartphone flashlight as the moving light source [5].

While most OLAT methods assume a coherent lighting distribution between training and testing, real-world scenarios often involve light sources that are randomly distributed around the scene, which may result in biased training data. This leads to an **out-of-distribution (OOD)** issue where test-time light sources deviate from the training distribution. This OOD relighting presents a more natural and challenging task, serving as a robust test for evaluating a model’s performance under “*truly*” novel lighting conditions. Compared with the standard OLAT setup, it complicates the learning of true object illumination properties and geometries due to a limited training corpus with entangled, time-varying lighting and viewing directions and, more critically, spatially biased lighting. As shown in Figure 1, existing OLAT methods struggle with the OOD relighting task, often producing unrealistic lighting effects, such as chaotic specular highlights and shadows, due to overfitting to the training samples. Also, because of the lack of generalizability, when new lighting distributions are introduced, they typically require full retraining or finetuning to adapt.

*Corresponding author.

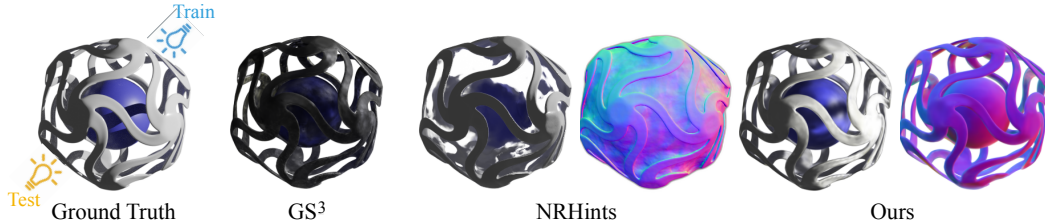


Figure 1: *Preliminary results.* NRHints [34] and GS^3 [1] struggle with out-of-distribution light positions (as specified in Section 2), as the blurry novel view synthesis results and erroneous normal predictions indicate. GS^3 does not require well-defined surface normals and does not directly output the normal image.

In this paper, we present two novel techniques to enable better OOD relighting. First, we provide a pilot study on **meta-learning-based 3DGS** and introduce MetaGS to mitigate overfitting and improve the generalizability of the OLAT models. This approach leverages bilevel optimization techniques to effectively address the challenges associated with limited training data and complex lighting variations. Specifically, we frame OLAT as a multi-task learning problem, treating rendering under specific lighting positions as distinct tasks. The core idea is to validate the learned Gaussian illumination properties under a given lighting condition using data from other lighting conditions. The resulting second-order gradients of 3DGS parameters can explicitly encourage the model to generalize to unseen illumination rather than overfitting to specific training samples, as is common with standard 3DGS loss functions.

Furthermore, MetaGS integrates simple yet fundamental physical priors into Gaussian splatting by incorporating a **learnable Blinn-Phong reflection model** [2]. This approach effectively decouples different shading components—diffuse, specular, and ambient—leading to a more physically grounded understanding of the interactions between objects and lighting in the training data. This design draws inspiration from prior research [39, 28, 13] indicating that modular and compositional representations enhance generalization capabilities. The decoupled rendering improves the model’s ability to generalize to novel lighting conditions, allowing it to independently adjust each shading component according to the specific lighting scenario.

We evaluate MetaGS on both synthetic and real-world datasets. It supports efficient point-light relighting under highly constrained training illuminations and significantly outperforms existing OLAT methods in OOD relighting tasks. We further demonstrate that both meta-learning and the differentiable Phong model individually contribute to improved generalization. Notably, MetaGS generalizes well to unseen environment maps, despite being trained exclusively in OLAT scenarios.

2 Preliminaries

Definition and challenges of OLAT. OLAT refers to a specialized setting for 3D reconstruction and relighting. This involves illuminating the subject with a single point light in sequential exposures, capturing one image at each light position. Each capture is represented as a tuple $\{O_t, V_t, P_t\}$, where O_t denotes the observed image, V_t represents the camera configuration, and P_t contains lighting information. In our context, P_t specifies the position of the point light. Unlike the “multiple lights, multiple cameras” setup, which requires multi-view images captured by multiple time-synchronized cameras for each lighting condition, OLAT relies on a single camera. This significantly simplifies data collection, particularly in scenarios with dynamic lighting. The rapidly changing light positions in OLAT introduce a new challenge for 3D reconstruction. The color of a 3D point can vary significantly, increasing ambiguity for estimating true object geometries and illumination properties.

Preliminary findings. We summarize existing OLAT relighting methods in Table 1. In the preliminary experiments, we evaluate NRHints [34], the prior art in NeRF-based relighting models. We observe that it underperforms in testing scenarios with out-of-distribution (OOD) lighting positions, a feature that we believe should be crucial for relighting methods. Specifically, in Figure 1, the lighting in the training set is arranged on one side of the hemisphere, while the lighting in the test set is arranged on the opposite side (cameras are located on both sides). In such cases, NRHints fails to generate reasonable rendering results, likely due to its implicit modeling of shadows and

Table 1: Comparison of the OLAT relighting methods.

