SCADE: NeRFs from Space Carving with Ambiguity-Aware Depth Estimates

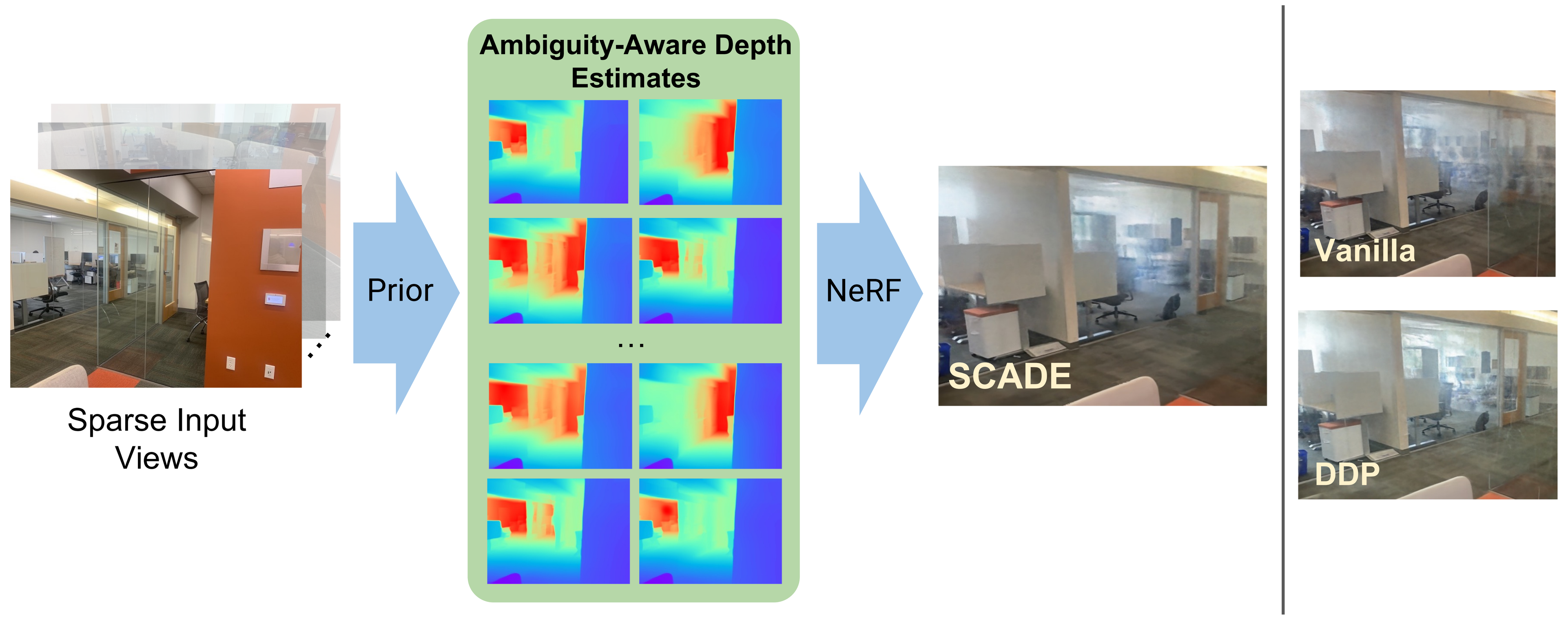
Mikaela Angelina Uy   Ricardo Martin-Brualla   Leonidas Guibas   Ke Li

SFU

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Visit our webpage!

