SharpDepth: Sharpening Metric Predictions using Diffusion Distillation

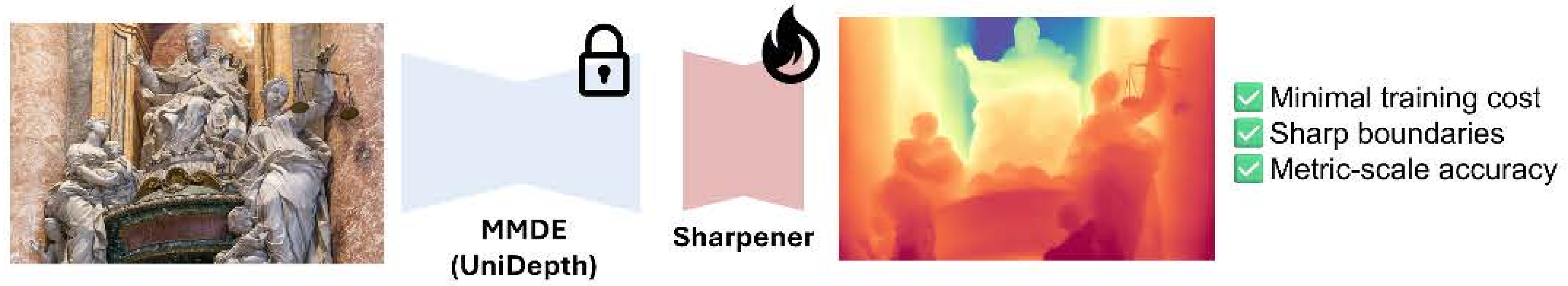
Duc-Hai Pham*, Tung Do*, Phong Nguyen, Binh-Son Hua, Khoi Nguyen, Rang Nguyen

Get accurate and sharp metric depth with minimal fine-tuning (no ground-truths needed)

Problem Statement:

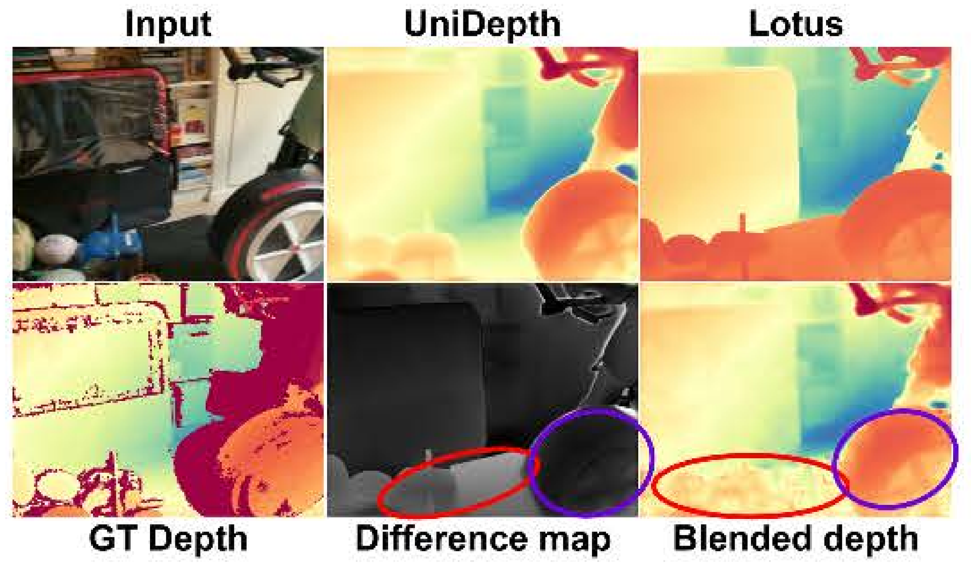
- Metric Depth Estimator are **metrically** accurate, but lack details
- Affine-invariant Diffusion Depth Estimator are **detailed**, **easy to fine-tuned**, but lack **metric** information and accuracy

→ **SharpDepth**: combining metric and diffusion models to get the best of both worlds



