HiSplat: Hierarchical 3D Gaussian Splatting for Generalizable Sparse-View Reconstruction

Shengji Tang[1,2], Weicai Ye[3], Peng Ye[2], Weihao Lin[1], Yang Zhou[1,2], Tao Chen[1], Wanli Ouyang[2]

[1]Fudan University, [2] Shanghai AI Laboratory, [3]State Key Lab of CAD&CG, Zhejiang University

Motivation & Key Point

Hierarchical 3D Gaussians help Reconstruction from bone to flesh

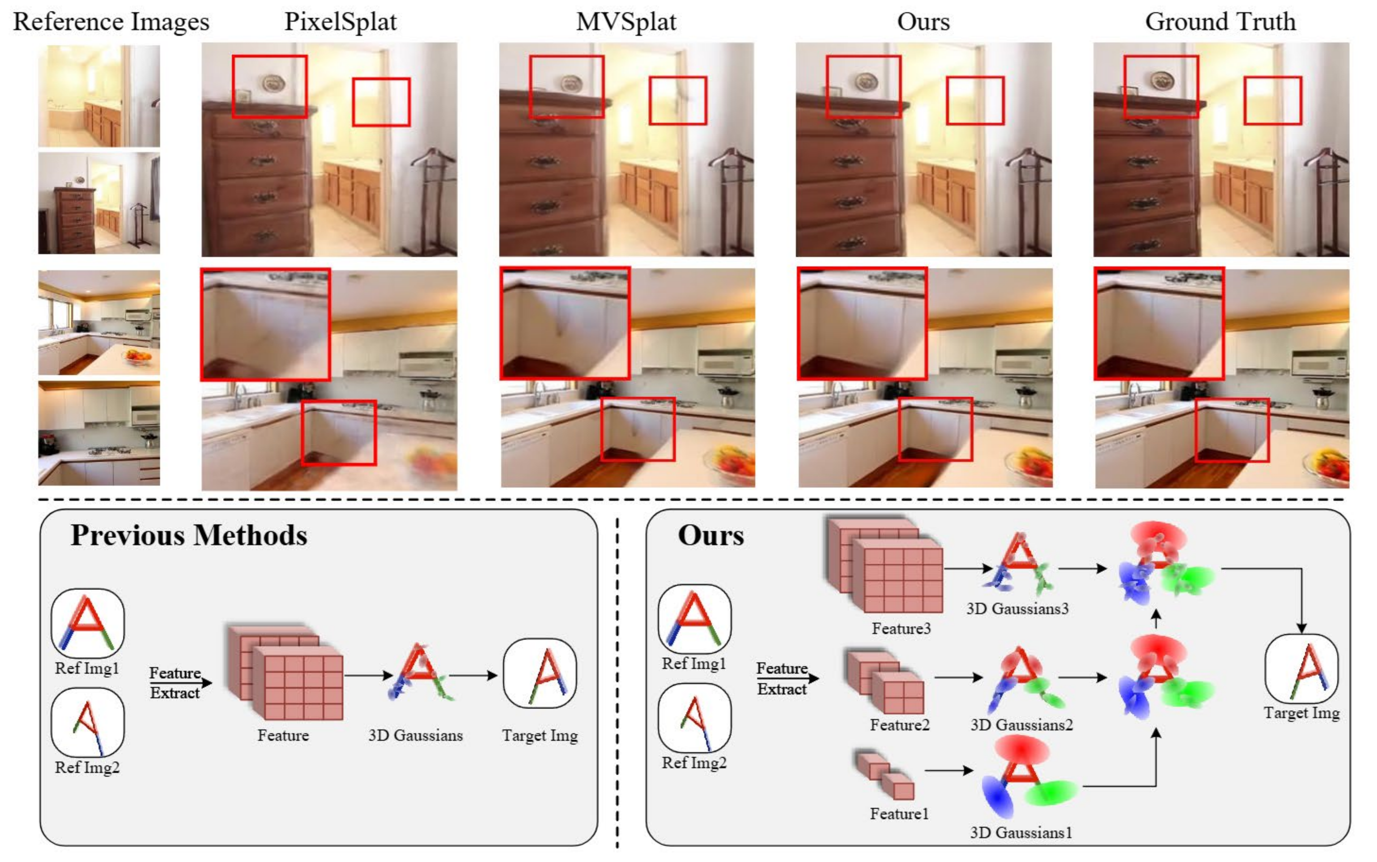


Figure 1. Comparison between HiSplat and previous methods. HiSplat constructs hierarchical 3D Gaussians which can better represent large-scale structures (more accurate location and less crack), and texture details (fewer artefacts and less blurriness).

