ROGR: Relightable 3D Objects using Generative Relighting

NeurIPS 2025 Highlight

Presented by: Akash Ghimire

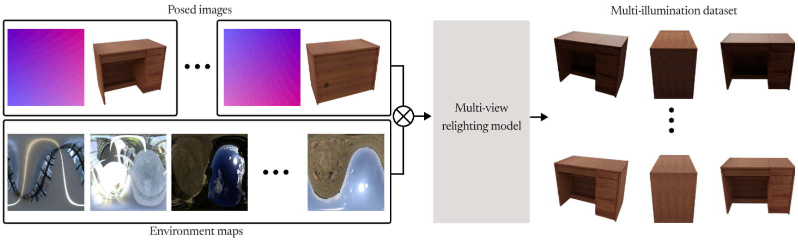
Overview

Input: posed images + target illumination

Output: 3D relit object under novel illumination

- Relightable 3D objects are useful for graphics, AR/VR, and editing.
- Traditionally, relighting requires accurate 3D geometry, material properties, environment lighting, and heavy computation.
- Learning relighting from only multi-view captures is an active research direction with strong practical benefits.

Pipeline and Contribution



- Train a **view-consistent diffusion relighting model**:

$$\left(\{I_i, \pi_i\}_{i=1}^{N}, E_j\right) \xrightarrow{g} \{\hat{I}_{i,j}\}_{i=1}^{N}$$

- Generate a **multi-illumination dataset** for each object:

$$\mathcal{D}_{\text{relit}} = \{g(I_i, \pi_i, E_j)\}_{i=1,\ldots,N,\ j=1,\ldots,M}$$

- Train a **relightable NeRF** for novel view synthesis and novel lighting:

$$F(\mathbf{x}, \mathbf{d}, E_{\text{general}}, E_{\text{specular}}, f_{\text{refl}}) \rightarrow (\mathbf{c}, \sigma)$$

Stage 2: Relightable NeRF

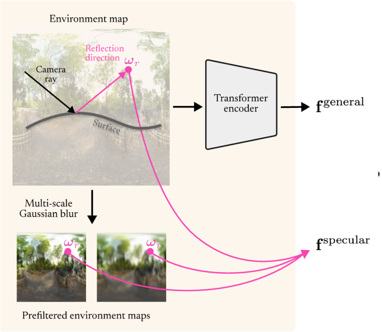
- Base backbone: NeRF-Casting-style radiance field for view-dependent appearance.
- New idea: add lighting conditioning to the NeRF.

$$F(\mathbf{x}, \mathbf{d}, E_{\text{general}}, E_{\text{specular}}, f_{\text{refl}}) \to (\mathbf{c}, \sigma)$$

- $\mathbf{x}, \mathbf{d}$: sample point and viewing direction
- E_{general}: global lighting embedding
- E_{specular}: light around the reflection direction
- f_{refl}: reflected feature used for specular color prediction

$$C(\mathbf{r}) = \sum_i T_i \alpha_i \mathbf{c}_i$$

Two Lighting Signals



- $\mathbf{E}_{\text{general}}$: encodes the whole environment map; captures global, low-frequency lighting.
- $\mathbf{E}_{\text{specular}}$: queries blurred environment-map values around the reflection direction; captures sharp highlights.
- Both signals are concatenated with the reflected feature before predicting specular color.

Experimental Setup

Training data

- 400k synthetic 3D objects, including 100k from Objaverse
- Each object rendered with 64 views $\times$ 16 HDR illuminations
- Environment maps sampled from 590 Polyhaven maps

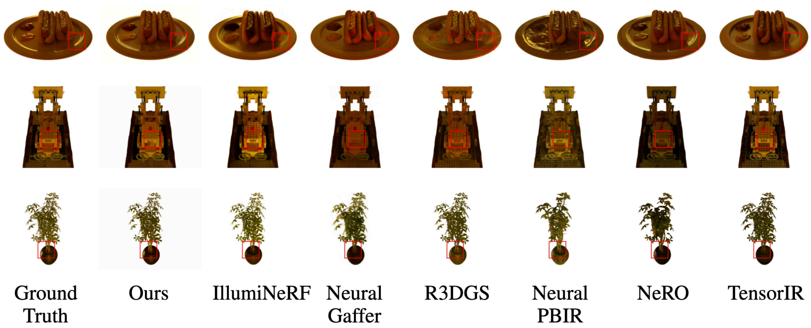
Diffusion model

- Fine-tuned latent diffusion model with CAT3D-style multi-view denoiser
- Input images: 512×512, encoded to $64 \times 64 \times 8$ latents
- 64 views, learning rate 10^{-4}, batch size 128, $360k$ iterations
- Trained on 128 TPU v5 chips

Relightable NeRF

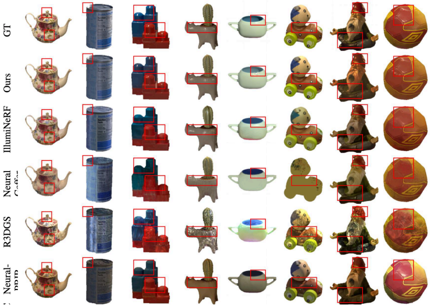
- Trained on 8 H100 GPUs for $500k$ steps
- Target environment map resolution: 512×512
- Specular conditioning uses 3 samples: 1 point sample + 2 Gaussian-blur scales

Qualitative Results: TensoIR



- More plausible specular highlights on reflective objects
- Better color consistency under novel illumination
- Strong gains on hard shiny scenes such as hot dog and ficus

Qualitative Results: Multi-Scene



- Improvements are not limited to one object category.
- ROGR handles both diffuse and reflective cases better than many baselines.
- The biggest visual gains appear around highlights, reflections, and material realism.

Quantitative Analysis

TensoIR benchmark

- Ours: **30.74** PSNR, **0.950** SSIM, **0.069** LPIPS
- Improves over IllumiNeRF: 29.71, 0.947, 0.072

Stanford-ORB benchmark

- Ours achieves best **PSNR-H = 26.21** and **SSIM = 0.980**
- Also stays competitive on **PSNR-L = 32.91** and **LPIPS = 0.027**

Takeaway: The numeric gains are not huge in every metric, but they are consistent and match the stronger visual quality on specular highlights and reflections.

Conclusion

- ROGR turns one multi-view capture under unknown lighting into a relightable 3D object.
- It first generates a multi-illumination relit dataset with a multi-view diffusion model.
- It then trains a lighting-conditioned NeRF that renders new views under new lights without per-light retraining.
- Main strength: better specular appearance while keeping fast feed-forward inference.

Limitations

- This is **not a strong relighting** model in the full physical sense.
- Proper relighting should first de-light, then re-light.
- ROGR does not explicitly learn that decomposition.
- So shadows, highlights, or other baked lighting from the original condition can remain under a new environment map.
- Because the model learns appearance in an MLP-style baked way, not with explicit physical geometry/material decomposition, scene-object composition is also difficult and not physically plausible.