

RGS: Reflection-aware Gaussian Splatting via Learning Geometry Continuity for Reflective Objects

Xiaobiao Du^{1,3} Yida Wang³ Cheng Bi³ Kun Zhan³ Xin Yu²

¹ University of Technology Sydney ² Adelaide University ³ Li Auto Inc.

Abstract—Gaussian Splatting has significantly improved the quality of novel view synthesis with explicit Gaussian representation. However, we observed that existing 3D Gaussian Splatting methods (3DGS) often suffer from surface collapse issues on reflective regions, and thus produce inferior geometry and low-quality specular. In this work, we propose a physically-based deferred rendering framework, named Reflection-aware Gaussian Splatting (RGS), that can accurately model specular regions and improve novel view synthesis performance. Specifically, we found that a powerful 3D foundation model can provide a strong 3D geometric prior to foster correct geometric modeling. Based on this, we propose a cross-view shape consistency regularization to regularize the geometry surface with the large model prior and cross-view constraints. In this manner, our RGS can produce smoother geometric surfaces on reflective regions while reducing geometric hollows. To further improve rendering results on reflective regions, we present a reflection-aware densification strategy that is designed to capture specular variations across various views. With this strategy, our RGS is able to render novel views of objects in higher quality. Extensive experiments demonstrate our method consistently renders high-quality reflective objects, achieving state-of-the-art performance. **Project Page:** <https://xiaobiaodu.github.io/reflectivegs/>

I. INTRODUCTION

Multi-view stereo (MVS) reconstruction toward 3D scenes and objects has benefited from advancements in volume rendering. In particular, neural radiance field [4] (NeRF) and neural surface [5], [6], [7] (NeuS) employ multi-layer perceptrons (MLP) to implicitly represent the 3D space with high-frequency details. Subsequently, Gaussian Splatting (3DGS) [8] and its 2D variants [9], explicitly represent the scene with learned Gaussian primitives, demonstrating real-time synthesis speed while the high-fidelity rendering [10] is guaranteed. It is demonstrated that high-quality 3D reconstruction can help robotic downstream tasks, such as robotic grasping [11], robotic manipulation [12], SLAM [13], and self-driving [14], [15]. However, reflective object novel view synthesis is very challenging because reflection on a 3D surface is view-dependent, violating cross-view consistency.

Specifically, we found three critical challenges for traditional methods to synthesize high-quality strong-specular objects. (1) We found that the previous methods, such as 3DGS-DR [1], tend to suffer from collapsed geometry on reflective surfaces due to overfitting and specular ambiguity as illustrated in Fig. 1 (d). Consequently, the quality of novel view synthesis is inferior; (2) Traditional vision

foundation models, such as DSINE [2], Depth Pro [3], and DepthAnything [16], fail to estimate normal and depth maps in highly reflective areas, as shown in Fig. 1 (a) and (b). Thus, it is very difficult to obtain accurate geometry in those areas, and this estimated inaccurate geometry cannot assist 3D Gaussians for correct specular mapping; (3) Although existing volume rendering methods [17], [18] can model environment light, they often struggle to capture strong specular reflections, leading to suboptimal rendering quality in reflective regions of novel views. Therefore, these methods fail to reconstruct high-quality reflective objects when the object is illuminated under strong light.

To solve the aforementioned challenges, we propose a physically-based deferred rendering framework, Reflection-aware Gaussian Splatting, named RGS, with three key components. (1) Since we found previous methods that tend to produce geometric hollows, we propose physically-based shading and BRDF modeling within deferred rendering. In this fashion, we can reconstruct geometry at a coarse level; (2) To reconstruct precise geometry, we introduce VGGT [19], a recent powerful 3D foundation model that can predict accurate depth information as shown in Fig. 1 (c). It can provide cross-view geometric consistency constraints for accurate geometry reconstruction. In this way, we can further refine the geometry of objects at a fine level; (3) To better decompose and learn specular information, we propose a reflection-guided densification strategy, enhancing the awareness of 3D Gaussians toward specular. With these enhanced 3D Gaussians, the specular effect can be more easily reconstructed, leading to more accurate novel view synthesis results. In this way, our RGS achieves high-quality geometry continuity and novel view synthesis for reflective objects as shown in Fig. 1 (e).

We conduct extensive experiments to qualitatively and quantitatively validate the efficacy of our proposed RGS. Notably, we visualize the surface geometry, as the normal map, demonstrating that our approach can reliably reconstruct precise and complete surface geometry without structural collapse. This enhancement in terms of geometric accuracy significantly improves the specular novel view synthesis (NVS) for reflective objects. Our method consistently outperforms previous approaches, achieving state-of-the-art results, particularly for reflective objects. In summary, our contributions can be summarized as follows:

- We propose Reflection-aware Gaussian Splatting (RGS)

Corresponding author: xin.yu@adelaide.edu.au

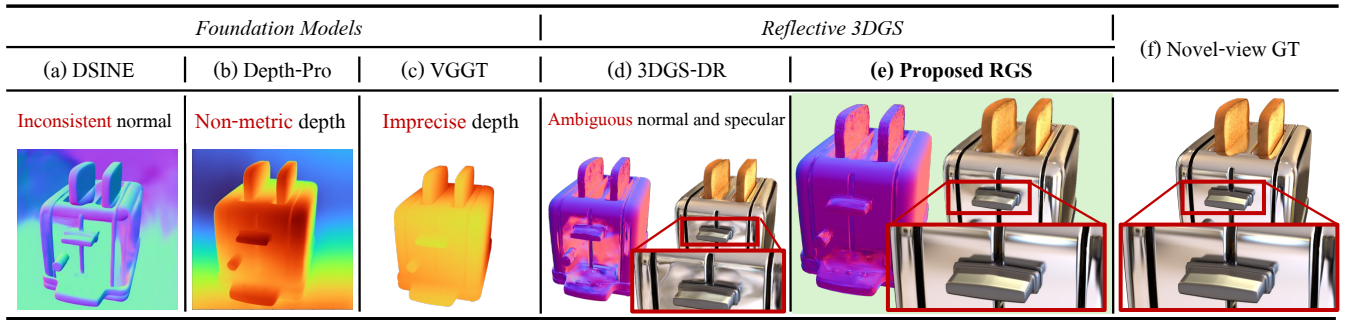


Fig. 1: **Our motivation and proposed method.** Existing state-of-the-art reflective 3DGS approaches, *e.g.* 3DGS-DR [1] in (d), struggle to precisely reconstruct the reflective surface, leading to unsatisfactory specular novel view synthesis. Traditional foundation models [2], [3] cannot estimate the correct normal in (a) and depth in (b) due to view-dependent ambiguity. We propose to use VGGT in (c) with cross-view geometry constraints for accurate reflective surface supervision. Thus, our proposed RGS in (e) is a simple yet effective method to regularize the specular surface for reflective objects. Our RGS achieves high-quality novel-view synthesis compared to its ground truth in (f).

with physically-based shading for reflective object NVS. RGS integrates a cross-view shape regularization strategy, providing cross-view consistency constraints and the 3D foundation model prior to foster accurate specular shading.

