JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

OmniObject3D: Large-Vocabulary 3D Object Dataset for Realistic Perception, Reconstruction and Generation


Tong Wu

Jiarui Zhang

Xiao Fu

Yuxin Wang

Jiawei Ren

Liang Pan

Wayne Wu

Lei Yang

Jiaqi Wang

Chen Qian

Dahua Lin ✉

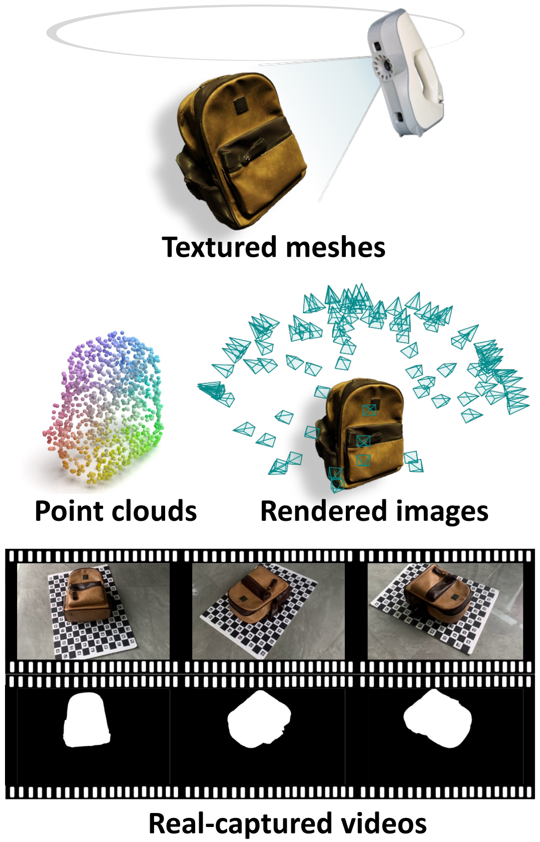
Ziwei Liu ✉

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Award Candidate

Paper tag: TUE-AM-076

Overview



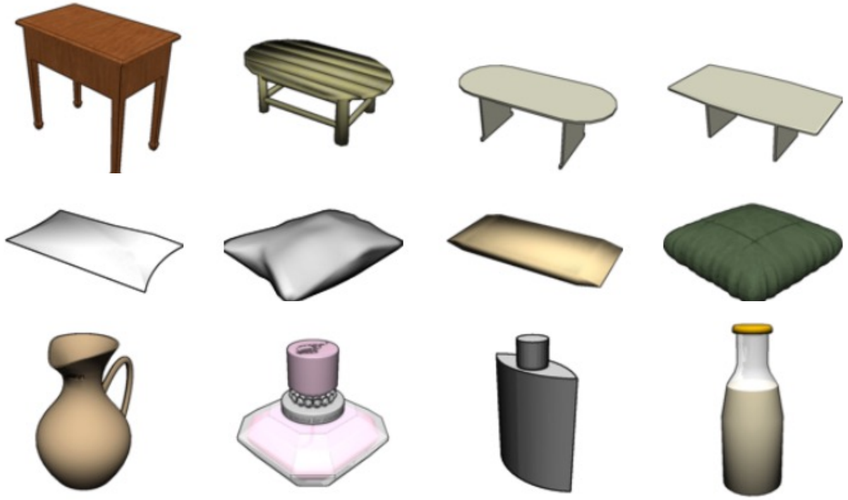
Perception | Novel View Synthesis | Surface Reconstruction | Generation

Background and motivation

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Synthetic data

ShapeNet
large in scale
low quality
not realistic



Multi-view images

CO3D
large in scale
No 3D GT

Real-world 3D scans

Google scanned objects
high quality
real-world scans
household objects

OmniObject3D

large-vocabulary
high quality
real-world scans

Statistics

	Dataset	Year	Real	Full 3D	Video	Num Objs	Num Cats
Synthetic data	ShapeNet	2015		√		51k	55
	ModelNet	2014		√		12k	40
	Objaverse	2023		√		818k	21k
	3D-Future	2020		√		16k	34
	ABO	2021		√		8k	63
	Toys4K	2021		√		4k	105
Multi-view image	CO3D V1/V2	2021	√		√	19k/40k	50
	MVImgNet	2023	√		√	219k	238
Real-world 3D scans	DTU	2014	√	√		124	NA
	GSO	2021	√	√		1k	17
	AKB-48	2022	√	√		2k	48
	Ours	2022	√	√	√	**6k**	**190**

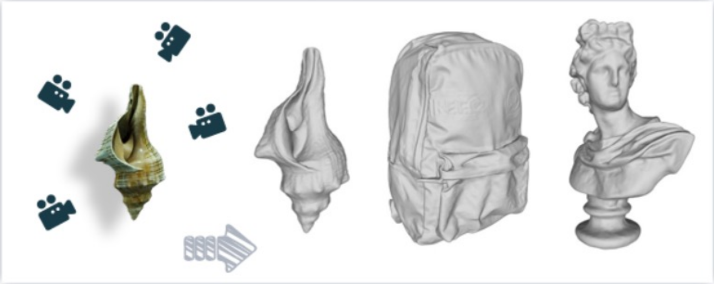
online assets with a variety of data types

