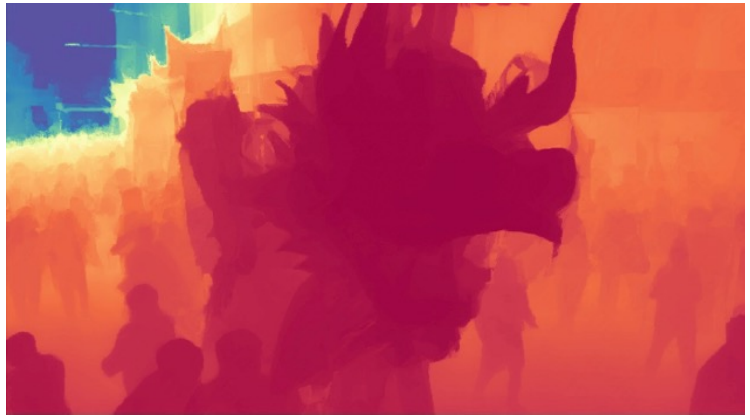
GeoWizard: Unleashing the Diffusion Priors for 3D Geometry Estimation from a Single Image

The 18th European Conference on Computer Vision (ECCV), 2024

Presenter : FU, Xiao

Geometry Attributes

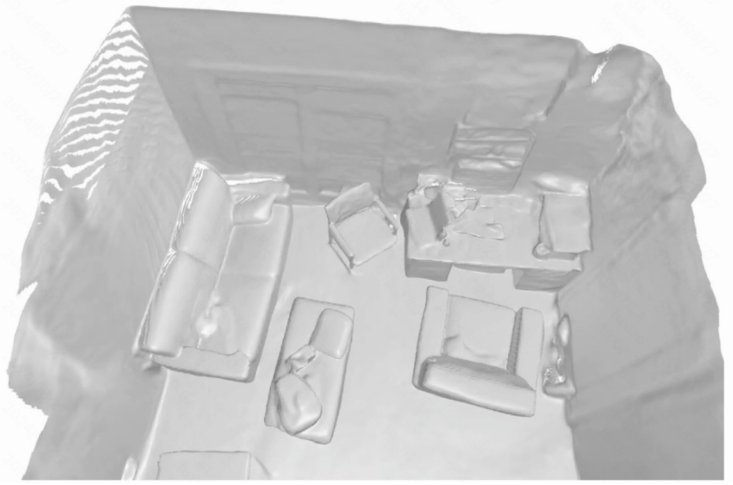


Depth

Surface Normal

Applications

Autonomous Driving



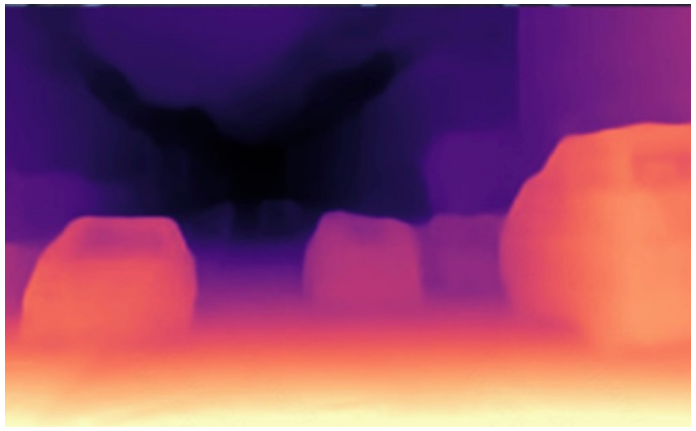
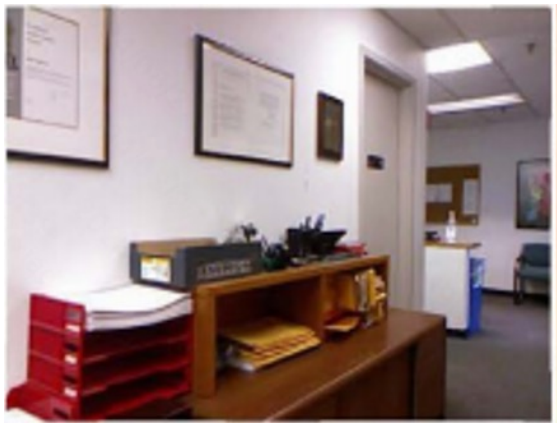
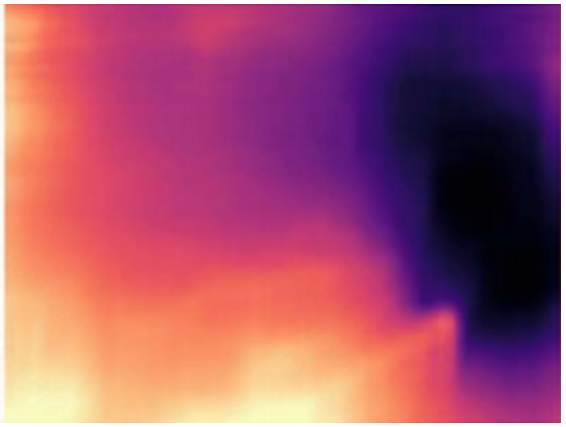
3D Reconstruction