- We also propose a reflection-guided densification strategy to enhance the 3D Gaussian primitives that contribute significantly to the rendering of specular, encouraging the learning of specular patterns.
- Extensive experiments in various datasets demonstrate the effectiveness of our method in reconstructing more accurate specular for high-quality novel views synthesis.

II. RELATED WORK

Volume Rendering. Unlike explicit 3D representations [20], pioneering neural approaches such as NeRF [21] and NeuS [18], [22] leverage MLPs with volumetric representations to model scenes implicitly. These methods synthesize high-quality novel views without explicit surface or illumination modeling. Subsequently, more and more works are proposed to improve the efficiency and quality of volume rendering. Mip-NeRF 360 [23] reparameterizes the scene for a more compact representation. Mip-NeRF [24] and Zip-NeRF [25] design a point sampling strategy for faster ray marching. In addition, Depth-NeRF [26] and Monosdf [27] introduce regularization terms to constrain the scene representation, leading to more accurate real-world geometry. Although their reconstruction performance is satisfying, their rendering speed is time-consuming due to the numerous ray sampling processes. To address this problem, multiple works have been proposed for the acceleration of the rendering process with novel scene representation. NSVF [28] proposes a neural sparse voxel field to replace MLP and improve the rendering efficiency. Instant-NGP [29] employs hash tables to speed up the rendering process and improve the reconstruction quality. TensoRF [30] reparameterizes the 4D scene tensor into a more compact low-rank representation with vector-matrices decomposition. Recently, Gaussian Splatting [1], a Gaussian-based explicit representation method, is proposed and achieves faster rendering speed with better

novel view synthesis results. MVGS [31] proposes a multi-view training strategy to improve 3D reconstruction quality and novel view synthesis results. However, these methods are struggling to deal with challenging reflective objects. In this work, we extend the Gaussian representation with reflective attributes and the learnable environmental map for reflective objects.

Inverse Rendering. It aims to decouple geometry, material, and lighting through a physically-based constraint model, given a set of images. The pioneering work, Ref-NeRF [17], achieves realistic specular highlights and reflections, addressing the limitations of NeRF in rendering shiny objects. By leveraging a novel training approach that incorporates both geometry and appearance, Neus [5] effectively captures fine details and complex shapes, enabling realistic surface reconstruction and rendering. By incorporating view-dependent effects directly into the neural implicit representation, Ref-Neus [18] achieves more accurate and visually appealing results for scenes featuring glossy and reflective materials. NeRO [32] utilizes a two-stage method to approximate the shading effects of both direct and indirect lights, achieving pleasing rendering results. GaussianShader [33] proposes a shading function and a normal extraction method for 3D Gaussian Splatting. GS-IR [34] proposes a baking-based occlusion to construct indirect lighting and a depth derivation-based regularization for normal estimation. 3DGS-DR [1] leverages a deferred rendering in pixel space to sample environmental effects and map them into the geometry surface. However, previous methods reconstruct collapsed geometry that considerably limits the mapping of reflection and the learning of the HDR environment map. Our proposed RGS provides a simple and effective geometry-regularized solution to reconstruct more accurate geometry for specular mapping and achieve better inverse rendering.

III. METHODOLOGY

We propose RGS, a novel deferred rendering-based reflection-aware Gaussian Splatting method that incorporates geometry regularization, specifically tailored for rendering reflective objects in novel views as shown in Fig 2. Our

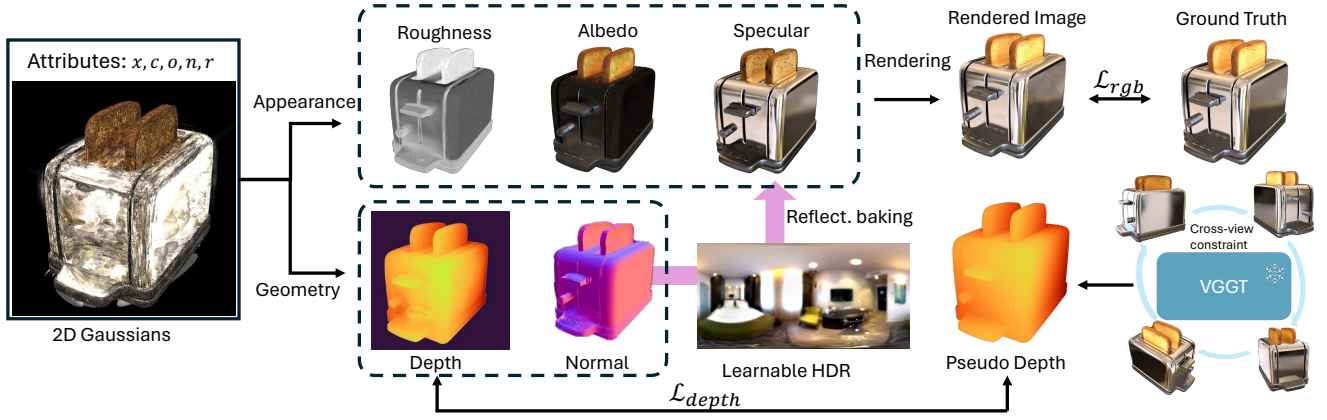


Fig. 2: **Pipeline of our proposed RGS.** Our method with the 2D Gaussian representation can render appearance and geometry maps with a learnable HDR map. We utilize VGGT [19] inherent with cross-view constraints to provide accurate geometry regularization. RGS typically focuses on the specular capture for high-quality novel view synthesis.

