PolarGuide-GSDR: 3D Gaussian Splatting Driven by Polarization Priors and Deferred Reflection for Real-World Reflective Scenes

Derui Shan[†,1] Qian Qiao[†,2] Hao Lu[3] Tao Du[*,1] Peng Lu[*,2]

[†]Equal contribution, [*]Corresponding authors. [1]North China University of Technology, Beijing, China

[2]Beijing University of Posts and Telecommunications, Beijing, China [3]University of Science and Technology Beijing, Beijing, China

1. Supplementary Material

To support the theoretical foundation of our proposed method, we include the detailed derivation of the polarization-based specular and diffuse reflection separation model in this section.

1.1. Polarization Model-Based Specular and Diffuse Reflection Separation

According to [4], the polarization characteristics of reflected light are primarily determined by the surface material's polarized bidirectional reflectance distribution function (polarized BRDF) [1, 3]. For most surfaces, the reflected light can typically be decomposed into three components: polarized specular reflection, polarized diffuse reflection, and unpolarized diffuse reflection.

We models the reflected light as two polarized components: the specular reflection component S_{sp} and the diffuse reflection component S_{dp}, corresponding to Fresnel reflection and subsurface scattering, respectively. Both components can be observed using a linear polarizer placed in front of the camera. The polarization camera to capture polarized images. It simultaneously captures images at four different polarization angles: $(0,\ \pi/4,\ \pi/2,\ 3\pi/4)$, denoted as $I_{0^\circ},\quad I_{45^\circ},\quad I_{90^\circ},\quad I_{135^\circ}$, respectively. Let the Stokes vector of the incident light be $S_{in} = [S_0, S_1, S_2]$, where

$$\begin{aligned} S_0 &= 0.5 \cdot (I_{0^\circ} + I_{45^\circ} + I_{90^\circ} + I_{135^\circ}) \\ S_1 &= I_{0^\circ} - I_{90^\circ} \\ S_2 &= I_{45^\circ} - I_{135^\circ} \end{aligned} \tag{1}$$

The Mueller matrix of a linear polarizer at angle θ is denoted as $M_{\text{pol}}(\theta)$, while the Mueller matrices for Fresnel reflection and transmission are denoted as M_R and M_T, respectively. The two polarization components can then be expressed as follows:

$$\begin{aligned} S_{sp} &= M_{pol}(\theta) M_R S_i \\ S_{dp} &= M_{pol}(\theta) M_T S_d \end{aligned} \tag{2}$$

Where S_d is the Stokes vector of the diffuse reflection. To achieve efficient reflection separation for rendering guidance, we assume $S_d \approx S_i - S_{sp}$. This assumption provides an initial cue for reflection separation, and the approximation error will be compensated during the subsequent intensity map fusion stage.

By inverting the DoLP formula for polarized diffuse reflection in the Nayar model [7], the zenith angle, which is the light incidence angle i on the object surface, can be computed. Combined with the surface refractive index ref_{id}, the refraction angle r is then obtained via Snell's law. The DoLP is calculated as follows:

$$DoLP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{3}$$

Let $d = \text{DoLP}$, $k = \text{ref}_{\text{id}}$, and $A = 2(1-d) - (1+d)\left(k^2 + \frac{1}{k^2}\right)$, $B = 4d$, $C = 1 + k^2$, $D = 1 - k^2$. Then, the incident angle i can be derived as:

$$i = \arcsin\left(\sqrt{\frac{-BC(A+B) - \sqrt{C^2(A+B)^2 - D^2(A-B)^2}}{2(A^2 - B^2)}}\right) \tag{4}$$

Subsequently, based on Snell's law, the refraction angle r is computed using the refractive index and the incident angle i derived in Eq. (4):

Specular reflection image is I_{sp}:

$$I_{sp}\left(\phi_{pol}\right) = \frac{I_{\max}^{sp} + I_{\min}^{sp}}{2} + \frac{I_{\max}^{sp} - I_{\min}^{sp}}{2}\cos(2\theta) \tag{5}$$

where

$$\begin{aligned} I_{\max}^{sp} &= \frac{S_0}{2}\left(\frac{\tan\alpha_-}{\sin\alpha_+}\right)^2 \cos^2\alpha_- \\ I_{\min}^{sp} &= \frac{S_0}{2}\left(\frac{\tan\alpha_-}{\sin\alpha_+}\right)^2 \cos^2\alpha_+ \\ \alpha_\pm &= i \pm r \end{aligned} \tag{6}$$

θ denotes the angle between the polarization direction and the x-axis and ϕ_{pol} denotes the channel angle of the polarizer.