Applications

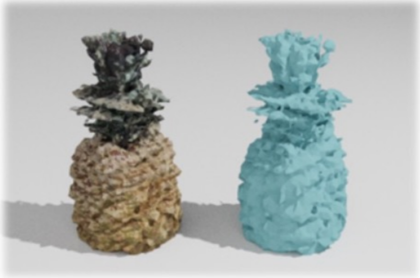
Perception

Novel View Synthesis

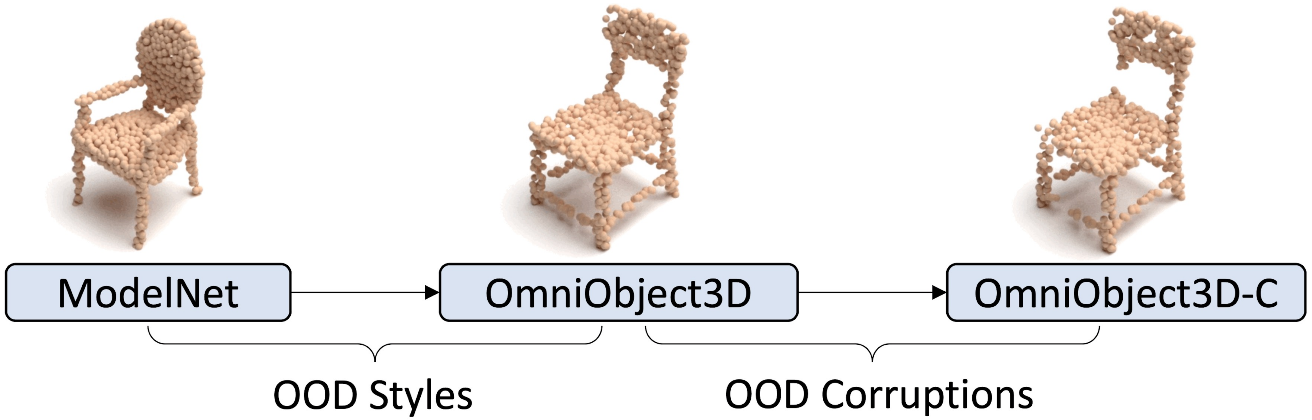
Surface Reconstruction



Generation

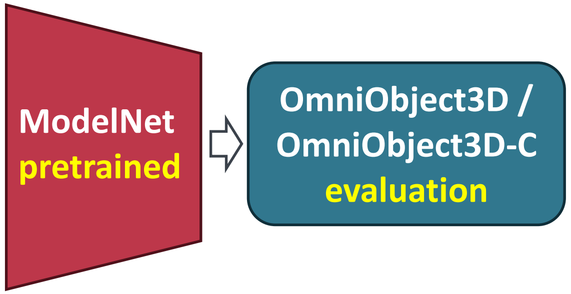


Robustness of point cloud classification



Differences between CAD models and real-scanned objects.

Common corruptions.



PointCloud-C (Ren et al. 2022)

Table 2. **Point cloud perception robustness analysis on OmniObject3D with different architecture designs.** Models are trained on the ModelNet-40 dataset, with OA_{Clean} to be their overall accuracy on the standard ModelNet-40 test set. OA_{Style} on OmniObject3D evaluates the robustness to OOD styles. mCE on the corrupted OmniObject3D-C evaluates the robustness to OOD corruptions. Blue shadings indicate rankings. †: results on ModelNet-C [75]. Full results are presented in the supplementary materials.

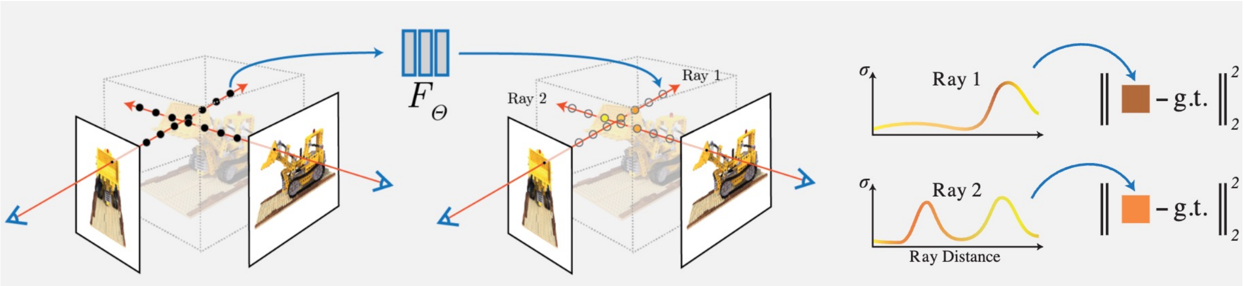
	$mCE^{\dagger}$ ↓	OA_{Clean} ↑	OA_{Style} ↑	mCE ↓
DGCNN [92]	1.000	0.926	0.448	1.000
PointNet [71]	1.422	0.907	0.466	0.969
PointNet++ [72]	1.072	0.930	0.407	1.066
RSCNN [51]	1.130	0.923	0.393	1.076
SimpleView [30]	1.047	**0.939**	0.476	0.990
GDANet [99]	<u>0.892</u>	0.934	<u>0.497</u>	**0.920**
PAConv [98]	1.104	0.936	0.403	1.073
CurveNet [97]	0.927	<u>0.938</u>	**0.500**	<u>0.929</u>
PCT [32]	0.925	0.930	0.459	0.940
RPC [75]	**0.863**	0.930	0.472	0.936

Novel view synthesis (two settings)