Overall Framework

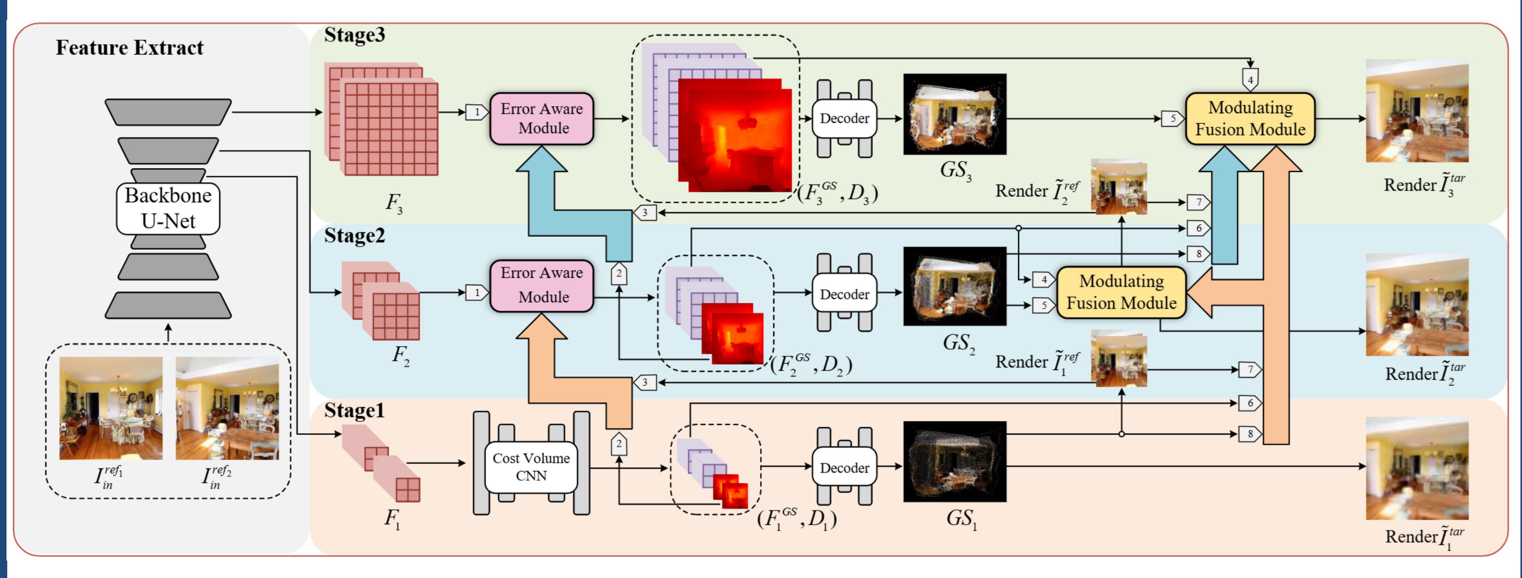


Figure 2. The overall framework of HiSplat. For simplicity, the situation with two input images is illustrated. HiSplat utilizes a shared U-Net backbone to extract different-scale features. With these features, three processing stages predict pixel-aligned Gaussian parameters with different scales, respectively. Error aware module and modulating fusion module perceive the errors in the early stages and guide the Gaussians in the later stages for compensation and repair. Finally, the fusing hierarchical Gaussians can reconstruct both the large-scale structure and texture details.