approach commences with the 2D Gaussian representation [9] along the ray with direction \mathbf{d}_i passing by N_d 2D Gaussians. Computed by a weighted sum of these N_d kernels, the per-pixel albedo color $\mathbf{C}(\mathbf{d}_i)$ can be obtained by integrating the colors $c_i(\mathbf{d}_i)$ with corresponding per-kernel spherical harmonics. Subsequently, we employ tile-based rasterization [8] to blend the colors c_i sorted by their respective depths z_i . The albedo color and depth can be represented as:

$$\mathbf{C}(\mathbf{d}_i) = \sum_{i=1}^{N_d} w_i c_i(\mathbf{d}_i), \quad \text{and} \quad \mathbf{D}(\mathbf{d}_i) = \sum_{i=1}^{N_d} w_i z_i(\mathbf{d}_i), \quad (1)$$

with the weight $w_i = \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j)$, where the α_i is calculated with Gaussian opacity o_i through:

$$\alpha_i = o_i \exp\left(-\frac{1}{2}(x_i)^T \Sigma_i^{-1}(x_i)\right), \quad (2)$$

where x_i denotes the distance between the Gaussian center and the camera following direction \mathbf{d}_i . In addition to these basic rendering attributes, we also render the normal $\mathbf{N}(\mathbf{d}_i)$ and the roughness map $\mathbf{R}(\mathbf{d}_i)$. Each Gaussian is associated with attributes of a normal n_i and roughness coefficient r_i , which contribute to $\mathbf{N}(\mathbf{d}_i)$ and $\mathbf{R}(\mathbf{d}_i)$ via

$$\mathbf{N}(\mathbf{d}_i) = \sum_{i=1}^{N_d} w_i (n_i + \Delta n_i), \quad \mathbf{R}(\mathbf{d}_i) = \sum_{i=1}^{N_d} w_i r_i(\mathbf{d}_i). \quad (3)$$

n_i is the normal vector defined in 2DGS [9]. Δn_i serves as a regularization term to encourage n_i closer to true normal.

A. Physically-Based Shading and BRDF Modeling

In this work, we utilize the deferred rendering framework to render RGB images following the physically-based shading and BRDF principles. The final rendered RGB image can be formulated as:

$$\begin{aligned} \mathbf{C} &= \mathbf{C}(\mathbf{d}_i) + f_\phi(\mathbf{C}_s(\mathbf{d}_o)), \\ \mathbf{d}_o &= 2 \frac{\langle \mathbf{d}_i, \mathbf{N}(\mathbf{d}_i) \rangle}{\|\mathbf{N}(\mathbf{d}_i)\|} \mathbf{N}(\mathbf{d}_i) - \mathbf{d}_i, \end{aligned} \quad (4)$$

where f_ϕ denotes a shader and $\mathbf{C}_s(\mathbf{d}_o)$ represents the learnable HDR, as outgoing radiance feature by reflection

direction \mathbf{d}_o . $\mathbf{C}_s(\mathbf{d}_o)$ also means the specular feature that can be computed through:

$$\mathbf{C}_s(\mathbf{d}_o) = \int_{\Omega^+} f_r(\mathbf{d}_i, \mathbf{d}_o; \mathbf{R}) \cdot L_i(\mathbf{d}_i) \cdot (\mathbf{d}_i \cdot \mathbf{N}) d\mathbf{d}_i, \quad (5)$$

which aims to compute the integral of reflection and incident light over the normal hemisphere. $f_r(\mathbf{d}_i, \mathbf{d}_o; \mathbf{R})$ means the bidirectional reflectance distribution function (BRDF), which can be represented via:

$$\int_{\Omega^+} f_r(\mathbf{d}_i, \mathbf{d}_o; \mathbf{R}) = \frac{D(\mathbf{h}; \mathbf{R}) \cdot F(\mathbf{d}_i, \mathbf{h}) \cdot G(\mathbf{d}_i, \mathbf{d}_o, \mathbf{h})}{4(\mathbf{n} \cdot \mathbf{d}_i)(\mathbf{n} \cdot \mathbf{d}_o)}, \quad (6)$$

where $\mathbf{h} = \frac{\mathbf{d}_i + \mathbf{d}_o}{\|\mathbf{d}_i + \mathbf{d}_o\|}$ means the half vector. D denotes the microfacet distribution function, controlled by the roughness. F and G represent the Fresnel term, and the shadowing-masking term, respectively. Therefore, the rendering color loss can be formulated as:

$$\mathcal{L}_{\text{rgb}} = \lambda \mathcal{L}_1(\mathbf{C}, \mathbf{C}_{\text{gt}}) + (1 - \lambda) \mathcal{L}_{\text{DSSIM}}(\mathbf{C}, \mathbf{C}_{\text{gt}}), \quad (7)$$

with a pre-defined weight λ balancing \mathcal{L}_1 and $\mathcal{L}_{\text{DSSIM}}$ as mentioned in 3DGS [8]. In addition, we also render Δn_i to obtain its 2D map for later constraint $\Delta \mathbf{N}(\mathbf{d}_i) = \sum_{i=1}^{N_d} w_i \Delta n_i$, where $\Delta \mathbf{N}(\mathbf{d}_i)$ is the 2D map of Δn_i . That can be seen as a residual form of the original normal map. To promote the accuracy of 2D Gaussian normal vectors, we regularize Δn_i through:

$$\mathcal{L}_{\Delta n} = \|\Delta \mathbf{N}(\mathbf{d}_i)\|_2, \quad (8)$$

where it is designed to encourage 2D Gaussian normal for more precise convergence, thereby reducing reliance on the residual normal for guidance.

