

GNERP: GAUSSIAN-GUIDED NEURAL RECONSTRUCTION OF REFLECTIVE OBJECTS WITH NOISY POLARIZATION PRIORS

Yang LI^{1,2,4}

Ruizheng WU⁴

Jiyong LI³

Yingcong CHEN^{1,2,†}

¹AI Thrust, HKUST(GZ), Nansha, Guangzhou, China

²Department of Computer Science & Engineering, HKUST, Clear Water Bay, HongKong SAR, China

³Department of Computer Science, Sun Yat-sen University, Panyu, Guangzhou, China

⁴R & D Center, SmartMore, Qianhai, Shenzhen, China

† Corresponding Author

yli803@connect.ust.hk, rzwu@cse.cuhk.hk, lijy373@mail2.sysu.edu.cn,
yingcongchen@ust.hk

ABSTRACT

Learning surfaces from neural radiance field (NeRF) became a rising topic in Multi-View Stereo (MVS). Recent Signed Distance Function (SDF)-based methods demonstrated their ability to reconstruct accurate 3D shapes of Lambertian scenes. However, their results on reflective scenes are unsatisfactory due to the entanglement of specular radiance and complicated geometry. To address the challenges, we propose a Gaussian-based representation of normals in SDF fields. Supervised by polarization priors, this representation guides the learning of geometry behind the specular reflection and captures more details than existing methods. Moreover, we propose a reweighting strategy in the optimization process to alleviate the noise issue of polarization priors. To validate the effectiveness of our design, we capture polarimetric information, and ground truth meshes in additional reflective scenes with various geometry. We also evaluated our framework on the PANDORA dataset. Comparisons prove our method outperforms existing neural 3D reconstruction methods in reflective scenes by a large margin. Supplemental materials can be found in this page.

1 INTRODUCTION

Reconstructing 3D shapes from 2D images (Furukawa et al., 2015) is a fundamental problem in computer vision and graphics, with downstream applications such as 3D printing (Chen & Yang, 2014), autonomous driving (Chen et al., 2017), and Computer Aided Design (Furukawa et al., 2010). Although diffuse objects are precisely reconstructed, reflective and textured-less scenes remain challenging. Traditional Multi-View Stereo (MVS) methods (Bregler et al., 2000) rely on stereo matching across views, which is hindered in the presence of specular surfaces and texture absence. Recent methods utilizing implicit neural representation learning for 3D reconstruction have shown promising accuracy (Mescheder et al., 2019; Yariv et al., 2021), yet they overlook the specular reflection between light rays and surfaces, failing to adequately handle glossy objects with high-frequency specular reflection.

Existing methods (Zhang et al., 2021; Liu et al., 2023; Dave et al., 2022) attempt to separate specular reflection components from radiance to improve the reconstruction process. These methods model the interaction of light rays and surfaces by Bidirectional Reflectance Distribution Functions (BRDFs) and estimate them by neural networks. However, the inverse problem posed by BRDFs formulation is highly ill-posed (Guo et al., 2014), and low-frequency bias (Xu et al., 2019) of neural BRDFs making the learned geometry over-smoothed (Liu et al., 2023). Therefore, high-frequency geometry with specular reflection shown in Fig. 1 (a) is intractable for them. Besides, a few methods employ polarization priors to facilitate the learning of specular reflection because they reveal information about surface normals. However, polarization information is always concentrated in

specular-dominant regions and noisy in diffuse regions (Kajiyama et al., 2023), making the reconstruction process in diffuse-dominant regions distorted.

Faced with the bias of neural BRDF and noise issues of polarization priors, we present a novel perspective for reconstructing the detailed geometry of reflective objects. Our key idea is to extend the geometry representation from scalar SDFs to Gaussian fields of normals supervised by polarization priors. Given a surface point, the normals within its neighborhood are approximated by a 3D Gaussian. And it's a more informative representation of geometry. The mean shows the overall (low-frequency) orientation of the surface, while the covariance captures high-frequency details. Coincidentally, the representation can be splatted into the image plane as 2D Gaussians, as illustrated in Fig. 1 (b). The splatting skips the disentangled specular radiance. Learning of the 2D Gaussians can be directly supervised by the polarization information about surface normals. Hence, it circumvents the separation of complex geometry and specular reflection and manages to learn detailed geometry.

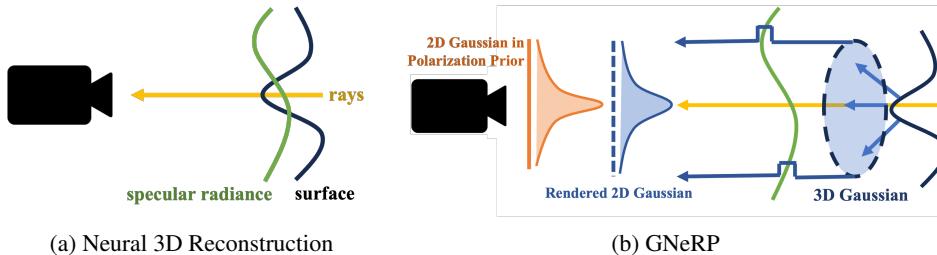


Figure 1: Visualization of Gaussians of normals in Neural Reconstruction pipelines. 2D Gaussians can be rendered from 3D Gaussians of learned normals.

Furthermore, to tackle the noise issues of polarization priors, we introduce a Degree of Polarization (DoP) based reweighting strategy. This strategy adaptively balances the supervision of radiance and polarization priors, enhancing the reconstructing accuracy in diffuse-dominant regions.

In summary, our contributions are as follows:

- We propose a novel polarization-based Gaussian representation of detailed geometry to guide the learning of geometry behind specular reflection.
- We propose a DoP reweighting strategy to alleviate noise and imbalance distribution problems of polarization priors.
- We collect a new challenging multi-view dataset consisting of both radiance and polarimetric images with more diverse and challenging scenes.