Diffuse reflection image is I_{dp}:

$$I_{dp}\left(\phi_{pol}\right) = \frac{I_{\max}^{dp} + I_{\min}^{dp}}{2} + \frac{I_{\max}^{dp} - I_{\min}^{dp}}{2}\cos(2(\theta - \pi/2)) \tag{7}$$

where

$$
\begin{aligned}
I_{\max}^{dp} &= \frac{S_d}{2} \frac{\sin 2i \sin 2r}{(\sin \alpha_+ \cos \alpha_-)^2} \\
I_{\min}^{dp} &= \frac{S_d}{2} \frac{\sin 2i \sin 2r}{(\sin \alpha_+ \cos \alpha_-)^2} \cos^2 \alpha_- \qquad (8)
\end{aligned}
$$

A more detailed derivation process can be found in [5].

Innovation in Polarization Normal Disambiguation: The work by [4] The pioneering work by [4] proposed using sparse SfM/MVS point clouds as geometric priors to resolve azimuth ambiguity in polarization-derived normals, laying the foundation for polarization-based 3D reconstruction. Subsequent studies have also employed geometric priors or multi-view consistency to resolve such ambiguities. However, these static geometric prior-based disambiguation methods heavily rely on initialization accuracy: when the initial geometry fails in highlights or motion-blurred regions, errors propagate persistently and are difficult to correct.

The pioneering work by [4] proposed using sparse SfM/MVS point clouds as geometric priors to resolve azimuth ambiguity in polarization-derived normals, laying the foundation for polarization-based 3D reconstruction. Subsequent studies have also employed geometric priors or multi-view consistency to resolve such ambiguities. However, these static geometric prior-based disambiguation methods heavily rely on initialization accuracy: when the initial geometry fails in highlights or motion-blurred regions, errors propagate persistently and are difficult to correct.

Moreover, unlike "polarization + NeRF" approaches that focus on inverse rendering, this paper introduces a physics-guided forward rendering pipeline, which directly uses polarization information to supervise the reconstruction process of 3DGS, while smoothing prior-induced errors through training strategies and reflection intensity maps. This enhances the physical interpretability of the results while preserving real-time rendering capability and reconstruction accuracy.

Therefore, this work is not merely a technical combination, but rather a systematic attempt to introduce physical priors into neural rendering, achieving a paradigm shift in polarization normal disambiguation from "static one-way constraints" to "dynamic two-way collaboration." As shown in Fig. 1, polarization normal disambiguation relying solely on initial 3DGS geometry (at 3k iterations) still exhibits clear geometric distortion in regions such as the windshield; through iterative optimization with our method, these errors are effectively eliminated, and the geometric structure is accurately recovered.

1.2. Rendering Model

The rendering model of this paper is built upon 3DGS-DR[10], and thus its core rendering pipeline follows the same approach.Following the 3DGS-DR approach, the deferred rendering model consists of two passes. The first pass renders the scene based on the original 3DGS method, which we will not elaborate on further here, and its formulation is as follows:

$$
C(\mathbf{v}) = \sum_i c_i(\mathbf{v}) G\left(\Theta_i, \mathbf{v}\right) \qquad (9)
$$

where Θ_i denotes the Gaussian parameters, $c_i(\mathbf{v})$ represents the per-Gaussian view-dependent spherical harmonics (SH) color, i is the Gaussian index, and $\mathbf{v}$ indicates the view direction.