	[35]	[5]	[6]	[34]	[1]	[12]	Ours
Point light	✓	✓	✓	✓	✓	✓	✓
Shadow computation	×	×	✓	✓	✓	✓	✓
OOD light	×	×	×	×	×	×	✓

specular reflections. We additionally evaluate a concurrent 3DGS-based approach [1] and observe similar degraded results under the OOD relighting setup. These findings highlight that understanding intrinsic illumination properties under arbitrary lighting variations remains a challenging task, as the model tends to overfit to perspective-constrained observations and fail to leverage enough physical principles when dealing with unseen lighting distribution.

3 Method

In this section, we present the details of MetaGS for addressing the OOD OLAT learning challenge:

- In Section 3.1, we incorporate physical shading priors within Gaussian splatting via a differentiable Phong reflection model to decouple mixed illumination components.
- In Section 3.2, we introduce a bilevel optimization scheme to estimate light-independent scene geometries and intrinsic illumination properties, marking an early effort in volume rendering.
- In Section 3.3, we discuss the entire training pipeline and the implementation details.

3.1 Differentiable Phong Model

Our *MetaGS* method leverages simple yet generalizable physical priors from the Blinn-Phong model [19], which captures three fundamental components of light transport: ambient, diffuse, and specular reflections. The core idea is to disentangle these illumination components by learning the interactions between (i) the normal vectors of the Gaussian points, (ii) the viewing directions, and (iii) the ray direction from the point light.

Specifically, the ambient component represents constant environmental illumination, simulating indirect scattering from surrounding surfaces to establish a baseline brightness level. The diffuse component, based on Lambertian law, describes light scattering in multiple directions from rough surfaces. The specular reflection is computed based on the angle between the light direction and the bisector ($\mathbf{h}$) between the viewing direction ($\mathbf{v}$) and the light direction ($\mathbf{l}$), with a shininess exponent that represents different degrees of glossiness.

In MetaGS, we extend the original 3DGS by introducing the decoupled computation for different shading components. As shown in Figure 2, in addition to the basic Gaussian attributes (*e.g.*, position $\mathbf{x}$, rotation R , scale S , opacity α , and spherical harmonic coefficients f), each Gaussian point is further associated with a normal vector $\mathbf{n}$, a 3-channel diffuse color k_d , and a 1-channel specular coefficient k_s . The newly added attributes enhance the model’s understanding of lighting effects, facilitating the learning process of light-independent object geometries. The overall color of a Gaussian point is determined by:

$$L_p = L_a + L_d + L_s = L_a + \sum_{|\text{lights}|} (k_d I_d + k_s I_s), \quad (1)$$

where $k_{\{d,s\}}$ and $I_{\{d,s\}}$ are the colors and intensities of the diffuse and specular components. $|\text{lights}|=1$ in OLAT. In graphics, zero-order spherical harmonics (SH) coefficients typically represent the basic, uniform component of a function defined over the sphere, essentially the average or constant part of a lighting environment. Therefore, we restrict Gaussians’ SH coefficients to zero order, denoted by f_0 , corresponding to the ambient color. We compute the diffuse and specular light transports by multiplying the color and intensity of each component, where the intensities I_d and I_s are defined as:

$$I_d = \frac{I}{r^2} \max(0, \mathbf{n} \cdot \mathbf{l}), \quad I_s = \frac{I}{r^2} \max(0, \mathbf{n} \cdot \mathbf{h})^p, \quad (2)$$

where I denotes the light emitted intensity, which is a global learnable parameter, $\mathbf{l}$ denotes the point-to-light normalized vector, and $\mathbf{h} = \frac{\mathbf{v} + \mathbf{l}}{\|\mathbf{v} + \mathbf{l}\|}$ denotes the bisector of the point-to-camera normalized

Algorithm 1 Meta-Training for OOD Relighting

```
1: Input: Training set  $\{\mathcal{D}_t\}_{1:T}$ , where  $D_t = \{\text{light position, camera parameters, RGB image}\}$ 
2: Hyperparameters: Learning rates  $\alpha, \beta$ 
3: Initialized parameters: Pretrained Gaussian attributes  $\{\theta_k\}$ , where  $\theta_k = (\mathbf{x}, \mathbf{n}, R, S, \alpha, f_0, k_d, k_s)$ 
4: while not converge do
5:   Sample disjoint data  $\mathcal{D}_{1:m}^{\text{sup}}$  and  $\mathcal{D}_{1:m}^{\text{query}}$  from  $\{\mathcal{D}_t\}_{1:T}$ 
6:   for task  $i$  in  $\{1 : m\}$  do ▷ Inner optimization loop
7:      $\theta'_i \leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}_{\text{final}}(\theta; \mathcal{D}_i^{\text{sup}})$ 
8:   end for
9:    $\{\theta\} \leftarrow \{\theta\} - \beta \sum_{i=1}^m \nabla_{\theta} \mathcal{L}_{\text{final}}(\theta'_i; \mathcal{D}_i^{\text{query}})$  ▷ Outer optimization step
10: end while
```

Inner optimization loop. In the inner loop, the model is trained on each support set to independently generate m sub-models, representing the hypothetical estimates θ'_i under each lighting condition. All sub-models start with the same parameters of θ . The light intensity I is a global parameter that is also learnable and optimized jointly with the Gaussian attributes. It is omitted here for clarity. We specify the loss function $\mathcal{L}_{\text{final}}$ in subsequent Section 3.3.

