Semantic-Guided Gaussian Splatting with Deferred Rendering

Nan Wang, Xiaohan Yan, Xiaowei Song, Zhicheng Wang

CAD Research Center, Tongji University, China

ABSTRACT

Revolutionizing novel view synthesis, 3D Gaussian Splatting has unlocked new horizons in 3D visual representation. Despite the efficiency and impressive rendering capabilities of GS, the accurate inverse rendering of reflective non-Lambertian surfaces remains a significant challenge, particularly in the context of diverse reflective materials settings, leading to inconsistent renderings and undermining the technology's potential in applications ranging from digital asserts production to virtual reality. We propose Semantic-Guided Gaussian Splatting (SGGS), which aims to address this challenge by leveraging the capabilities of semantic features derived from cutting-edge 2D foundation models, revolutionizing material properties optimization for Gaussians. By integrating this high-level understanding, we enhance the model's resilience against reflective surfaces and significantly improve multi-view consistency, which is a crucial step towards seamless immersive experiences. Our experiments systematically demonstrate that SGGS outperforms previous methods in terms of both rendering quality and geometry.

OBJECTIVES

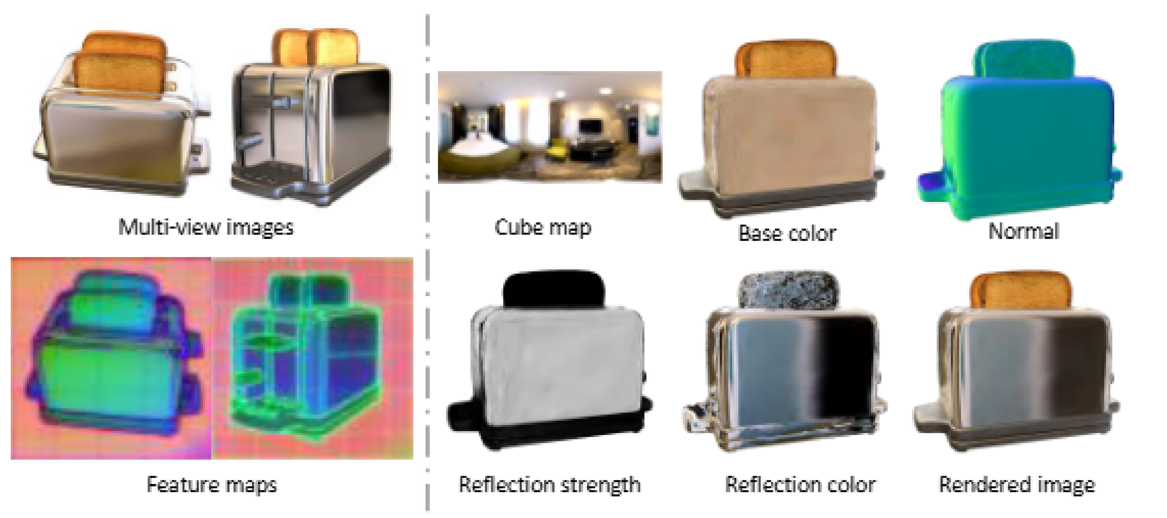


Fig. 1. Our model initiates ingesting multi-view images in conjunction with their corresponding feature maps and a trainable environment map. Ultimately, it constructs a Gaussian splatting representation that decouples geometry, texture and environmental lighting.

Rendering equation

$$C = \sum_{i \in \mathcal{N}} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (2)$$

$$C_d(v) = (1 - R(v))C(v) + R(v)E(\frac{2v \cdot N(v)N(v)}{\|N(v)\|} - v) \quad (16)$$

Regularization Terms

$$\mathcal{L}_{semantic} = \left\| F_t(I) - \theta_{decoder}(F_s(\hat{I}) \right\|_1 \quad (6)$$

$$F_s(\hat{I}) = rasterize(\theta_{encoder}(f_i)), i \in \mathcal{N} \quad (7)$$

$$\mathcal{L}_{geo} = \frac{1}{V} \sum_{p_r \in V} \|M(p_r) \odot (p_r - H_{sr} H_{rs} p_r)\| \quad (11)$$

$$M(p_r) = [\|D(p_r) - D(p_r - H_{sr} H_{rs} p_r)\| < \tau] \quad (12)$$

$$\mathcal{L}_{normal} = \frac{1}{W} \sum_{p \in W} |\overline{\nabla I}|^5 \nu(N_d(p), N(p)) \quad (15)$$