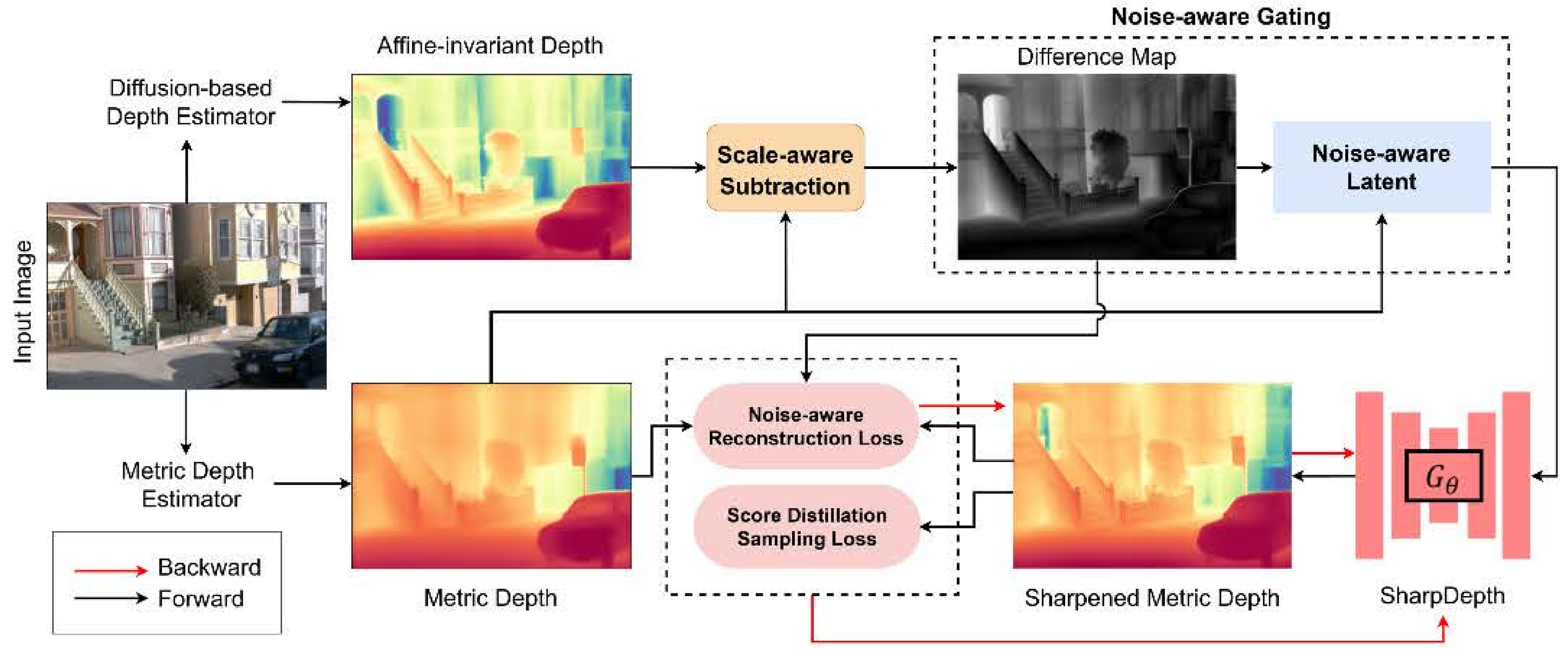
Intuition:

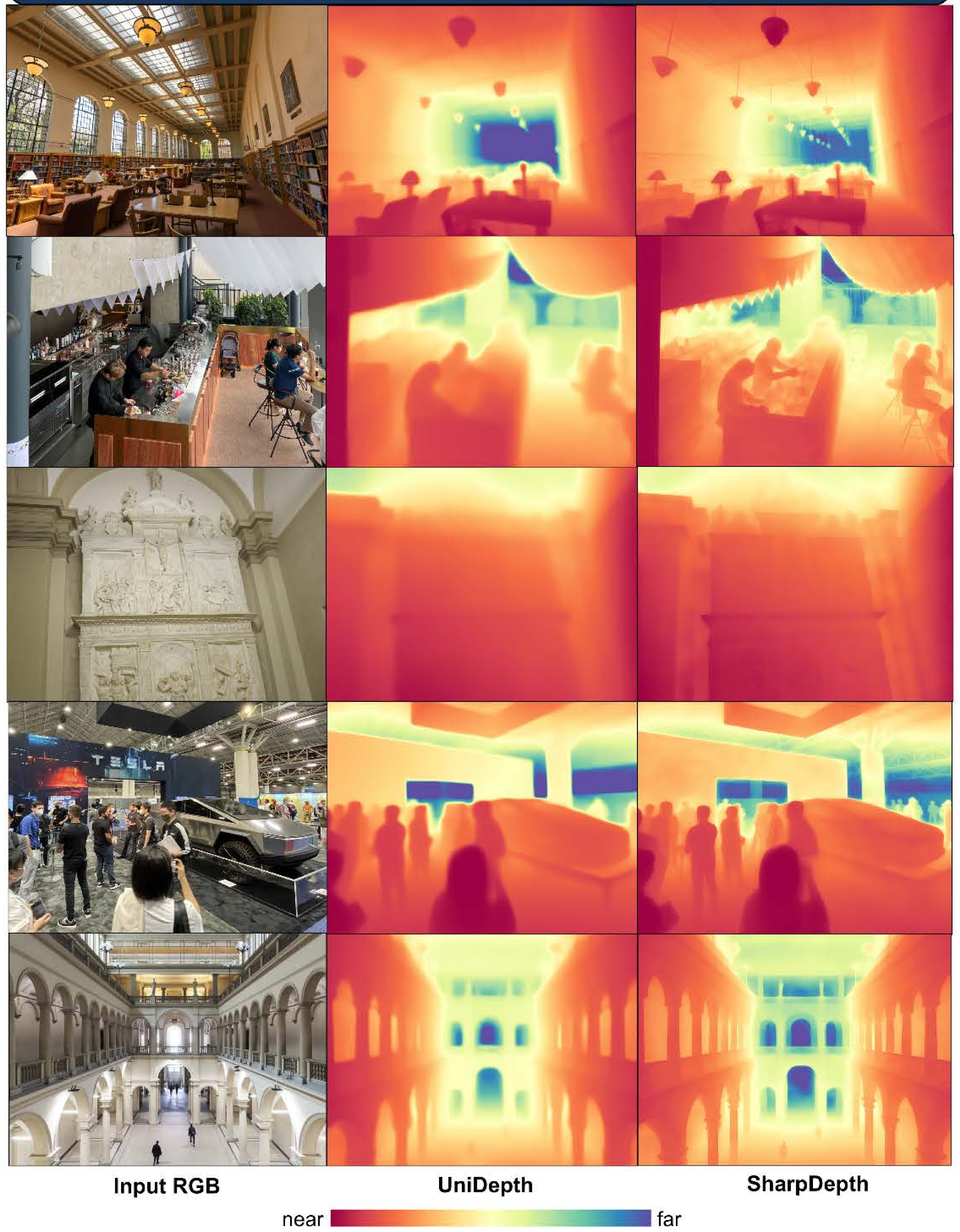
Per-pixel difference between two depth estimators → A measure of boundary sharpness



Methodology:

- SDS loss: adding diffusion-like details to predictions
- Reconstruction loss: maintaining accuracy of original metric predictions