Novel View Synthesis

Bottleneck

Poor generalization on in-the-wild data

1) **Low Diversity** : most models are confined to specific scenarios

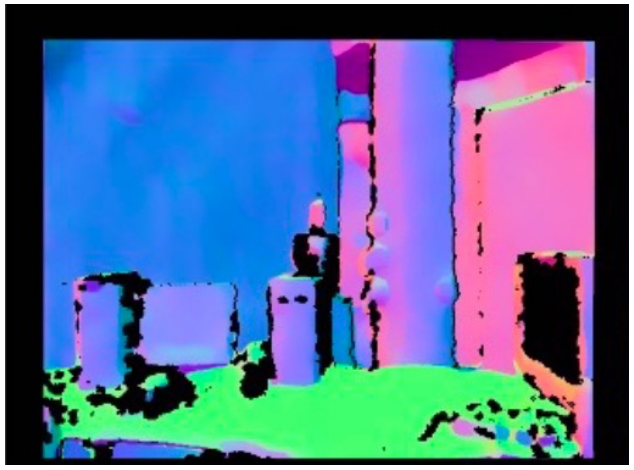





Autonomous Driving

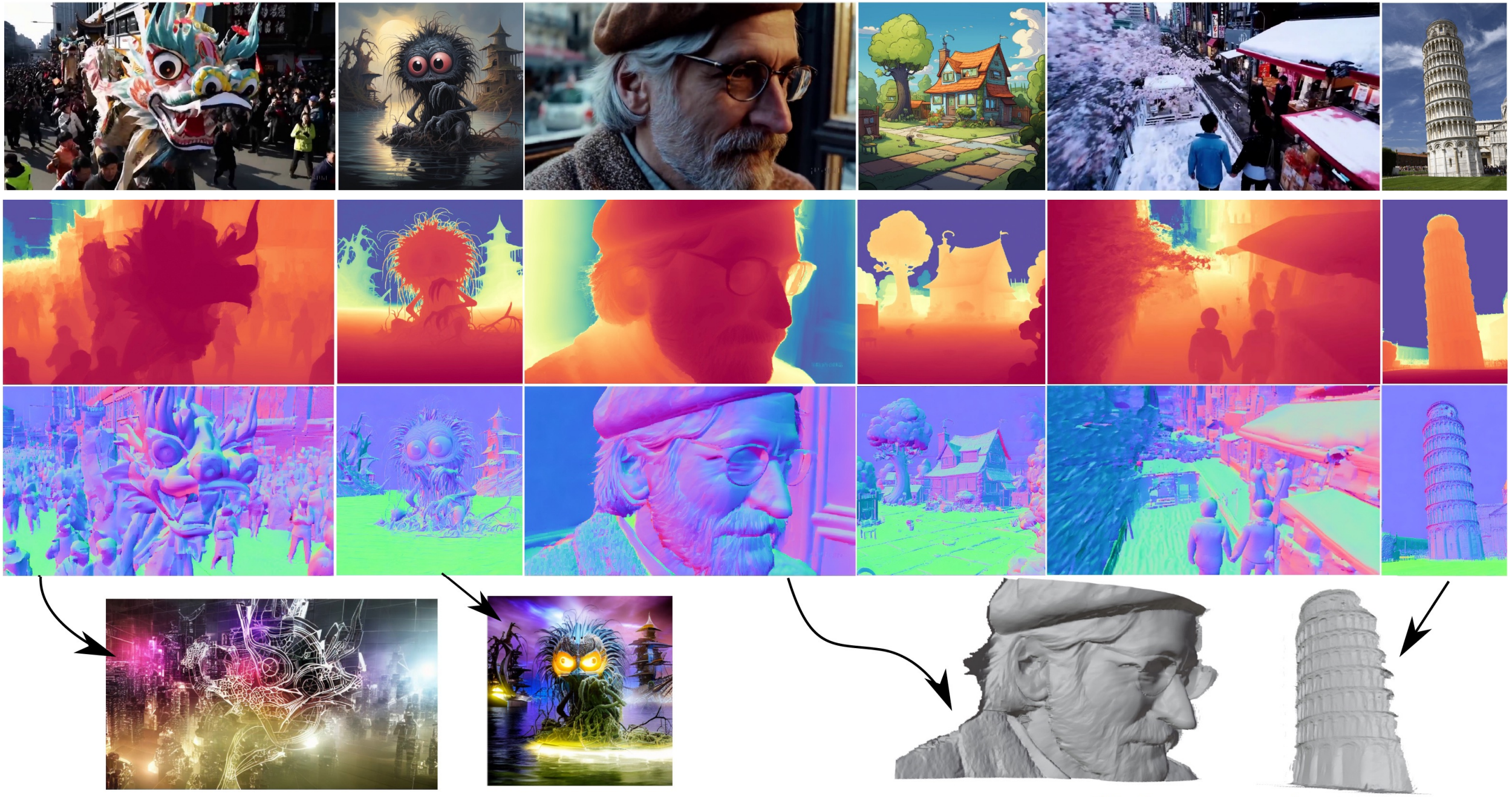
Indoor Scene

2) **Poor accuracy** : generate pseudo data using MVS reconstruction or self-training techniques