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

❑ *Single-scene optimization models*

Multi-view images from **one scene** ⇨ **Train** ⇨ **The same scene Inference**

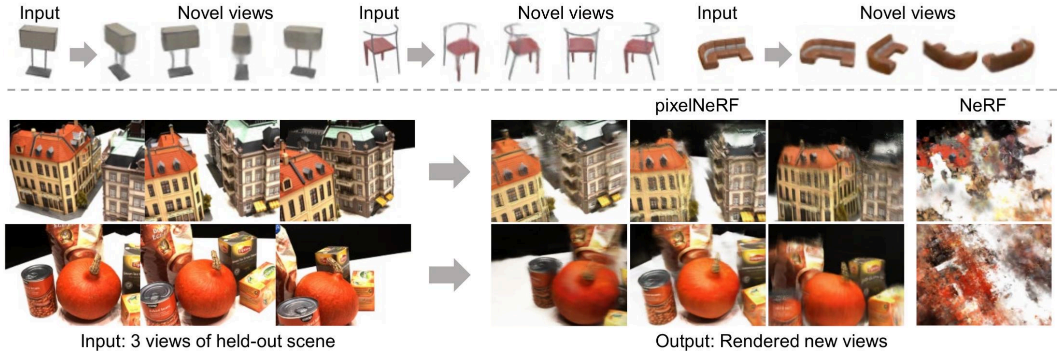


⇧

- NeRF (Mildenhall et al., 2021)
- Mip-NeRF (Barron et al., 2021)
- Plenoxels (Yu et al., 2021)

❑ *Generalizable models*

Multi-view images from scenes of **one category** or **across different categories**. ⇨ **Train** ⇨ **New scene with one or few views Inference**

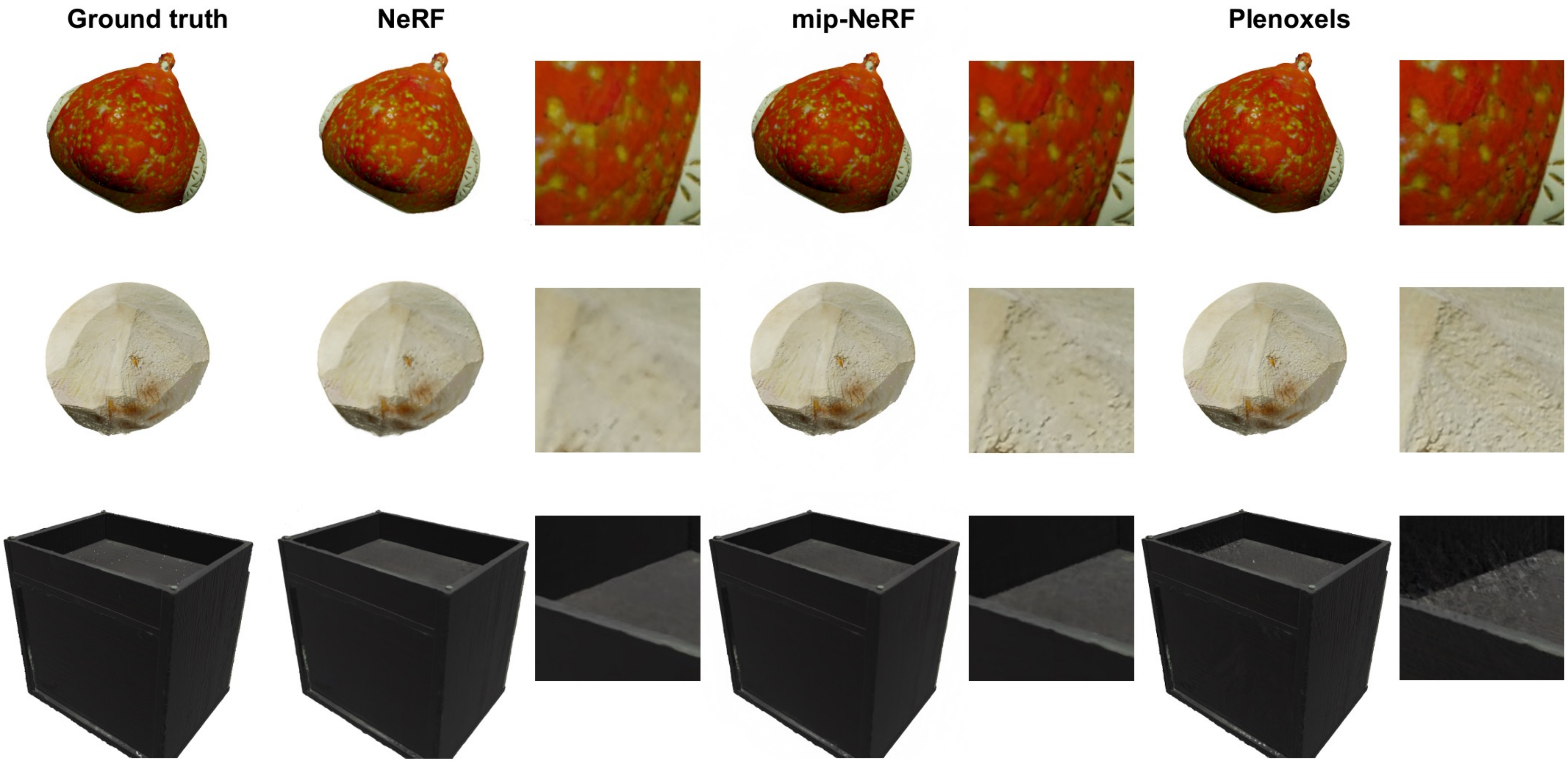


⇧

- pixelNeRF (Yu et al., 2021)
- MVSNeRF (Chen et al., 2021)
- IBRNet (Wang et al., 2021)