Detailed Approach

➢ Hierarchical Feature Extraction and Depth Estimation

- **Hierarchical Feature Extraction:** To extract hierarchical multi-scale features from input multi-view 2D images, we construct a CNN and Transformer mixed feature extraction backbone network with a U-Net architecture.
- **Hierarchical Depth Estimation:** For hierarchical Gaussians, the larger-scale Gaussians form the skeleton, while smaller Gaussians are near larger-scale Gaussians as a supplement. Therefore, the depth prediction principle is estimating the largest-scale Gaussian depth accurately, and using it as a reference to predict the depth of smaller-scale Gaussians

➢ Error Aware Module

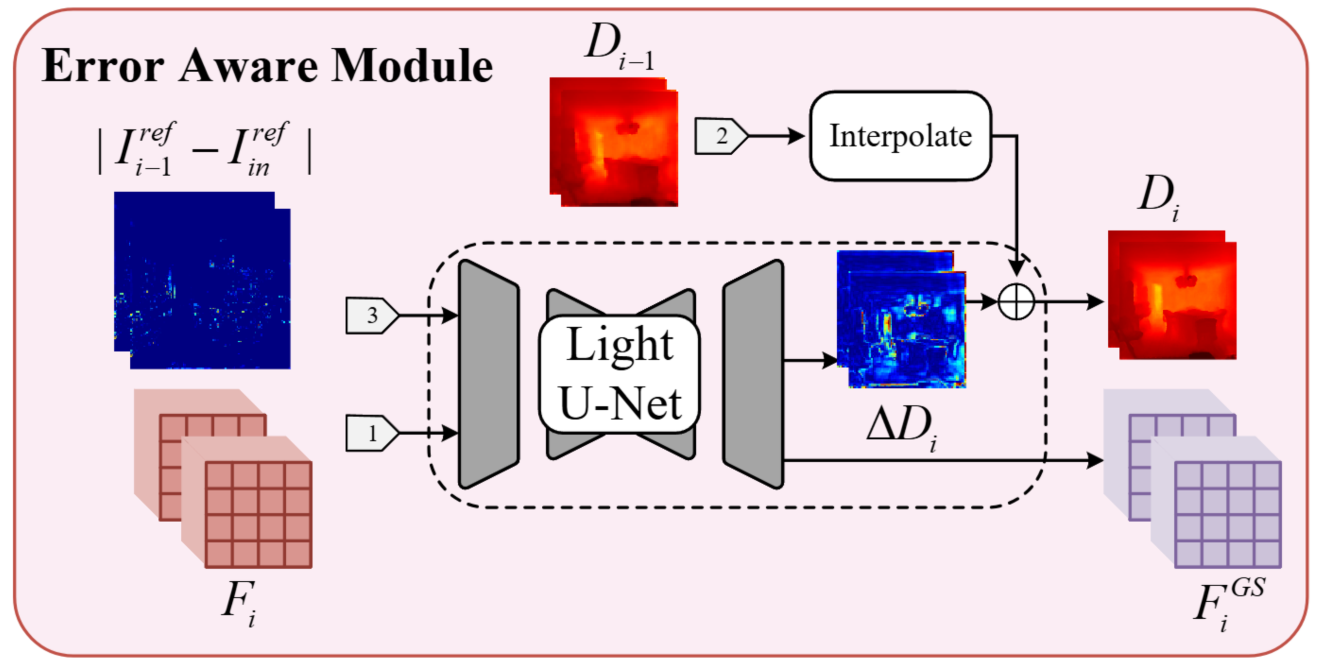


Figure 3. Illustration of Error Aware Module. To enable small-scale Gaussians to supplement the lacking details and correct structural errors of the large-scale Gaussians, we render the mixed Gaussians from the previous stage from input views and compute an error map with the input images. A lightweight 2D U-Net with two predictors is used to aggregate and generate the depth offsets and Gaussian features.

