香港中文大學
The Chinese University of Hong Kong

NEURAL INFORMATION PROCESSING SYSTEMS

無人系統研究團隊
Unmanned Systems Research Group

FHGS: Feature-Homogenized Gaussian Splatting

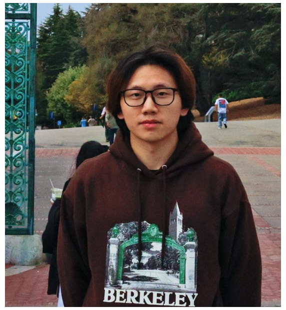
——Resolving the Anisotropic Contradiction in 3D Semantic Fields with Physics-Inspired Non-Differentiable Feature Guidance

December 2025

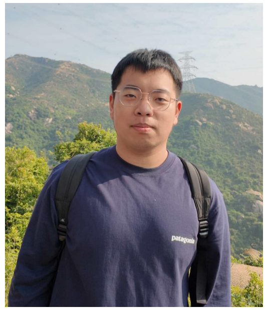
Department of Mechanical and Automation Engineering

The Chinese University of Hong Kong

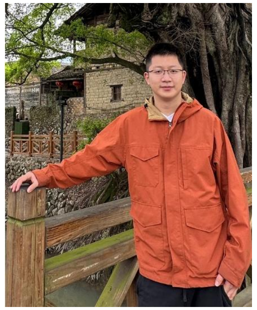
Q. G. Duan



Benyun Zhao



Mingqiao Han



Yijun Huang

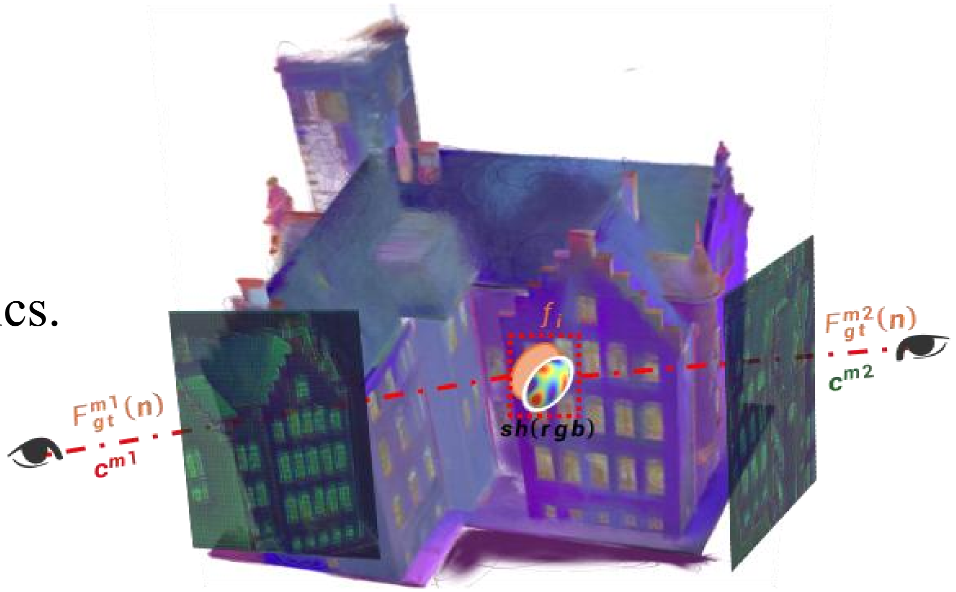
Ben M. Chen

Background

- 3D Gaussian Splatting enables real-time, high-fidelity 3D scene rendering, while ***2D-Vision-Large-Models*** (e.g., SAM, CLIP, DINOv3) excel at image-level semantic understanding.
- Robot tasks (e.g., VLN and task planning) rely on expressive ***3D scene representation*** that supports reliable semantic understanding.
- Methods like PointNet require costly dataset creation and training, whereas ***2D-Vision-Large-Models*** are already mature.
- ***Gaussian Splatting*** provides a ***differentiable*** 3D framework, enabling semantic distillation from 2D to 3D.

