

The Less You Depend, The More You Learn: Synthesizing Novel Views from Sparse, Unposed Images without Any 3D Knowledge

Haoru Wang^{*1}, Kai Ye^{*1}, Minghan Qin², Yangyan Li³, Wenzheng Chen¹, Baoquan Chen¹

¹ Peking University, ² Bytedance Seed, ³ Ant Group



Code

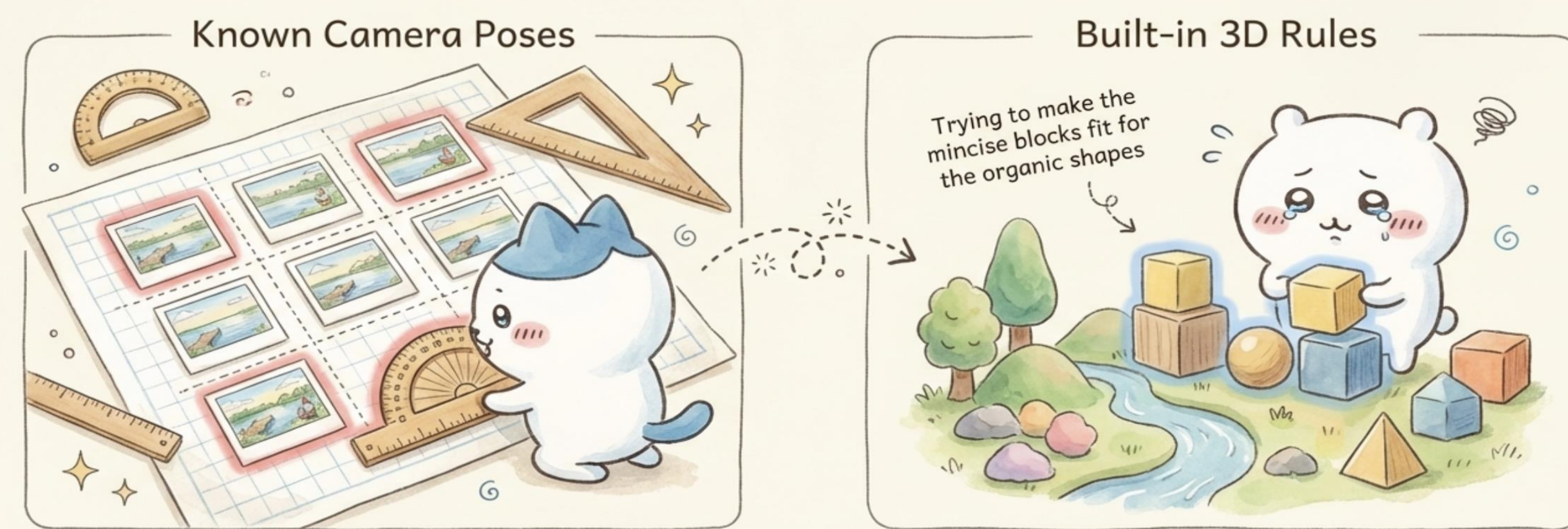


ICLR

Observation: *Unleash Scalability*

3D Knowledge:

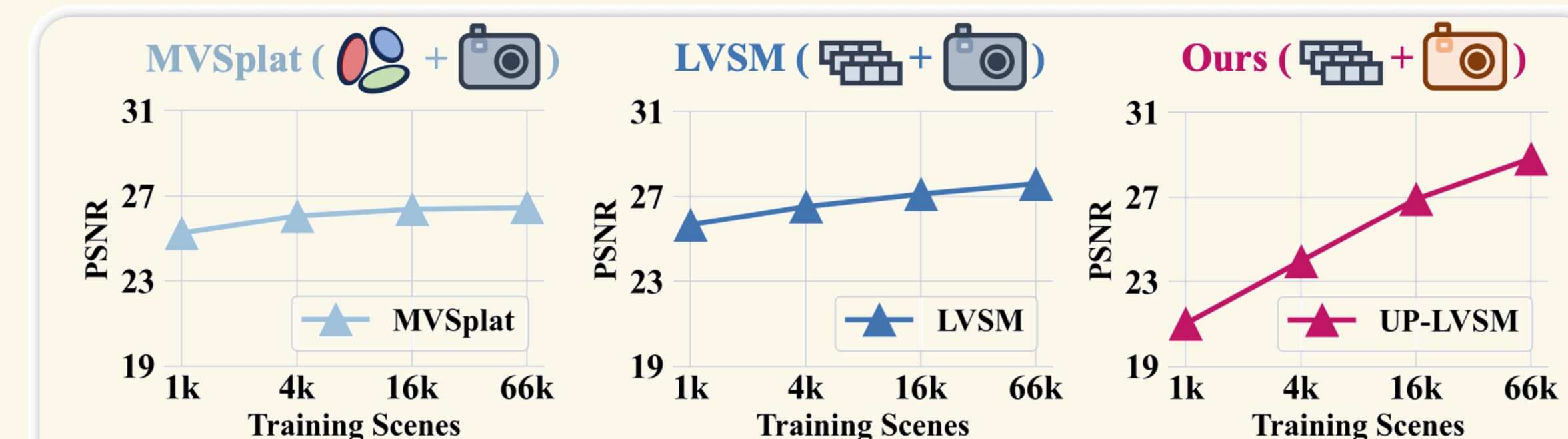
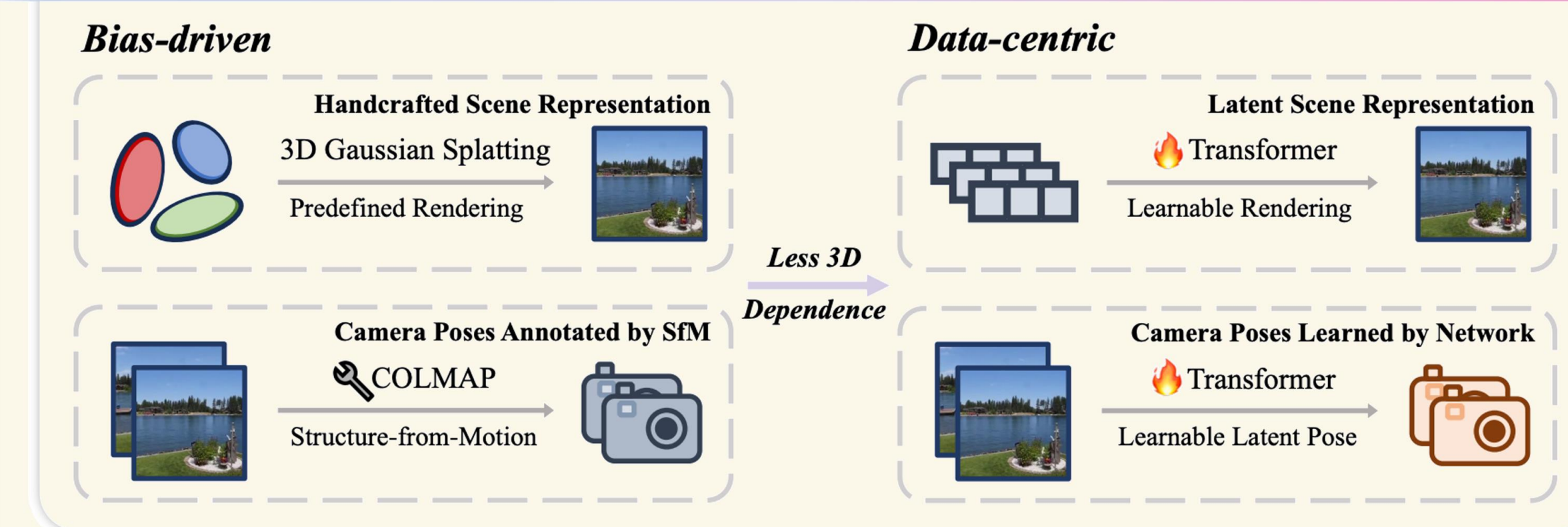
- Existing NVS methods depend on **3D Knowledge**, e.g. camera & 3D representations



Scalability:

- Methods with *less dependence* on **3D Knowledge** derive *greater scalability*