B. Cross-view Shape Consistency Regularization

To reconstruct fine-grain geometry, we introduce a powerful foundation model, VGGT [19], that possesses a strong geometry prior and cross-view constraint for the estimation of depth. As shown in Fig. 2, we leverage VGGT to estimate the depth given all training views. In this way, we can obtain a pseudo depth map \mathbf{D}' since VGGT is not the metric scale method. Even though these depth images do

not possess physical meaning, they can provide cross-view geometry consistency constraints to regularize Gaussian Splatting. Due to the scale ambiguity, we use the Pearson loss [35], a correlation-based criterion, to regularize the depth map rendered by 2D Gaussians:

$$\mathcal{L}_{\text{depth}}(\mathbf{D}', \mathbf{D}) = 1 - \frac{\text{Cov}(\mathbf{D}', \mathbf{D})}{\sigma(\mathbf{D}')\sigma(\mathbf{D})}, \quad (9)$$

where $\text{Cov}(\cdot)$ calculates covariance of two variables and σ means a variance operator. \mathbf{D} means the depth map rendered by Gaussian Splatting.

In our proposed cross-view depth regularization strategy, the obtained pseudo depth map is estimated from the powerful 3D foundation model, VGGT. Note that this strategy benefits from the multiple views to impose cross-view consistency that can refine geometry at a fine-grain level. In particular, we employ a Pearson loss to regularize the depth map rendered by 2D Gaussians since VGGT cannot estimate real depth. The pseudo depth maps essentially contain accurate depth information, providing effective cross-view consistency information for the regularization of 2D Gaussians. In this way, our RGS can reconstruct more accurate geometry for better specular effect mapping.

Reflection-guided Normal Regularization. In our experiments, we find that reconstructed reflective regions always collapse since Gaussian-based methods are prone to overfitting specular effects and reconstructing cavernous geometry on reflective regions. Therefore, it is necessary to regularize surface geometry for better mapping of the specular effects. Inspired by Total Variation (TV) loss [36], we find that it has a powerful regularization capacity for surface geometry. Consequently, we employ the rendered roughness map \mathbf{R} to pinpoint reflective regions $\hat{\beta}$ through:

$$\hat{\beta} = \begin{cases} 1, & \text{if } \mathbf{R} \leq \mathbf{R}_p \\ 0, & \text{otherwise} \end{cases}, \quad (10)$$

where \mathbf{R}_p means the p -th percentile of \mathbf{R} and $\hat{\beta}$ can be seen as a mask, identifying strongly reflective regions. Empirically, \mathbf{R}_p is enough to capture reflective regions. Therefore, we leverage TV loss to regularize the normal maps \mathbf{N} within strongly reflective regions via:

$$\mathcal{L}_{\text{normal}} = \hat{\beta} \odot \sum_{i,j} \|\mathbf{N}_{i+1,j} - \mathbf{N}_{i,j}\|^2 + \|\mathbf{N}_{i,j+1} - \mathbf{N}_{i,j}\|^2, \quad (11)$$

where \odot denotes Hadamard product. In this way, the surface collapse problem in reflective regions would be largely mitigated, leading to better specular mapping.

C. Reflection-guided Densification

Although some previous methods can reconstruct plausible geometry, they fail to capture strong specular. Specular, as a kind of view-dependent information, is challenging for a limited number of explicit 2D Gaussians to capture its patterns. Inspired by MVGS [31], which leverages a multi-view densification technique to improve the densification effect, we propose a reflection-guided densification strategy

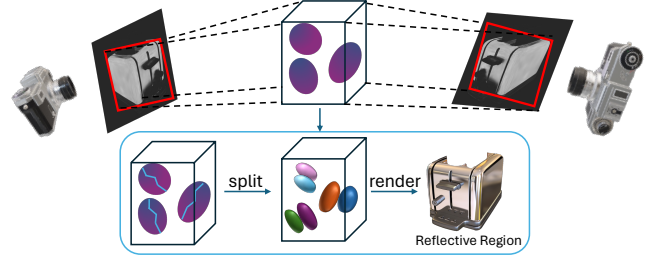


Fig. 3: **Illustration of the proposed reflection-guided densification.** We utilize a multi-view ray tracing technique to cast rays from the vertices of the red rectangle for localizing significant 2D Gaussians and enhancing their densification for better specular pattern capture.

Algorithm 1: Reflection-guided Densification

Input: Multi-view camera parameters $\{\mathcal{C}_v\}_{v=1}^V$, current 3D Gaussians \mathcal{G}
Output: Updated 3D Gaussians \mathcal{G}' with enhanced specular fitting

- 1 **for each** view $v \in \{1, \dots, V\}$ **do**
- 2 Extract specular intensity map $\hat{\beta}$; Compute 2D bounding boxes \mathcal{B}_v on $\hat{\beta}$;
- 3 Emit rays from vertices of \mathcal{B}_v using camera parameters \mathcal{C}_v ;
- 4 **end**
- 5 **for each pair of views** (v_i, v_j) **do**
- 6 Intersected planes formed by rays from \mathcal{B}_{v_i} and \mathcal{B}_{v_j} to localize 3D regions \mathcal{R} ;
- 7 Identify Gaussians $\mathcal{G}_{\text{spec}} \subset \mathcal{G}$ located in regions \mathcal{R} ;
- 8 **end**
- 9 **for each Gaussian** $g \in \mathcal{G}_{\text{spec}}$ **do**
- 10 Decrease densification threshold $\tau_g \leftarrow \tau_g/2$;
- 11 **end**
- 12 Update \mathcal{G}' by performing densification using updated thresholds;
- 13 **return** \mathcal{G}'

to foster 2D Gaussians learning toward reflective 3D regions and facilitate the fitting convergence for solving this problem.

As shown in Fig. 3, we start by locating the 2D Gaussians that contribute significantly to this kind of specular information, and densify these 2D Gaussians. To achieve this, we extract specular intensity maps from roughness maps, which highlight 2D reflective regions. Then we employ a cross-view ray tracing technique, emitting rays from 2D reflective regions and locating significant Gaussians in 3D space. Finally, we enhance the densification effect of these 2D Gaussians.