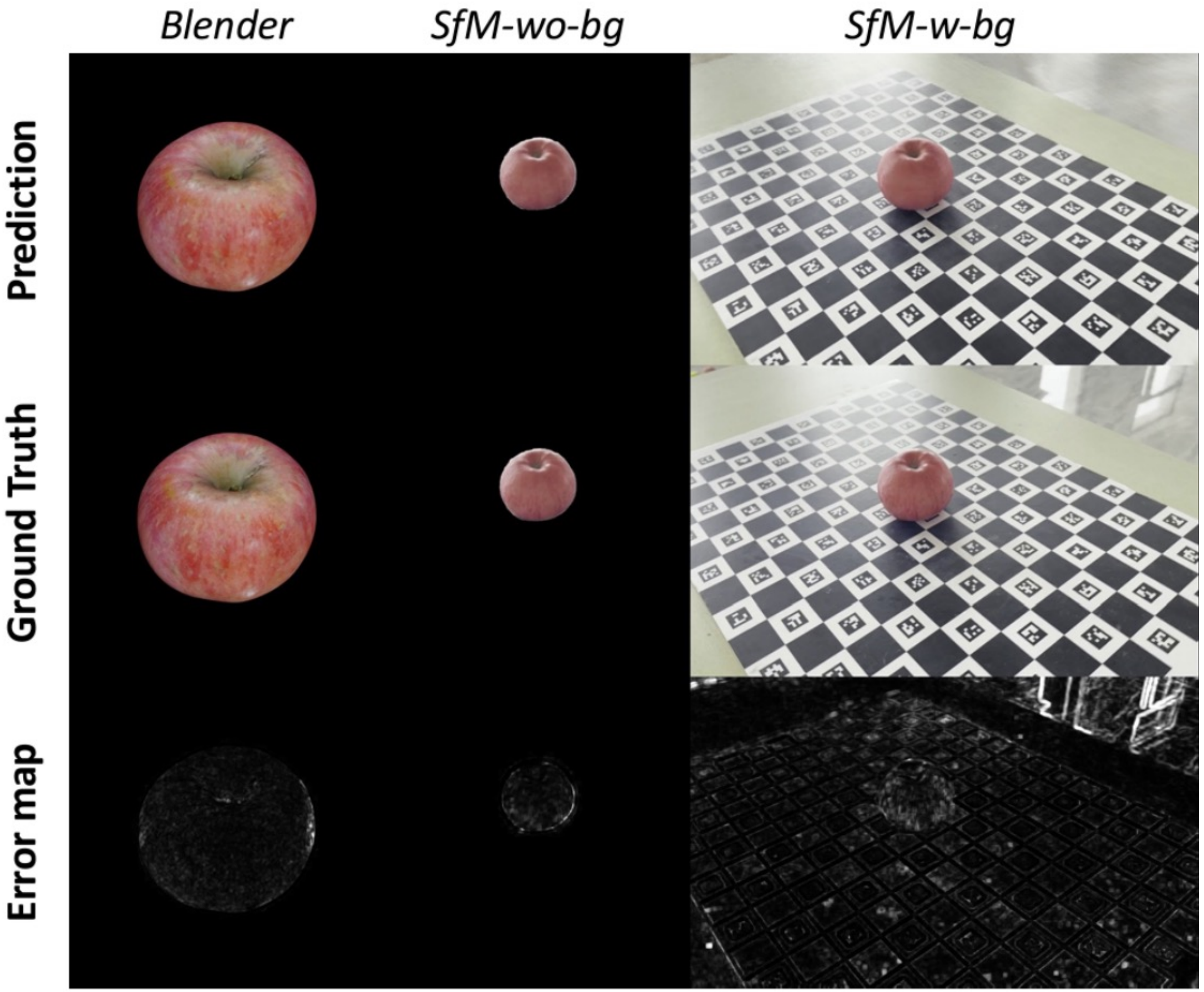
Novel view synthesis

❑ *Single-scene optimization models*



Novel view synthesis

- *Single-scene optimization models*



Novel view synthesis

- ❑ *Single-scene optimization models -> Results by mip-NeRF*

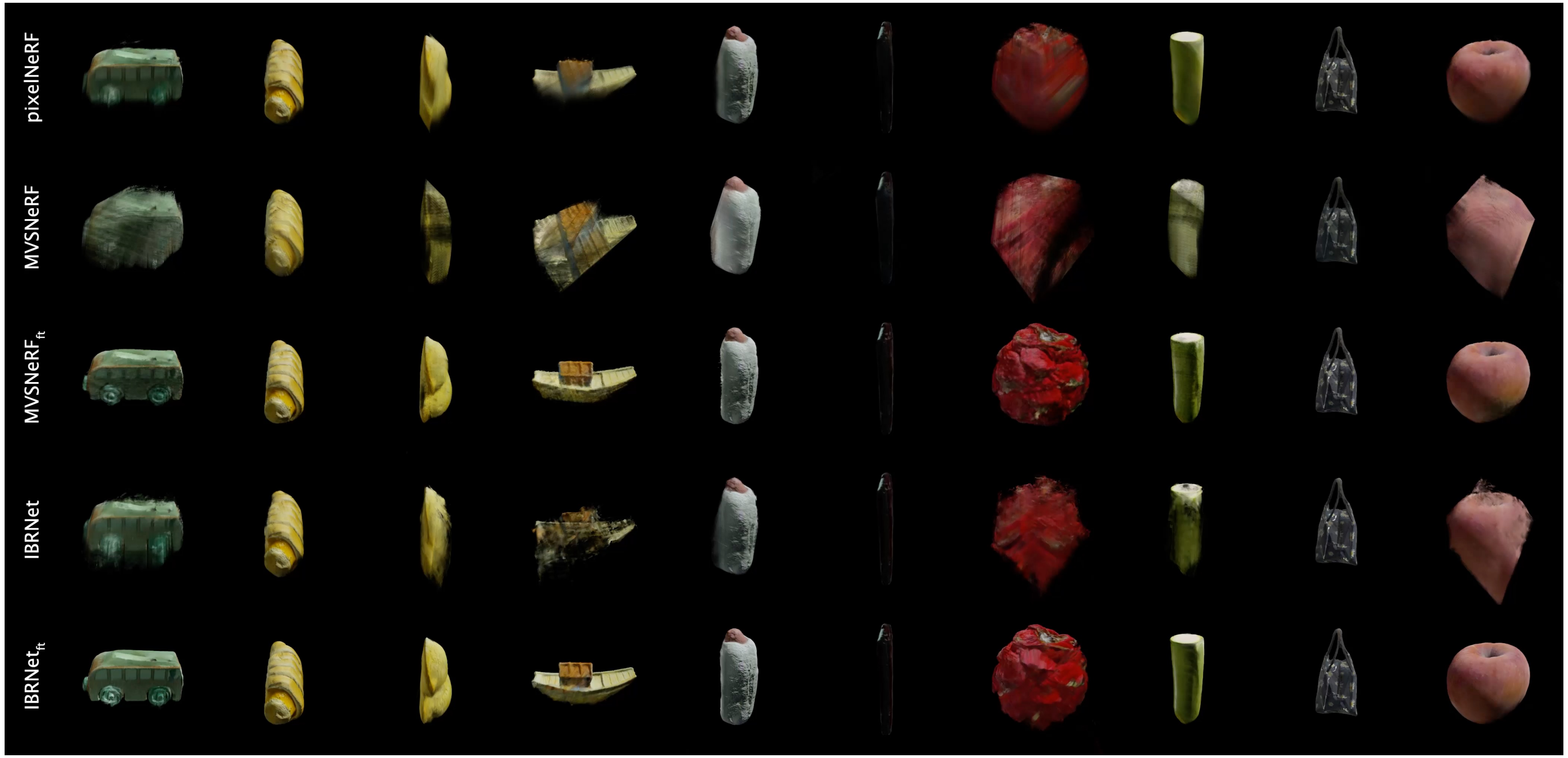
Novel view synthesis

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Generalizable models

Table 4. **Cross-scene novel view synthesis results on 10 categories.** 'Cat.' and 'All*' denote training on each category and training on all categories except the 10 test ones, respectively.