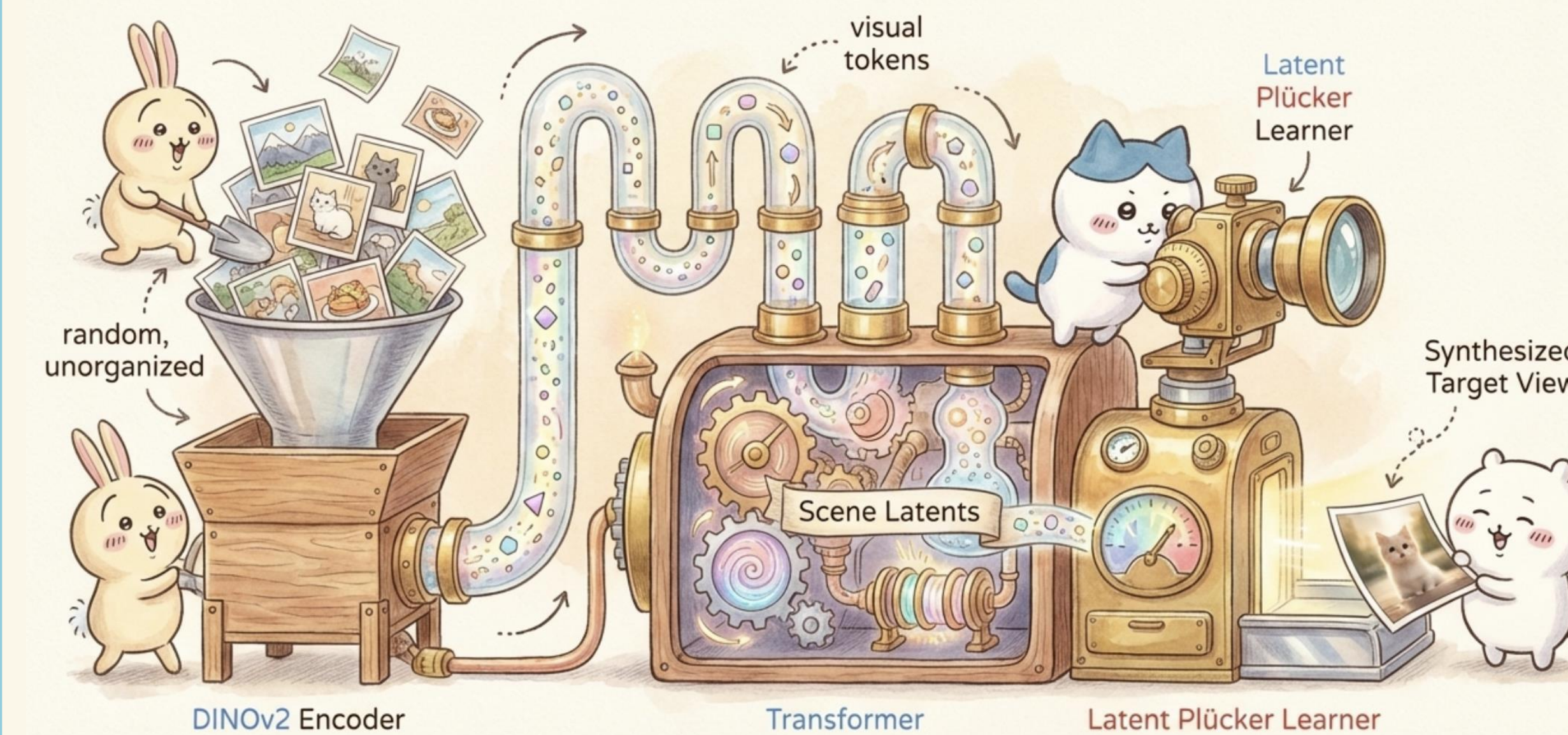
The Less You Depend on Explicit 3D Knowledge