Then, following the 3DGS-DR approach, we integrate polarization information with the deferred reflection model to yield the final pixel color $C'(\mathbf{v})$.

$$
\begin{aligned}
C'(\mathbf{v}) = 2\Big[&\big(1 - R(\mathbf{v})\big) C(\mathbf{v}) \\
&+ R(\mathbf{v}) E\Big(\frac{2\mathbf{v} \cdot N(\mathbf{v}) N(\mathbf{v})}{\|N(\mathbf{v})\|} - \mathbf{v}\Big)\Big], \qquad (10)
\end{aligned}
$$

$$
N(\mathrm{v}) = \sum_i n_i G\left(\Theta_i, \mathrm{v}\right), \qquad (11)
$$

$$
R(\mathrm{v}) = \sum_i r_i G\left(\Theta_i, \mathrm{v}\right), \qquad (12)
$$

where $R(\mathbf{v})$ represents the intensity map, and E refers to the learned environment map queried in the reflection direction using a bilinear filter. n_i denotes the shortest axis of each Gaussian ellipsoid, interpreted as its normal vector, and r_i represents a per-Gaussian scalar controlling its specular reflection strength.

Since our method leverages polarization priors to guide the reconstruction of both diffuse and specular reflection spaces, we scale the original 3DGS-DR rendering formulation by a factor of 2. This encourages the model to learn and separate the scene representation into these two distinct spaces.

1.3. Implementation Details of PolarGuide-GSDR

Experimental Details. We use COLMAP [9] for Gaussian initialization. All experiments in this study were conducted on the AutoDL cloud computing platform. Except for the GNeRP method, all other methods used a standard hardware configuration consisting of 30 GB of system RAM, a 50 GB data disk, a single NVIDIA GeForce RTX 4090 (24 GB) GPU, and a CPU—either an Intel Xeon Platinum 8352V with 16 vCPUs or an Intel Xeon Platinum 8481C with 25 vCPUs. Minor variations in CPU models across instances arose from parallel experiments on multiple datasets; however, all experiments for a given dataset

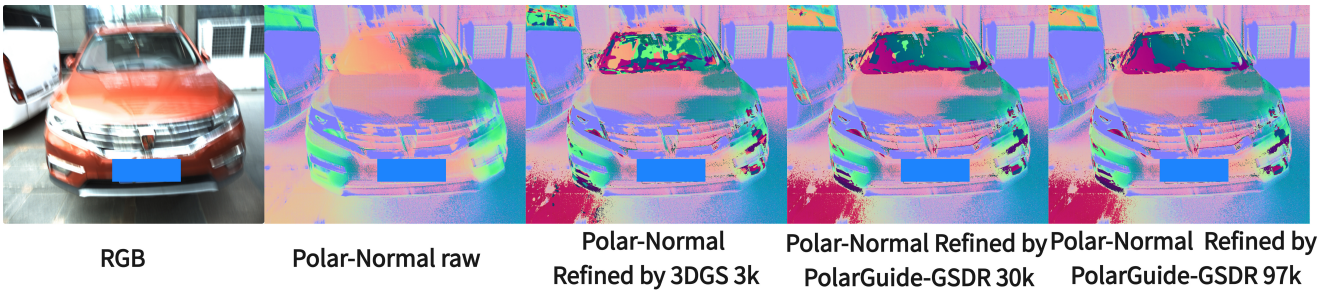

Figure 1. Iterative Disambiguation between 3DGS-Polarization

were performed on the same instance to ensure consistency. Due to its higher GPU memory requirements, the GNeRP method was trained on two RTX 4090 (24 GB) GPUs, with a similar CPU configuration. The Python environment for all methods remained consistent with that used in the 3DGS baseline.

COLMAP initialization based on RGB images may fail in highly reflective scenes due to feature matching failures, thereby affecting pose estimation. Since the current experimental datasets do not contain extreme reflective regions, this issue is not the main focus of our study. We attempted to use the diffuse reflection images separated through polarization as input for COLMAP (taking the Automotive and Glass datasets as examples). This approach improved 3DGS-DR by approximately 0.4 dB, while the enhancement in our method was not significant. The analysis suggests that, on the one hand, the datasets did not excessively include highly reflective scenes; on the other hand, the multi-source constraints such as polarization-derived surface normals introduced in our method already played a similar role in suppressing errors in reflective regions during training. We will further explore the robustness of SLAM in highly reflective areas in future work.