Motivation

- **Fundamental conflict**: *anisotropic* color vs. *isotropic* semantics.
- **Efficiency bottleneck**: slow multi-view optimization.
- **Feature distortion**: degradation of pretrained knowledge.



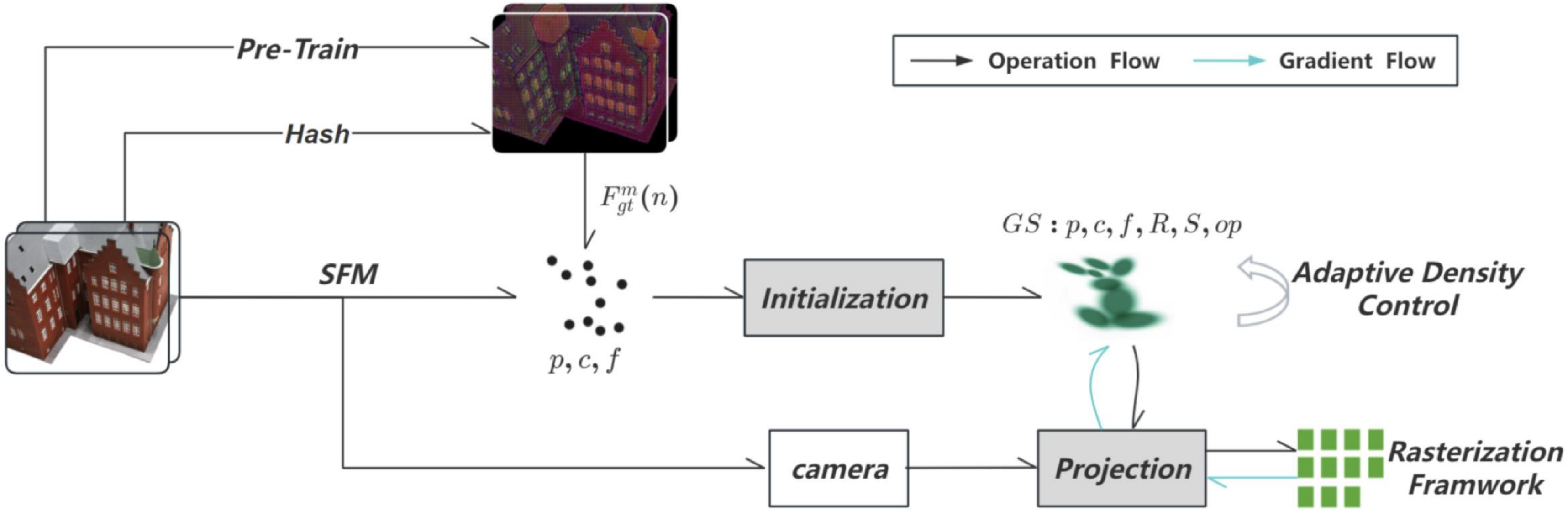
Feature Field: $F^{m1}_{gt}(n) = F^{m2}_{gt}(n) = f_i$

RGB Field: $c^{m1} \neq c^{m2}$

The contradiction between the ***anisotropic*** color of gaussian primitives in RGB field and the ***isotropic*** requirement of semantic features

Our Method: Feature-Homogenized Gaussian Splatting (FHGS)

To resolve the anisotropy-isotropy conflict, we propose FHGS, a framework built on three core innovations:



Pipeline of the General Feature Fusion Architecture

- **Universal Fusion Architecture:** Fusing semantic features from 2D models into 3D Gaussians.
- **Non-Differentiable Feature Driving (NDFD):** Using feature loss to exclusively optimize Gaussian geometry and opacity.
- **Physics-Inspired Dual-Drive Optimization:** Applying a dual-force loss for both global alignment and local coherence.