METHODS

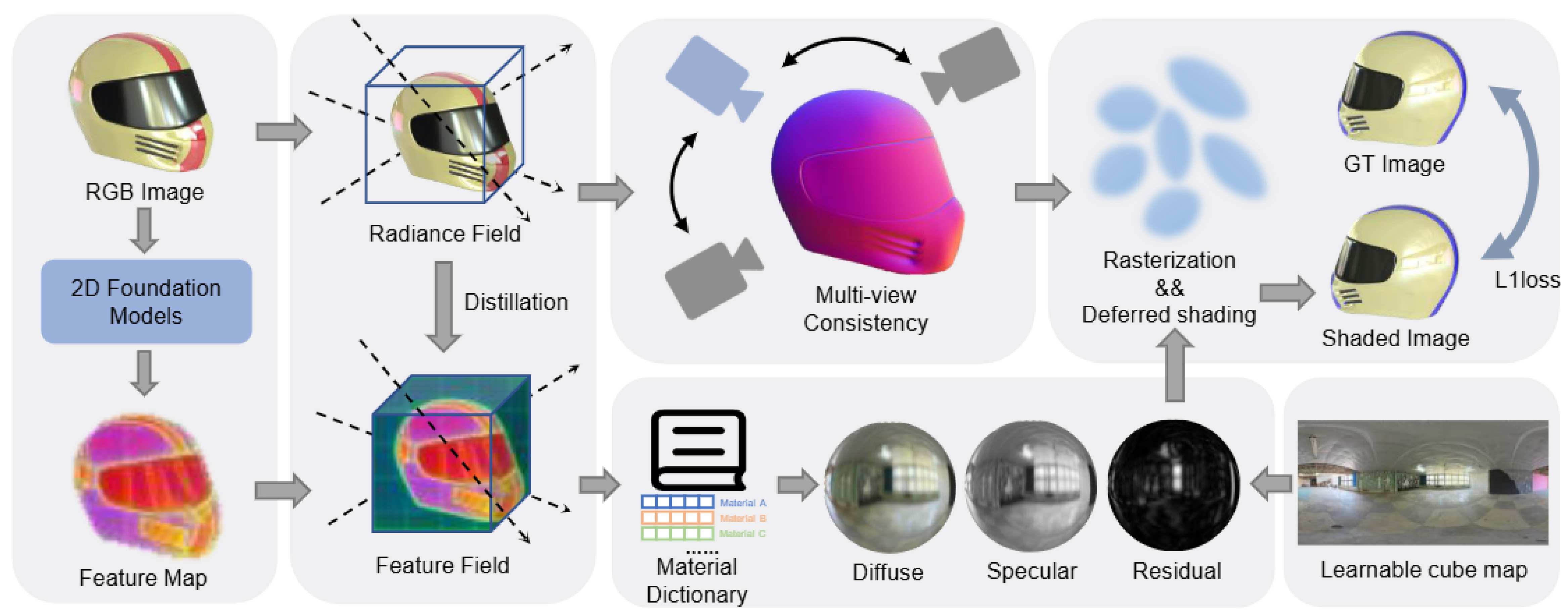


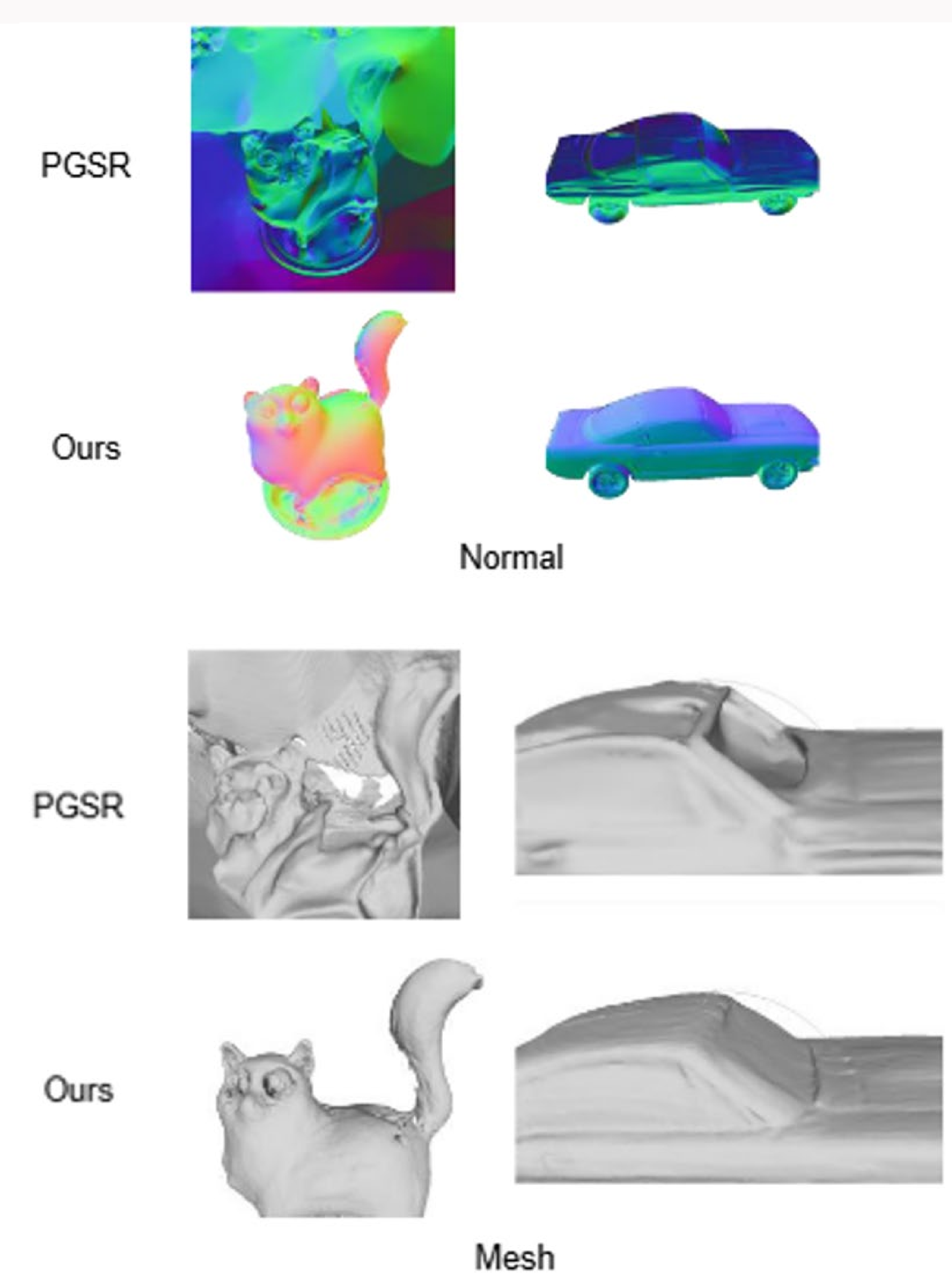
Fig. 2. **Overview of our method.** Our model begins by feeding source images into the LSeg encoder[24], which then produces corresponding semantic maps. Following this, we adopt distillation techniques to construct a feature field. Subsequently, we establish a material dictionary that transforms this feature field into material attributes for every individual Gaussian. These attributes, combined with a learnable environment map and the scene geometry after consistency constraints, facilitate splatting and deferred shading procedure that ultimately yields the final rendered image.

RESULTS

Our experiments systematically demonstrate that SGGS outperforms previous methods in terms of both rendering quality and geometry.

TABLE I
RENDERING PERFORMANCE COMPARISON ON SHINY BLENDER.