Outer optimization loop. In the outer loop, a global gradient update is performed across all sampled tasks, updating the Gaussian attributes. We first compute the loss function for each query sample, $\mathcal{L}_{\text{final}}(\cdot, \cdot; \mathcal{D}_i^{\text{sup}})$, using the corresponding inner-loop model hypotheses θ'_i . We then aggregate the m losses to update the initial θ . This approach, involving task-specific adaptation followed by a global update, enables the model to learn generalizable representations across varying lighting conditions.

3.3 Entire Training Pipeline

We employ a three-stage training scheme. The first stage, similar to 3DGS, focuses on training the core Gaussian attributes, including position $\mathbf{x}$, rotation R , scale S , opacity α , and spherical harmonic coefficients f . After this stage, the model tends to capture a rough geometry and an average color across different illuminations, serving as an effective initialization for further refinement. In the second stage, we incorporate the normal attributes into the optimization process, acknowledging that learning a Phong model heavily relies on accurate normal estimations. In the final stage, we integrate the diffuse and specular components, training all parameters concurrently through meta-learning.

Objective functions. We follow 3DGS [10] to compute the deviation of the predicted image from the ground truth, using an L1 regularization and a D-SSIM term as the RGB loss. We also employ a sparse loss [9] to encourage the opacity values α of the Gaussian spheres to approach either 0 or 1, thereby facilitating learning with opaque objects. For normal estimation, we follow GaussianShader [9] to compute the normal vectors using Gaussian’s shortest axis direction, and progressively align them with the depth-inferred pseudo-normals.

Implementation details. We implement MetaGS using PyTorch [18]. For optimization, we use the Adam optimizer [11] with the same parameters as those specified by [10]. Our training procedure contains three stages: the first stage last for $10k$ iterations, the second stage last for $5k$ iterations, followed by the final meta-learning stage of $10k$ iterations. All experiments are conducted on a single NVIDIA RTX 3090 GPU.

4 Experiments

4.1 Experimental Settings

We perform three types of OLAT relighting experiments:

- *OOD relighting:* We introduce a novel view synthesis setup to assess the generalizability of 3D relighting models under OOD lighting conditions. The point lights in the training set are positioned on one side of the upper hemisphere, while the lighting in the test set is placed on the opposite side.
- *Camera-light-colocated relighting:* This setup mirrors the one used in IRON [35], where the camera and point light are co-located for each image. Similar to the OOD setting, the colocated