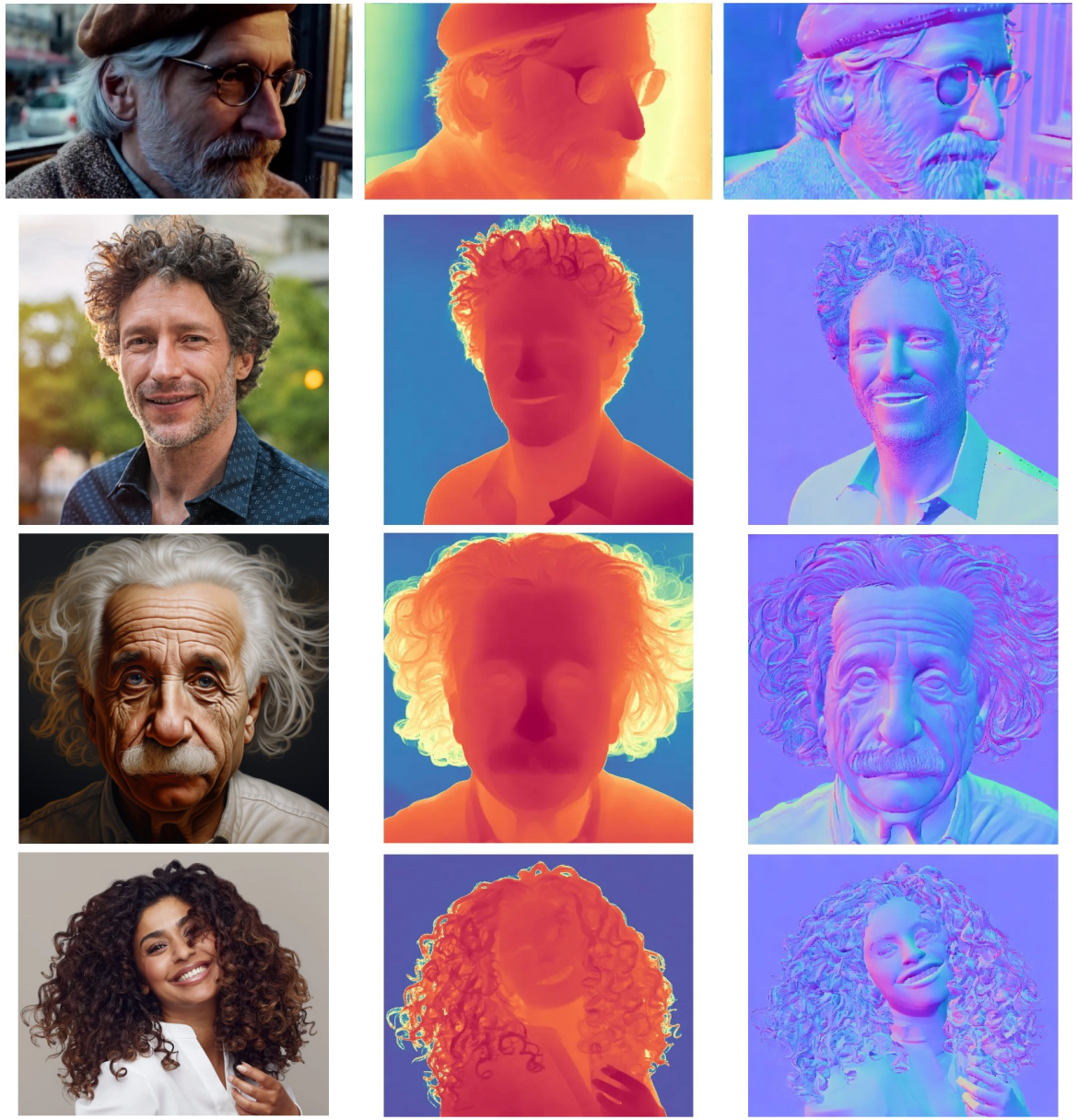
Incomplete / Low quality



Depth/Normal Guided Image Generation

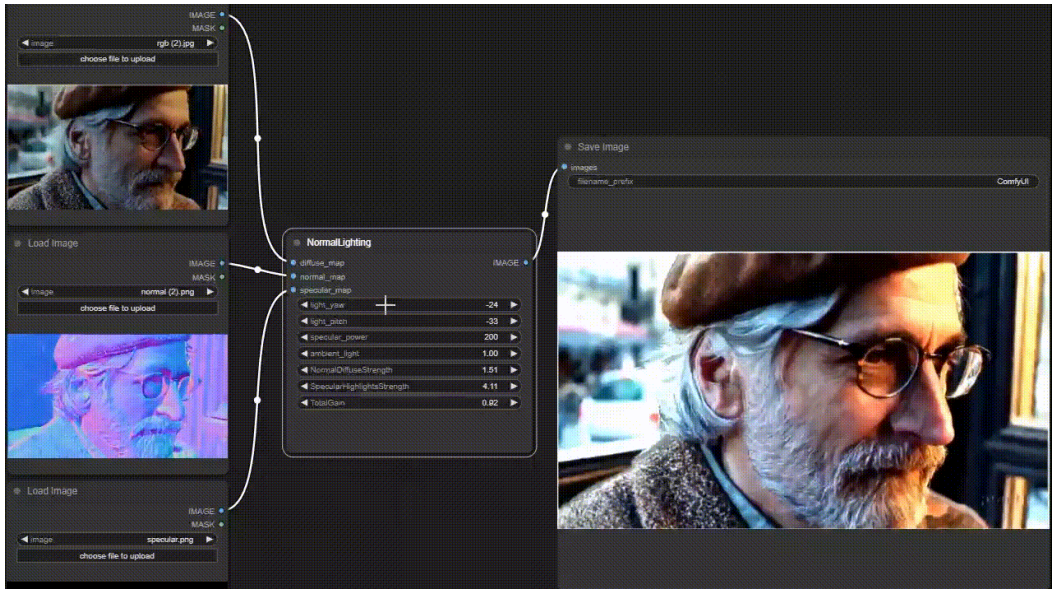
3D Reconstruction

Hair-level Detail



Point Cloud Rendering w/ Depth

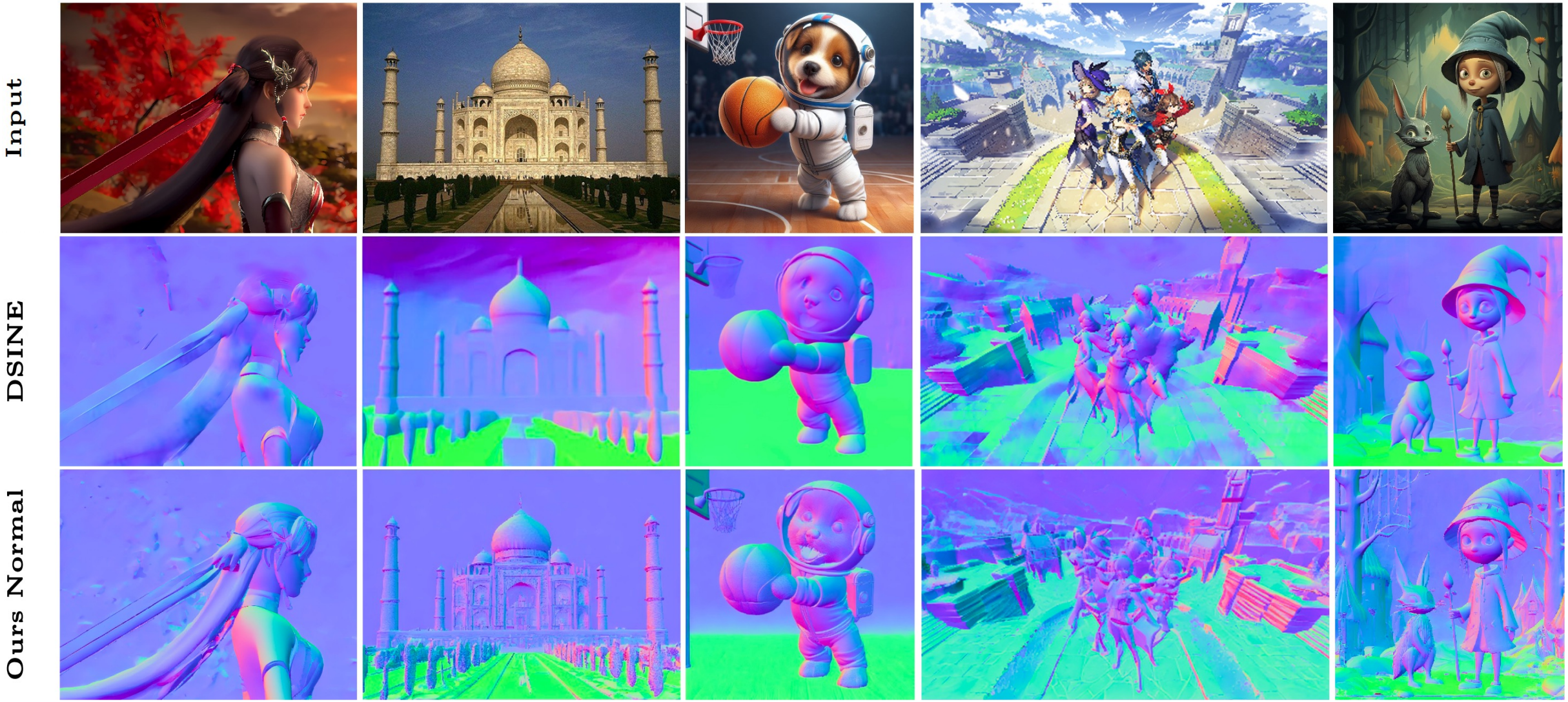
Image Relighting w/ Normal



Motivation : Latent diffusion models are better to capture **high-frequency