To be specific, we first calculate the bounding boxes of 2D reflective maps, representing the highly reflective regions of an object, as shown in the red rectangle of Fig. 3. Instead of emitting rays from these reflective regions, we leverage the camera parameters to emit rays in the vertices of bounding boxes per view, avoiding huge computational costs. These rays forming planes would have intersections with others emitted from other views. The intersected planes

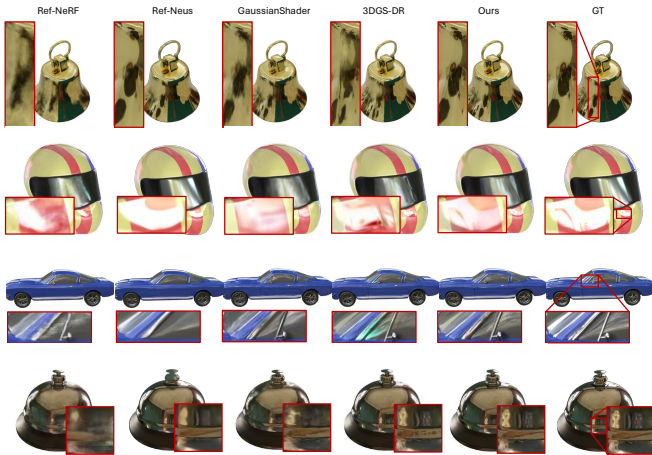


Fig. 4: **Visual comparisons of our method with state-of-the-art methods for novel view synthesis toward challenging reflective objects.** We demonstrate that our method reconstructs specular and reflective surfaces more precisely. We display the close-up for better differentiation.

can locate 3D areas containing a set of important 2D Gaussians contributing significantly to the specular. We enhance these 2D Gaussians’ densification by decreasing their densification threshold by half. In this way, the 2D Gaussians in these regions are easier to learn and capture the specular. It leads to a better fitting process and a more compact structure for 2D Gaussians. Therefore, our RGS can be more effective in capturing specular effects and rendering specular highlights. The proposed reflection-guided densification process is described in the Algorithm 1.

D. Losses and Implementation

Our method is designed in a simple and streamlined manner. In summary, the total loss of RGS can be represented as:

$$\mathcal{L} = \lambda_{\text{rgb}}\mathcal{L}_{\text{rgb}} + \lambda_{\text{depth}}\mathcal{L}_{\text{depth}} + \lambda_{\text{normal}}\mathcal{L}_{\text{normal}} + \lambda_{\Delta n}\mathcal{L}_{\Delta n}, \quad (12)$$

where \mathcal{L}_{rgb} is our main loss function that contributes considerably to the whole learning process. $\mathcal{L}_{\text{normal}}$ and $\mathcal{L}_{\text{depth}}$ are the regularization terms that regularize the surface to be smooth, avoiding overfitting phenomena and collapsed hollows. $\mathcal{L}_{\Delta n}$ is a regularization function that facilitates more accurate learning of the normal vectors from 2D Gaussians.

Implementation. First, we set the weight λ_{rgb} as 1 for the main RGB loss in our experiments. Then, the weights for geometry-regularized losses are set as $\lambda_{\text{normal}} = 1e^{-2}$ and $\lambda_{\text{depth}} = 1e^{-3}$. For the residual normal loss weight, we set $\lambda_{\Delta n} = 1e^{-3}$. To obtain the reflection intensity map, we set the percentile p as 0.9. As for other parameter settings, we follow 3DGS-DR [1] for fair comparisons. We also leverage multi-view training [31] to enhance multi-view constraints and further improve performance. We conduct experiments on an RTX 3090 Ti GPU.

IV. EXPERIMENTS

Datasets and Metrics. To demonstrate the effectiveness of our method to reflective objects and applicability to general objects, we conduct extensive experiments on four datasets

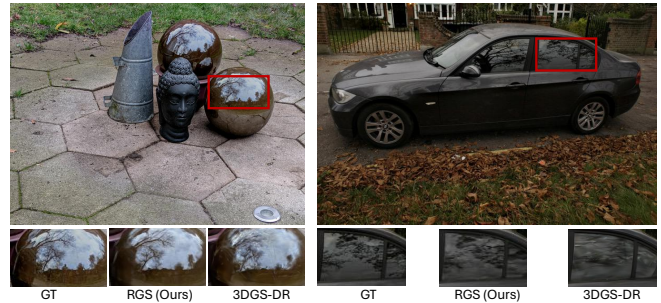


Fig. 5: **Qualitative comparisons of our method with existing state-of-the-art methods on Ref-NeRF Real [17].** We demonstrate that our method not only improves novel view synthesis results of synthetic reflective objects but also real-world reflective objects.

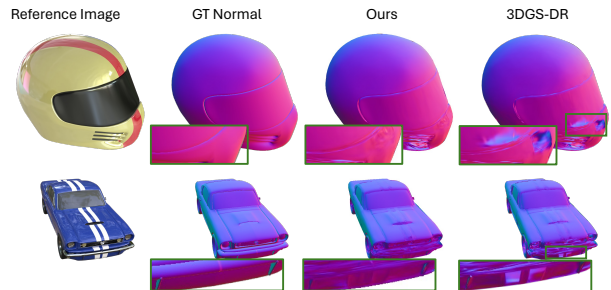


Fig. 6: **Comparisons of rendered normal between our proposed method and 3DGS-DR.** Better normal means more accurate geometry reconstruction.

with 23 scenes, including reflective and general object datasets, such as Shiny Blender [17], Glossy Synthetic [32], Ref-NeRF Real [17], and NeRF Synthetic [4]. For the evaluation metrics, we follow previous works, utilizing PSNR, SSIM, and LPIPS to evaluate the performance of novel view synthesis.

A. Quantitative and Qualitative Results

In our experiments, we evaluate our method for novel view synthesis quantitatively and qualitatively, especially for challenging reflective objects.

Novel View Synthesis. Novel view synthesis measures the performance of a reconstructed 3D model synthesizing various views distinct from training views. In our experiments, we conduct extensive comparisons from reflective objects to general objects, demonstrating the generalization of our method. As shown in Table I, we compare our method with existing state-of-the-art novel view synthesis methods for reflective objects, including Ref-NeRF [17], NPC [37], 3DGS [8], GaussianShader [33], ENVIDR [39], 3DGS-DR [1], Spec-Gaussian [40], 3iGS [38], EnvGS [41], Ref-GS [42], and Ref-Gaussian [43] on three representative datasets, such as Shiny Blender [17], Glossy Synthetic [32], and Ref-NeRF Real [17]. Note that these datasets contain reflective objects with glossy surfaces. Ref-NeRF Real [17] is a real dataset containing reflective objects in real scenes, so the real lighting condition would be more challenging. As we can see from Table I, our proposed RGS achieves the best quantitative results compared with other methods, demonstrating that

TABLE I: **Quantitative comparisons of state-of-the-art reflective object novel view synthesis methods.** We demonstrate our method effectively improves novel view synthesis performance for challenging reflective objects even in real scenes. We display results on Glossy Synthetic datasets [32], Shiny Blender [17], and Ref-NeRF Real [17]. The **best**, **second best**, and **third best** results are highlighted respectively for better visibility.