Training Process As described in the methodology section of the main text, surface normals estimated from polarization cues exhibit ambiguity. To mitigate this uncertainty, we first perform initial training of the 3DGS and deferred reflectance model for approximately 10,000 iterations. Based on this initialization, the polarized surface normals are disambiguated. Subsequently, the corrected polarized surface normals, specular reflection images, and diffuse reflection images are used to guide and supervise the model training. Specifically, the diffuse reflection images supervise the diffuse component in both 3DGS and the deferred shading model, while the specular reflection images and surface normals primarily supervise the reconstruction of the specular reflection space in the deferred shading model. Constrained by the inherent difficulty in fully eliminating polarization ambiguity and separation inaccuracies near vertical reflection angles, we adopt a phased training strategy to guide the model toward gradual convergence. The detailed procedure is as follows:

- Initial reconstruction using 3DGS for approximately 10,000 iterations.
- Introduction of corrected polarization priors to jointly supervise 3DGS and the deferred reflectance model for about 30,000–80,000 iterations.
- Final optimization using RGB images and Polarization normal for end-to-end refinement of the rendered output, lasting roughly 10,000 iterations.

The total number of training iterations is set with reference to the 97,000 iterations used by 3DGS-DR in real-world scenes. For datasets with sparse viewpoints, we moderately reduce the total iteration count to prevent overfitting, while maintaining consistency with the iteration settings of 3DGS-DR. The hyperparameters λ and η mentioned in the main text are adjusted slightly across datasets; their specific values are provided in the configuration files included in the supplementary material. Leveraging the intensity image and adopting a training strategy that ultimately relies solely on single RGB images and polarization normals for optimization effectively minimizes errors introduced by eliminating the unpolarized diffuse component.

Parameters reflect each component's physical role and optimization goal. λ_{refl} and λ_{base} are raised above the 3DGS default ($\lambda = 0.2$) to enhance texture in reflective regions. For rendering loss, $\eta_{\text{rgb}} = 1$ ensures RGB priority; due to the inherent error in reflection separation, $\eta_{\text{refl}} = \eta_{\text{base}} \in [0.1, 0.2]$; due to the inherent ambiguity in polarized normal, $\eta_{\text{normal}} \in [0.1, 0.2]$. Parameters are easily determined via grid search (step size: 0.05). The refractive index is fixed at 1.5 (approximating common non-conductive materials) following NeISF; Since the separation prior is only guidance, the model is insensitive to this value. A final pure-RGB phase corrects accumulated prior errors, reducing parameter sensitivity and boosting robustness.

Thus our method demonstrates significant advantages over prior works, such as NeISF++ [8], particularly in its handling of material properties and its applicability to complex, real-world scenes.

The key distinction lies in our treatment of the refractive index. While methods like NeISF++ require the precise