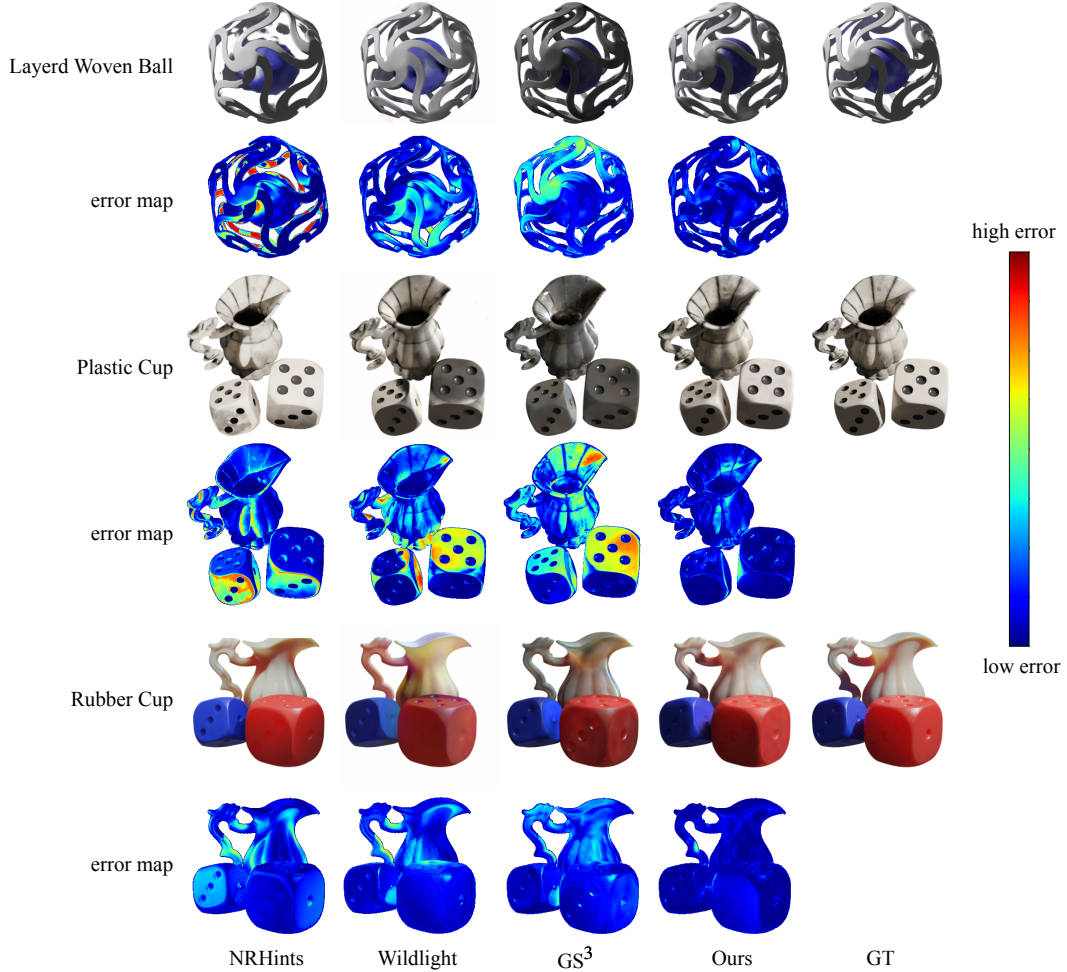


Figure 3: *OOD relighting results on synthetic data.* We present rendered novel views and error maps. While baselines often misrepresent shadows or light-dependent effects (*e.g.*, incorrect shadows in *Plastic Cup*), our model better infers surface appearance.

Table 2: *PSNR results for OOD relighting.* See our Appendix for full comparisons.

Method	Synthetic			Real-world						
	Ball	PlasCup	RubCup	Cat	CatSmall	CupFabric	Fish	FurScene	Pikachu	Pixiu
NRHints [34]	17.25	23.92	27.44	18.04	24.63	24.65	22.57	21.55	24.00	23.03
WildLight [5]	21.73	20.95	24.02	18.65	22.53	24.03	21.47	20.33	19.09	20.22
GS ³ [1]	18.84	20.30	24.37	17.66	23.34	25.04	21.12	17.34	24.11	19.63
Ours	26.76	27.54	27.95	26.45	26.44	27.29	24.68	24.82	25.54	25.65

configuration also represents a constrained lighting regime, as the light source is fixed to the camera pose. For fair comparison, we generate colocated versions of the synthetic scenes accordingly.

- *Environment map relighting:* We evaluate the generalization of the models exclusively trained in the OLAT setup to novel view synthesis with unseen environment lighting maps.

We evaluate MetaGS on 3 synthetic scenes and 7 real captured scenes from NRHints [34], featuring a diverse range of materials, complex object shapes, significant self-occlusion, and intricate shadow effects. Each image is rendered (or captured) with a unique camera pose and point-light position. Full details of the datasets are provided in the Appendix.

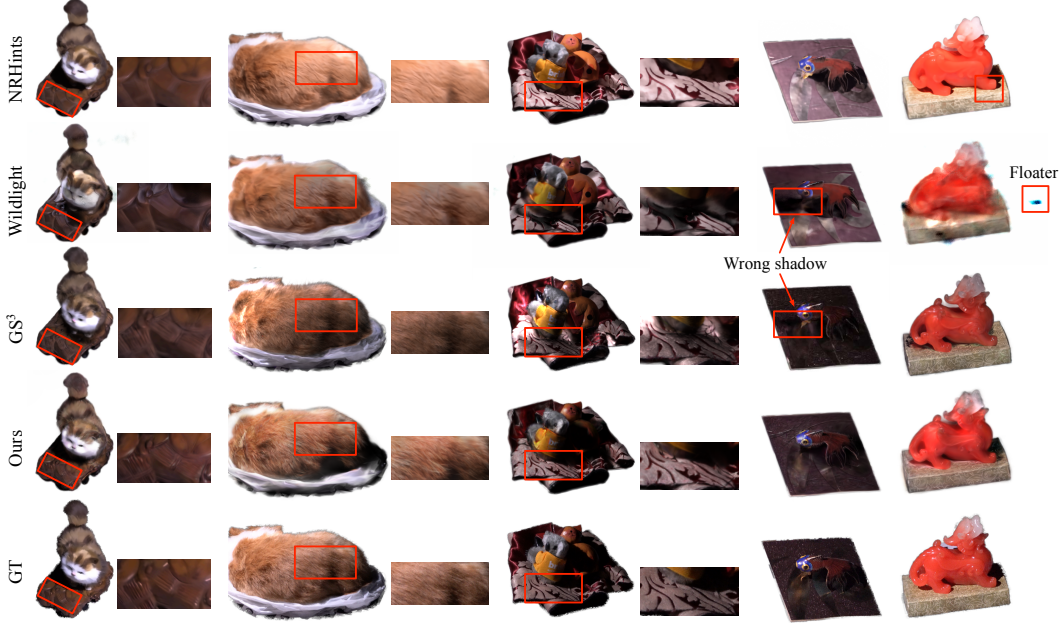


Figure 4: *OOD relighting results on real-world data.* As highlighted with the red boxes, baseline models struggle with some level of global shading consistency, including color shifts, incorrect shadows, and floating artifacts. Our approach presents physically plausible specular highlights and geometrically consistent shadows that closely match ground truths.