details** and spatial 3D layouts than traditional CNN/transformer archtectures with **limited training data**.

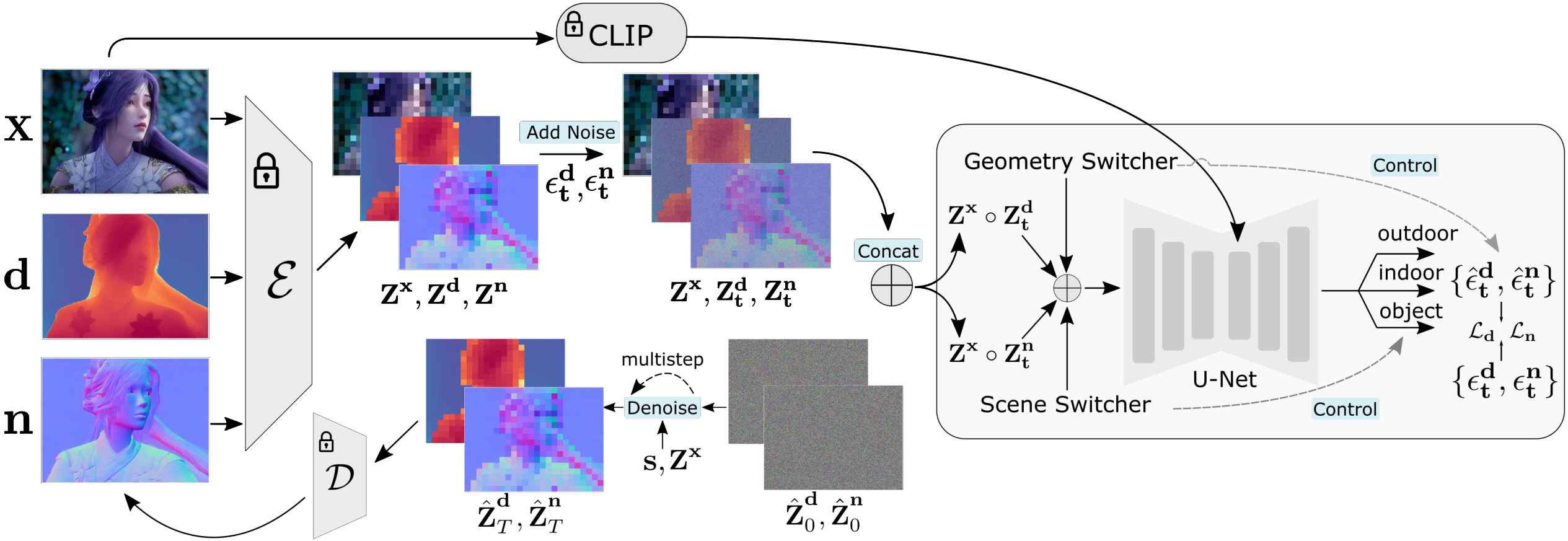
DepthAnything : CVPR'24

Motivation : Latent diffusion models are better to capture **high-frequency details** and spatial 3D layouts than traditional CNN/transformer archtectures with **limited training data**.