2 RELATED WORK

2.1 MULTI-VIEW 3D RECONSTRUCTION

Traditional Multi-view Stereo focuses on the extraction of cross-view features to generate 3D points. (Schönberger et al., 2016; Galliani et al., 2015) try to estimate the depth map of the observed scene with multi-view consistency and fuse the depth maps into dense point clouds. These methods suffer from accumulating errors due to complex pipelines, and features are hard to be extracted from reflective objects. (Mescheder et al., 2019) explicitly models the objects’ occupancy in a voxel grid to guarantee a complete object model is created. However, the resolution of the voxel limits the accuracy of the reconstructed surface. Recently, the success of NeRF (Mildenhall et al., 2020), which uses a simple MLP to encode the color and density information for a scene, inspired researchers to resort to implicit representation for multi-view 3D reconstruction. The representative works are Unisurf (Oechsle et al., 2021), NeuS (Wang et al., 2021), and VolSDF (Yariv et al., 2021), which exploit an MLP to model a Signed Distance Function (SDF) for a target scene. These methods optimize the implicit representation, i.e., SDF, by minimizing the MSE loss between the rendered pixel’s radiance value and the corresponding pixel’s radiance value in GT images. Such a paradigm works

well with Lambertian surfaces. However, only view-conditioned radiance fields fail in reflective scenes.

2.2 BRDF FOR REFLECTIVE OBJECTS RECONSTRUCTION

In the regions with complex geometry, BRDFs always exhibit high-frequency variations due to the normals terms, while the low-frequency implicit bias of neural networks (Xu et al., 2019) disables neural BRDFs from predicting these abrupt changes. It always results in over-smoothed geometry. For example, NeRO (Liu et al., 2023) adopts Micro-facet BRDF (Cook & Torrance, 1981) parameterized by material and normal distribution terms. Although its results of smooth mirror-like objects are excellent, the spatial continuity of neural BRDF is a barrier to the combination of complex geometry and specular reflection. In the regions with complex geometry, sole multi-view images with disentangled radiance result in severe ill-posedness of the inverse problem, as is shown in Fig. 1 (a). Moreover, explicit estimation of anisotropic normals distribution has been used in rendering delicate objects, such as anisotropy shading of hairs (Banks, 1994), to improve the perception of orientation and shapes (Ament & Dachsbacher, 2015). However, anisotropic normals distribution in neural SDFs for 3D reconstruction remains under-defined and non-trivial. Our method proposes 3D Gaussians, of which anisotropic 3D covariance is more informative than the scalar normals distribution term in NeRO. The latter only measures the concentration of normals at a surface point.

2.3 MULTI-VIEW 3D RECONSTRUCTION WITH POLARIZATION

Polarization prior reveals the azimuth angle of the surface normal, i.e., the angle between the normal projection onto the image plane and the positive x-axis of the image. Shape-from-polarization has been investigated by other papers (Atkinson & Hancock, 2006; Foster et al., 2018; Fukao et al., 2021; Cui et al., 2017; Kadambi et al., 2015; Zhao et al., 2020) before the invention of neural 3D reconstruction. But most of them are focused on common scenes. For example, PMVIR (Zhao et al., 2020) exploits the relation of the polarization angle and the azimuth angle of normals but with only Lambert shading, and thus it cannot treat reflective objects at all. Neural 3D Reconstruction with polarization priors has also been explored. Sparse Ellipsometry (Hwang et al., 2022) develops a device to capture polarimetric information and 3D shapes concurrently. However, their reconstruction is always disturbed by the noise in diffuse-dominant regions. For example, PANDORA (Dave et al., 2022) extends radiance in BRDF into polarimetric dimensions while the geometry of diffuse regions cannot be learned properly.

2.4 GAUSSIANS IN 3D SCENE REPRESENTATION

Gaussians are used to represent the attributes of 3D scenes. Mip-NeRF (Barron et al., 2021) encodes Gaussian regions of space rather than infinitesimal points for anti-aliasing. (Zwicker et al., 2001) proposes Gaussian splatting that taking volume data as 3D Gaussians and nearly projects the 3D Gaussian to the 2D one (Kerbl et al., 2023). (Kerbl et al., 2023) implements the splatting pipeline on the NeRF for real-time rendering. In numerical geometry, (Berkmann & Caelli, 1994) calculates the covariance matrix from the projections of the normal vectors to highlight the edges and local geometry of surfaces. Inspired by them, we demonstrate a further fact that taking surface normals as 3D Gaussians and going through a similar splatting pipeline would exactly be transformed into 2D Gaussians. Our 2D Gaussians are coincidentally available for polarization priors. Thus, supervised by polarization priors, the learned 3D Gaussians capture more details, which represent the average orientation of normals by means and the changes within the neighborhood by covariance matrices.

3 METHODS

3.1 PRELIMINARY OF POLARIZATION

Here, we introduce the concept of polarization and its mathematical relation to surface normals projected to the captured images. The prior contributes to the disentanglement of specular radiance and geometry.

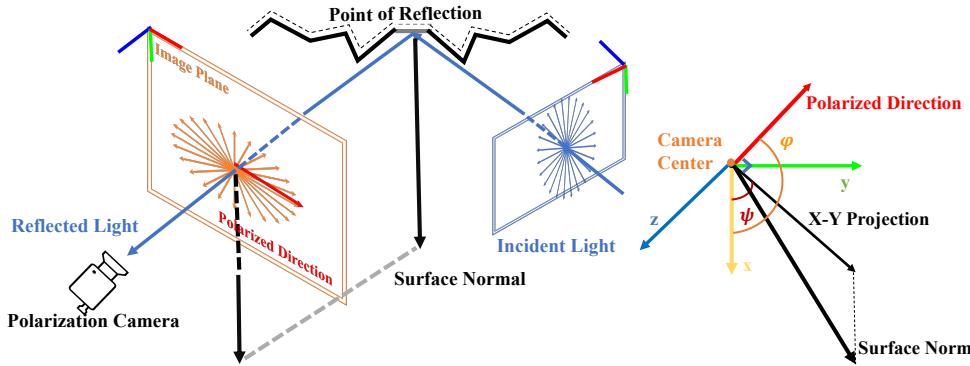


Figure 2: Illustration of polarization shift in specular reflection. The right figure is a detailed description of the geometry relation between AoP and the surface normal. ψ is the azimuth angle. φ is the AoP, which is the angle from the positive x-axis to the polarized direction.