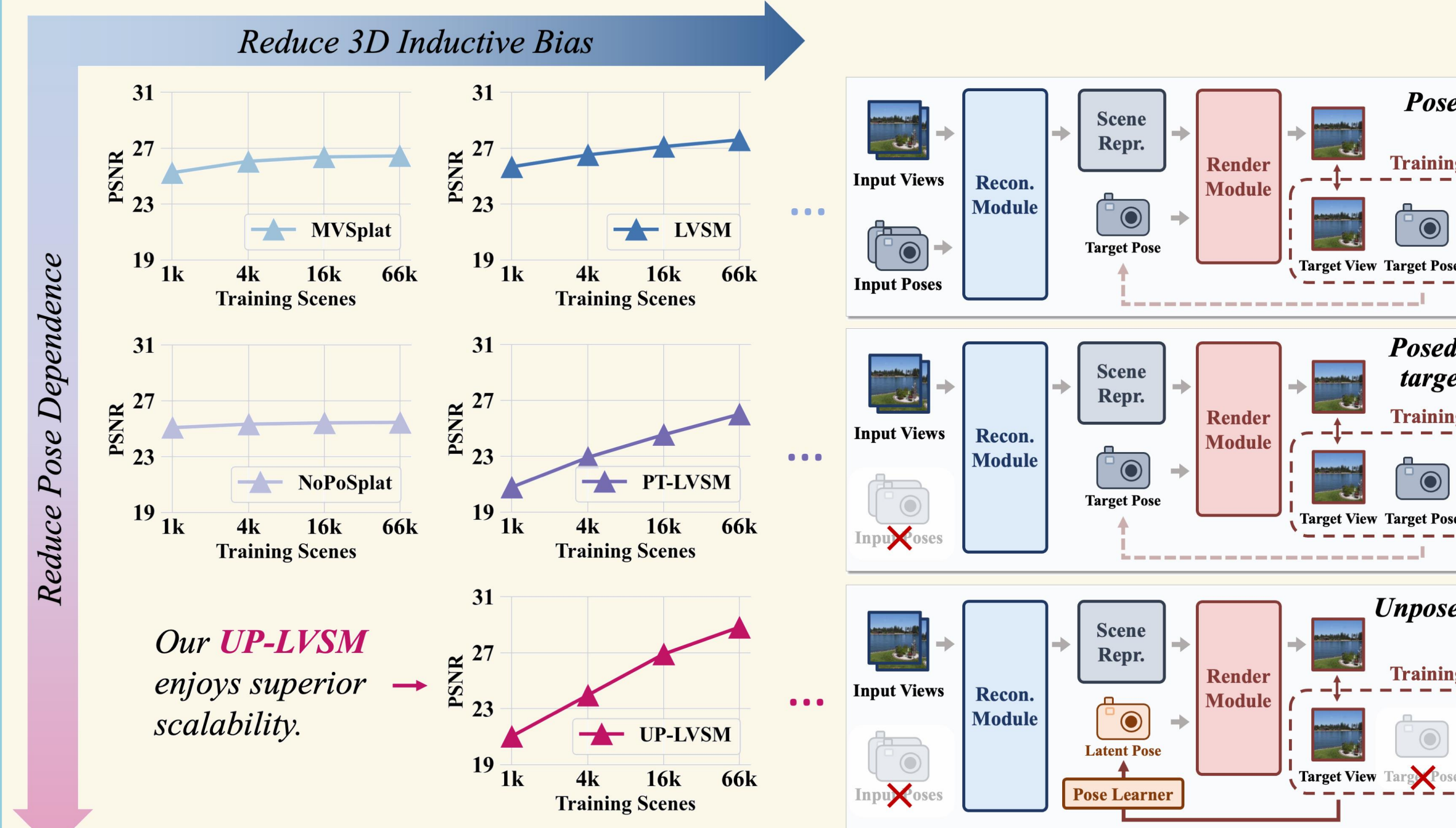
The More You Learn from Abundant Data

Method: *Build with SSL*

Based on our observation, we propose **UP-LVSM** (*Unposed Large View Synthesis Model*), a novel **Self-Supervised** framework achieves SOTA NVS performance.



Experiments validate our method with SOTA performance & scaling.



Analysis: *Probe Implicit 3D*

3D Awareness

- Correspondence & GTA metrics show strong 3D awareness.