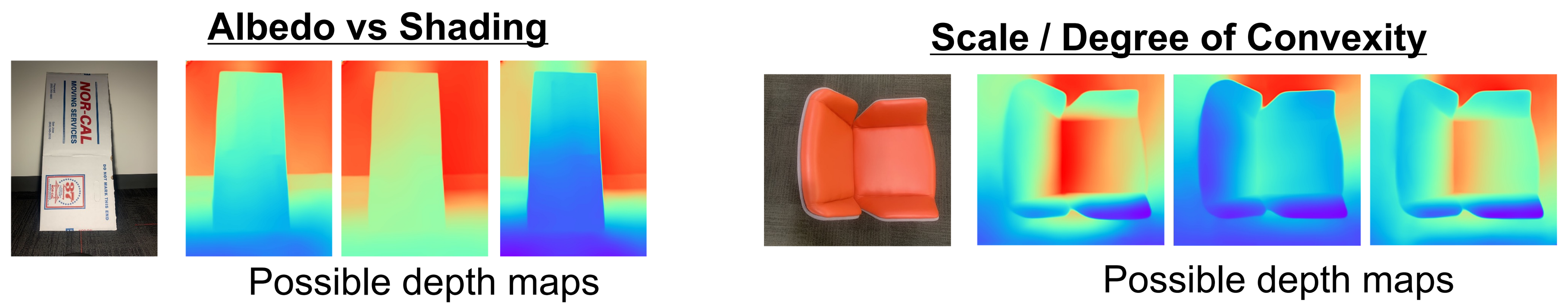
PROBLEM OVERVIEW



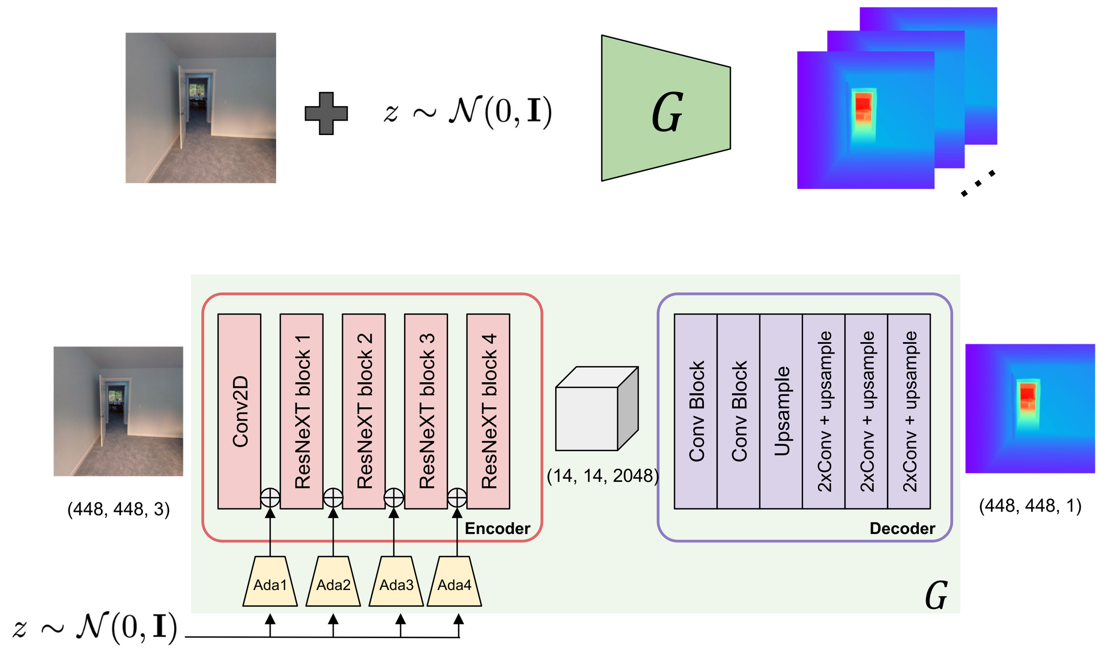
- We tackle the problem of NeRF reconstruction under **sparse**, **unconstrained** views for **in-the-wild** indoor scenes by leveraging on a **generalizable prior** to constrain the NeRF optimization.