Polarimetry describes the vibration status of light waves. Since light is a type of transverse wave that only oscillates in the plane perpendicular to the light path (Collett, 2005), the full polarimetric cues of rays are always represented by planar ellipses (Smith & Ward, 1974). The magnitude of vectors inside these ellipses alludes to the amplitude of the light wave vibration along the vectors, as shown in Fig. 2. Common light sources, such as sunlight and LED spotlights, emit unpolarized light, i.e., the light vibrates equally in all directions. In our captured scenes, objects are mostly illuminated directly by light sources, so we assume that the incident light is unpolarized. During reflection, the vibration in each direction is absorbed unequally, and unpolarized incident light turns into partially polarized reflected light captured by polarization cameras. The Angle of Polarization (AoP) and Degree of Polarization (DoP) are two cues of the polarization ellipse functionally related to projected surface normals at the points of reflection, which can be formulated as:

$$\varphi(i, j) = \frac{1}{2} \arctan \frac{\mathbf{s}_2(i, j)}{\mathbf{s}_1(i, j)}, \quad \rho(i, j) = \frac{\sqrt{\mathbf{s}_1^2(i, j) + \mathbf{s}_2^2(i, j)}}{\mathbf{s}_0(i, j)}, \quad \{\varphi, \rho\} \in \mathbb{R}^{H \times W}, \quad (1)$$

where φ, ρ are AoP and DoP, (i, j) is the pixel index, and $\mathbf{s} = [\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$ is Stokes vector directly calculated from polarization capture. Generally, in specularity-dominant regions, the relation between projected normals and AoP is fixed as Fig 2(b) and the equation $\psi + \frac{\pi}{2} \equiv \varphi \bmod \pi$. Moreover, DoP is significantly higher in these regions. Details of polarization analysis are shown in the Appendix.

3.2 GAUSSIAN GUIDED POLARIMETRIC NEURAL 3D RECONSTRUCTION PIPELINE

Polarimetric neural 3D reconstruction refers to reconstructing surfaces by neural implicit surface learning, given N calibrated multi-view images $\mathcal{X} = \{\mathbf{C}_i\}_{i=1}^N$ with pixel-aligned polarization priors $\mathcal{Y} = \{\varphi_i, \rho_i\}_{i=1}^N$. First, we introduce a general pipeline of learning surface by volume rendering, taking NeuS (Wang et al., 2021) as an example. Sec. 3.2.2 introduces the 3D Gaussian of surface normals and its transforms to 2D Gaussian in the image plane. It represents the geometry of surface points more precisely and thus can separate detailed geometry from high-frequency specular radiance. 3.2.3 presents our full optimization containing radiance loss and Gaussian loss, which measures the gap between these 2D Gaussians and polarization priors. We propose a DoP reweighting strategy to alleviate the aforementioned noise and imbalanced distribution of polarization priors. It balances the influence of radiance and polarimetric cues adaptively. Finally, Sec. An overview of the entire framework is shown in Fig. 3.

3.2.1 LEARNING SURFACE BY VOLUME RENDERING

NeRF (Mildenhall et al., 2020) proposed a novel render pipeline with a combination of spatial neural radiance fields and volume rendering (Kajiya & Von Herzen, 1984) to synthesize high-quality novel view images. Unlike traditional explicit meshes, the representation of 3D scenes in NeRF

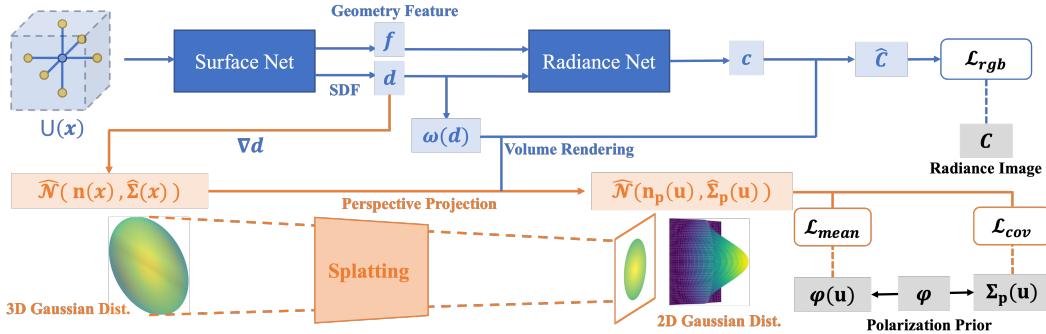


Figure 3: Illustration of our method.

is decomposed into spatial-dependent density fields and radiance fields depending on both spatial position and viewing direction. Then, the color of an arbitrary pixel with a ray $\mathbf{r} = \mathbf{o} + t\mathbf{d}$ passed through can be rendered by volume composition along the ray:

$$\hat{\mathbf{C}}(\mathbf{r}) = \sum_{k=1}^K T_i \alpha_i \mathbf{c}_i(\mathbf{r}_i, \mathbf{d}), \quad T_i = \exp\left(-\sum_{j=1}^{i-1} \alpha_j \delta_j\right), \quad \alpha_j = 1 - \exp(-\sigma_j(\mathbf{r}_j) \delta_j), \quad (2)$$

, where K points $\{\mathbf{x}|\mathbf{o}+t_i\mathbf{d}\}_{i=1}^K$ on the ray are sampled. σ_i and \mathbf{c}_i are approximated volume density and radiance predicted by neural networks with position \mathbf{x} and viewing direction \mathbf{d} as inputs. δ_i is the length of sampled interval $[t_{i-1}, t_i]$. α_i and T_i denote the transmittance and alpha value of points, and by them the final color is alpha composited (Max, 1995). The neural network is optimized by the mean square error between the ground truth color $\mathbf{C}(\mathbf{r})$ in the image and the rendered color $\hat{\mathbf{C}}(\mathbf{r})$.

