Supplementary Material for OLAT Gaussians for Generic Relightable Appearance Acquisition

This supplementary document provides implementation details of OLAT Gaussians (Sec. 1), a comparison between learnable normals and reference mesh normals (Sec. 2), additional ablation studies (Sec. 3, including camera pose residuals, learnable feature $\boldsymbol{f}_{\mathrm{g}}$, and number of training views), and additional limitations and discussion (Sec. 4, including environmental lighting and material editing).

1 IMPLEMENTATION DETAILS

Architecture. We implement Instant-NeuS+ based on Instant-NSR-pl [Guo 2022] and OLAT Gaussians based on the 3DGS project using PyTorch [Paszke et al. 2019] to build MLPs. We initialize 3D Gaussians for all scenes with 100k points sampled from corresponding proxy meshes. The feature vector $\boldsymbol{f}_{\mathrm{g}}$ of each Gaussian has 64 nodes. Both the incident illumination and scattering MLPs consist of 4 hidden layers with 128 nodes. We use the sigmoid function to activate the incident illumination value and the exponential function to activate the scattering value. Since the datasets do not explicitly calibrate the point light intensity, we empirically set $L_{\mathrm{e}} = 0.8 \cdot max(\{\|\boldsymbol{x}_{\mathrm{l}}\|^2\})$ for all scenes.

Camera Pose Residuals. Camera calibration for multi-view real-world images unavoidably incorporates errors, which cause inconsistency among views. To mitigate this issue, we assign each camera view a trainable camera pose residual, containing a rotation quaternion $\boldsymbol{\Delta q}$ and a translation vector $\boldsymbol{\Delta t}$, which are jointly optimized to refine camera poses. Assuming minor camera calibration errors, we set 1e-5 as the learning rate of $\{\boldsymbol{\Delta_q}, \boldsymbol{\Delta_t}\}$.

Training and Rendering. We train Instant-NeuS+ for 30k iterations to get each proxy mesh using the AdamW optimizer [Kingma and Ba 2014] with β_1=0.9 and β_2=0.99. We train OLAT Gaussians of each scene for 30k iterations using the Adam optimizer [Kingma and Ba 2014] with β_1=0.9 and β_2=0.999. We use the densification strategy of Pixel-GS [Zhang et al. 2024] to refine areas covered by large Gaussians. Our models are trained on a single NVIDIA 3090 GPU. Depending on the scene's complexity, it typically takes around 30 minutes to train Instant-NeuS+ and 25 minutes to train OLAT Gaussians. Tested on a PC with a single NVIDIA 2080ti GPU, a trained model typically renders at around 30 FPS.

2 QUALITY OF LEARNABLE NORMALS

Besides contributing to better specular highlights, the learnable normals of Gaussians are trained to closely match the reference mesh normals by 32.7982 PSNR, 0.9769 SSIM, and 0.0388 LPIPS on average. Fig. 3 shows qualitative comparisons, where the learnable normals are visually close to the reference mesh normals.

3 ADDITIONAL ABLATION STUDIES

Camera Pose Residuals. OLAT Gaussians assign each camera view a pose residual to be jointly optimized. As shown in Tab. 1 and Fig. 1, although visually similar, our method produces more errors at contours (see the L1 error maps in insets) without the camera pose residuals.

Learnable Features. The learnable features $\boldsymbol{f}_{\mathrm{g}}$ assigned to Gaussians embed most material properties that cannot be conveyed by other inputs to F_{incident} and F_{scatter}. As shown in Tab. 1 and Fig. 2,

Table 1. Quantitative results of ablation studies.

Ablation Variants	PSNR ↑	SSIM ↑	LPIPS ↓
w/o camera pose residuals	34.1611	0.9577	0.0646
w/o learnable features	27.4776	0.9179	0.1118
50 training views	28.6642	0.9102	0.0899
100 training views	30.6024	0.9321	0.0760
250 training views	33.0895	0.9530	0.0648
500 training views	34.3669	0.9606	0.0631
Ours	**34.6826**	**0.9617**	**0.0629**

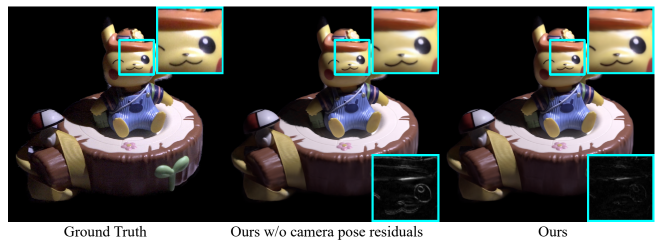

Fig. 1. Ablation study of camera pose residuals. L1 error maps are visualized in insets at the bottom-right corners.