Quantitative Results

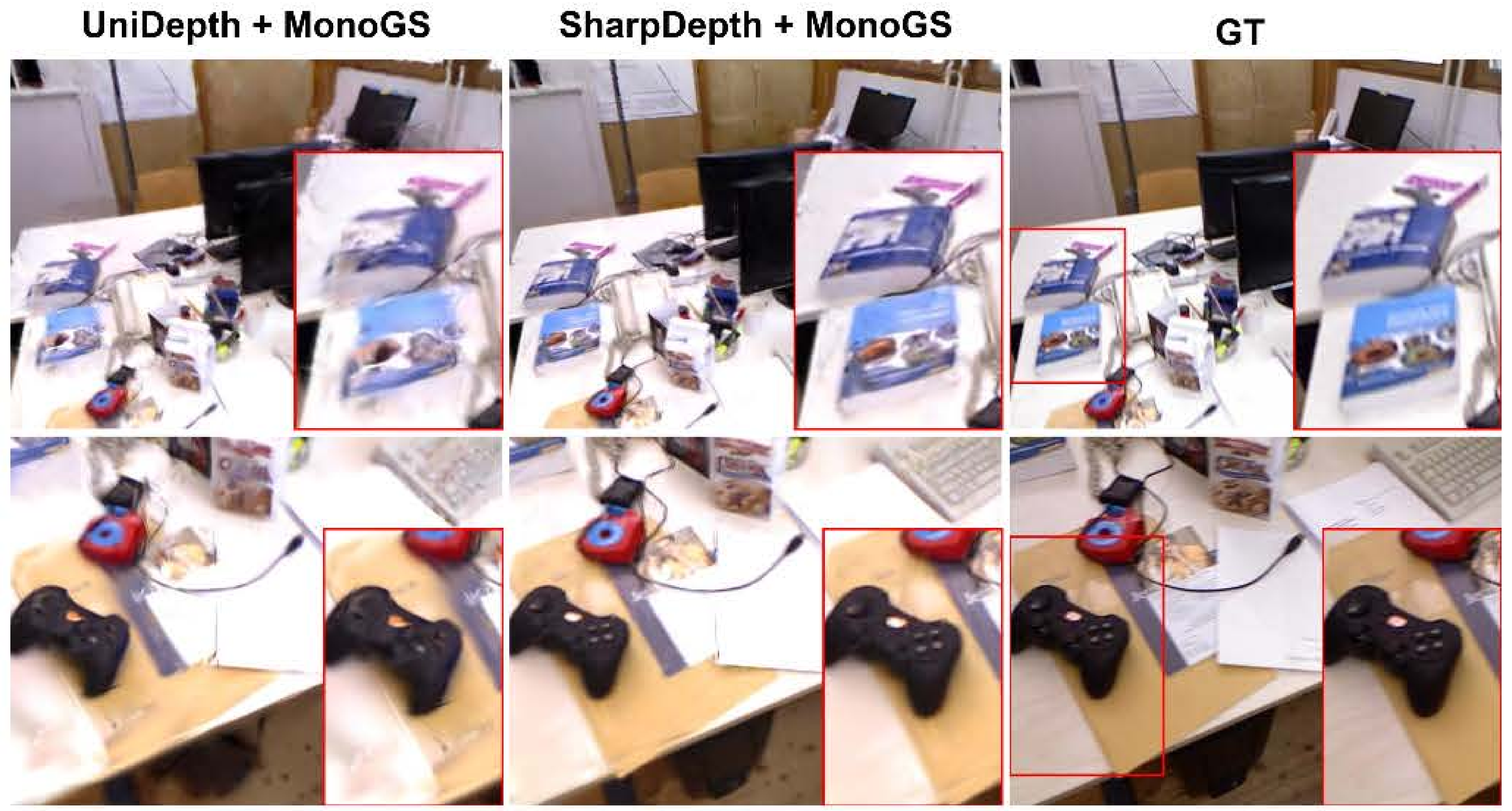
Method	KITTI	NYUv2	Diode	Booster	nuScenes
UniDepth [21]	97.9	98.4	66.0	28.0	83.8
ZocDepth [2]	96.5	95.2	30.2	20.8	21.8
Metric3Dv2 [14]	97.6	97.4	88.8	15.5	84.2
DAv2-Metric [35]	85.3	1.0	22.0	1.0	19.1
SharpDepth (ours)	97.3	96.9	61.5	27.7	78.5

Table 1. **Comparison for percentage of inlier pixels (δ_1) - higher percentage represents better accuracy.** We ranked methods as best, second-best, and third-best.

Method	Sintel	Unreal4K	Spring
UniDepth [21]	3.73	8.65	5.29
ZoeDepth [2]	3.35	5.39	4.05
Metric3Dv2 [14]	1.92	2.75	1.78
DAv2-Metric [35]	1.51	2.47	1.36
SharpDepth (ours)	1.94	1.37	1.24

Table 2. **Comparison for depth boundary error $\epsilon_{\mathrm{DBE}}^{\mathrm{acc}}$ - lower errors represents better sharpness.** We ranked methods as best, second-best, and third-best.

Applications to Gaussian Splatting SLAM



In-the-wild 3D point-clouds