Despite realistic novel view images, the geometry of scenes extracted from learned density fields is inaccurate with floating artifacts since the shape is not defined in the density field. NeuS (Wang et al., 2021) defines surfaces as the zero-level set of Signed Distance Field (SDF), and density is derived from SDF:

$$[d(\mathbf{x}_i), \mathbf{f}(\mathbf{x}_i)] = f(\mathbf{x}_i), \quad \alpha_i = \max\left(\frac{\Phi_s(d(\mathbf{x}_i)) - \Phi_s(d(\mathbf{x}_{i+1}))}{\Phi_s(d(\mathbf{x}_i))}, 0\right), \quad \mathbf{c}_i = c(\mathbf{x}_i, \mathbf{n}_i, \mathbf{d}, \mathbf{f}_i), \quad (3)$$

where c and f are the geometry network and radiance network, $d(\mathbf{x}_i)$ is the signed distance to the surface and $\mathbf{f}_i = \mathbf{f}(\mathbf{x}_i)$ is the geometry feature. α_i is defined by SDF with Laplace distribution $\Phi_s(x) = (1 + e^{-sx})^{-1}$, where the variance s is a trainable parameter. The volume rendering process is analogous to NeRF, while the radiance network takes normals $\mathbf{n}_i = \nabla_{\mathbf{x}} d(\mathbf{x}_i)$ and geometry feature \mathbf{f}_i as additional inputs.

3.2.2 GAUSSIAN SPLATTING OF NORMALS

Neural SDF-based 3D reconstruction excels at smooth Lambertian objects. With neural BRDF defining the specular reflection between rays and surfaces, smooth surfaces of reflective objects can also be properly learned. However, the low-frequency implicit bias of neural networks (Xu et al., 2019) is a barrier for both of them to recover delicate geometry behind specular reflection, such as abrupt normal changes in NeRO (Liu et al., 2023). Thus, we propose a 3D Gaussian estimation of distributions of normals as an additional representation of geometry details. We show how it can be splatted to the image plane, making it available for 2D polarization supervision.

Instead of separate vectors assigned to each point, the normal within the neighborhood of an arbitrary position \mathbf{x}_i is assumed as a Gaussian:

$$\mathcal{G}(\mathbf{x}|\mathbf{x}_i) = \mathcal{N}(\mathbf{n}(\mathbf{x}_i), \Sigma(\mathbf{x}_i)) = \frac{1}{(2\pi)^{\frac{3}{2}} |\Sigma(\mathbf{x}_i)|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{n}(\mathbf{x}_i))^T \Sigma(\mathbf{x}_i)^{-1} (\mathbf{x} - \mathbf{n}(\mathbf{x}_i))\right), \quad (4)$$

where $\mathbf{n} \in \mathbb{R}^3$ is the normal, and $\Sigma \in \mathbb{R}^{3 \times 3}$ is the covariance of the Gaussian. Given a ray with discretization $\{\mathbf{x}_i|\mathbf{x} + t_i\mathbf{d}\}_{i=1}^K$, additional M positions within the neighborhood are super-sampled

to estimate the covariance. In this paper, M is 6 containing \mathbf{x}_{i-1} , \mathbf{x}_{i+1} and four positions around the ray. Hence, the unbiased estimation of Gaussian can be formulated as:

$$\hat{\mathcal{G}}(\mathbf{x}|\mathbf{x}_i) = \mathcal{N}(\mathbf{n}(\mathbf{x}_i), \hat{\Sigma}(\mathbf{x}_i)) = \mathcal{N}\left(\mathbf{n}(\mathbf{x}_i), \frac{1}{M-1} \sum_{j=1}^M (\mathbf{n}(\mathbf{x}_i^j) - \mathbf{n}(\mathbf{x}_i)) (\mathbf{n}(\mathbf{x}_i^j) - \mathbf{n}(\mathbf{x}_i))^T\right), \quad (5)$$

where $\mathbf{n}(\mathbf{x}_i^j) = \nabla_{\mathbf{x}} d(\mathbf{x}_i^j)$, $\mathbf{n}(\mathbf{x}_i^j) \in \mathbb{R}^3$. However, those 3D Gaussians are not accessible in captured 2D images, and volume rendering in Sec. 3.2.1 only takes 3D scalar fields into account, making the projection of 3D Gaussians non-trivial. Alternatively, (Zwicker et al., 2001) presents a splatting approach treating colors in 3D space as Gaussian kernels and visualizing them on the image plane. We apply analogous transforms and further prove our normal-based 3D Gaussians are exactly splatted to 2D Gaussians. Given a viewpoint, the transform can be formulated as:

$$\hat{\mathcal{G}}(\mathbf{x}|\mathbf{x}_i)_p = \mathcal{N}(\mathbf{JWn}(\mathbf{x}_i), \mathbf{JW}\hat{\Sigma}(\mathbf{x}_i)\mathbf{W}^T\mathbf{J}^T) = \mathcal{N}\left(\begin{bmatrix} \mathbf{n}_p(\mathbf{x}_i) \\ 0 \end{bmatrix}, \begin{bmatrix} \hat{\Sigma}_p(\mathbf{x}_i) & 0 \end{bmatrix}\right), \quad (6)$$