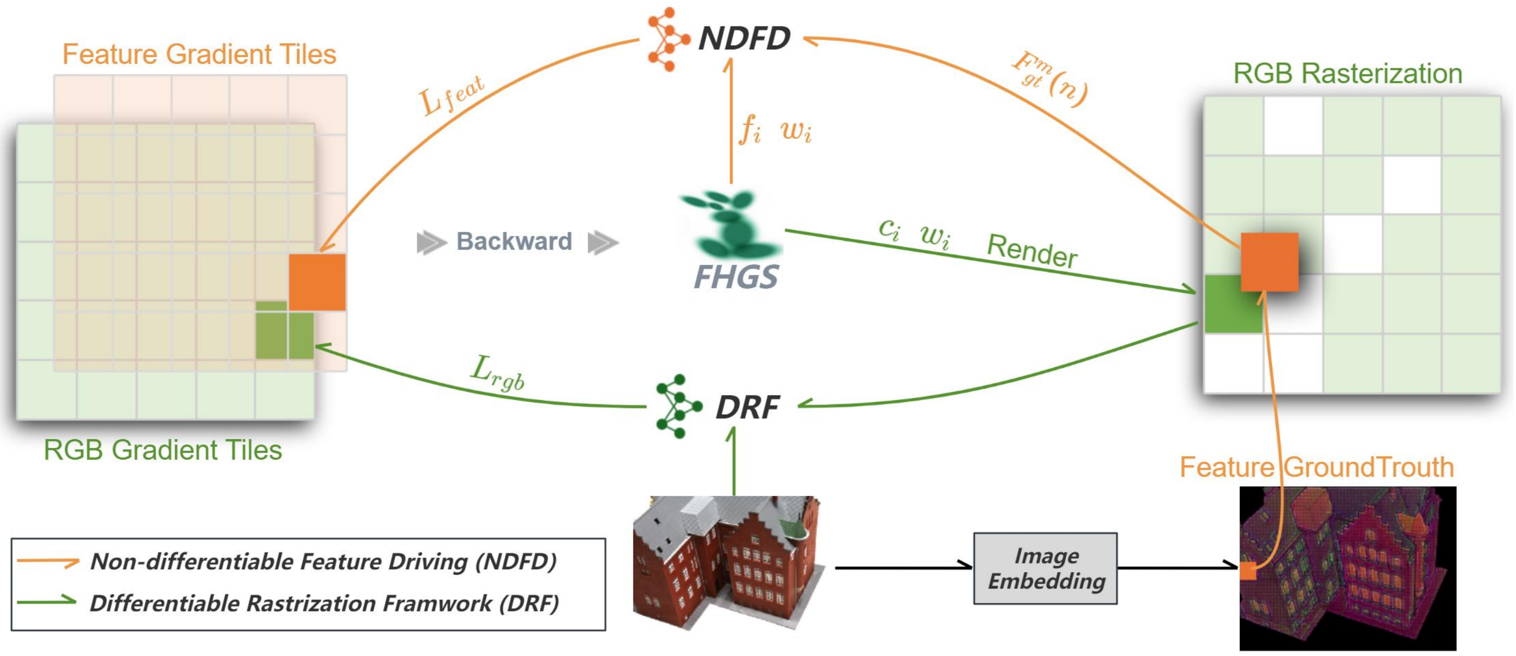
➢ **Differentiable Rasterization Framework (DRF - Green Path):**

❖ Optimizes **Color** (c_i) using a standard, differentiable renderer and RGB loss.

➢ **Non-Differentiable Feature Driving (NDFD - Orange Path):**

❖ Optimizes ***Geometry & Opacity*** (f_i) using a direct loss in the feature space.