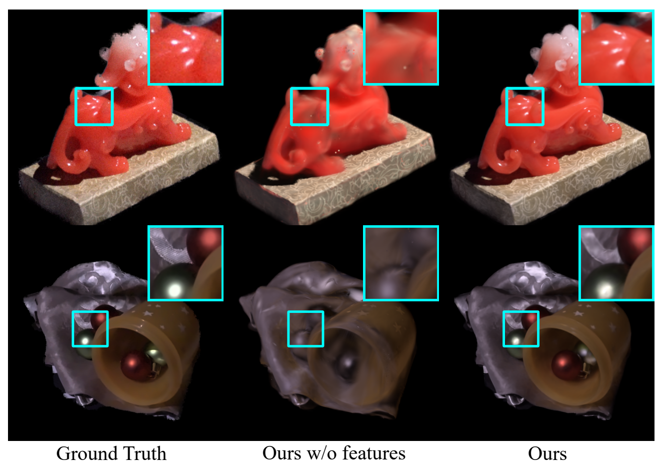

Fig. 2. Ablation study of learnable features. Without per-Gaussian learnable features, the results of OLAT Gaussians show homogeneous appearances lacking local high-frequency details.

Author's Contact Information:

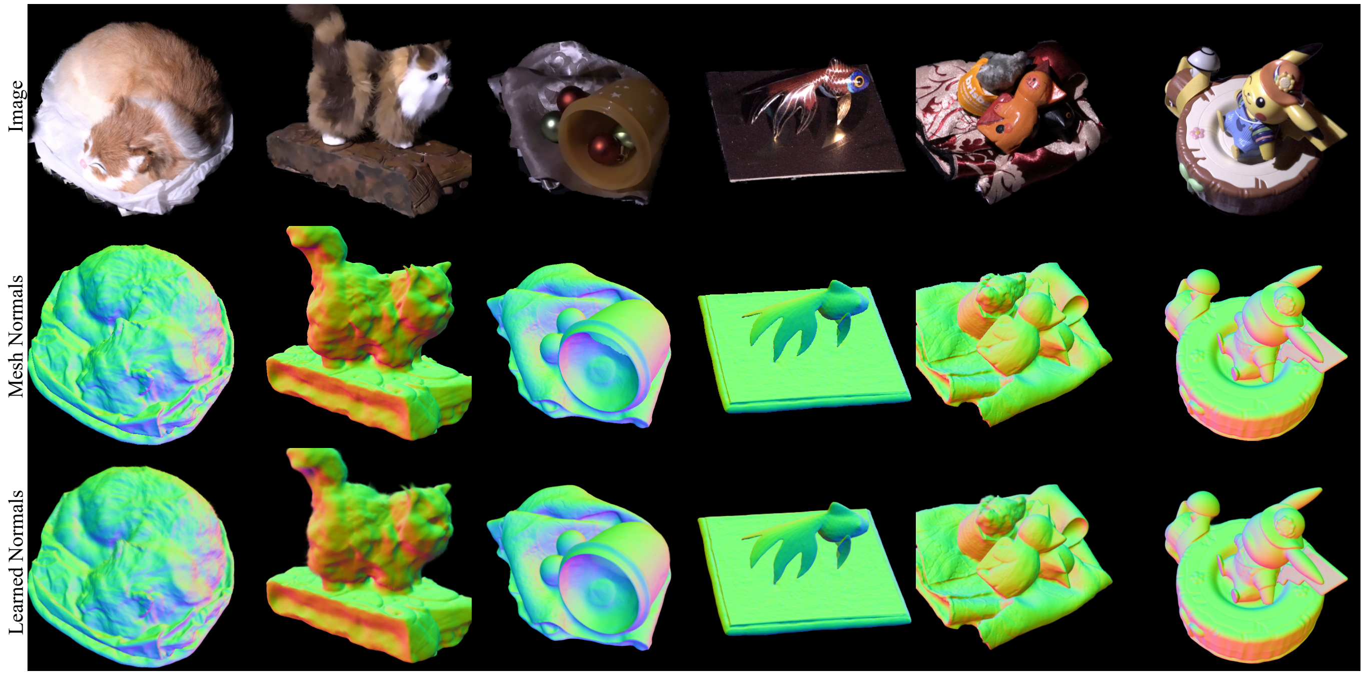

Fig. 3. Qualitative comparisons between the learned normals and reference mesh normals, which are visually close.