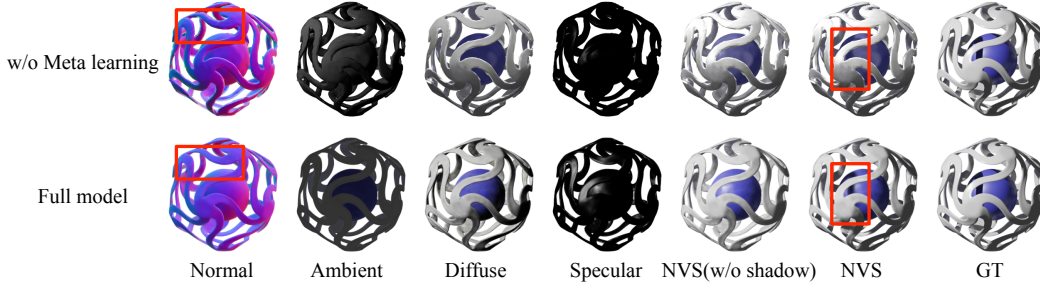


Figure 5: *Ablation studies with qualitative results.* As shown in the decoupling components, the meta-learning scheme is essential for learning object illumination properties and geometries. Without it, the model generates plain lighting effects, particularly for the diffuse and specular components.

For each synthetic scene, we generate a total of 600 OLAT images using Blender, with 500 for training and 100 for testing, following out-of-distribution or colocated data patterns. For the real-world scenes, we split the data according to their point light positions to construct out-of-distribution datasets, and use 600 training images per scene—a significantly smaller subset than the full NRHints dataset.

We compare MetaGS against both NeRF-based and Gaussian-based models, with a particular focus on the OOD relighting setup. The baseline models include state-of-the-art approaches from the past two years: GS³ [1], NRHints [34], WildLight [5], Relightable 3DGS [8], GaussianShader [9], and IRON [35]. The evaluation metrics include PSNR, SSIM [25], and LPIPS [38].

4.2 Results for Out-of-distribution OLAT Relighting

We evaluate MetaGS in scenarios with OOD point light positions to explore its generalizability to unseen illuminations. The qualitative and quantitative results are respectively presented in Figure 3, Figure 4, and Table 2. Our model remarkably outperforms the baselines in handling complex light interactions. For example, in Figure 4, the lighting and shading effects on the cat’s fur are rendered with high fidelity, while baseline models tend to produce incorrect color tones. Furthermore, as indicated by the error maps in Figure 3, all of three OLAT baselines tend to provide unrealistic results,



Figure 6: *Relighting with environment maps*: Our method, trained under OLAT settings, successfully generalizes to unseen environmental lighting; while the compared method is trained in an all-light-on setup that provides multiple lighting conditions per training view, it still fails to generalize to perform relighting with novel environment maps and exhibits visible artifacts.

especially for regions with significant specular reflections or shadows. Please see our Appendix for video visualizations on OOD relighting, where baselines’ performance drops as the light enters the OOD region, while our model provides better relighting results.

The generalizability of MetaGS to OOD illumination stems from two key factors. First, it is due to the Phong model’s ability to capture the underlying principles of light interaction. For example, instead of simply viewing specular highlights as the view-dependent color of a point (which is not generalizable), our method explicitly calculates the interactions of rays and surface normals, enabling the model to generate accurate highlights even in areas not directly illuminated during training. Second, meta-learning further enhances generalization by simulating test-time conditions during training, effectively reducing overfitting to specific illuminations. In the Appendix, we empirically show that the benefit of meta-learning does not come from a larger batch size.

Ablation studies. We conduct a series of ablation studies to validate the effectiveness of the meta-learning scheme as well as the shadow computing. From Table 3 and Figure 5, we observe that the proposed meta-learning training scheme strongly impacts performance. Without it, the model struggles to converge smoothly and make unrealistic component estimations, resulting in plain rendering quality.

Table 3: *Ablation studies of each model component for OOD relighting*. We show the average results of all three synthetic scenes.

Method	PSNR [↑]	SSIM [↑]	LPIPS [↓]
Full model	27.42	0.9546	0.0505
w/o Meta-learning	19.14	0.8781	0.0892
w/o Shadow	21.53	0.9105	0.0735

4.3 Generalization to Other Relighting Setups with Constrained OLAT Training

Camera-light-located relighting. We evaluate MetaGS in the camera-light-located setup, as introduced by IRON [35]. Like the OOD scenario, the colocated setup also imposes limited lighting diversity, as the point light is tied to the camera’s position throughout training. Quantitative results in Table 4 show that MetaGS outperforms IRON, achieving more accurate illumination inference under new point positions, and generating fewer rendering artifacts. Visualizations are presented in the Appendix.

Table 4: *Novel view synthesis results in PSNR under camera-light-located relighting setup*. Full results in all evaluation metrics are included in the Appendix.

Method	Ball	PlaCup	RubCup
IRON [35]	26.99	34.43	36.22
Ours	38.72	36.90	38.89

Few-light and Fixed-elevation-light Relighting. To evaluate MetaGS under constrained illumination conditions, we additionally design the following settings that vary in light distribution, coverage, and sparsity, simulating more realistic capture scenarios. In the few-light setting, only three training lights are sampled, each paired with 200 camera views in a one-light-multi-camera configuration. Two variants are considered: (1) *interpolation*, where training lights are uniformly distributed over

Table 5: Novel view synthesis results in the **few-light** and **fixed-elevation** relighting setup.