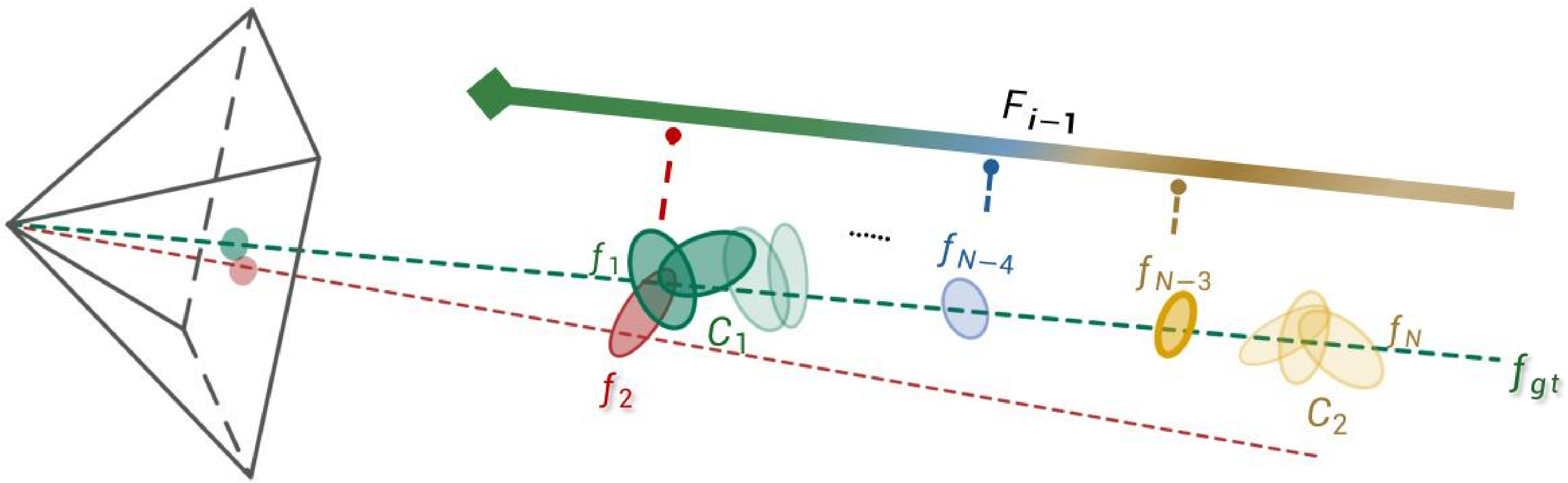
❖ This path **bypasses the renderer**, using abstract features to guide the scene's structure.



Schematic representation of the two mechanisms of FHGS: NDFD and DRF

This separation leads to more stable and meaningful geometry optimization, driven by high-level features, while the appearance is refined for photorealism.

Dual-Drive Mechanism



The illustration of proposed Dual-Drive Mechanism

$$Loss_{gt} = \sum_{i=1}^{N} \sigma_i w_i$$

Featrue-Field–Driven

$$Loss_{cf} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \left(1 - cos\langle f_i \cdot f_j \rangle\right)$$

Intercharge Force(F_{i-1}*)*

(implemented as the accumulated force exerted by all charges on the current charge)

- Inspired by electric fields driving charges
- Feature-field driving + principle of "***like attracts, unlike repels***"
- Minimize similarity error to obtain a multi-view–stable semantic feature field