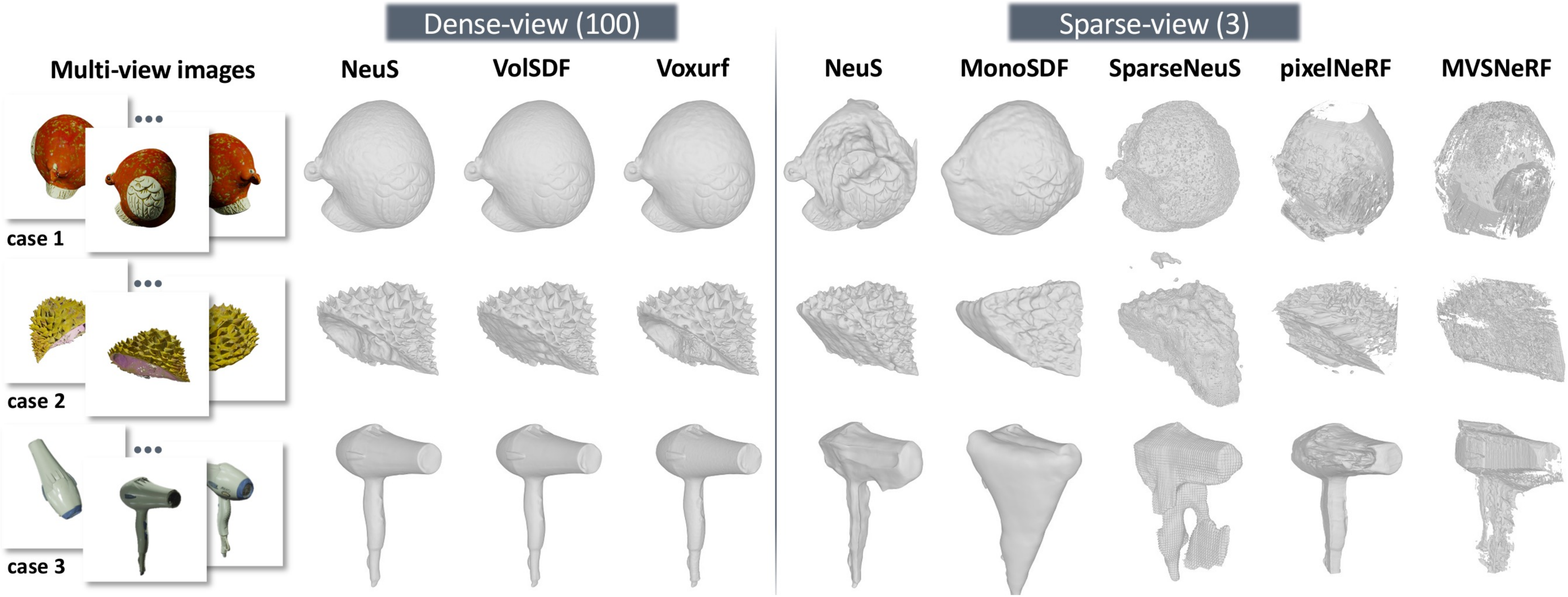
Method	Train	PSNR (↑)	SSIM (↑)	LPIPS (↓)	$\mathcal{L}_1^{\text{depth}}$ (↓)
MVSNeRF [11]	All*	17.49	0.544	0.442	0.193
	Cat.	17.54	0.542	0.448	0.230
	All*-ft.	25.70	0.754	0.251	0.081
	Cat.-ft.	25.52	0.750	0.264	**0.076**
IBRNet [91]	All*	19.39	0.569	0.399	0.423
	Cat.	19.03	0.551	0.415	0.290
	All*-ft.	**26.89**	**0.792**	**0.215**	0.081
	Cat.-ft.	25.67	0.760	0.238	0.099
pixelNeRF [105]	All*	22.16	0.692	0.342	0.109
	Cat.	20.65	0.676	0.348	0.195