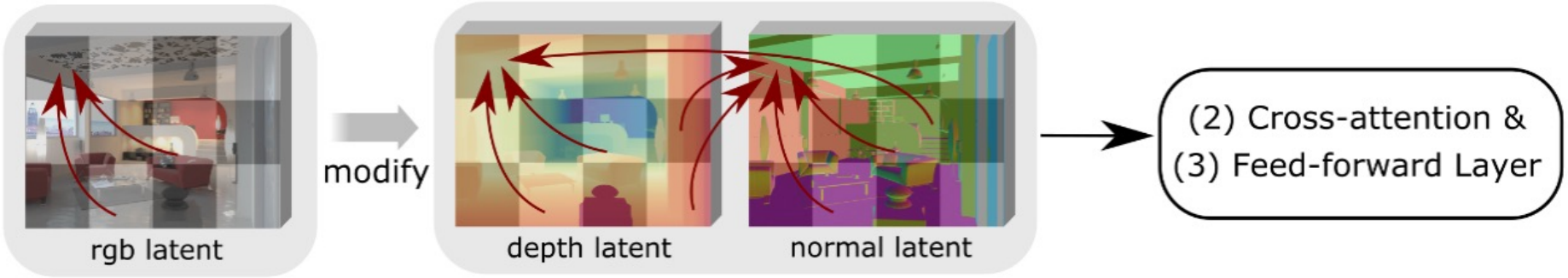
DSINE : CVPR'24

Methodology

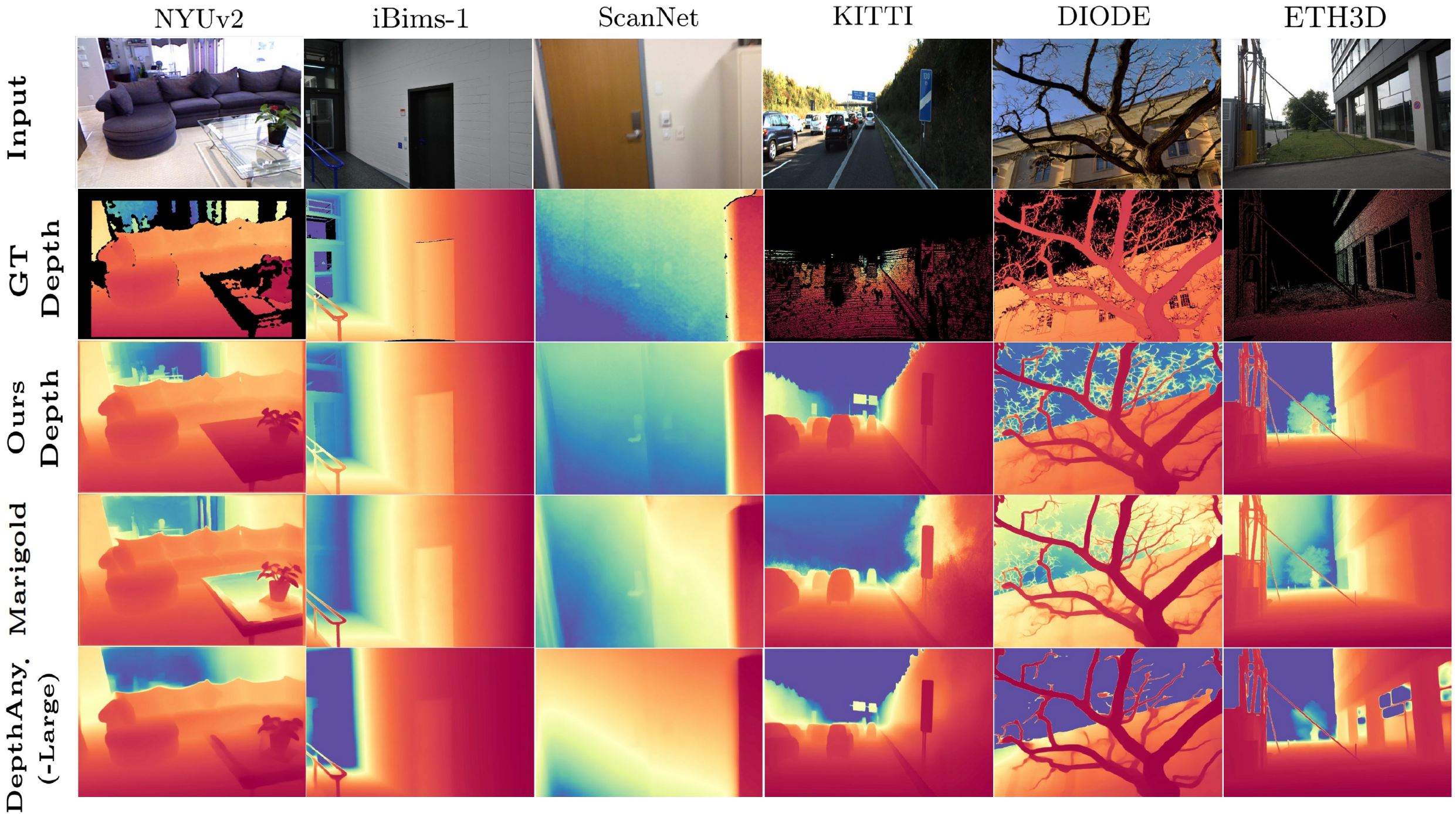


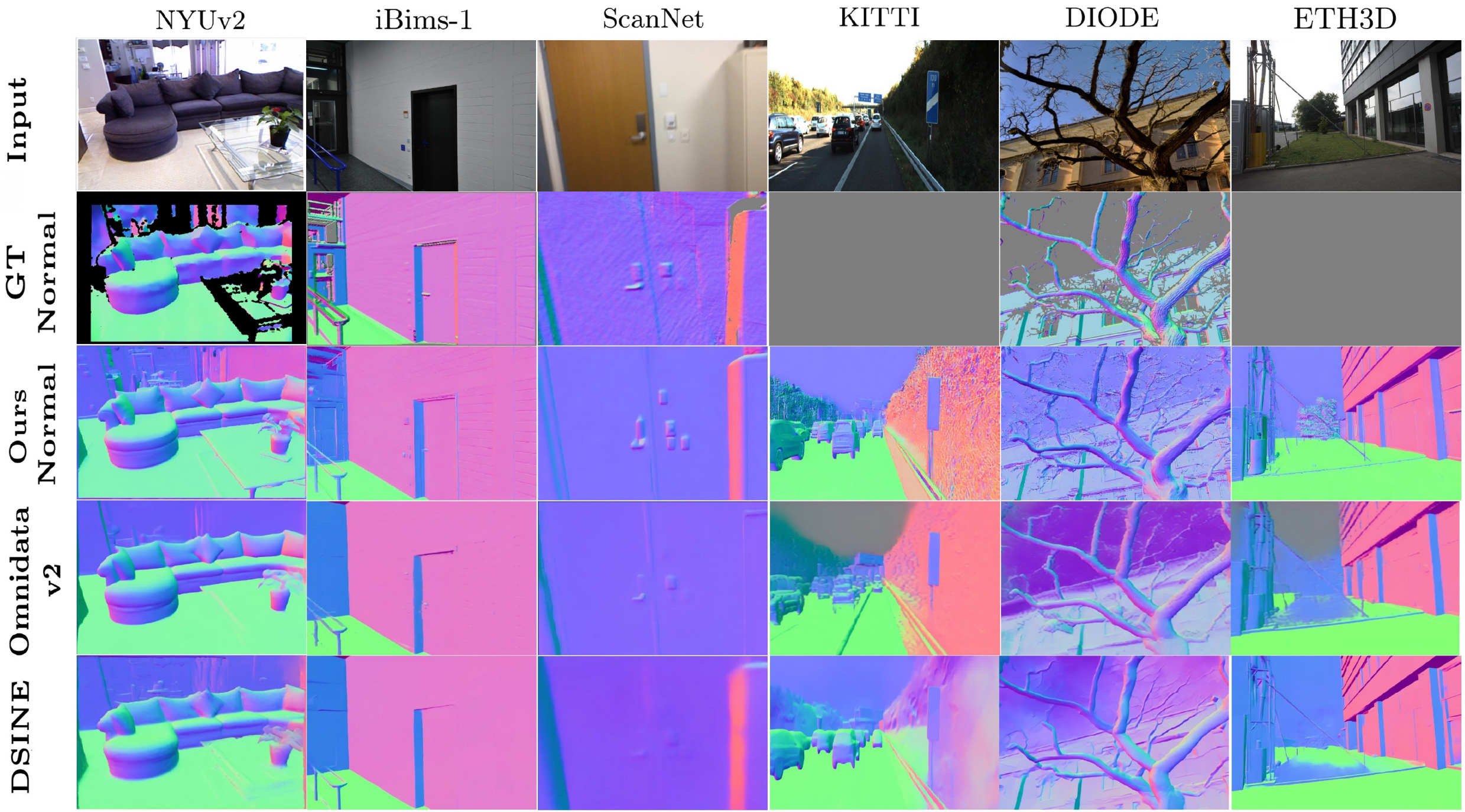
Contribution

(1) Geometric Domain Switcher

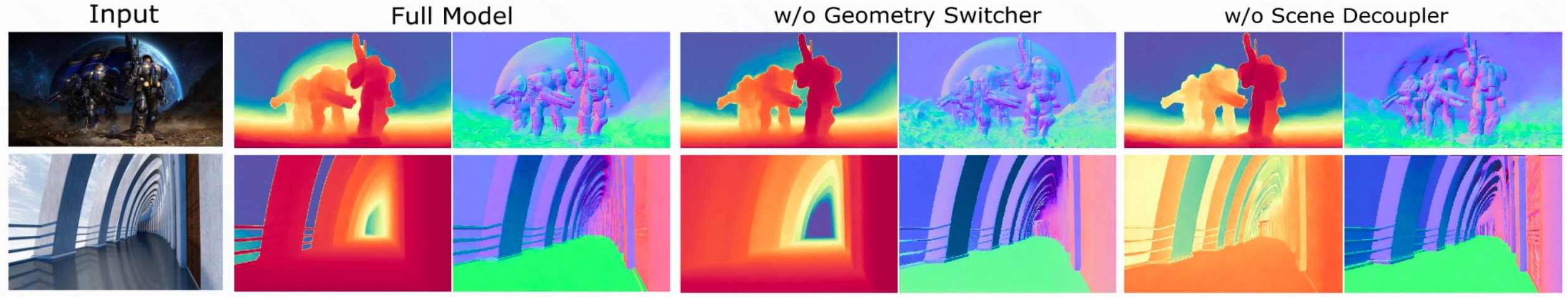


Dataset	Synthetic/Real	domain	Original Num.	Filtered Num.
Hypersim	synthetic	Indoor	~74,000	25,463
Replica	synthetic	Indoor	~104,000	50,884