AMBIGUITY-AWARE PRIOR

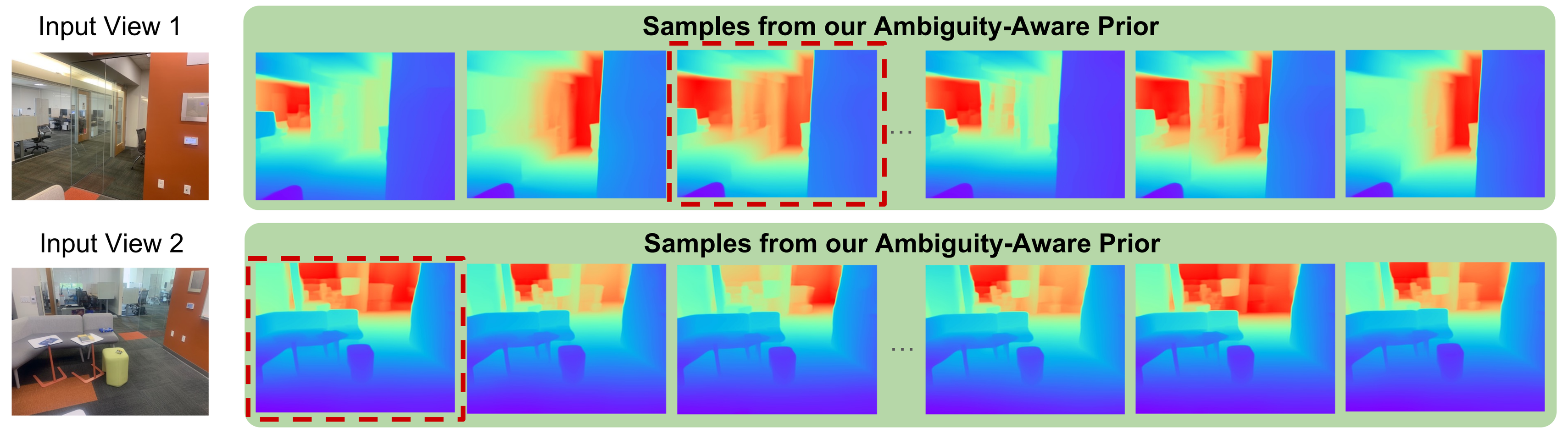
- Monocular depth [1] is generalizable, but is inherently **ambiguous**:



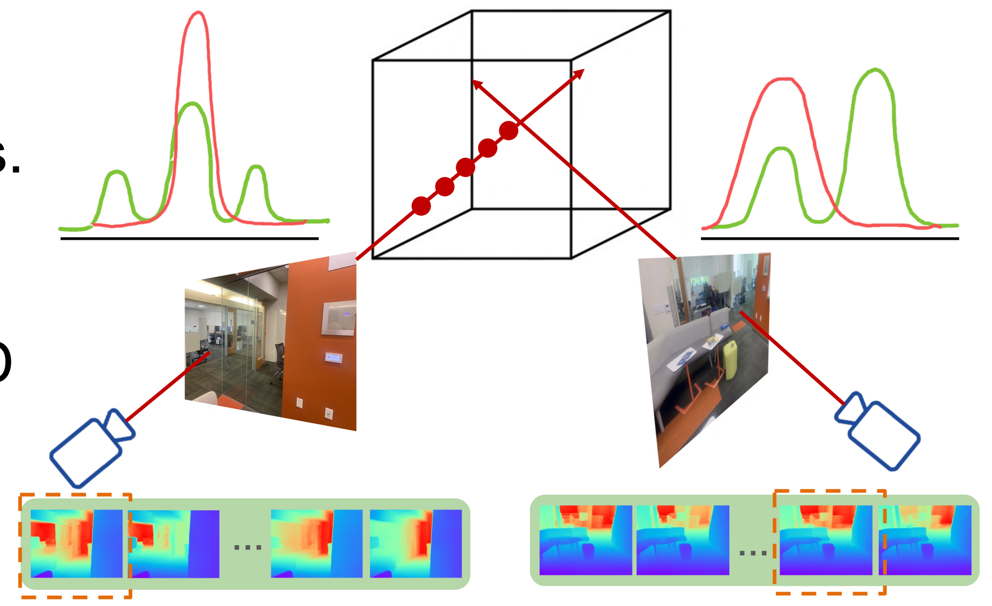
- To handle the ambiguity, we represent depth as a **distribution**, which can be multimodal, by leveraging on **conditional implicit maximum likelihood estimation (cIMLE)** [2].