➢ Modulating Fusion Module

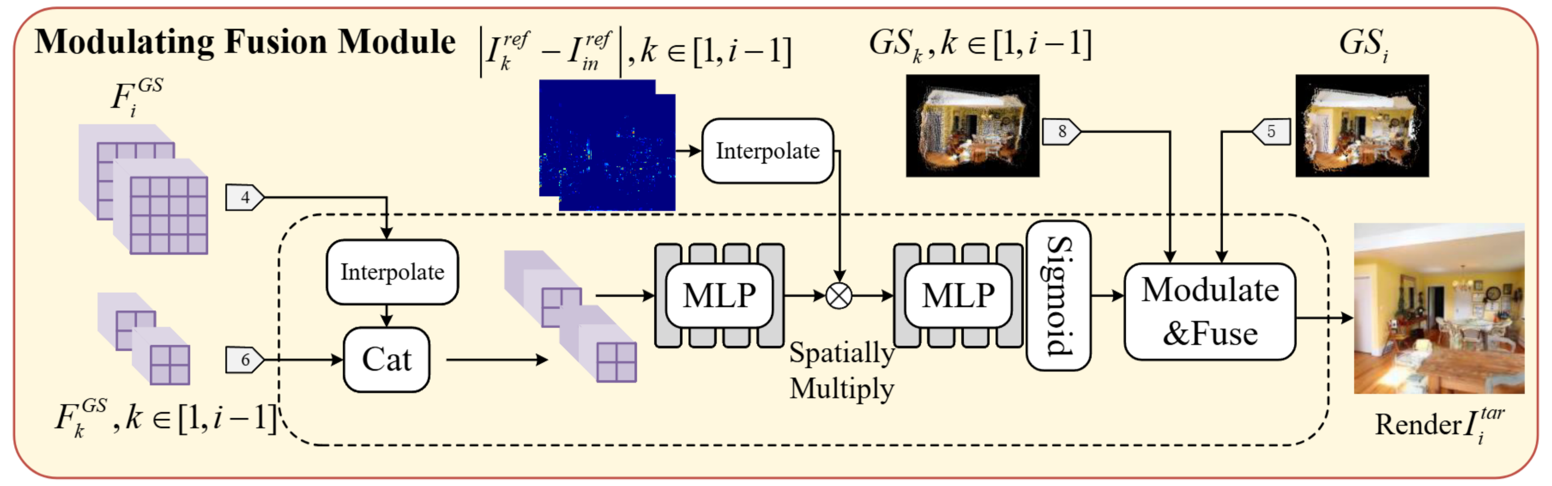


Figure 4. Illustration of Modulating Fusion Module. Inspired by spatial attention, Modulating Fusion Module focuses on modulating and reweighting the Gaussians' opacity in the areas with significant errors.