3D-ken-burns	synthetic	Outdoor	~76,000	76,048
Simulation_disparity	synthetic	Outdoor	~ 40,000	39,630
G-buffer Objaverse	synthetic	Object	~32,00,000	85,997


NYUv2
iBims-1
ScanNet
KITTI
DIODE
ETH3D
Input
GT Depth
Ours Depth
Marigold
DepthAny. (-Large)

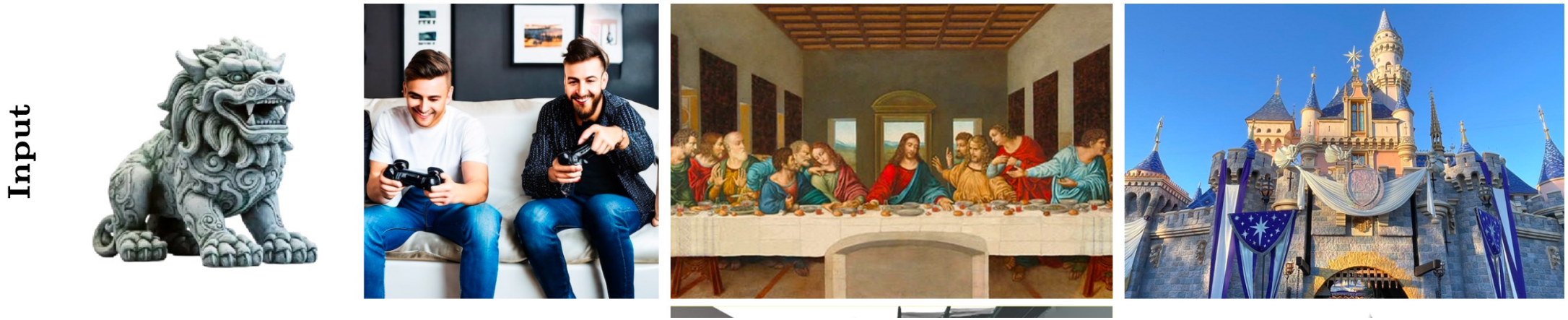

NYUv2
iBims-1
ScanNet
KITTI
DIODE
ETH3D
Input
GT Normal
Ours Normal
Omnidata v2
DSINE