Method	Fixed-elevation			Few lights (Interp)			Few lights (Extrap)		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
NRHints [34]	25.76	0.9633	0.0402	18.05	0.8823	0.0911	18.62	0.8960	0.0681
GS ³ [1]	22.93	0.9393	0.0580	19.92	0.9165	0.0693	18.06	0.8781	0.0920
Ours	27.48	0.9590	0.0371	24.90	0.9476	0.0462	23.94	0.9369	0.0533

Table 6: *Results of the proposed meta-learning scheme integrated with various relighting models.* PSNR values demonstrate the generalization performance, with Δ (Meta-Base) indicating the improvement from applying meta-learning.

Method	Ball	PlasCup	RubCup	Cat	Catsmall	CupFabric	Fish	FurScene	Pikachu	Pixiu	Avg.	Δ
Ours	26.76	27.54	<u>27.95</u>	26.45	26.44	27.29	24.68	24.82	25.54	25.65	26.31	–
NRHint	17.25	23.92	27.44	18.04	24.63	24.65	22.57	21.55	24.00	23.03	22.71	–
+meta	17.70	<u>24.57</u>	28.18	19.75	<u>25.44</u>	25.83	24.04	<u>21.90</u>	24.51	<u>24.18</u>	23.61	0.90
RNG	20.22	22.72	24.94	NaN	24.20	25.09	NaN	20.81	23.55	NaN	23.08	–
+meta	<u>22.65</u>	23.93	25.88	<u>21.94</u>	25.07	26.52	23.66	21.75	24.90	22.75	23.91	1.31
GS ³	18.84	20.30	24.37	17.66	23.34	25.04	21.12	17.34	24.11	19.63	21.18	–
+meta	21.54	23.18	26.15	21.55	25.16	<u>26.96</u>	<u>24.38</u>	21.22	<u>25.02</u>	22.98	<u>23.81</u>	2.63

the hemisphere and test lights are randomly sampled from the full hemisphere, and (2) *extrapolation*, where training lights are restricted to one side forming a triangular configuration, and test lights are sampled from the opposite side. In the fixed-elevation-light setting, all training lights are placed at a fixed elevation of 45° , uniformly distributed along a horizontal ring, while test lights are drawn from elevation bands outside this range (i.e., $\leq 30^\circ$ or $\geq 55^\circ$), simulating fixed-height capture trajectories. We report average results on three synthetic scenes in Table 5. Our method consistently achieves strong performance across various lighting settings. In contrast, both baselines degrade significantly under sparse or biased lighting. In particular, NRHints fails to reconstruct the Ball scene under the Three Lights setting, producing a collapsed geometry, which is similar to the failure case described in the paper’s OOD setting. These results demonstrate that MetaGS not only generalizes well to unseen lighting conditions, but also remains robust under extremely limited or clustered illumination. This highlights its practical applicability in real-world relighting scenarios where lighting is sparse, biased, or expensive to capture.

Environment map generalization. Our model generalizes effectively to environment map relighting. We approximate global illumination by importance sampling point lights from the environment map based on their pixel intensities. Notably, for our model in particular (unlike the compared methods), the sampled light directions at test time may follow a distribution different from those encountered during training, presenting a significant challenge for OOD generalization. Instead of assuming infinite light source distance, we simulate distant lighting by placing point lights along environment directions at a fixed distance (twice the scene radius). We compare our method with R3G [8] and GaussianShader [9]. Both methods fail to converge when trained directly on out-of-distribution (OOD) data, so we construct training data that covers the full hemispherical illumination space. Even with this setup, GaussianShader still fails to converge, while Relightable 3DGS struggles with the OLAT learning task and exhibits noticeable artifacts. As shown in Figure 6, MetaGS produces diverse and visually plausible relighting results under complex environment maps, demonstrating its robustness and generalization capability.

To further validate MetaGS’s capability in handling complex lighting conditions, we conduct experiments with test sets containing 2–3 light sources. We also present the results for free-viewpoint relighting with in-distribution point light positions. Please refer to the Appendix for details.

4.4 Meta-Learning with Different Relighting Models

The proposed meta-learning scheme is inherently applicable to other rendering frameworks. To evaluate the generality of the proposed meta-learning scheme, we integrate it into several alternative relighting models across diverse families, including both NeRF-based and 3DGS-based methods,

and spanning from BRDF-based to fully implicit formulations, including NRHints [34], GS³ [1] and RNG [7]. The PSNR results are summarized in Table 6.

While meta-learning consistently enhances OOD generalization across various rendering models, its benefits are most pronounced when paired with explicit geometric modeling and shading methods. Notably, the meta-learning approach yields the largest improvement on our Gaussian-Phong rendering model. This is because the Phong model’s simple yet physically grounded design introduces strong inductive biases that enable meta-learning to capture generalized scene structural information even under limited lighting conditions.

5 Related Work

Recent differentiable volume rendering techniques, including NeRF-based [17], SDF-based [32, 33], and 3DGS-based [10] methods, have significantly improved the quality and efficiency of novel view synthesis for 3D scenes. NeRF-based methods utilize deep neural networks to model volumetric scene functions, encoding color and density to synthesize high-quality images from sparse viewpoints [17, 15, 37]. However, these approaches typically require substantial training time. 3DGS significantly reduces both training and inference time. It converts point cloud data into a continuous volumetric representation by applying Gaussian kernels to point cloud data, facilitating rendering and further processing. This technique is now widely adopted for efficient 3D and 4D reconstructions across varied data types [28, 9, 16, 26, 31].