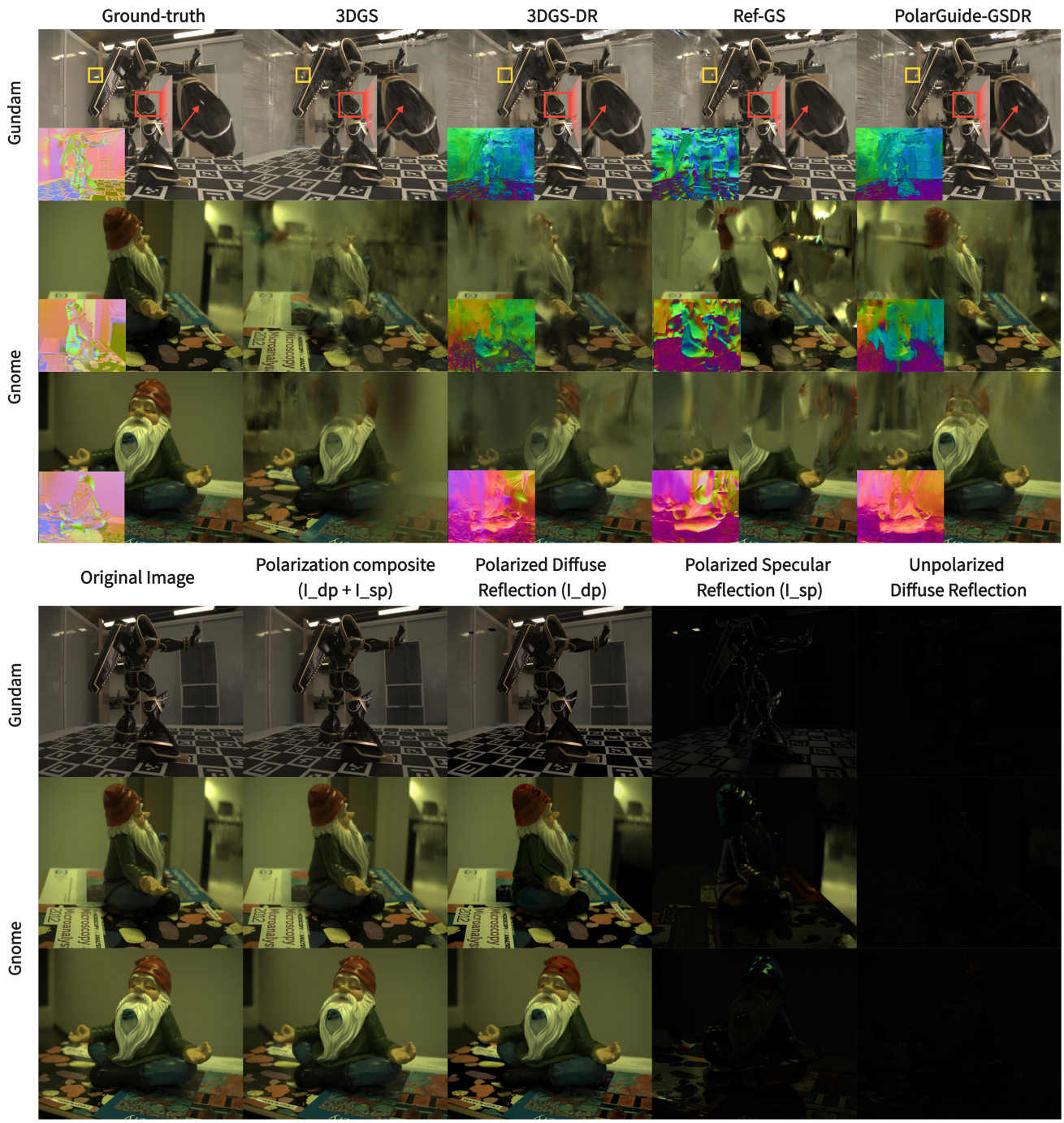

Figure 2. Comparison Test Views of Reconstruction Results and Reflection Separation on the Gnome [6] and Gundam [2] Datasets)

estimation of spatially-varying refractive indices, often involving additional neural fields like SDFs and typically limiting application to single-object, uniform-material scenes, we employ a fixed refractive index (e.g., 1.5) as part of a robust, phased training strategy.

This strategy renders our method largely agnostic to the specific value of the refractive index. The fixed refractive index is used only for an initial, coarse separation of specular and diffuse reflections. These coarsely separated components then serve as stable guidance to bootstrap the 3D reconstruction process in the early stages. Crucially, the final reconstruction quality is refined through end-to-end op-