We show examples of cross-scene NVS by pixelNeRF, MVSNeRF, and IBRNet ***given 3 views*** *(ft denotes* ***fine-tuned with 10 views****).*

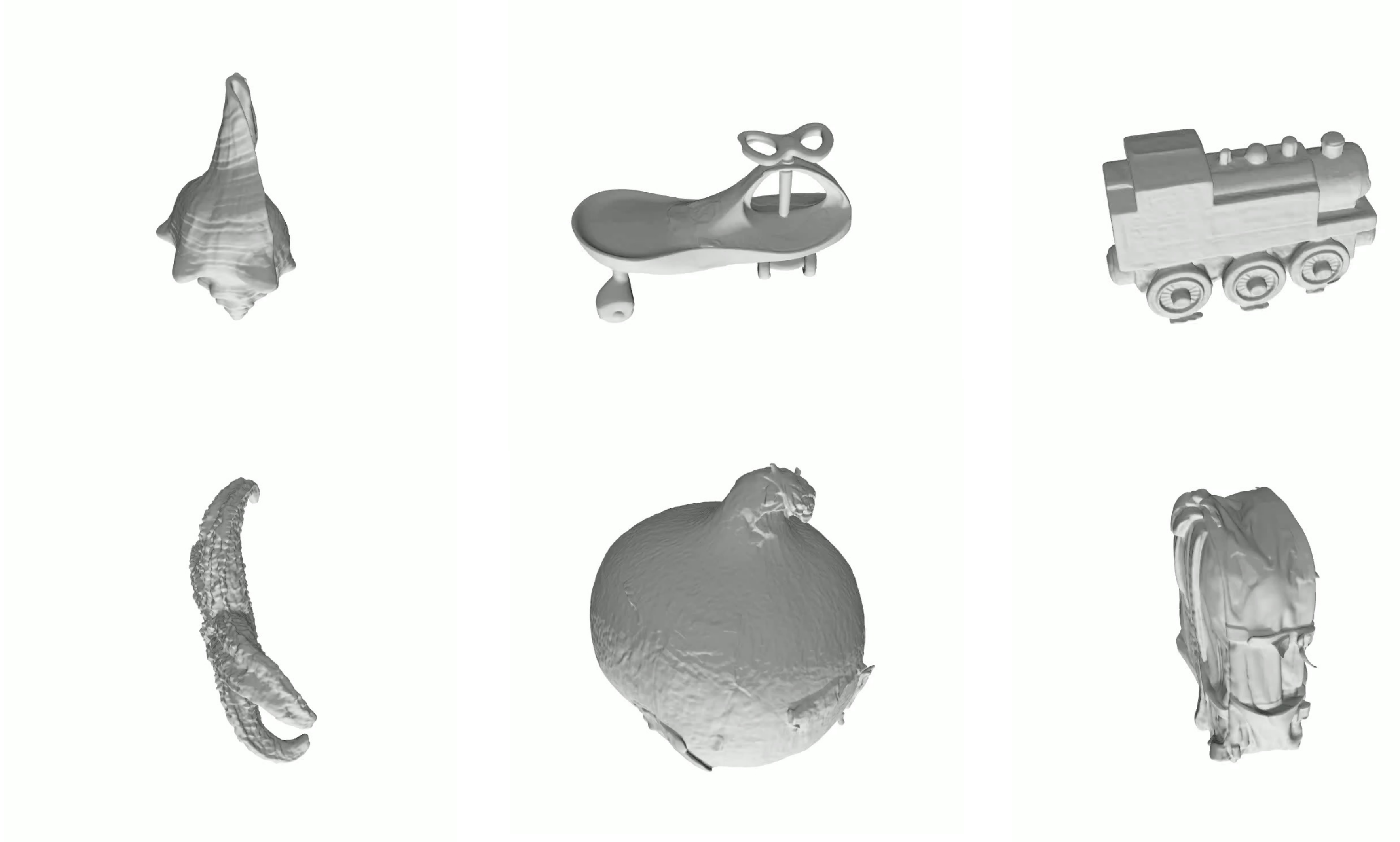
Surface reconstruction (two settings)