The task of 3D relighting involves altering the illumination in a 3D scene while maintaining its geometry. It requires decomposing materials and lighting of a scene from multiple images, which is challenging due to its high-dimensional nature. Recent advances in volume rendering have introduced various solutions for this task, including those based on neural fields [39, 27, 22, 20, 21, 13, 23, 22, 30] and the methods based on Gaussian point clouds [28, 9, 8, 36]. A limitation of these methods is that they require a substantial volume of multi-view images captured under individual lighting conditions, making them impractical in real-world scenarios. Several studies have proposed NeRF-based 3D relighting techniques within the OLAT framework [29, 39, 34], which greatly reduce the demands on training data. Nonetheless, these techniques tend to be computationally intensive because of NeRF’s inherent complexity in the volume rendering process. In parallel, concurrent 3DGS-based approaches [6, 1] have also explored the OLAT relighting challenge. Despite their efforts, these approaches still face difficulties when dealing with lighting conditions not encountered during training, similar to the limitations observed in NeRF-based methods.

In contrast, our method addresses this issue by incorporating specific physical priors, represented by the Phong reflection model, into the Gaussian splatting framework. We have also designed a meta-learning approach to enhance the method’s generalizability to OOD relighting scenarios.

6 Conclusions and Limitations

In this paper, we explored a novel and challenging problem: out-of-distribution (OOD) 3D relighting. To address this, we proposed MetaGS, which builds on Gaussian splatting and presented a novel bilevel optimization-based meta-learning framework that explicitly promotes generalizable Gaussian geometry and appearance learning. This meta-learning formulation allows the model to adapt to varying light sources and viewpoints, even when the training data is biased or sparsely sampled in the lighting space. Furthermore, MetaGS incorporates a differentiable Blinn-Phong reflection model within Gaussian splatting, effectively disentangling lighting effects into ambient, diffuse, and specular shading components, thereby improving the physical realism and reconstruction fidelity under diverse lighting. Extensive experiments on both synthetic and real-world datasets demonstrate that MetaGS significantly outperforms existing OLAT relighting approaches under OOD relighting conditions. It also achieves strong generalization to complex environment maps.

Despite these advances, several limitations remain. An open challenge is ensuring the robustness of our approach under more complex lighting conditions. Moreover, our current framework only accounts for direct illumination; incorporating indirect lighting could further enhance the quality of relighting. Additionally, the current model assumes simple Phong reflection, which may limit fidelity when modeling materials with strong subsurface scattering or anisotropic reflectance.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant 62250062), the Shanghai Municipal Science and Technology Major Project (Grant 2021SHZDZX0102), and the Fundamental Research Funds for the Central Universities.

References

- [1] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. GS³: Efficient relighting with triple gaussian splatting. In *ACM SIGGRAPH Asia*, 2024.
- [2] James F Blinn. Models of light reflection for computer synthesized pictures. In *ACM SIGGRAPH*, pages 192–198, 1977.
- [3] Yeonjin Chang, Yearim Kim, Seunghyeon Seo, Jung Yi, and Nojun Kwak. Fast sun-aligned outdoor scene relighting based on TensorRF. In *WACV*, pages 3626–3636, January 2024.
- [4] Jiaxin Chen, Xiao-Ming Wu, Yanke Li, Qimai Li, Li-Ming Zhan, and Fu-lai Chung. A closer look at the training strategy for modern meta-learning. *NeurIPS*, 33:396–406, 2020.
- [5] Ziang Cheng, Junxuan Li, and Hongdong Li. WildLight: In-the-wild inverse rendering with a flashlight. In *CVPR*, pages 4305–4314, 2023.
- [6] Jiahui Fan, Fujun Luan, Jian Yang, Miloš Hašan, and Beibei Wang. RNG: Relightable neural Gaussians. *arXiv preprint arXiv:2409.19702*, 2024.
- [7] 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.
- [8] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3D Gaussian: Real-time point cloud relighting with BRDF decomposition and ray tracing. *arXiv preprint arXiv:2311.16043*, 2023.
- [9] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. GaussianShader: 3D Gaussian splatting with shading functions for reflective surfaces. In *CVPR*, pages 5322–5332, 2024.
- [10] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *TOG*, 42(4):1–14, 2023.
- [11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [12] Zhiyi Kuang, Yanchao Yang, Siyan Dong, Jiayue Ma, Hongbo Fu, and Youyi Zheng. Olat gaussians for generic relightable appearance acquisition. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024.
- [13] Ruofan Liang, Huiting Chen, Chunlin Li, Fan Chen, Selvakumar Panneer, and Nandita Vijaykumar. ENVDR: Implicit differentiable renderer with neural environment lighting. In *ICCV*, pages 79–89, 2023.
- [14] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. GS-IR: 3D Gaussian splatting for inverse rendering. In *CVPR*, pages 21644–21653, 2024.
- [15] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *NeurIPS*, 33:15651–15663, 2020.
- [16] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3D Gaussians: Tracking by persistent dynamic view synthesis. In *3DV*, pages 800–809. IEEE, 2024.
- [17] 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.
- [18] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimeshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 32, 2019.

