10 Experiments II: Monocular Depth & Feed-Forward 3DGS

Monocular Depth Estimation

vs Depth Anything 2

Method	KITTI (AbsRel ↓)	Sintel (AbsRel ↓)	NYUv2 (δ1 ↑)
MiDaS v3.1	0.076	0.245	0.892
Depth Anything 2	0.058	0.198	0.965
DA3-Mono (Ours)	**0.054**	**0.185**	**0.971**

Monocular Depth: +7% on KITTI vs DA2

Improvements **LIKELY** stem from (not individually ablated): (1) **depth (not disparity) target** → better for downstream 3D tasks, (2) **expanded synthetic teacher data** → broader geometry coverage, (3) **exponential encoding** → enhanced near-field discrimination.

Feed-Forward 3DGS (GS-DPT Head)

Fine-tuning Strategy: `Frozen DA3 Backbone`

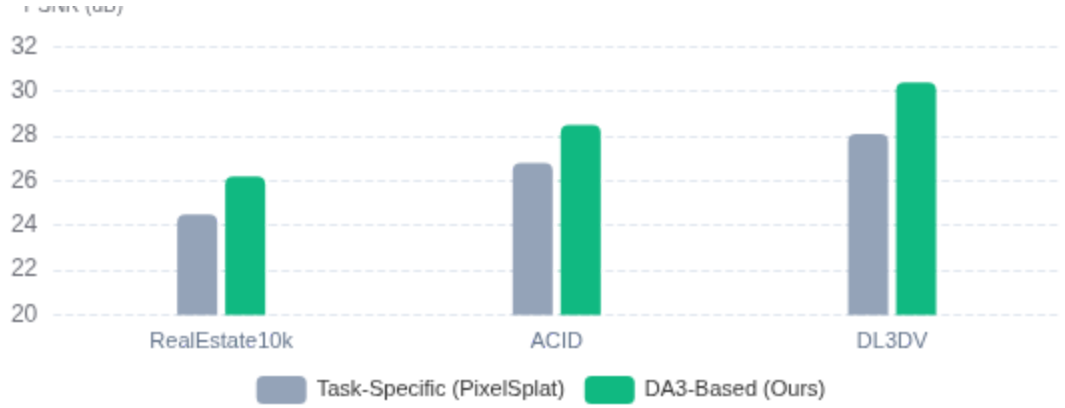
Input: Images + (Optional) Poses

↓

GS-DPT Prediction: Per-pixel Gaussians (σ, q, s, c)

Novel View Synthesis Quality (PSNR)

Pose-Adaptive



Core Findings

Geometry

FOUNDATION > TASK-SPECIFIC

Generalist backbone outperforms specialized NVS models

Adaptivity

WORKS W/ OR W/O POSE

Single model handles both settings seamlessly