where $\mathbf{W} \in \mathbb{R}^{3 \times 3}$, $\mathbf{J} \in \mathbb{R}^{3 \times 3}$ are viewing transform matrix and normal projection matrix (Chen et al., 2022), respectively. Derivation of them is shown in the Appendix. It shows that only the first two rows of the transformed mean vector and the upper 2×2 square block of the transformed covariance matrix remain non-zero, splatting 3D Gaussians to 2D Gaussians in the image plane. For simplification, 2D Gaussians are also denoted by $\hat{\mathcal{G}}(\mathbf{x}|\mathbf{x}_i)_p = \mathcal{N}(\mathbf{n}_p(\mathbf{x}_i), \hat{\Sigma}_p(\mathbf{x}_i))$, where $\mathbf{n}_p \in \mathbb{R}^2$, $\hat{\Sigma}_p(\mathbf{x}_i) \in \mathbb{R}^{2 \times 2}$. Moreover, the SVD of the covariance matrix $\hat{\Sigma}_p = \hat{\mathbf{V}}\hat{\Lambda}\hat{\mathbf{V}}^T$ circumvents the ill-posedness of the covariance matrix and reveals its relation to anisotropic normal distribution. Intuitively, if the geometry appears smooth from the imaging perspective, then the corresponding normals of the neighborhood will be projected to similar vectors, resulting in an insignificant deviation of the eigenvalues. Otherwise, the deviation would be significant. Eigenvectors also show the local shape of the position, as shown in Fig. 4 (e). Finally, all 2D Gaussians on the ray passing through the pixel \mathbf{u} is composited by volume rendering:

$$\hat{\mathcal{G}}(\mathbf{u}) = \mathcal{N}\left(\sum_{k=1}^K T_i \alpha_i \mathbf{n}_p(\mathbf{x}_i), \sum_{k=1}^K T_i \alpha_i \hat{\Sigma}_p(\mathbf{x}_i)\right) = \mathcal{N}(\mathbf{n}_p(\mathbf{u}), \hat{\Sigma}_p(\mathbf{u})), \quad (7)$$

where T_i and α_i are in Eq. 2. $\mathbf{n}_p(\mathbf{u}) \in \mathbb{R}^2$, $\hat{\Sigma}_p(\mathbf{u}) \in \mathbb{R}^{2 \times 2}$. The mean of 3D Gaussians $\mathbf{n}(\mathbf{x}_i)$, which is splatted to $\mathbf{n}_p(\mathbf{u})$, represents the overall orientation of \mathbf{x}_i . And the covariance $\hat{\Sigma}(\mathbf{x}_i)$ and splatted $\hat{\Sigma}_p(\mathbf{u})$ in the image model the high-frequency details. In this way, our representation captures more details than NeuS and other methods based on the neural BRDF parameterized by isotropic normals distribution. Another main strength of those 2D Gaussians is direct supervision by polarization priors, which is introduced in Sec. 3.2.3.

3.2.3 OPTIMIZATION WITH REWEIGHTED POLARIZATION PRIORS

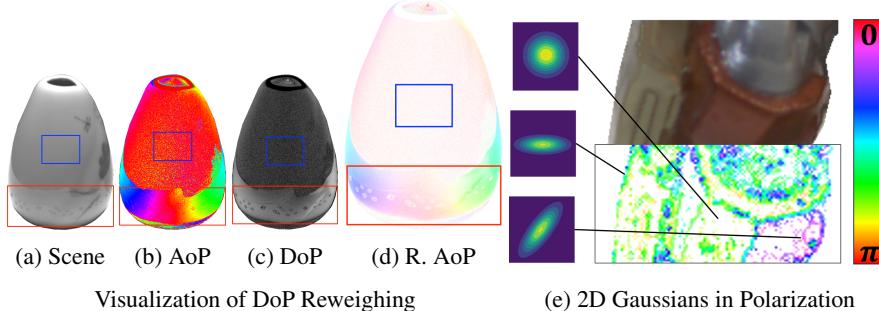


Figure 4: Visualization of Reweighted AoP Priors. Red boxes bound specular reflection dominant regions, and the blue boxes bound diffuse ones. (d) is the AoP map reweighted by DoP. Saturation in (e) indicates the degree of anisotropy, and color represents the direction of the singular vector of 2D Gaussians’ covariance. A few 2D Gaussians are drawn as ellipses for intuition.

The 2D Gaussian in Sec. 3.2.2 can be directly extracted from AoP priors $\{\varphi_i\}_{i=1}^N$ in Eq. 1 by:

$$\begin{aligned}\tilde{\mathcal{G}}_{\mathbf{p}}(\mathbf{u}|\mathbf{u}_i) &= \mathcal{N}(\tilde{\mathbf{n}}_{\mathbf{p}}(\mathbf{u}_i), \tilde{\Sigma}_{\mathbf{p}}(\mathbf{u}_i)) \\ &= \mathcal{N}\left(s\mathbf{v}(\psi(\mathbf{u}_i)), \frac{1}{M'-1} \sum_{j=1}^{M'} (\mathbf{v}(\psi(\mathbf{u}_i^j)) - \mathbf{v}(\psi(\mathbf{u}_i))) (\mathbf{v}(\psi(\mathbf{u}_i^j)) - \mathbf{v}(\psi(\mathbf{u}_i)))^T\right),\end{aligned}\quad (8)$$

where $\mathbf{u}_i^{(j)} = (\mathbf{x}_i^{(j)})_{\mathbf{p}}$ is the corresponding pixel index of (super-sampled) points on the ray. Therefore, $M' = 4$ since \mathbf{x}_{i-1} and \mathbf{x}_{i+1} are on the same ray as \mathbf{x}_i . $\mathbf{v}(\theta)$ represents a 2D unitary vector rotated by θ . $\Psi \equiv \varphi + \frac{\pi}{2} \bmod \pi$ is the azimuth angle of normals, derived from the AoP in Eq. 1. And s is a scale factor. Similar to Sec. 3.2.2, the estimated covariance matrix is decomposed into $\tilde{\Sigma} = \tilde{\mathbf{V}}\tilde{\Lambda}\tilde{\mathbf{V}}^T$. We define the degree of anisotropy (DoA) of those 2D Gaussians as $\frac{\Lambda_0}{\Lambda_1}$. 2D Gaussians saturated by DoA are visualized in Fig. 4 (e). In this Fig, color is concentrated to and coherent along the edges of the scene. It shows DoA is higher in the region with complicated geometry and surface changes most dramatically along singular vectors of covariance. Before optimization, the polarization prior AoP is reweighted by DoP to alleviate the aforementioned observational noise and imbalanced distribution problem in Sec. 1. The noise of AoP is mainly generated by diffuse reflection because it's always weakly polarized (Kajiyama et al., 2023). The DoP in diffuse-dominant regions is significantly lower than specular-dominant ones, as shown in Fig. 4 (c) and (d). Thus, the reweighted AoP defined as $\varphi \cdot \rho$ is proposed as an alternative supervision with less noise. Meanwhile, radiance is disentangled with surroundings in specular reflection dominant areas. To adaptively balance radiance and polarization priors, our full loss function during reconstructing is defined as:

$$\begin{aligned}\mathcal{L} &= \alpha(1 - \rho)\mathcal{L}_{\text{color}} + \beta\rho(\mathcal{L}_{\text{mean}} + \mathcal{L}_{\text{cov}}) + \gamma\mathcal{L}_{\text{eik}} + \delta\mathcal{L}_{\text{mask}}, \\ \mathcal{L}_{\text{color}} &= \|\hat{\mathbf{C}}(\mathbf{u}) - \mathbf{C}(\mathbf{u})\|_2, \quad \mathcal{L}_{\text{mean}} = \|\hat{\varphi}(\mathbf{n}_{\mathbf{p}}(\mathbf{u})) - \varphi(\mathbf{u})\|_1, \\ \mathcal{L}_{\text{cov}} &= \left(\left\|\frac{\hat{\Lambda}_1}{\hat{\Lambda}_0} - \frac{\tilde{\Lambda}_1}{\tilde{\Lambda}_0}\right\|_1 + \beta' \langle \hat{\mathbf{V}}, \tilde{\mathbf{V}} \rangle\right)(\mathbf{u}), \quad \mathcal{L}_{\text{eik}} = \frac{1}{K} \sum_{i=1}^K (\|\nabla_{\mathbf{x}} d(\mathbf{x}_i)\|_2 - 1)^2,\end{aligned}\quad (9)$$

where $\mathcal{L}_{\text{color}}$ and $\mathcal{L}_{\text{mask}}$ are the radiance rendering loss and the BCE loss of object masks in NeuS (Wang et al., 2021). Splatted 2D Gaussians is supervised by $\psi(\mathbf{u})$ and $\tilde{\Sigma}_{\mathbf{p}}(\mathbf{u})$ in Eq. 8. $\hat{\varphi}(\mathbf{n}_{\mathbf{p}}(\mathbf{u})) \equiv \psi(\mathbf{u}) + \frac{\pi}{2} \bmod \pi$ and ψ is the azimuth angle of normals. The supervision of radiance and polarization priors are reweighted by the DoP ρ . Especially, only Anisotropy (ratio of singular values) and eigenvectors are supervised to avoid scaling and numerical issues. If the local shape is like a plane, normals will change smoothly in all directions, and the Anisotropy approaches 1. If there are some details like edges, normals tend to change abruptly and exhibit directionality, represented by eigenvectors. \mathcal{L}_{eik} is a regularization term of the gradient of SDF widely used (Gropp et al., 2020). α, β, γ and δ are hyper-parameters.

4 EXPERIMENTS

To evaluate the effectiveness of our method, we tested GNeRP on objects from multiple scenes and compared them with existing state-of-the-art neural 3D reconstruction methods.

PolRef Dataset The methods are evaluated on the PANDORA dataset (Dave et al., 2022) and captured scenes by ourselves. The PANDORA includes 3 reflective objects (Owl, Blackvase, and Gnome) with polarization priors. However, their ground truth shapes are unavailable for quantitative evaluation. Moreover, the diversity of materials, geometry, and illumination is not enough for overall comparisons. Only the geometry of the Gnome scene is complicated but less reflective. Only a mirror-like ball in the Blackvase reflects surroundings other than highlights. Other common datasets, including Shiny Blender (Verbin et al., 2022), lack polarization priors for our method. To comprehensively evaluate the performance of 3D reconstruction methods, a new challenging multi-view dataset named PolRef was collected, consisting of objects with reflective and less-textured surfaces captured with various illumination. Radiance images and aligned polarization priors were captured in one shot using polarization cameras. To obtain precise and complete ground truth shapes, objects were produced using SLA 3D printers, with an accuracy tolerance of $\pm 0.1mm$. Detailed descrip-

tions are shown in the Appendix. The dataset will be released to facilitate further research on 3D reconstruction in more challenging scenes in the future.

Experimental Settings GNeRP is built upon NeuS (Wang et al., 2021). The geometry network and radiance network in Fig. 3 is the same as that of NeuS. Since the covariance loss \mathcal{L}_{cov} in Eq. 9 refines the details of the geometry, it will not be activated during the initial 50K steps. The model is trained for 200k iterations and takes about 6 hours on a server with 4 NVIDIA RTX 3090 Ti GPUs for the reconstruction. After optimization, the meshes are extracted from learned SDF by Marching Cubes (Lorensen & Cline, 1998) with a resolution of 512^3 . The hyper-parameter settings are shown in the Appendix.

4.1 COMPARISON WITH STATE-OF-THE-ART METHODS

We conducted the comparison of reconstruction accuracy between our methods and several state-of-the-art methods, including baseline methods for neural 3D reconstruction (Unisurf (Oechsle et al., 2021), VolSDF (Yariv et al., 2021), and NeuS (Wang et al., 2021)), two view-consistency based methods (NeuralWarp (Darmon et al., 2022) and Geo-NeuS (Fu et al., 2022)), two new methods for reconstruction of reflective objects (NeRO (Liu et al., 2023) and Ref-NeuS (Ge et al., 2023)), and a polarization-based method (PANDORA (Dave et al., 2022)).