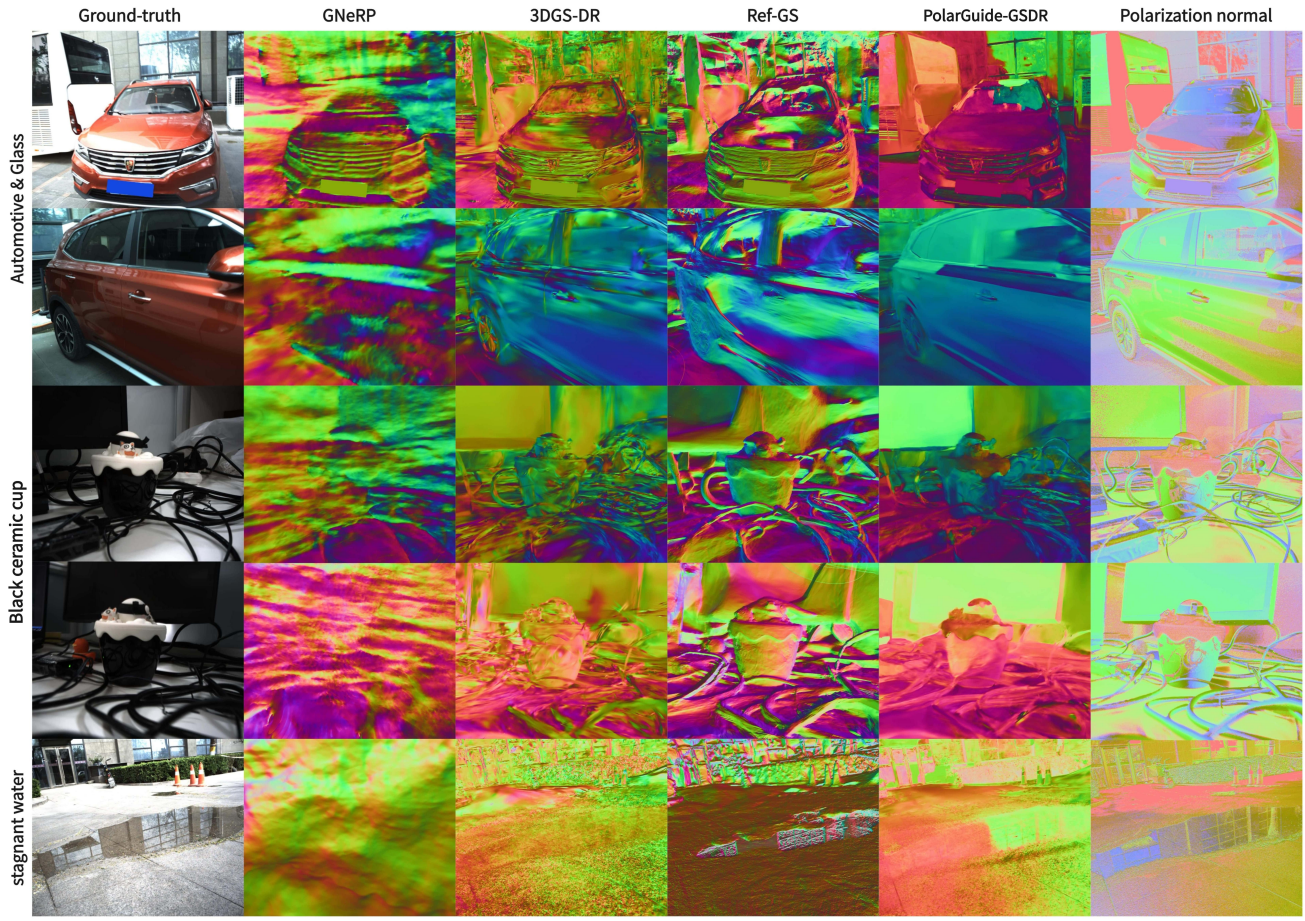

Figure 3. Surface Normals Comparison Chart of the Test Set in the Self-Collected Dataset

timization supervised primarily by RGB images and polarization normals. This phased approach effectively decouples the final output from the inaccuracies of the initial coarse separation.

Consequently, our framework bypasses the challenging and often ill-posed problem of per-pixel refractive index estimation. This not only simplifies the system architecture by eliminating the need for complex networks to represent refractive indices but also enables the successful reconstruction of full, complex scenes containing objects with diverse materials—a scenario where methods requiring precise refractive index estimation typically fail.

Self-Collected Dataset Construction For the self-collected dataset used in this study, image acquisition was performed with a Sony color polarization camera. In the absence of such specialized equipment, researchers can alternatively capture data by attaching a polarizing filter to a standard color RGB camera. Raw data for polarization information calculation can be obtained by rotating the filter and capturing images at four specific angles: 0°, 45°, 90°, and 135°.