Ablation Study



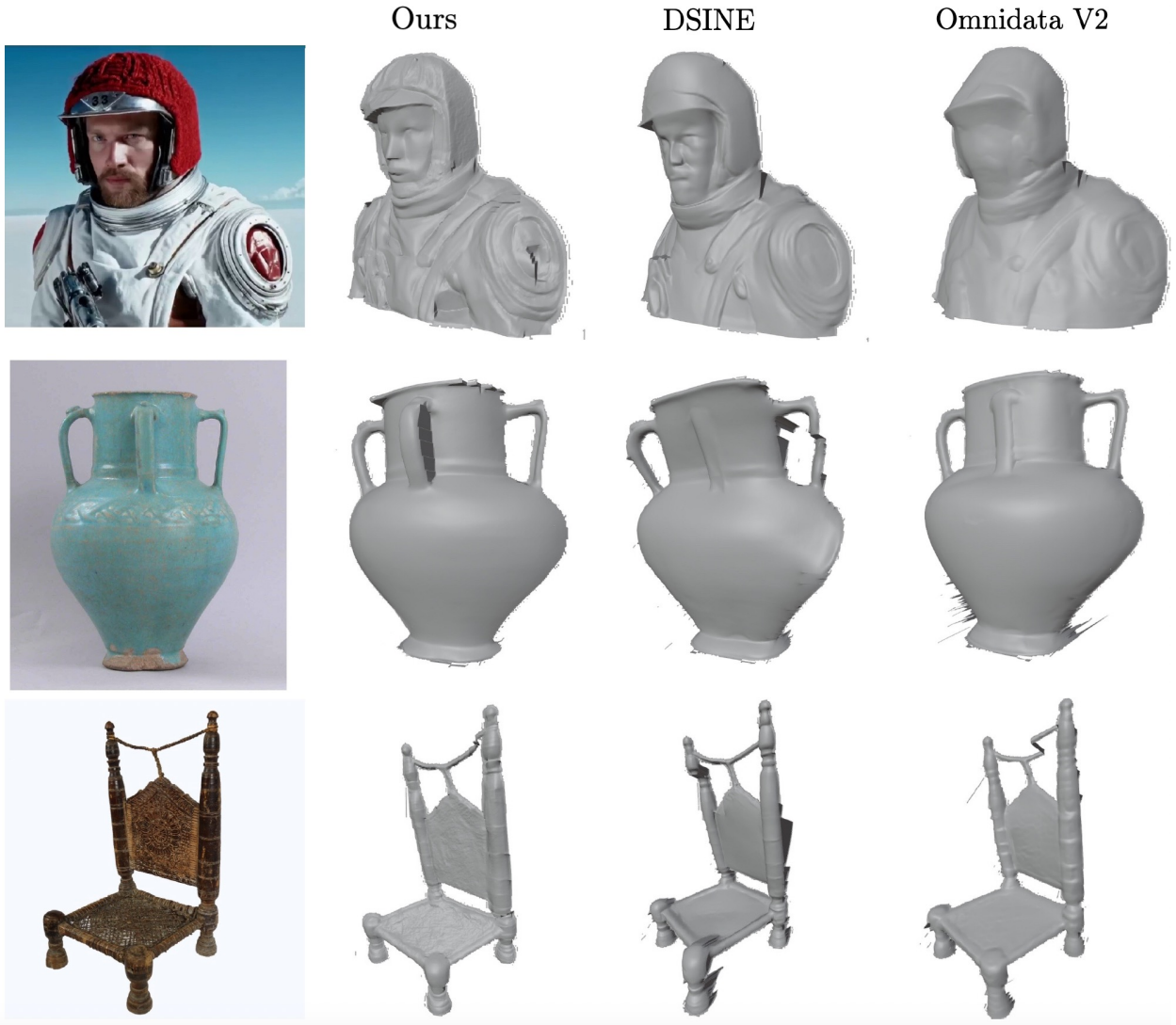
Method	Indoor			Outdoor			Object			Overall		
	AbsRel ↓	Mean ↓	GC ↓	AbsRel ↓	Mean ↓	GC ↓	AbsRel ↓	Mean ↓	GC ↓	AbsRel ↓	Mean ↓	GC ↓
Separate models	7.4	15.1	18.2	12.5	26.2	27.9	5.2	18.2	20.1	8.5	16.9	19.1
w/o Geometry Switcher	5.7	13.1	17.3	9.8	22.3	27.1	**3.3**	15.8	18.5	6.9	15.0	18.1
w/o Scene Decoupler	5.8	13.8	15.4	10.5	24.7	24.5	3.7	15.5	17.9	7.5	16.1	16.5
Full Model	**5.5**	**12.6**	14.7	**9.6**	**22.1**	**23.5**	3.5	**15.4**	**17.6**	**6.7**	**14.8**	**16.2**

Application 1 : 3D Reconstruction w/ Geometric Cues



Geometric Cues	Acc↓	Comp↓	C-$\mathcal{L}_1$ ↓	Prec↑	Recall ↑	F-score↑
Omnidata v2	0.035	0.048	0.042	79.9	68.1	73.3
DSINE	0.036	0.045	0.040	**80.1**	70.2	74.7
GeoWizard (Ours)	**0.033**	**0.042**	**0.038**	80.0	**70.7**	**75.1**