Ablation Study

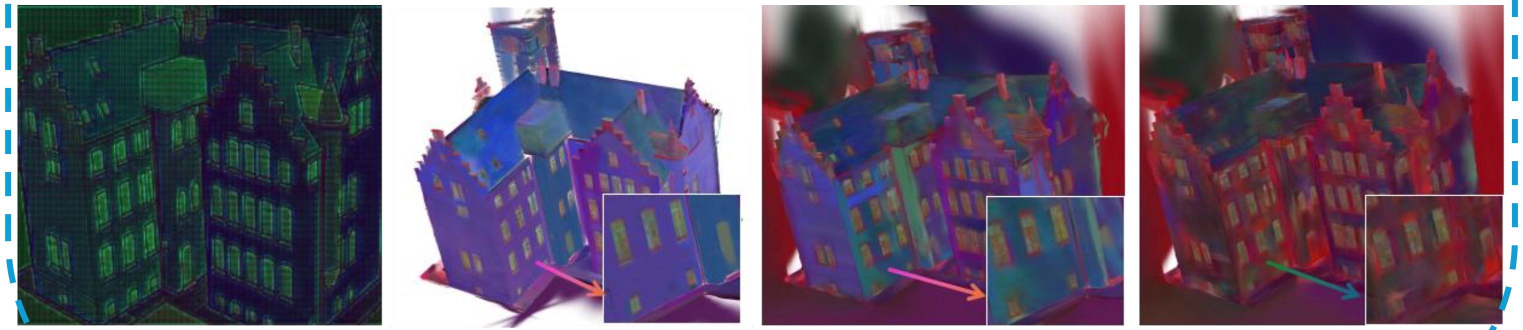
Ablation on Feature Reconstruction

(a) Ground Truth: The ***target high-fidelity*** features.

(b) Ours (FHGS): Reconstructs ***sharp & accurate*** features, closely matching the ground truth.

(c) Ablation (w/o L_cf): ***Performance degrades*** without our feature consistency loss, proving its importance.

(d) Baseline (2DGS): Produces ***blurry and inaccurate*** features.

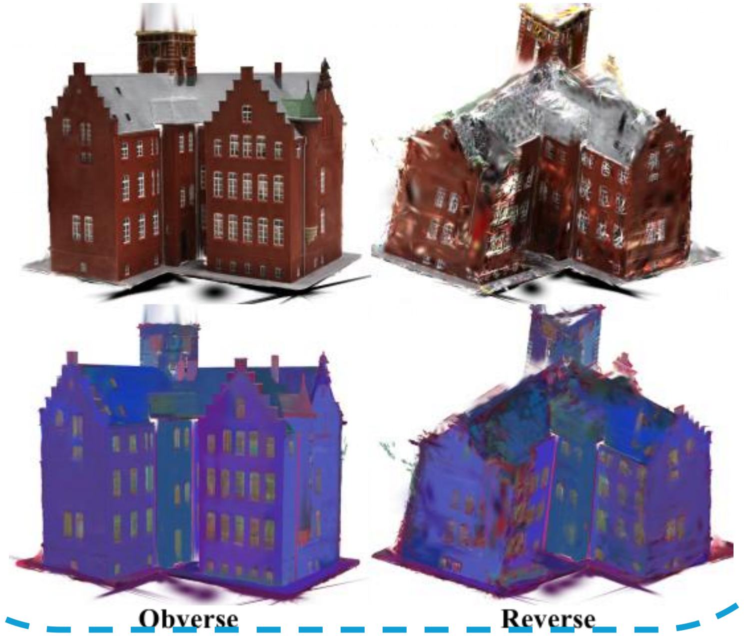


(a) Feature ground truth (b) FHGS (**ours**) (c) FHGS (w/o L_{cf}) (d) 2DGS (baseline)