OUR APPROACH: SCADE



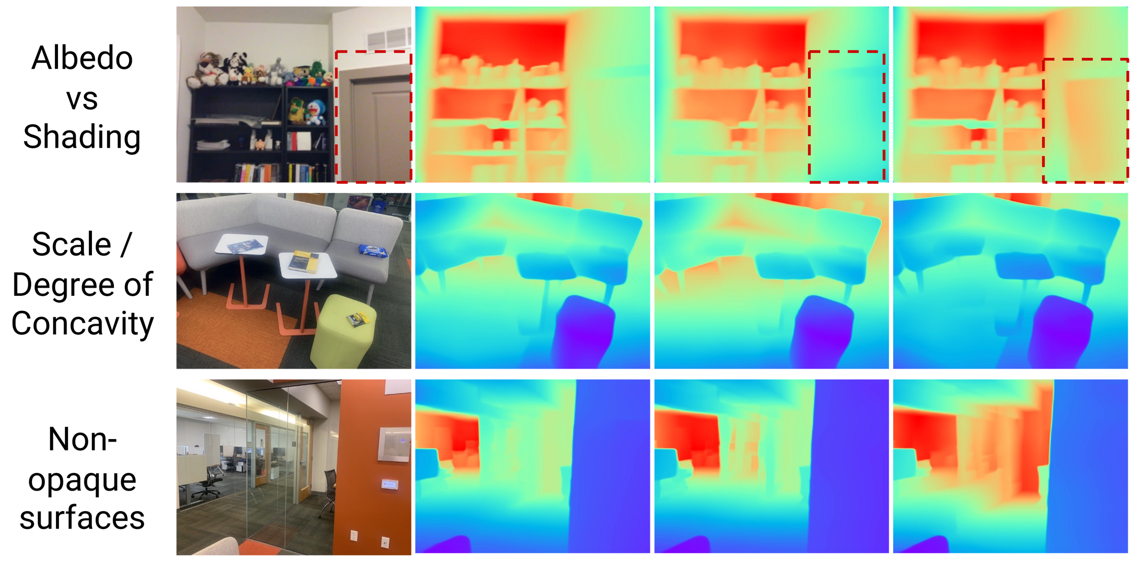
- Inspired by classical **space carving** [3], our loss **fuses** information from multiple views.
- It is **mode seeking** and **sample-based**, leading to supervision in 3D instead of 2D moments.



$$\mathcal{L}_{\text{space\_carving}}(\mathbf{r}) = \sum_{i\in[N]} \min_{j\in[M]} \| x_i - y_j \|_2^2$$

OUR AMBIGUITY-AWARE DEPTH ESTIMATES

- We represent ambiguities and capture **variable modes** through **samples** from our ambiguity-aware prior.