❑ *Multi-view image surface reconstruction*



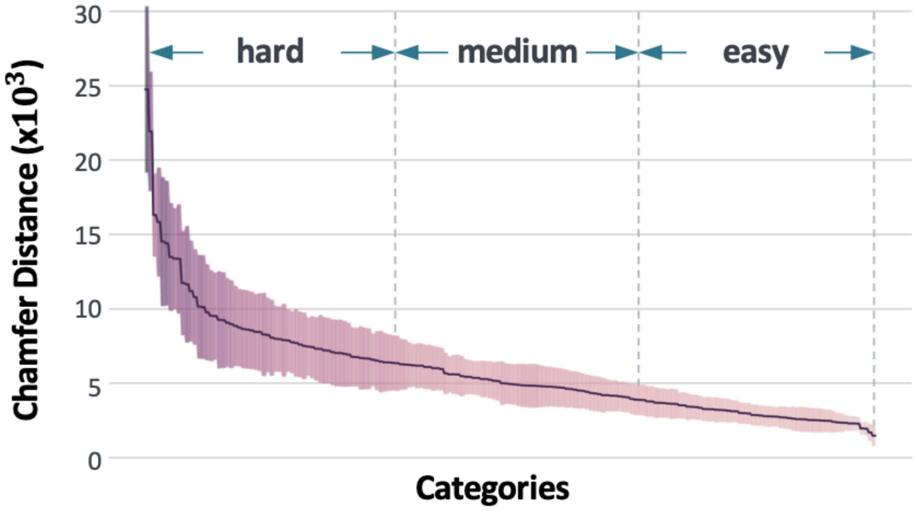
Surface reconstruction

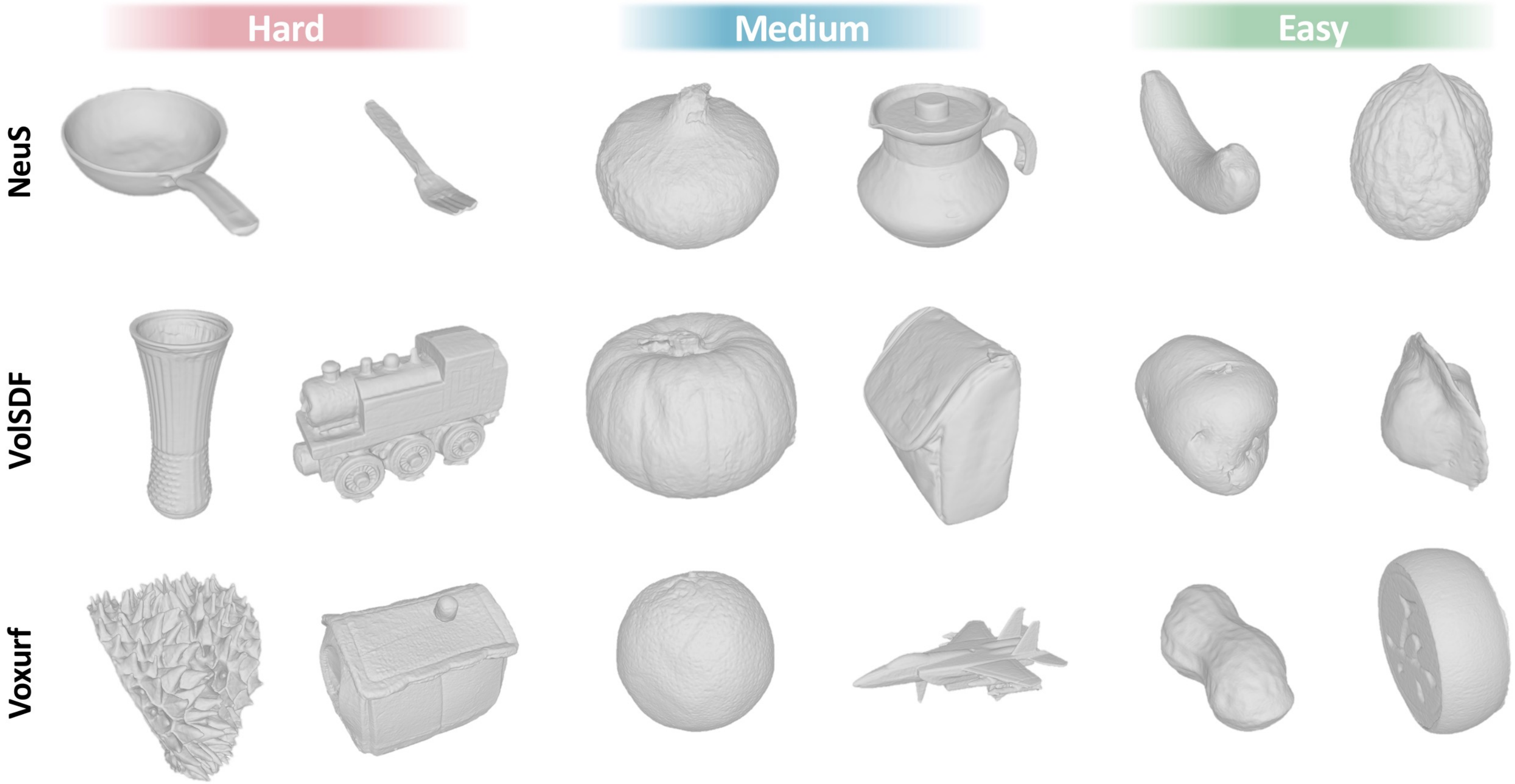
❑ *Multi-view image surface reconstruction (dense-view)*