Code: https://github.com/Open3DVLab/HiSplat

Experimental Results

➢ Quantitative Results

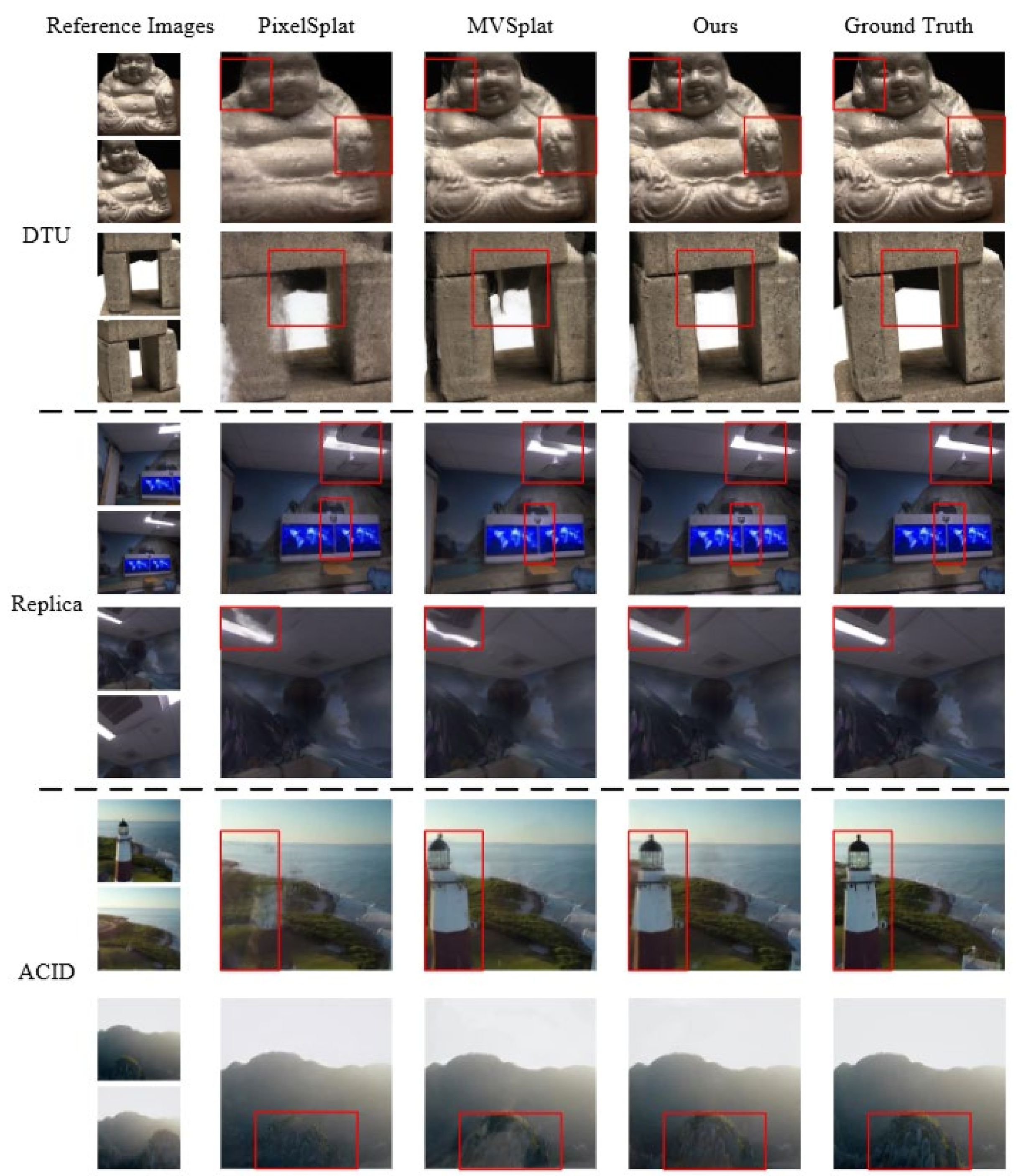
- **Novel Views Synthesis**

Method	RealEstate10K			ACID		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
pixelNeRF (Yu et al., 2021)	20.43	0.589	0.550	20.97	0.547	0.533
GPNR (Suhail et al., 2022)	24.11	0.793	0.255	25.28	0.764	0.332
AttnRend (Du et al., 2023)	24.78	0.820	0.213	26.88	0.799	0.218
MuRF (Xu et al., 2024)	26.10	0.858	0.143	28.09	0.841	0.155
PixelSplat (Charatan et al., 2024)	25.89	0.858	0.142	28.14	0.839	0.150
MVSplat (Chen et al., 2024b)	26.39	0.869	0.128	28.25	0.843	0.144
TranSplat (Zhang et al., 2024)	26.69	0.875	0.125	28.35	0.845	0.143
HiSplat(Ours)	27.21	0.881	0.117	28.75	0.853	0.133

- **Cross-Dataset Generalization**

Method	RealEstate10K → DTU			RealEstate10K→ACID			RealEstate10K→Replica		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
PixelSplat (Charatan et al., 2024)	12.89	0.382	0.560	27.64	0.830	0.160	23.98	0.821	0.202
MVSplat (Chen et al., 2024b)	13.94	0.473	0.385	28.15	0.841	0.147	23.79	0.817	0.165
TranSplat (Zhang et al., 2024)	14.93	0.531	0.326	28.17	0.842	0.146	-	-	-
HiSplat(Ours)	16.05	0.671	0.277	28.66	0.850	0.137	27.17	0.899	0.113