- [19] Bui Tuong Phong. Illumination for computer generated pictures. In *Seminal graphics: pioneering efforts that shaped the field*, pages 95–101. 1998.
- [20] Kripasindhu Sarkar, Marcel C Bühler, Gengyan Li, Daoye Wang, Delio Vicini, Jérémy Riviere, Yinda Zhang, Sergio Orts-Escolano, Paulo Gotardo, Thabo Beeler, et al. LitNeRF: Intrinsic radiance decomposition for high-quality view synthesis and relighting of faces. In *ACM SIGGRAPH Asia*, pages 1–11, 2023.
- [21] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. NeRV: Neural reflectance and visibility fields for relighting and view synthesis. In *CVPR*, pages 7495–7504, June 2021.
- [22] Jia-Mu Sun, Tong Wu, Yong-Liang Yang, Yu-Kun Lai, and Lin Gao. SOL-NeRF: Sunlight modeling for outdoor scene decomposition and relighting. In *ACM SIGGRAPH Asia*, pages 1–11, 2023.
- [23] 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 *CVPR*, pages 20762–20772, June 2023.
- [24] Dongqing Wang, Tong Zhang, and Sabine Süssstrunk. NEMTO: Neural environment matting for novel view and relighting synthesis of transparent objects. In *ICCV*, pages 317–327, October 2023.
- [25] 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.
- [26] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4D Gaussian splatting for real-time dynamic scene rendering. In *CVPR*, pages 20310–20320, 2024.
- [27] Tong Wu, Jia-Mu Sun, Yu-Kun Lai, and Lin Gao. DE-NeRF: Decoupled neural radiance fields for view-consistent appearance editing and high-frequency environmental relighting. In *ACM SIGGRAPH*, pages 1–11, 2023.
- [28] Tong Wu, Jia-Mu Sun, Yu-Kun Lai, Yuewen Ma, Leif Kobbelt, and Lin Gao. DeferredGS: Decoupled and editable Gaussian splatting with deferred shading. *arXiv preprint arXiv:2404.09412*, 2024.
- [29] Yingyan Xu, Gaspard Zoss, Prashanth Chandran, Markus Gross, Derek Bradley, and Paulo Gotardo. ReNeRF: Relightable neural radiance fields with nearfield lighting. In *ICCV*, pages 22581–22591, 2023.
- [30] Siqi Yang, Xuanning Cui, Yongjie Zhu, Jiajun Tang, Si Li, Zhaofei Yu, and Boxin Shi. Complementary intrinsics from neural radiance fields and CNNs for outdoor scene relighting. In *CVPR*, pages 16600–16609, June 2023.
- [31] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction. In *CVPR*, pages 20331–20341, 2024.
- [32] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *NeurIPS*, 34:4805–4815, 2021.
- [33] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. MonoSDF: Exploring monocular geometric cues for neural implicit surface reconstruction. *NeurIPS*, 35:25018–25032, 2022.
- [34] 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*, pages 1–11, 2023.
- [35] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. IRON: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *CVPR*, pages 5565–5574, 2022.
- [36] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. PhysSG: Inverse rendering with spherical Gaussians for physics-based material editing and relighting. In *CVPR*, pages 5453–5462, June 2021.
- [37] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. NeRF++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020.
- [38] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018.

- [39] 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. *TOG*, 40(6):1–18, 2021.
- [40] Ruichen Zheng, Peng Li, Haoqian Wang, and Tao Yu. Learning visibility field for detailed 3D human reconstruction and relighting. In *CVPR*, pages 216–226, June 2023.
- [41] Zuo-Liang Zhu, Beibei Wang, and Jian Yang. Gs-ror: 3d gaussian splatting for reflective object relighting via sdf priors. *arXiv preprint arXiv:2406.18544*, 2024.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction clearly summarize the main contributions, which are consistent with the methods and results presented in the paper.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: See Section 6.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[NA\]](#)

Justification: See Section 4.1 and 3.3 for the setup of the setting and experimental details.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: We give a detailed explanation of our model in Section 3 and plan to release our code.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We include the code in the supplementary in an anonymous pattern, and we will soon release our latest version of the code publicly before the decision of the paper acceptance.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: See Section 3.3, Section 4 and supplementary materials.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: See 4 for the discussion of statistical significance.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: See Section 3.3.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The research conducted in the paper adheres to the NeurIPS Code of Ethics in all aspects.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The ability to synthesize realistic lighting effects and relight scenes from novel viewpoints could enhance the quality of visual media, including movies, video games, and virtual reality experiences. This could lead to more immersive and engaging entertainment, thereby benefiting society by providing richer visual experiences. However, the ability to manipulate lighting and scenes convincingly could be exploited to create deceptive or misleading content, including deepfakes or falsified evidence in legal or political contexts. This misuse could contribute to the spread of misinformation and undermine trust in visual media.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This paper does not involve the responsible release of data or models that have a high risk for misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: Our dataset includes assets modified from licensed assets acquired from Adobe Stock. According to Adobe's Terms of License, we have obtained the full Adobe Stock license for our dataset.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This paper does not involve crowdsourcing experiments and research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLM is only used for writing, editing, or formatting purposes.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.