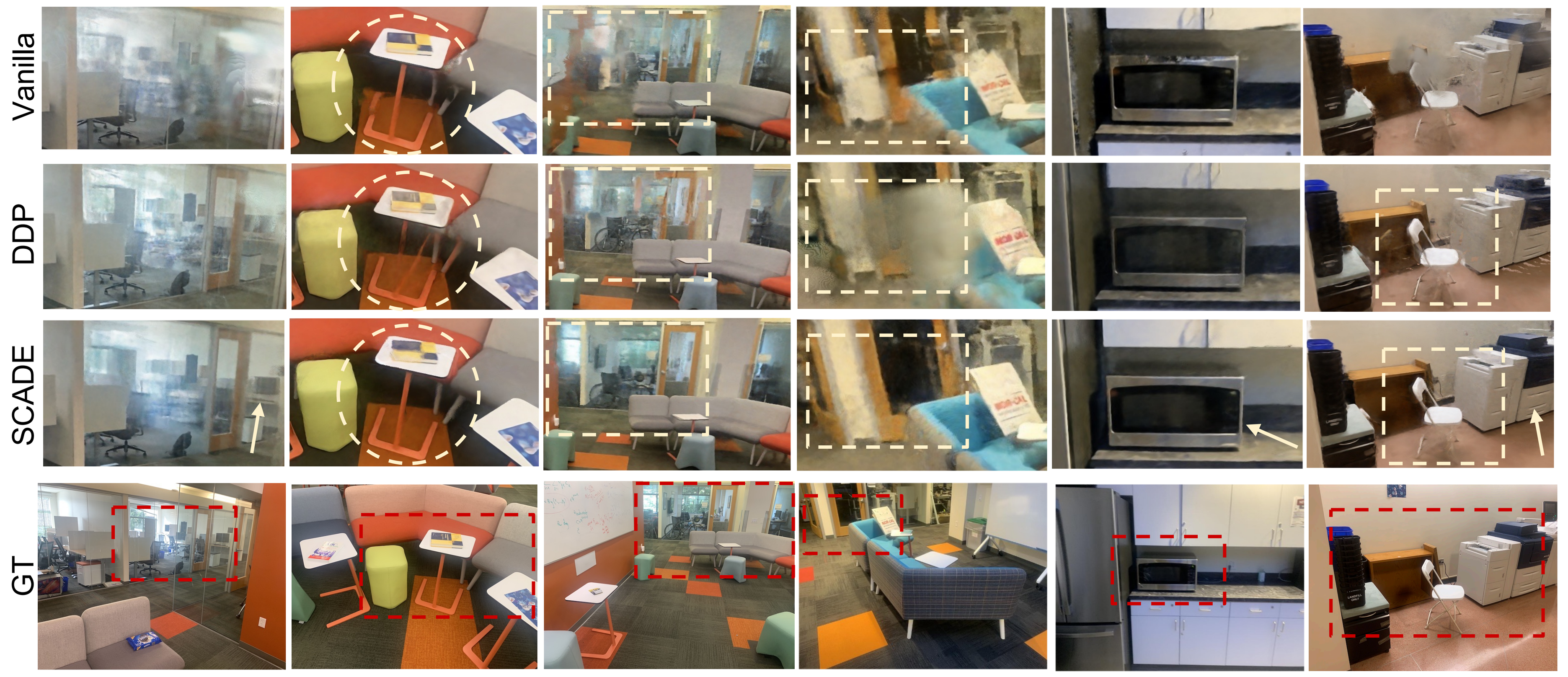


Why does it work?