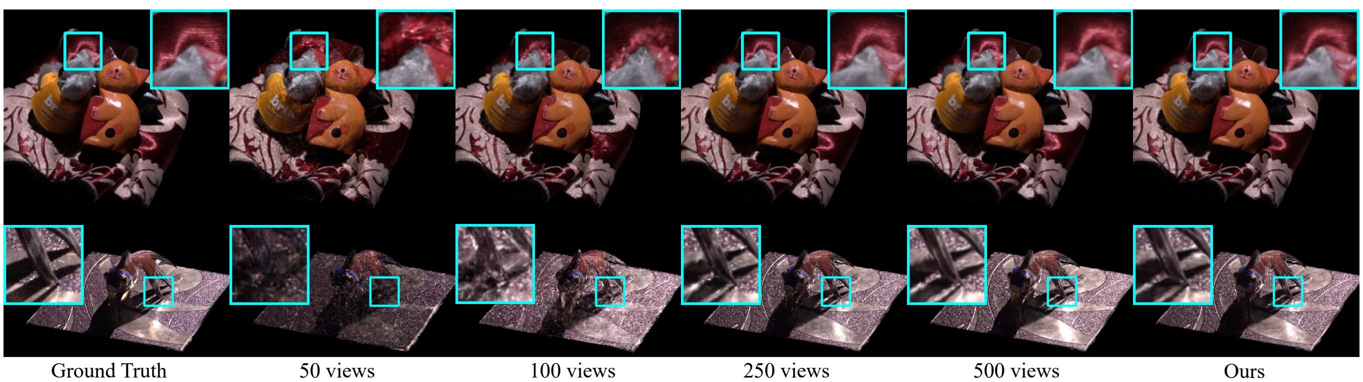

Fig. 4. Ablation study of using fewer training views. OLAT Gaussians produce reasonable results with 250 input views and benefit from more input views.

OLAT Gaussians without learnable features produce homogeneous appearances that lack local high-frequency details.

Number of Training Views. Although OLAT Gaussians have reduced the required number of training views from 10,000 in DNL to 500~1,000, a broad sampling of both views and lights is still necessary to model complex light effects. As shown in Tab. 1 and Fig. 4, 250 training views are sufficient for OLAT Gaussians to produce plausible results, and fewer training views lead to noticeable artifacts. While the results of OLAT Gaussians are improved by more training views, the magnitude of improvement decreases as the number of training views increases.

4 ADDITIONAL LIMITATIONS AND DISCUSSION

Extension to Area and Environmental Light. Without additional training, one plausible approach to this extension is to use an environment map's pixels as queries to get OLAT values, which are accumulated as the Gaussians' colors. Example results of this approach are shown in Fig. 5, where OLAT Gaussians produce photorealistic results under various environmental maps. However, since the processing time increases as the number of pixel samples grows, considerable further adaptation is still required to run OLAT Gaussians with real-time performance for environmental light.

Material Editing. Although our shading formulation demonstrates better relighting quality than parametric PBR shading, the learnable feature f_g does not support semantically meaningful editing

Environment Map | Relighted View 1 | Relighted View 2 | Environment Map | Relighted View 1 | Relighted View 2

Fig. 5. Environmental relighting results. Environment maps and two corresponding relighted views are shown. OLAT Gaussians produce natural relighted appearances under various environment maps that are unseen during training.

like common material parameters. Augmenting $\boldsymbol{f}_{\mathrm{g}}$ with semantic meaning will promote our method to broader usage.

REFERENCES

Yuan-Chen Guo. 2022. Instant Neural Surface Reconstruction. https://github.com/bennyguo/instant-nsr-pl.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).

Zheng Zhang, Wenbo Hu, Yixing Lao, Tong He, and Hengshuang Zhao. 2024. Pixel-GS: Density Control with Pixel-aware Gradient for 3D Gaussian Splatting. *arXiv preprint arXiv:2403.15530* (2024).