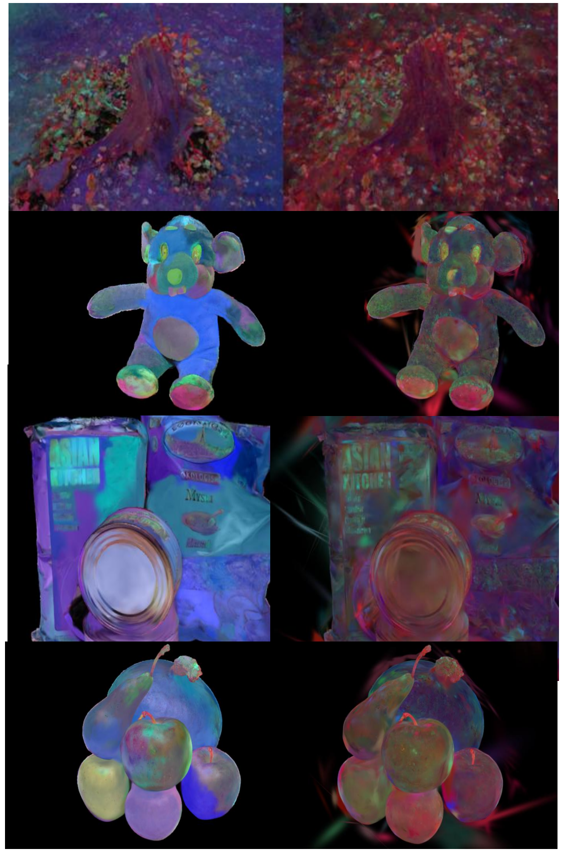
Resulting 3D Geometry

High-Fidelity Geometry:
Our features produce a ***clean and detailed*** 3D model

Novel View Consistency:
Our model maintains a ***robust and coherent*** structure even from new viewpoints (e.g., Reverse view).



Experiment



We conducted extensive experiments on multiple public datasets, including indoor DTU and outdoor Mip-NeRF 360. FHGS shows clear advantages on several fronts.

- **Feature fusion quality** — ***cleaner, crisper***
  - **Sharper** semantic boundaries.
  - **Significantly suppressed** background noise compared to Feature3DGS.
- **Geometry reconstruction & denoising** — ***semantics-guided geometry***
  - Injecting semantic cues results in **smoother, more complete** surfaces.
  - **Effectively removes** artifacts like speckles and "floaters".
- **Training efficiency** — orders-of-magnitude ***faster***

Powered by **NDFD** and an efficient **physics-based model**, FHGS trains far faster than competing methods. On DTU, Feature3DGS takes over 24 hours, whereas FHGS completes in about ***5 minutes***—surpassing even our 2DGS baseline and paving the way for real-time applications.

Table 1 –Indoor scenes

FHGS matches Feature3DGS in PSNR, achieves the ***best FE/FL1***, and is ***>10× faster***, even slightly faster than 2DGS.

Table 1: Quantitative results comparison on indoor scenes

Method	DTU-24 [29]				DTU-37 [29]				MN360-kitchen [15]			
	PSNR↑	FE↓	FL1↓	Time↓	PSNR↑	FE↓	FL1↓	Time↓	PSNR↑	FE↓	FL1↓	Time↓
2DGS	30.1	1.35	0.61	6.1m	30.5	1.31	0.52	6.3m	30.2	1.32	0.79	6.5m
Feature3DGS	**31.5**	0.52	0.24	82.2m	**31.1**	0.88	0.31	73.2m	**31.7**	0.63	0.31	113.2m
FHGS (**ours**)	30.9	**0.15**	**0.22**	**5.2m**	30.8	**0.21**	**0.18**	**5.7m**	30.8	**0.23**	**0.21**	5.1m

Table 2 – Outdoor scenes

FHGS keeps the ***strongest feature consistency*** (lowest FE/FL1) with much ***shorter training*** time than Feature3DGS, at the cost of a small PSNR drop on shadowed, low-semantic regions.