Model	0° ~ 15° ↑	15° ~ 30° ↑	30° ~ 60° ↑	60° ~ 180° ↑
CLIP (Radford et al., 2021)	6.6	5.2	4.7	3.0
MAE (He et al., 2022)	10.8	7.8	6.0	3.5
DINOv2 (Oquab et al., 2023)	36.8	27.5	17.9	8.0
UP-LVSM (Ours)	31.9	25.4	18.0	8.2

Model	PSNR@All ↑	SSIM@All ↑	LPIPS@All ↓
DUS3R (Wang et al., 2024)	19.28	0.630	0.391
LVSM (Jin et al., 2025)	18.73	0.590	0.415
UP-LVSM (Ours)	18.81	0.601	0.409



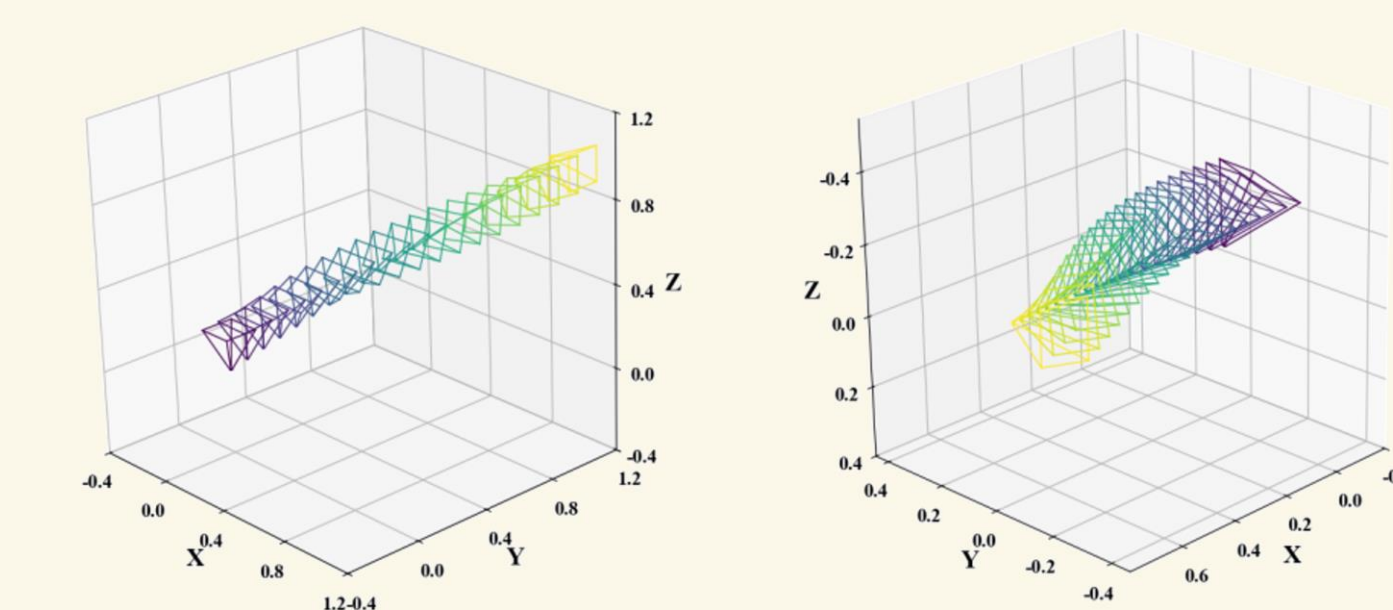
Implicit Geometry

- Latent Plucker Learner* forces meaningful camera manifold.

Model	Trans.@0.1 ↑	Trans.@0.2 ↑	Trans.@0.3 ↑
Rayzer (Jiang et al., 2025)	61.2	84.2	92.8
UP-LVSM (Ours)	71.3	89.2	96.4

Model	Rot.@10° ↑	Rot.@20° ↑	Rot.@30° ↑
Rayzer (Jiang et al., 2025)	99.6	99.9	100
UP-LVSM (Ours)	98.4	99.6	99.8

Input Camera Sequence Mapped in Latent Space



Controllable Rendering

- Camera mapper* makes camera controllable, *Self-Supervised* NVS useful.

Method	Train→Test	PSNR ↑	SSIM ↑	LPIPS ↓
LVSM	ACID→ACID	27.03	0.774	0.201
UP-LVSM	ACID→ACID	27.21	0.787	0.194
UP-LVSM	RE10K→ACID	27.33	0.792	0.182



Generalizable Scalability

- Scaling with only images, effective validated across diverse datasets.