Dataset	Glossy Synthetic			Shiny Blender			Ref-NeRF-Real		
Method & Metrics	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Ref-NeRF [17]	27.49	0.927	0.100	33.12	0.961	0.079	23.62	0.645	0.238
NPC [37]	21.96	0.841	0.181	27.48	0.921	0.145	22.87	0.605	0.320
3DGS [8]	26.49	0.917	0.092	30.35	0.946	0.083	23.85	0.659	0.230
3iGS [38]	26.39	0.913	0.089	30.64	0.948	0.077	23.95	0.657	0.258
GaussianShader [33]	27.53	0.921	0.086	31.96	0.957	0.067	23.46	0.647	0.257
ENVIDR [39]	29.56	0.952	0.059	33.46	0.967	0.045	23.00	0.605	0.331
Spec-Gaussian [40]	27.54	0.931	0.079	31.00	0.950	0.075	23.71	0.652	0.259
3DGS-DR [1]	30.13	0.953	0.058	34.08	0.971	0.052	23.99	0.658	0.264
EnvGS [41]	29.78	0.952	0.061	33.82	0.968	0.065	24.15	0.659	0.259
Ref-GS [42]	30.59	0.957	0.058	34.80	0.973	0.056	24.21	0.661	0.258
Ref-Gaussian [43]	30.36	0.958	0.051	35.01	0.972	0.056	24.13	0.657	0.258
RGS (Ours)	31.24	0.965	0.043	35.38	0.976	0.050	24.31	0.670	0.257

TABLE II: **Quantitative evaluation of our method compared with previous methods on the NeRF Synthetic dataset [4].** We demonstrate our method not only performs well for reflective objects but also has a good generalization to general objects. GShader* means that we report the results with the official implementation repository of GaussianShader.

Method	NeRF[4]	VolSDF[44]	Ref-NeRF[17]	ENVIDR[39]	3DGS[8]	GShader*[33]	RGS (Ours)
PSNR	31.01	27.96	31.29	28.13	33.30	31.46	33.51
SSIM	0.947	0.932	0.947	0.956	0.969	0.960	0.968
LPIPS	0.081	0.096	0.058	0.067	0.030	0.042	0.028

TABLE III: **Quantitative normal comparisons among different methods on the Shiny Blender dataset [17].** We demonstrate that our method can reconstruct better normal maps for more precise geometry.

Method	ENVIDR	Ref-Gaussian	3DGS-DR	EnvGS	RGS (Ours)
MAE \downarrow	2.74	2.18	2.62	2.66	2.04
SSIM \uparrow	0.948	0.952	0.947	0.945	0.954
LPIPS \downarrow	0.072	0.053	0.065	0.062	0.051

our method achieves state-of-the-art performance. We also display visual comparisons in Fig. 4 and Fig. 5. These visual comparisons demonstrate that our proposed RGS can reconstruct better specular and clearer reflective appearance close to the ground truth. To general objects, we present our results in Table II. Table II demonstrates our method achieving state-of-the-art results compared with other methods. It also indicates our method not only performs well for reflective objects but also for general objects. It is attributed to the proposed RGS focusing on geometry regularization and 2D Gaussian densification for specular capture.

Geometry Measurement. In addition to the RGB renderings, we also show geometry, as the normal map in Fig. 6 and Table III. As we can see in Fig. 6, 3DGS-DR [1] fails to reconstruct the accurate normal. The inaccurate normal leads to the project error of specular and affects the learning of the environment light. In contrast, our proposed

TABLE IV: **Budget comparisons among different methods on Shiny Blender dataset.** We demonstrate that our method with fewer parameters achieves state-of-the-art performance.

Method	3DGS	3DGS-DR	EnvGS	RGS (Ours)
Parameters \downarrow	3.3M	1.1M	2.9M	0.1M
Train. Time \downarrow	0.6h	2h	3h	3h
FPS \uparrow	604	71	21	146
PSNR \uparrow	30.35	34.08	33.82	35.38

RGS achieves high-quality normal reconstruction. We also show the quantitative results in Table III. It objectively demonstrates that our proposed RGS can achieve better normal reconstruction compared with previous methods, in terms of MAE, SSIM, and LPIPS. The accurate normal suggests better geometry. It also indicates that our method can reconstruct more accurate surfaces to achieve better specular mapping for reflective object novel view synthesis.

Budget Comparisons. We also conduct budget comparisons as shown in Table IV. Compared to 3DGS, our method achieves a better trade-off between rendering speed and performance. Note that our method can still reach the real-time rendering speed. Even though our method utilizes the enhanced densification technique, the multi-view constraints encourage the structure of 3D Gaussians to be more compact. Therefore, our method can achieve state-of-the-art results with fewer parameters compared with previous methods.

B. Ablation Study

To demonstrate the effectiveness of the proposed components, we conduct ablation studies on the Shiny Blender Dataset [17]. In particular, we leverage 2DGS [1] integrated with the deferred rendering framework as our baseline and progressively add our proposed components into it. The ablation results are displayed in Fig 7 and Table V. It is obvious that the baseline produces a low-quality normal map with a collapsed surface geometry and also renders an

TABLE V: **Ablation study of our proposed RGS on the Shiny Blender dataset [17].** We demonstrate the effectiveness of the proposed each component.