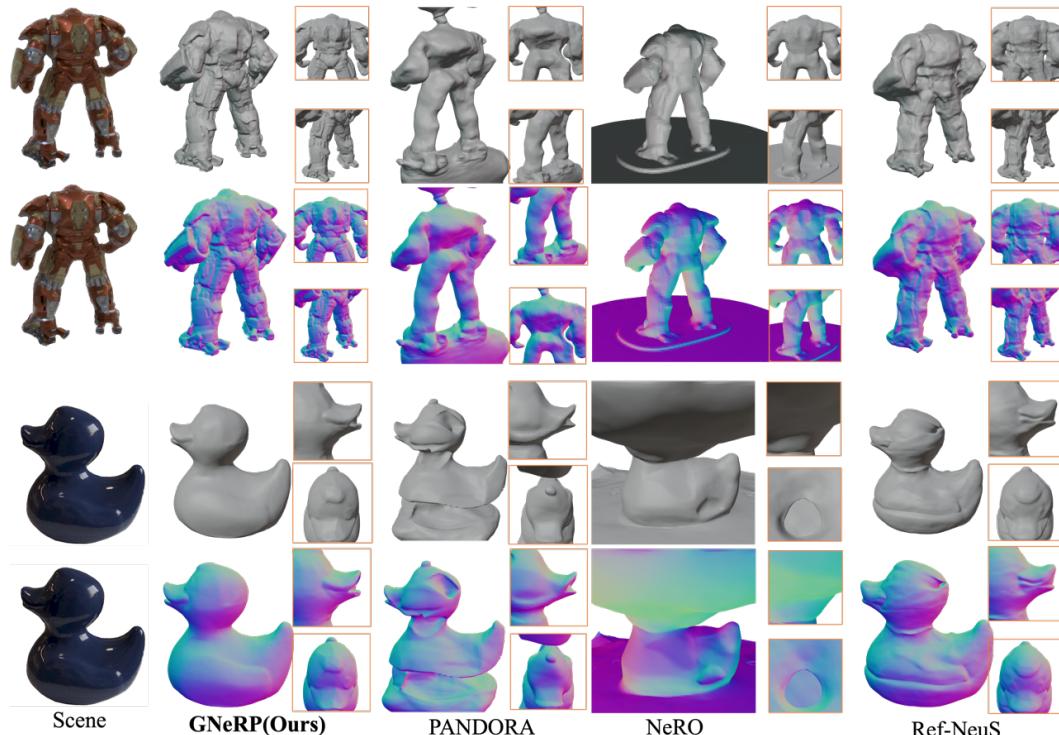


Figure 5: Visual comparison of our method and state-of-the-art methods.

A qualitative comparison between our method and state-of-the-art methods specially designed for reflective objects is shown in Fig. 5, which demonstrates that our method significantly improves the geometry details and accuracy of normals. In the Ironman scene, NeRO reconstructed an over-smoothed geometry. Due to the spatial continuity of neural BRDF, it failed to reconstruct the high-frequency armor details with abrupt normal changes. The shape of Ref-NeuS is more accurate, but the sole scalar SDF is not able to predict the geometry details. The duck scene is more reflective with a combination of highlights and reflection of surroundings. Although Ref-NeuS detected the reflective regions, it was still misled by the environment radiance and reconstructing concave holes. The results of PANDORA are over-smoothed in Ironman and disturbed by noise in polarization priors in Duck. Additional comparisons of different scenes are shown in the Appendix. We conduct

Methods	Ironman	Duck	Cow	Snorlax	Mean
Unisurf* (Oechsle et al., 2021)	3.97	10.83	14.43	14.33	10.89
VolSDF (Yariv et al., 2021)	2.72	5.16	5.95	3.20	4.26
NeuS (Wang et al., 2021)	<u>2.28</u>	<u>2.12</u>	3.82	<u>2.11</u>	<u>2.58</u>
Geo-NeuS (Fu et al., 2022)	4.77	10.12	17.48	10.39	10.82
NeuralWarp (Darmon et al., 2022)	12.44	19.78	5.41	20.57	14.55
NeRO* (Liu et al., 2023)	2.29	23.75	<u>2.95</u>	24.30	13.32
Ref-NeuS (Ge et al., 2023)	<u>1.88</u>	<u>1.93</u>	<u>3.66</u>	<u>1.99</u>	<u>2.34</u>
PANDORA † (Dave et al., 2022)	4.61	5.28	7.96	5.73	5.90
GNeRP	1.34	1.63	1.39	1.05	1.35

Table 1: Quantitative comparison with state-of-the-art methods. The lower is better. * indicates the method doesn't use object masks. † refers to the use of polarization priors. The best scores are **bold**, the second best scores are double underlined, and the third best scores are underlined.

quantitative comparisons on the four scenes with ground truth meshes in our dataset. The evaluation metric is Chamfer Distance according to NeuS (Wang et al., 2021) and Unisurf (Oechsle et al., 2021). Scores are reported in Table 1, which shows our method reconstructs more precise meshes in all four scenes. NeRO (Liu et al., 2023) needs environment information to calculate occlusion loss, and Unisurf also learns occupancy from backgrounds. Training them with masks failed directly, so we report the scores without masks in Tab. 1 denoted by *. Geo-NeuS needs sparse points from Structure-from-Motion (SfM) Schönberger et al. (2016) to calculate SDF loss and select the pairs based on SfM for the warping process. We did the sparse reconstruction in COLMAP and followed the pairs selection method in NeuralWarp. Full polarimetric acquisition (Stokes vector $[s_0, s_1, s_2]$, see in the Appendix) is required by PANDORA. We processed the raw polarization capture to follow its data conventions. Sparse reconstruction of reflective scenes was noisy and incomplete, resulting in the worst accuracy by Geo-NeuS and NeuralWarp. Ref-NeuS demonstrated comparable scores on all scenes, but our method still outperformed.

4.2 ABLATION STUDY

Scene	NeuS	w/ $\mathcal{L}_{\text{mean}}$	w/ \mathcal{L}_{cov}	w/ ReW. $\mathcal{L}_{\text{mean}}$	w/ ReW. \mathcal{L}_{cov}	Full
Snorlax	2.11	2.03	3.01	1.81	2.07	1.05
Cow	3.82	2.72	5.54	1.94	2.29	1.39

Table 2: Ablation Study. $\mathcal{L}_{\text{mean}}, \mathcal{L}_{\text{cov}}$ are in Eq. 9.