Application 1 : 3D Reconstruction w/ Geometric Cues



Application 2 : Novel View Synthesis

Application 3 : Image-based Generation

Ours Depth

Ours Normal

DepthAnything Depth

DSINE Normal

GeoWizard v2

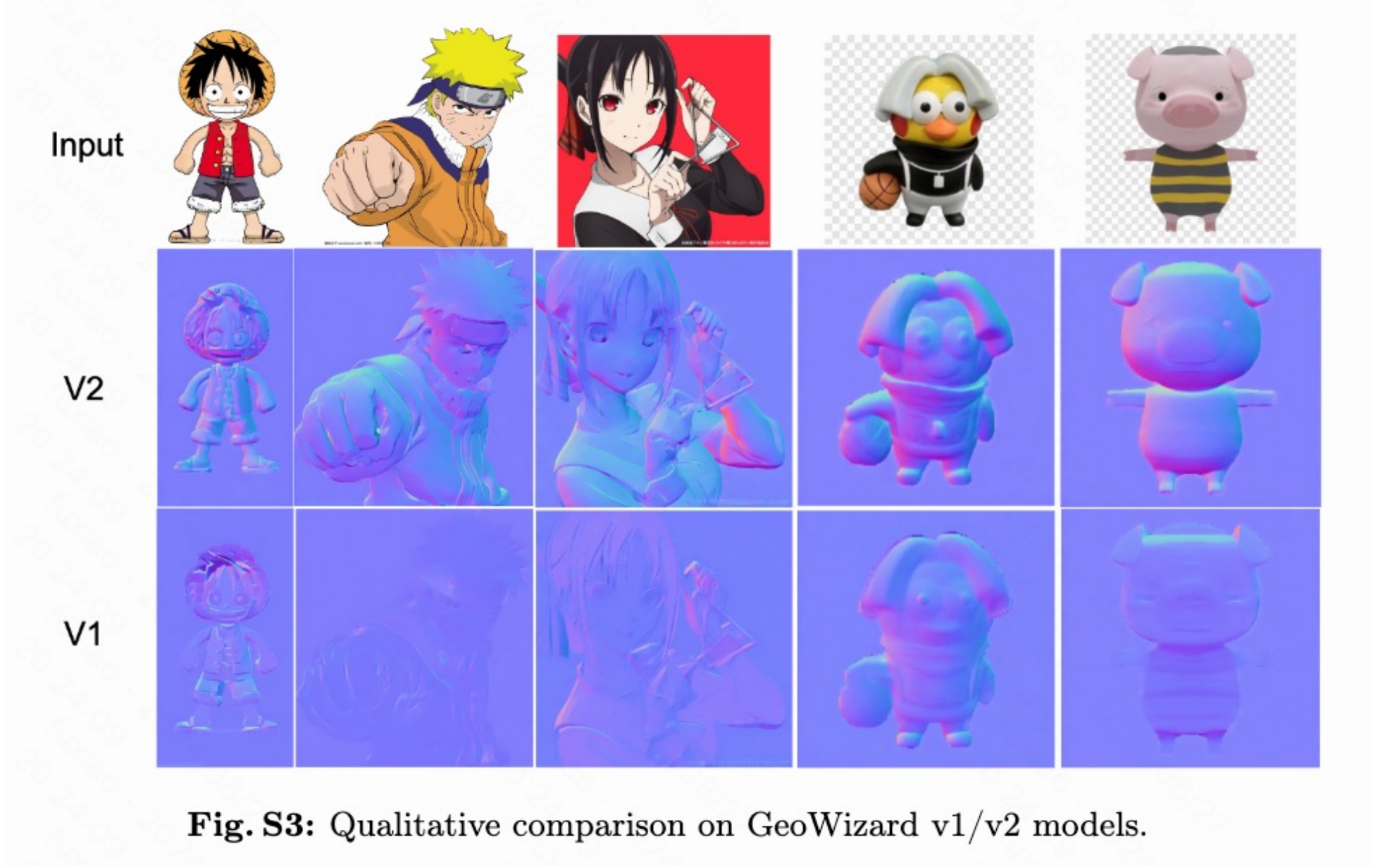


Fig. S3: Qualitative comparison on GeoWizard v1/v2 models.

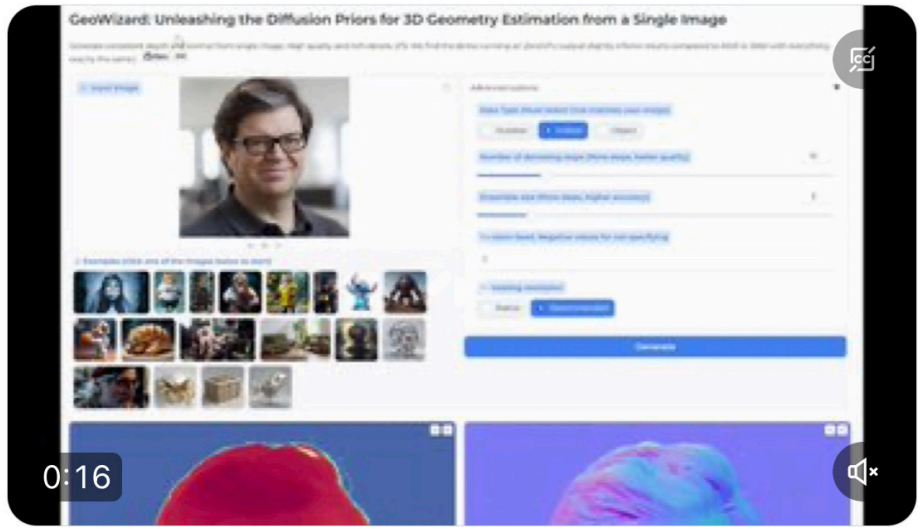
Social Impact

Code : https://github.com/fuxiao0719/GeoWizard

Gradio
@Gradio

🤩 GeoWizard is the coolest!

👉 Have you tried the free app on @huggingface Spaces yet?



0 views

07:06 · 27/03/2024 · **12,9K** Views

54 Likes **11** Retweets **1** Quote