CSCR	RD	PSNR	SSIM	FPS	Parameters
		33.29	0.967	132	1.4M
✓		34.54	0.974	137	0.4M
	✓	34.63	0.975	143	0.2M
✓	✓	35.38	0.976	146	0.1M

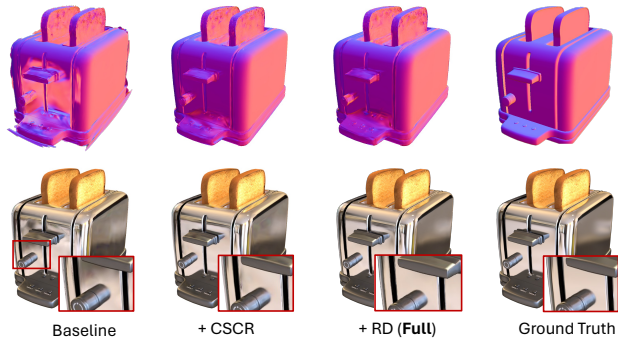


Fig. 7: **Ablation studies of our proposed components in RGS.** We show qualitative and quantitative results to fully demonstrate the effectiveness of RGS. CSCR represents the proposed cross-view shape consistency regularization. RD denotes the proposed reflection-guided densification. We use red rectangles to highlight the details for better differentiation.

inferior novel view. On the contrary, when we integrate our proposed cross-view shape consistency regularization (CSCR) into it, the normal map becomes smooth without collapse, and the novel view is approaching the Ground Truth. It is attributed to the proposed CSCR exploiting the foundation model prior and cross-view constraints to regularize the fitting process of 2DGS. Finally, when we embed our reflection-guided densification (RD) into the method, our method is completed and achieves high-quality results compared with previous methods. It makes the specular learning process easier with the ray-based densification strategy. In particular, the normal map is very similar to Ground Truth, which indicates that our method achieves highly accurate geometry. These ablation studies demonstrate the effectiveness of the proposed components.

Decomposition Results. To further analyze our method, we visualize our rendering attributes, including albedo, roughness, specular, depth, normal, and the final image as shown in Fig. 8. As we can see from the top of Fig. 8, our RGS can successfully decouple albedo and specular, which facilitates for the learning and mapping of the environmental lighting effect. In addition, the bottom of Fig. 8 demonstrates that rendered depth and normal maps are accurate in representing the learned geometry under our proposed restriction.

Foundation Model Assistance. We also conduct an ablation study in Fig. 9 to demonstrate the necessity of the involvement of the foundation model, VGGT. As displayed in Fig. 9, when the depth assistance supervision is involved, our

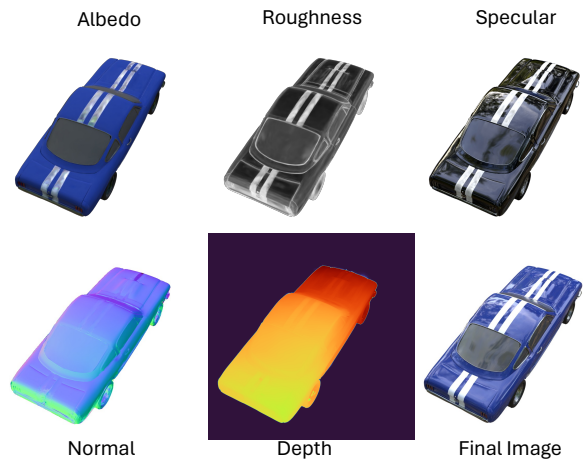


Fig. 8: **Decomposition results of our proposed RGS.** The final image is the combination of albedo and specular maps.



Fig. 9: **Ablation study of foundation model assistance.** We demonstrate that VGGT can provide helpful geometry supervision for Gaussian Splatting to improve performance.

RGS can reconstruct more accurate geometry for specular mapping and improve novel view synthesis performance. It is attributed to the accurately estimated depth from VGGT, which contains powerful cross-view constraints and the foundation model prior. These results highlight the potential of foundation models to improve geometry reconstruction, providing stronger constraints for training 3D Gaussians and enabling high-quality novel view synthesis.

V. CONCLUSION

In conclusion, existing Gaussian-based methods face challenges in rendering glossy objects due to specular ambiguity across camera views. In this work, we proposed a reflection-aware Gaussian Splatting method, named RGS, with a 3D foundation model prior to regularize the geometry for high-quality specular novel view synthesis. We first propose a cross-view shape consistency regularization strategy that exploits VGGT to impose cross-view consistency constraints for the refinement of the geometry. To better capture specular patterns, we propose a reflection-guided densification strategy that enhances specular capture. Our method achieves state-of-the-art novel view synthesis for reflective objects in terms of quantitative metrics and qualitative evaluation.

Acknowledgments This research is funded in part by ARC-Discovery grant (DP220100800 to XY) and ARC-DECRA grant (DE230100477 to XY). We thank all anonymous reviewers and ACs for their constructive suggestions.