Training Time Cost Among all compared algorithms, 3DGS requires the shortest training time, completing in under 30 minutes on most datasets. Due to an increased number of training iterations, 3DGS-DR takes approximately twice as long as 3DGS, typically around one hour. Although Ref-GS is set to 30,000 iterations, its training time is slightly longer than that of 3DGS-DR. Our proposed PolarGuide-GSDR method employs polarization cues to jointly guide and supervise the 3DGS and deferred reflection models. The incorporation of additional polarization-based supervision introduces computational overhead, leading to a training time approximately 30% longer than that of 3DGS-DR, yet it remains comparable to the time required by Ref-GS. Given that GNeRP demands significantly longer training—approximately 10 times that of 3DGS—it is excluded from detailed comparison here.

1.4. Supplementary Comparison of Results on Public Datasets

As shown in Figure 2, our method achieves a significant improvement in reconstruction quality on the Gun-

dam dataset, particularly in restoring complex lighting and material reflections. This improvement is largely attributed to the effective separation of specular and diffuse reflections achieved through polarization imaging. This prior guides the model to separately reconstruct and fuse the two reflection components, leading to the synthesis of more photorealistic novel views. (Note: Since GNeRP produces severely blurred results when input masks are unavailable, we have excluded its corresponding quantitative comparisons to maintain valid analysis.)

However, the sparse nature of both datasets inevitably introduces some artifacts in the reconstructions. For instance, the Gnome dataset contains only 35 images in total (with 5 reserved for testing). To mitigate these artifacts, we separately reconstruct the diffuse and specular reflection spaces using polarization priors and introduce polarization-derived surface normal constraints. This approach significantly reduces artifacts and enhances the robustness of the reconstruction.

Due to space limitations in the main text and to facilitate intuitive correlation between the reconstructed surface normals and the original images, we have integrated them into the same figure. To more clearly demonstrate the comparative effectiveness of the normal reconstruction, Figure 3 provides supplementary multi-view detailed illustrations. As shown, our proposed PolarGuide-GSDR method effectively suppresses noise in surface normal reconstruction while maintaining excellent detail characteristics: in complex curved areas, it avoids over-smoothing or over-sharpening of normals, restoring authentic geometric continuity; in weak-texture regions (such as stagnant water surfaces), it still reconstructs reasonable normal variations. PolarGuide-GSDR demonstrates significant performance improvements in both visual quality and geometric accuracy, fully validating the effectiveness of the polarization normal prior information.

1.5. about code

Please follow our platform:
https://github.com/junxiaofanchen/PolarGuide-GSDR

References

[1] David Brayford, Martin J Turner, and W Terry Hewitt. A physical model for the polarized scattering of light. In *TPCG*, pages 17–26, 2008. 1

[2] Guanying Chen, Yujie He, Le He, Qian Wu, and Yebin Yu. Pisr: Polarimetric neural implicit surface reconstruction for textureless and specular objects. In *European Conference on Computer Vision (ECCV)*, pages 205–222. Springer Nature Switzerland, 2024. 4

[3] Charly Collin, Sumanta Pattanaik, Patrick LiKamWa, and Kadi Bouatouch. Computation of polarized subsurface brdf for rendering. In *Graphics Interface 2014*, pages 201–208. AK Peters/CRC Press, 2020. 1

[4] Zhan Cui, Jinwei Gu, Baoquan Shi, Yasuyuki Matsushita, and Katsushi Ikeuchi. Polarimetric multi-view stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1558–1567, 2017. 1, 2

[5] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo supplementary material, 2017. Supplementary material. 2

[6] Ankit Dave, Yujia Zhao, and Ashok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. In *European Conference on Computer Vision (ECCV)*, pages 538–556. Springer Nature Switzerland, 2022. 4

[7] Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun, and Qifeng Chen. Shape from polarization for complex scenes in the wild. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 12632–12641, 2022. 1

[8] Chengcheng Li, Takahiro Ono, Takashi Uemori, and Yebin Yu. Neisf++: Neural incident stokes field for polarized inverse rendering of conductors and dielectrics. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 26493–26503, 2025. 3

[9] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 2

[10] Kun Ye, Qi Hou, and Kun Zhou. 3d gaussian splatting with deferred reflection. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024. 2