To validate the effectiveness of the proposed modules, we test the following three settings as shown in Tab. 2. W/ $\mathcal{L}_{\text{mean}}$ refers to the naive supervision of φ and azimuth angle of normals in SDF. Due to the noise, the results are worse. W/ $\mathcal{L}_{\text{mean}}$ refers to the polarization supervision with only covariance. The reconstruction is focused on details and results in the worst scores. W/ ReW. $\mathcal{L}_{\text{mean}}$ indicates the reweighted losses $(1 - \rho)\mathcal{L}_{\text{color}} + \rho\mathcal{L}_{\text{mean}}$. Similarly, w/ReW. \mathcal{L}_{cov} represents $(1 - \rho)\mathcal{L}_{\text{color}} + \rho\mathcal{L}_{\text{cov}}$. The reweighting does improve the efficiency of polarization priors. Finally, the full setting shows the best scores. Additional visualization is shown in the Appendix.

5 CONCLUSION

We propose GNeRP to reconstruct the detailed geometry of reflective scenes. In GNeRP, we propose a new Gaussian-based representation of normals and introduce polarization priors to supervise it. We propose a DoP reweighing strategy to resolve noise issues in polarization priors. We collect a new, challenging multi-view dataset with non-Lambertian scenes to evaluate existing methods more comprehensively. Experimental results demonstrate the superiority of our method.

REFERENCES

- Marco Ament and Carsten Dachsbacher. Anisotropic ambient volume shading. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):1015–1024, 2015.
- Gary A Atkinson and Edwin R Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Transactions on Image Processing*, 15(6):1653–1664, 2006.
- David C Banks. Illumination in diverse codimensions. In *SIGGRAPH*, pp. 327–334, 1994.
- Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. *ICCV*, 2021.
- J. Berkmann and T. Caelli. Computation of surface geometry and segmentation using covariance techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1114–1116, Jan 1994.
- Christoph Bregler, Aaron Hertzmann, and Henning Biermann. Recovering non-rigid 3d shape from image streams. In *CVPR*, pp. 690–696, 2000.
- Guangcheng Chen, Li He, Yisheng Guan, and Hong Zhang. Perspective phase angle model for polarimetric 3d reconstruction. In *ECCV*, pp. 398–414, 2022.
- Xiaozhi Chen, Huimin Ma, Ji Wan, Bo Li, and Tian Xia. Multi-view 3d object detection network for autonomous driving. In *CVPR*, pp. 1907–1915, 2017.
- Yi Ping Chen and Ming Der Yang. Micro-scale manufacture of 3d printing. In *Applied Mechanics and Materials*, volume 670, pp. 936–941, 2014.
- Edward Collett. Field guide to polarization. In *SPIE*, 2005.
- Robert L. Cook and Kenneth E. Torrance. A reflectance model for computer graphics. In *SIGGRAPH*, 1981.
- Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In *CVPR*, pp. 1558–1567, 2017.
- François Darmon, Bénédicte Basclé, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *CVPR*, pp. 6260–6269, 2022.
- Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. *arXiv preprint arXiv:2203.13458*, 2022.
- James J Foster, Shelby E Temple, Martin J How, Ilse M Daly, Camilla R Sharkey, David Wilby, and Nicholas W Roberts. Polarisation vision: overcoming challenges of working with a property of light we barely see. *The Science of Nature*, 105(3):1–26, 2018.
- Qiancheng Fu, Qingshan Xu, Yew-Soon Ong, and Wenbing Tao. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *NIPS*, 2022.
- Yoshiki Fukao, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino. Polarimetric normal stereo. In *CVPR*, pp. 682–690, 2021.
- Yasutaka Furukawa, Brian Curless, Steven M Seitz, and Richard Szeliski. Towards internet-scale multi-view stereo. In *CVPR*, pp. 1434–1441, 2010.
- Yasutaka Furukawa, Carlos Hernández, et al. Multi-view stereo: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 9(1-2):1–148, 2015.
- Silvano Galliani, Katrin Lasinger, and Konrad Schindler. Massively parallel multiview stereopsis by surface normal diffusion. In *ICCV*, pp. 873–881, 2015.
- Wenhang Ge, Tao Hu, Haoyu Zhao, Shu Liu, and Ying-Cong Chen. Ref-neus: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection. *arXiv preprint arXiv:2303.10840*, 2023.

- Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020.
- Xiaojie Guo, Xiaochun Cao, and Yi Ma. Robust separation of reflection from multiple images. In *CVPR*, pp. 2195–2202, 2014.
- Inseung Hwang, Daniel S. Jeon, Adolfo Muñoz, Diego Gutierrez, Xin Tong, and Min H. Kim. Sparse ellipsometry: Portable acquisition of polarimetric svbrdf and shape with unstructured flash photography. *TOG*, 41(4), 2022.
- Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *ICCV*, pp. 3370–3378, 2015.
- James T Kajiya and Brian P Von Herzen. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3):165–174, 1984.
- Soma Kajiyama, Taihe Piao, Ryo Kawahara, and Takahiro Okabe. Separating partially-polarized diffuse and specular reflection components under unpolarized light sources. In *WACV*, pp. 2549–2558, 2023.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023.
- Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. Nero: Neural geometry and brdf reconstruction of reflective objects from multi-view images. In *SIGGRAPH*, 2023.
- William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pp. 347–353. 1998.
- Nelson Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.
- Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, pp. 4460–4470, 2019.
- 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. In *ECCV*, 2020.
- Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *ICCV*, pp. 5589–5599, 2021.
- Johannes L Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for unstructured multi-view stereo. In *ECCV*, pp. 501–518, 2016.
- Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *ECCV*, 2016.
- Bruce D Smith and Stan H Ward. On the computation of polarization ellipse parameters. *Geophysics*, 39(6):867–869, 1974.
- Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *CVPR*, pp. 5481–5490, 2022.
- Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021.
- Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019.
- Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *NIPS*, 34:4805–4815, 2021.

Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. PhySG: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *CVPR*, 2021.

Jinyu Zhao, Yusuke Monno, and Masatoshi Okutomi. Polarimetric multi-view inverse rendering. In *ECCV*, pp. 85–102, 2020.

Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa volume splatting. In *VIS*, pp. 29–538, 2001.