REFERENCES

- [1] K. Ye, Q. Hou, and K. Zhou, “3d gaussian splatting with deferred reflection,” *arXiv preprint arXiv:2404.18454*, 2024.
- [2] G. Bae and A. J. Davison, “Rethinking inductive biases for surface normal estimation,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [3] A. Bochkovskii, A. Delaunoy, H. Germain, M. Santos, Y. Zhou, S. R. Richter, and V. Koltun, “Depth pro: Sharp monocular metric depth in less than a second,” *arXiv preprint arXiv:2410.02073*, 2024.
- [4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [5] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, “Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction,” *arXiv preprint arXiv:2106.10689*, 2021.
- [6] Y. Wang, Q. Han, M. Habermann, K. Daniilidis, C. Theobalt, and L. Liu, “Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3295–3306, 2023.
- [7] Y. Wang, I. Skorokhodov, I. Skorokhodov, and P. Wonka, “Hf-neus: Improved surface reconstruction using high-frequency details,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 1966–1978, 2022.
- [8] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3d gaussian splatting for real-time radiance field rendering,” *ACM Transactions on Graphics (ToG)*, vol. 42, no. 4, pp. 1–14, 2023.
- [9] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao, “2d gaussian splatting for geometrically accurate radiance fields,” in *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–11, 2024.
- [10] Z. Yu, A. Chen, B. Huang, T. Sattler, and A. Geiger, “Mip-splatting: Alias-free 3d gaussian splatting,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19447–19456, June 2024.
- [11] S. Iwase, M. Z. Irshad, K. Liu, V. Guizilini, R. Lee, T. Ikeda, A. Amma, K. Nishiwaki, K. Kitani, R. Ambrus, *et al.*, “ZeroGrasp: Zero-shot shape reconstruction enabled robotic grasping,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17405–17415, 2025.
- [12] W. Cui, C. Zhao, Y. Chen, H. Li, Z. Zhang, D. Zhao, and H. Wang, “Cl3r: 3d reconstruction and contrastive learning for enhanced robotic manipulation representations,” *arXiv preprint arXiv:2507.08262*, 2025.
- [13] H. Matsuki, R. Murai, P. H. Kelly, and A. J. Davison, “Gaussian splatting slam,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18039–18048, 2024.
- [14] X. Du, H. Sun, M. Lu, T. Zhu, and X. Yu, “Dreamcar: Leveraging car-specific prior for in-the-wild 3d car reconstruction,” *arXiv preprint arXiv:2407.16988*, 2024.
- [15] X. Du, H. Sun, S. Wang, Z. Wu, H. Sheng, J. Ying, M. Lu, T. Zhu, K. Zhan, and X. Yu, “3drealcar: An in-the-wild rgb-d car dataset with 360-degree views,” *arXiv preprint arXiv:2406.04875*, 2024.
- [16] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, “Depth anything: Unleashing the power of large-scale unlabeled data,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10371–10381, 2024.
- [17] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, “Ref-nerf: Structured view-dependent appearance for neural radiance fields,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5481–5490, IEEE, 2022.
- [18] W. Ge, T. Hu, H. Zhao, S. Liu, and Y.-C. Chen, “Ref-neus: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4251–4260, 2023.
- [19] J. Wang, M. Chen, N. Karaev, A. Vedaldi, C. Rupprecht, and D. Novotny, “Vggt: Visual geometry grounded transformer,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5294–5306, 2025.
- [20] Y. Wang, D. J. Tan, N. Navab, and F. Tombari, “Self-supervised latent space optimization with nebula variational coding,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, p. 1397–1411, Mar. 2024.
- [21] P. Wang, Y. Liu, Z. Chen, L. Liu, Z. Liu, T. Komura, C. Theobalt, and W. Wang, “F2-nerf: Fast neural radiance field training with free camera trajectories,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4150–4159, 2023.
- [22] Y. Wang, X. Zhang, K. Zhan, P. Jia, and X. Lang, “Hineus: High-fidelity neural surface mitigating low-texture and reflective ambiguity,” *arXiv preprint arXiv:2506.23854*, 2025.
- [23] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, “Mip-nerf 360: Unbounded anti-aliased neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5470–5479, 2022.
- [24] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, “Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5855–5864, 2021.
- [25] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, “Zip-nerf: Anti-aliased grid-based neural radiance fields,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19697–19705, 2023.
- [26] K. Deng, A. Liu, J.-Y. Zhu, and D. Ramanan, “Depth-supervised nerf: Fewer views and faster training for free,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12882–12891, 2022.
- [27] Z. Yu, S. Peng, M. Niemeyer, T. Sattler, and A. Geiger, “Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction,” *Advances in neural information processing systems*, vol. 35, pp. 25018–25032, 2022.
- [28] L. Liu, J. Gu, K. Zaw Lin, T.-S. Chua, and C. Theobalt, “Neural sparse voxel fields,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 15651–15663, 2020.
- [29] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *arXiv:2201.05989*, Jan. 2022.
- [30] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, “Tensorf: Tensorial radiance fields,” *arXiv preprint arXiv:2203.09517*, 2022.
- [31] X. Du, Y. Wang, and X. Yu, “Mvgs: Multi-view regulated gaussian splatting for novel view synthesis,” 2026.
- [32] Y. Liu, P. Wang, C. Lin, X. Long, J. Wang, L. Liu, T. Komura, and W. Wang, “Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images,” in *SIGGRAPH*, 2023.
- [33] Y. Jiang, J. Tu, Y. Liu, X. Gao, X. Long, W. Wang, and Y. Ma, “Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5322–5332, 2024.
- [34] Z. Liang, Q. Zhang, Y. Feng, Y. Shan, and K. Jia, “Gs-ir: 3d gaussian splatting for inverse rendering,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21644–21653, 2024.
- [35] J. Benesty, J. Chen, Y. Huang, and I. Cohen, “Pearson correlation coefficient,” in *Noise reduction in speech processing*, pp. 1–4, Springer, 2009.
- [36] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: nonlinear phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [37] G. Kopanas, T. Leimkühler, G. Rainer, C. Jambon, and G. Drettakis, “Neural point catacaustics for novel-view synthesis of reflections,” *ACM Transactions on Graphics (TOG)*, vol. 41, no. 6, pp. 1–15, 2022.
- [38] Z. J. Tang and T.-J. Cham, “3igs: Factorised tensorial illumination for 3d gaussian splatting,” in *European Conference on Computer Vision*, pp. 143–159, Springer, 2024.
- [39] R. Liang, H. Chen, C. Li, F. Chen, S. Panneer, and N. Vijaykumar, “Envirdr: Implicit differentiable renderer with neural environment lighting,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 79–89, 2023.
- [40] Z. Yang, X. Gao, Y.-T. Sun, Y. Huang, X. Lyu, W. Zhou, S. Jiao, X. Qi, and X. Jin, “Spec-gaussian: Anisotropic view-dependent appearance for 3d gaussian splatting,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 61192–61216, 2024.
- [41] T. Xie, X. Chen, Z. Xu, Y. Xie, Y. Jin, Y. Shen, S. Peng, H. Bao, and X. Zhou, “Envgs: Modeling view-dependent appearance with environment gaussian,” *arXiv preprint arXiv:2412.15215*, 2024.
- [42] Y. Zhang, A. Chen, Y. Wan, Z. Song, J. Yu, Y. Luo, and W. Yang, “Ref-gs: Directional factorization for 2d gaussian splatting,” *arXiv preprint arXiv:2412.00905*, 2024.
- [43] Y. Yao, Z. Zeng, C. Gu, X. Zhu, and L. Zhang, “Reflective gaussian splatting,” *arXiv preprint arXiv:2412.19282*, 2024.
- [44] L. Yariv, J. Gu, Y. Kasten, and Y. Lipman, “Volume rendering of neural implicit surfaces,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 4805–4815, 2021.