➢ Qualitative Comparison



More Discussions

➢ Analysis of 3D Gaussian Primitives

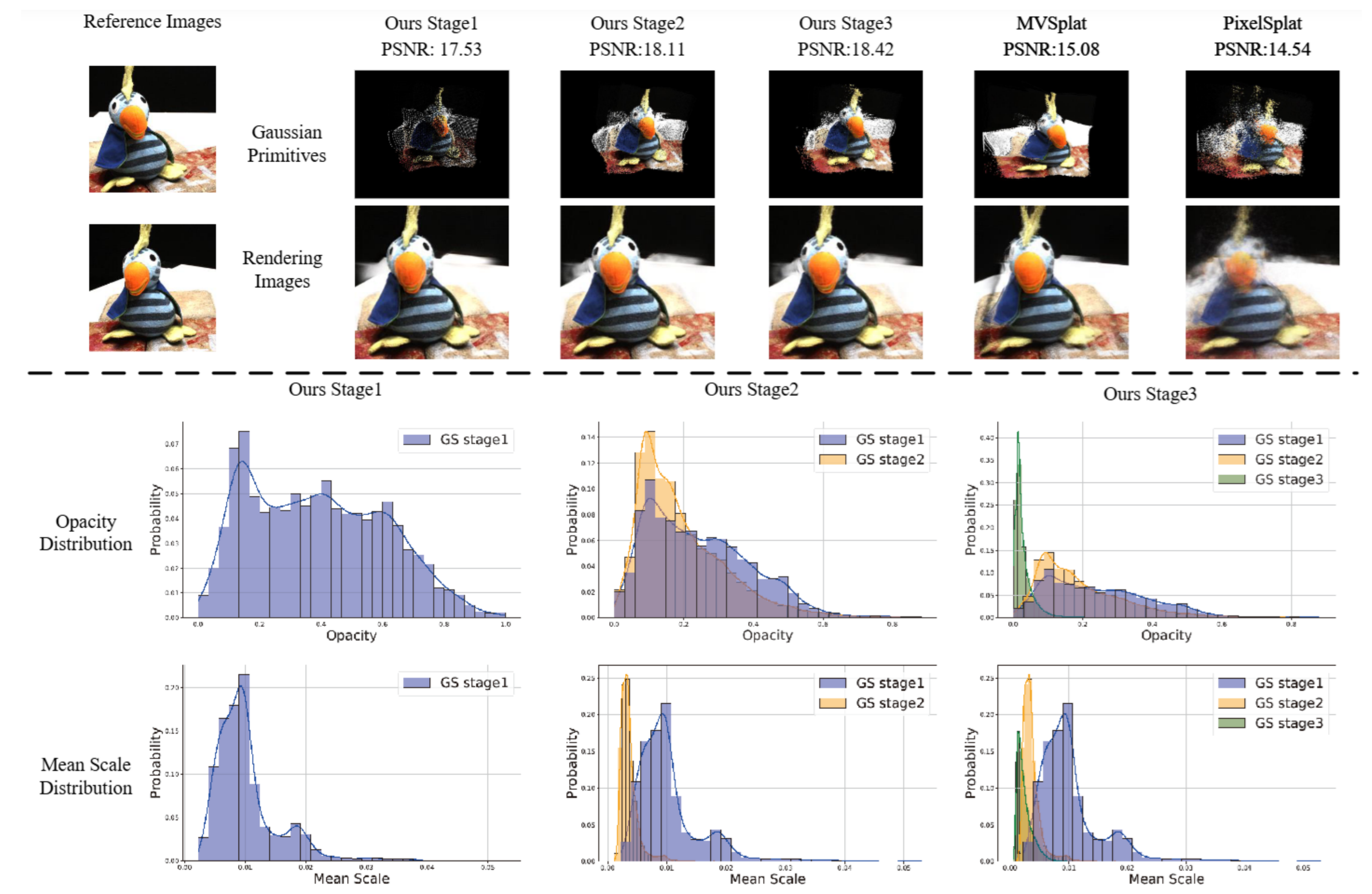


Figure 5. Comparison of Gaussian primitives in different stages on DTU. HiSplat can gradually generate large-scale solid Gaussians as "bone" and small-scale transparent Gaussians as "flesh", confirming better rendering quality and geometry.