	Shiny Blender[22]					
	Car	Ball	Helmet	Teapot	Toaster	Coffee
PSNR↑						
Ref-NeRF[22]	30.41	29.14	29.92	45.19	25.29	33.99
NeRO[23]	25.53	30.26	29.20	38.70	26.46	28.89
3DGS[15]	27.24	27.69	28.32	45.68	20.99	32.32
GShader[17]	27.90	30.98	28.32	45.86	26.21	32.39
3DGS-DR[21]	30.39	33.66	31.69	47.12	27.02	34.65
Ours	**30.55**	**33.81**	**32.01**	**47.13**	**27.34**	**34.85**
SSIM↑						
Ref-NeRF[22]	0.949	0.956	0.955	0.995	0.910	0.972
NeRO [23]	0.949	0.974	0.971	0.995	0.929	0.956
3DGS[15]	0.930	0.937	0.951	0.996	0.895	0.971
GShader[17]	0.930	0.937	0.951	0.996	0.895	0.971
3DGS-DR[21]	**0.962**	0.979	0.971	**0.997**	0.943	0.976
Ours	0.961	**0.979**	**0.975**	0.995	**0.951**	**0.978**
LPIPS↓						
Ref-NeRF[22]	0.051	0.307	0.087	0.013	0.118	0.082
NeRO[23]	0.074	0.094	0.050	0.012	0.089	0.110
3DGS[15]	0.047	0.161	0.079	0.007	0.126	0.078
GShader[17]	0.045	0.121	0.076	0.007	0.079	0.078
3DGS-DR[21]	**0.033**	**0.098**	0.049	**0.005**	0.081	0.076
Ours	0.034	0.111	**0.046**	0.007	**0.067**	**0.050**



ABLATION & OTHER EXPERIMENTS

Geometry evaluation

TABLE II
RECONSTRUCTION PERFORMANCE COMPARISON ON GLOSSY SYNTHETIC.

	Glossy Synthetic[23]					
	bell	cat	luyu	potion	tbell	teapot
chamfer distance↓						
PGSR[20]	20.23	6.00	6.76	5.32	7.48	5.69
Ours	**3.12**	**2.66**	**2.42**	**1.27**	**1.94**	**2.31**

Ablation study

TABLE III
ABLATION STUDY OF SGGS.

Model Setting	PNSR↑	chamfer distance↓
w/o $\mathcal{L}_{semantic}$	32.02	1.85
w/o $\mathcal{L}_{mvs}$	33.62	1.96
w/o $\mathcal{L}_{normal}$	34.11	1.94
Full Model	34.28	1.81

Convergence speed

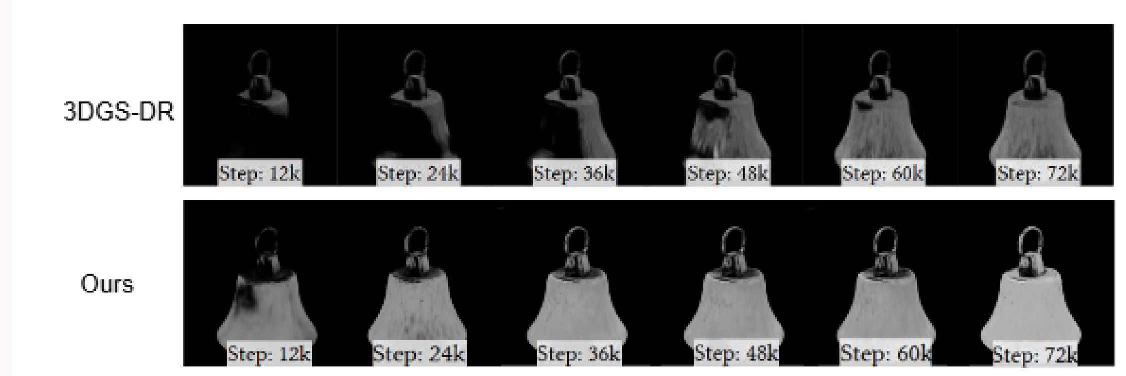


Fig. 3. Our method exhibits a rapid convergence rate, which is achieved through its sophisticated semantic comprehension of scenes.

Contributions

- Semantic-Guided Gaussian Splatting

Our method enhances 3D Gaussian Splatting by incorporating high-level understanding, derived from 2D foundation models.

- Reflective Surface Rendering

Our method addresses the difficulty in accurately rendering reflective, non-Lambertian surfaces.

- Multi-View Consistency

SGGS enhances multi-view consistency, which is important for immersive experiences.

- 2D Foundation Model Integration

SGGS leverages the capabilities of cutting-edge 2D foundation models to provide semantic guidance.

- Material Properties Optimization

SGGS revolutionizes the optimization of material properties of Gaussians, leading to more accurate and realistic rendering.