Table 2: Quantitative results comparison on outdoor scenes

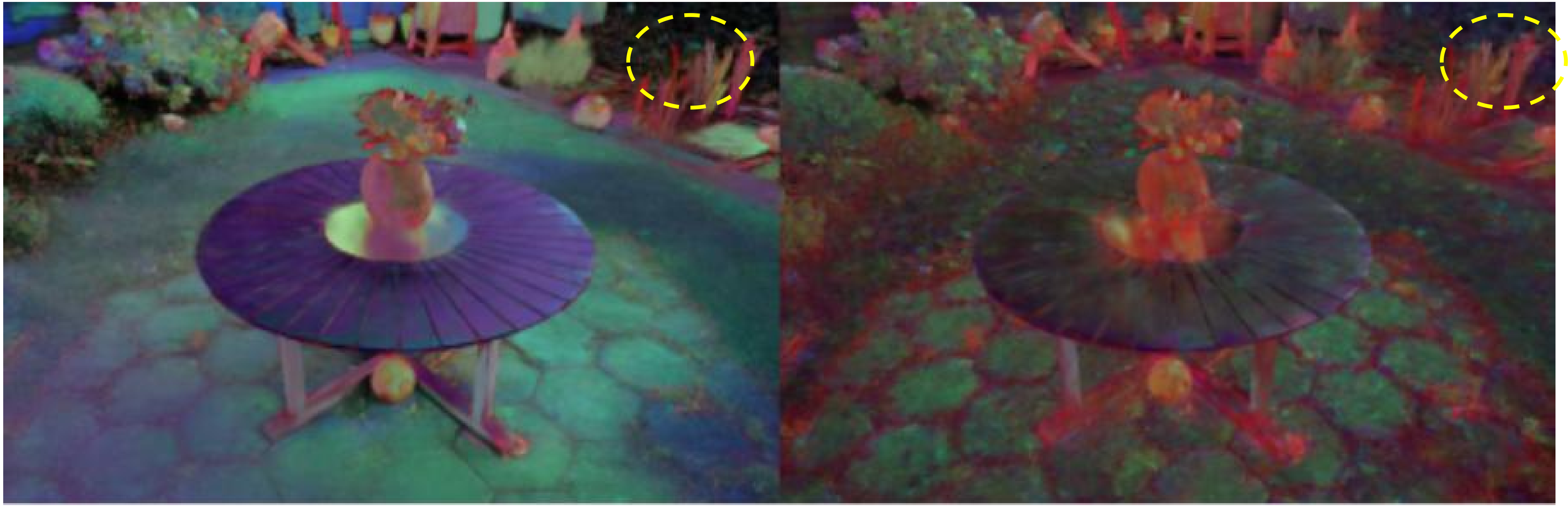
Method	COLMAP [31]				MN360-Garden [15]				TnT-Caterpillar [30]			
	PSNR↑	FE↓	FL1↓	Time↓	PSNR↑	FE↓	FL1↓	Time↓	PSNR↑	FE↓	FL1↓	Time↓
2DGS	27.4	1.73	0.83	10.16m	31.3	1.67	0.75	6.3m	26.8	1.72	0.76	5.2m
Feature3DGS	**28.2**	0.55	0.42	181m	**31.6**	0.65	0.33	155.4m	-	-	-	-
FHGS (**ours**)	26.5	**0.25**	**0.24**	**7.8m**	30.6	**0.25**	**0.18**	**6.1m**	**26.6**	**0.21**	**0.41**	**5.2m**

Table 3 – DTU reconstruction

FHGS attains the ***lowest chamfer distance*** with competitive PSNR, using fewer Gaussians and <5 min training—far faster than Feature3DGS.

Table 3: Quantitative results between *FHGS*, 3DGS, 2DGS and Feature3DGS on the DTU [29], we report chamfer distance, PSNR (training-set view), reconstruction time, model size and point number.

Methods	CD↑	PSNR↑	Time↓	PN↓	MB (Storage)
3DGS	1.96	**35.76**	11.2m	532k	113
2DGS	0.83	33.42	5.5m	342k	**52**
Feature3DGS	1.85	35.25	>24h	642k	745
FHGS (**ours**)	**0.75**	34.21	**4.8m**	**196k**	183



PSNR* of rendered images** — ***indoor gains, outdoor trade-offs

FHGS outperforms our 2DGS baseline on indoor datasets, but falls below it outdoors. The drop is mainly due to shadows *(see yellow circles)*, fragmented soil, and other low-semantic textures that reduce numeric fidelity in PSNR.