➢ Analysis of 2D Error Map

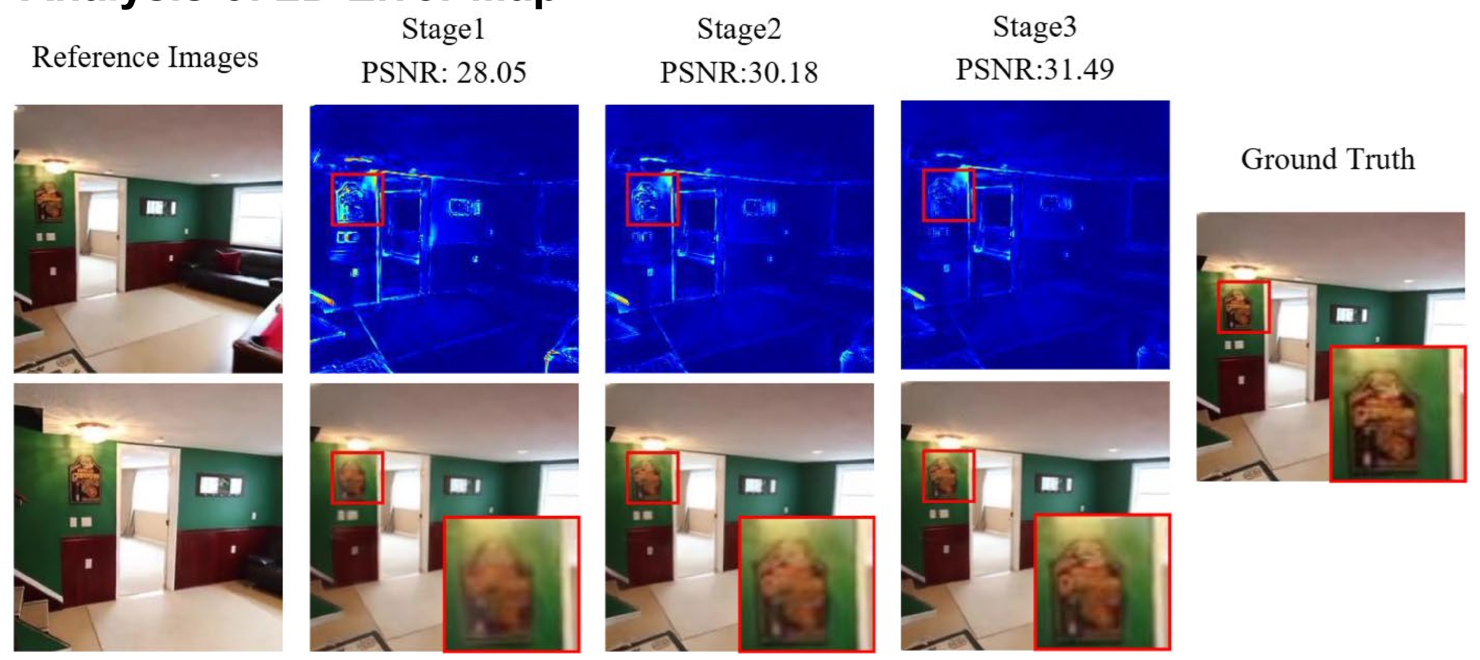


Figure 6. Comparison of rendering images from different stages on RealEstate10K. HiSplat can perceive the error, and utilize Gaussians in the later stages to add details and correct errors gradually.

➢ Analysis of Efficiency and Effectiveness

Method	Peak Memory/GB↓	Inference Time/s↓	PSNR↑	SSIM↑	lPIPS↓
PixelSplat (Charatan et al., 2024)	24.17	1.47	25.89	0.858	0.142
MVSplat (Chen et al., 2024)	14.08	0.27	26.39	0.869	0.128
HiSplat-Stage 1	12.91	0.24	26.13	0.858	0.141
HiSplat-Stage 2	14.74	0.36	26.99	0.879	0.120
HiSplat-Stage 3	23.34	0.51	27.21	0.881	0.117