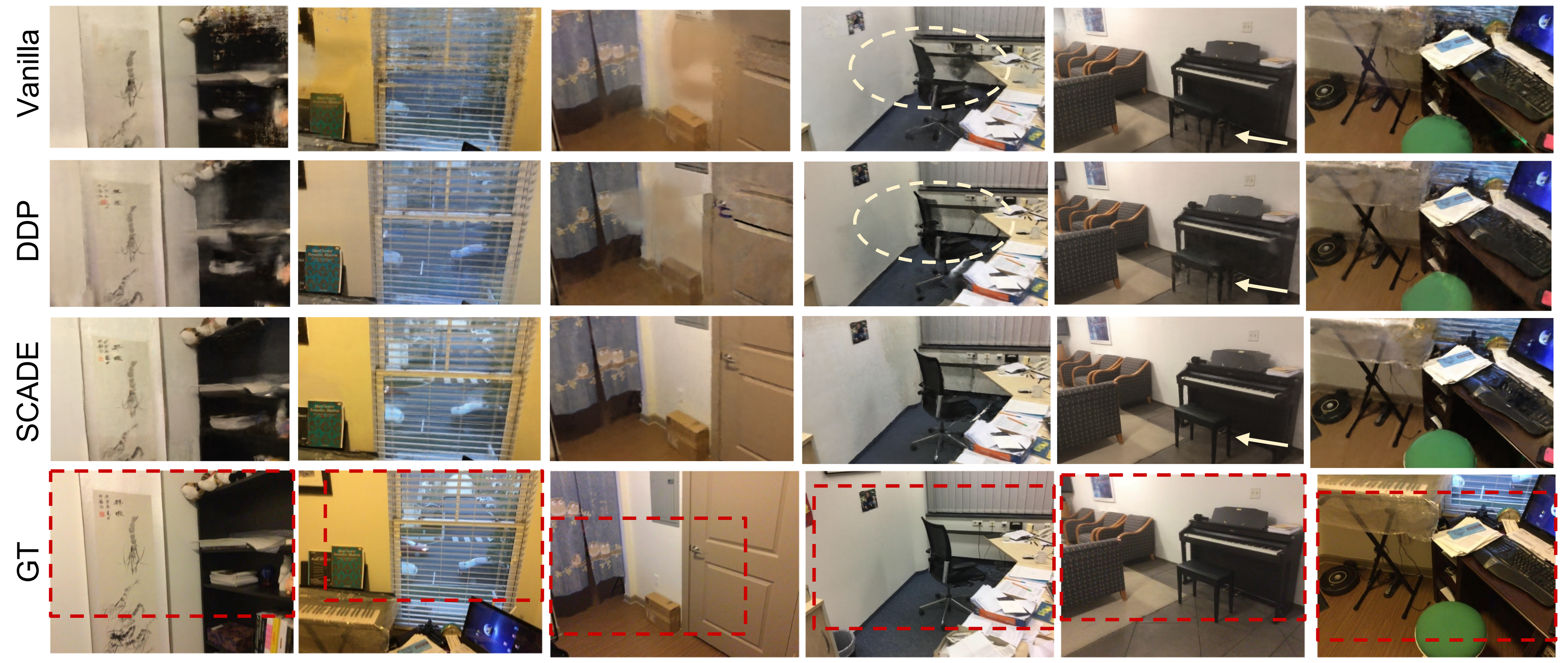
- Training images are **labelled differently**.
- Also works on **reflective surfaces**.

RESULTS

In-the-Wild



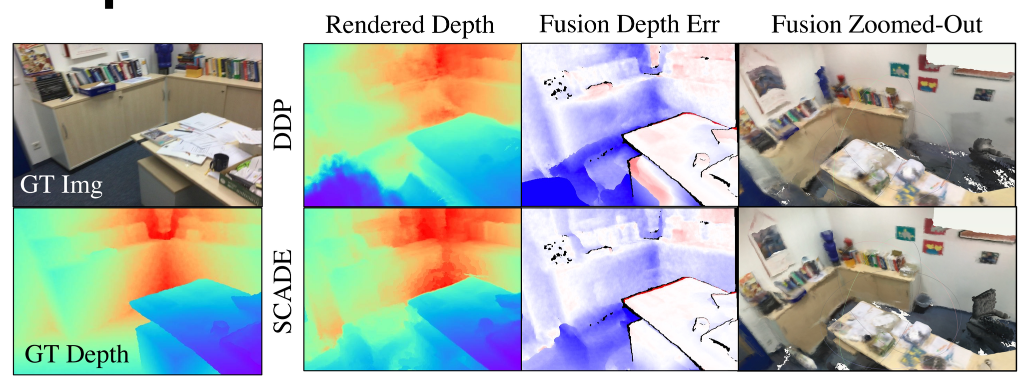
Scannet



Ablation

	PSNR ↑	SSIM ↑	LPIPS ↓
MonoSDF supervision	20.13	0.710	0.332
DDP prior - single sample	20.85	0.712	0.320
DDP prior - multiple samples	21.00	0.718	0.316
Our prior - single sample	21.22	0.714	0.318
SCADE (Ours)	**21.54**	**0.732**	**0.292**

Depth and Fusion



References:
[1] Learning to Recover 3D shape from a Single Image. W. Yin, et. al., CVPR 2021.
[2] Multimodal Image Synthesis with Conditional Implicit Maximum Likelihood Estimation. K. Li, et. al., IJCV 2020.
[3] A Theory of Shape by Space Carving. K. Kutulakos and S. Seitz, IJCV 2000.