Surface reconstruction

❑ *Multi-view image surface reconstruction (dense-views)*





Surface reconstruction

❑ *Multi-view image surface reconstruction (sparse-view)*

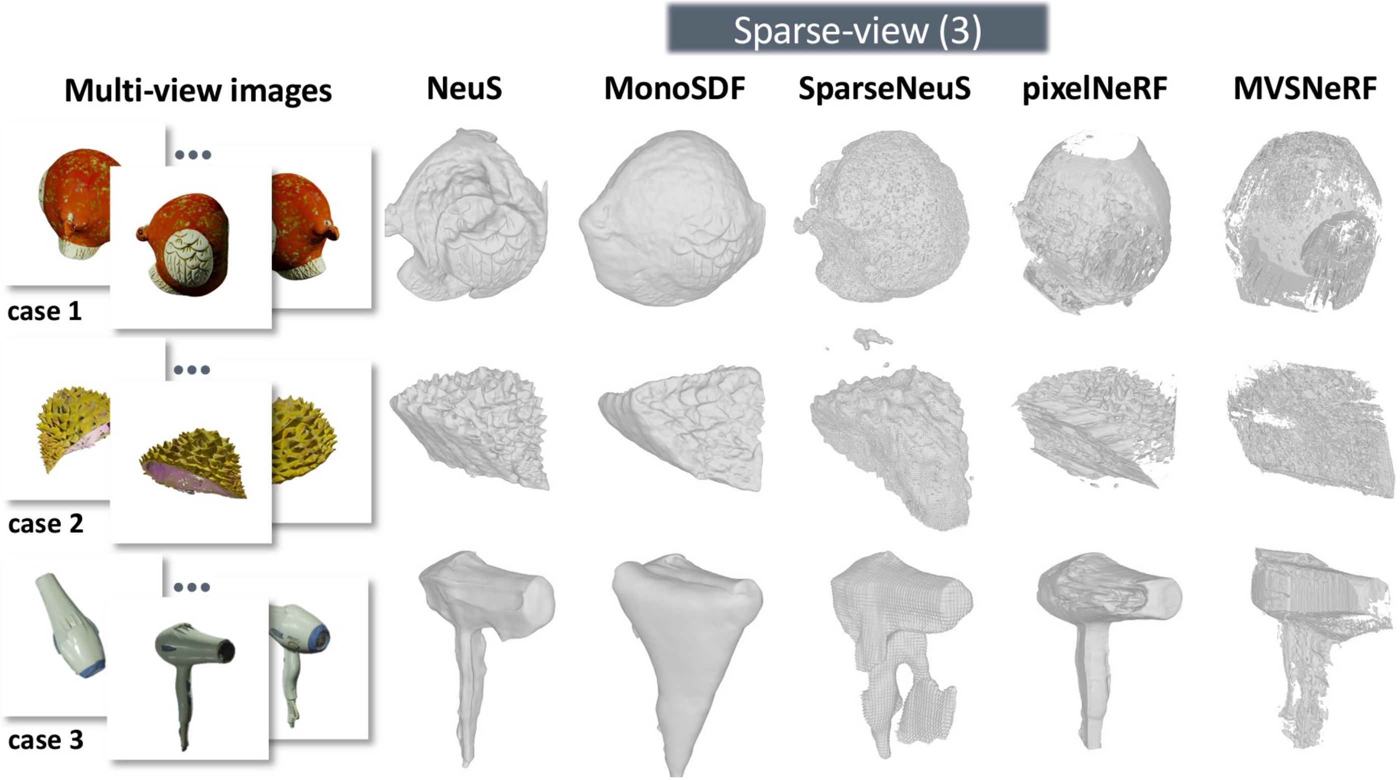


Table 6. **Sparse-view (3-view) surface reconstruction results.**

Method	Train	Chamfer Distance $\times 10^3$ ($\downarrow$) Hard	Medium	Easy	Avg
NeuS [90]	Single	29.35	27.62	24.79	27.3
MonoSDF [106]	Single	35.14	35.35	32.76	34.68
SparseNeuS [54]	1 cat.	34.05	31.32	31.14	32.36
	10 cats.	30.75	30.11	28.37	29.87
	All cats.	**26.13**	**26.08**	**22.13**	**25.00**
	Easy	28.39	26.65	23.76	26.48
	Medium	27.38	26.66	23.08	25.87
	Hard	27.42	26.95	24.63	26.47
MVSNeRF [11]	All cats.	56.68	48.09	48.70	51.16
pixelNeRF [105]	All cats.	63.31	59.91	61.47	61.56

3D Object Generation

❑ *3D object generation with textures*

Training data:
Multi-view images rendered by models from multiple categories

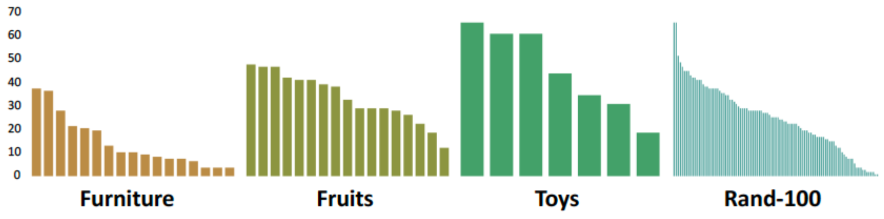
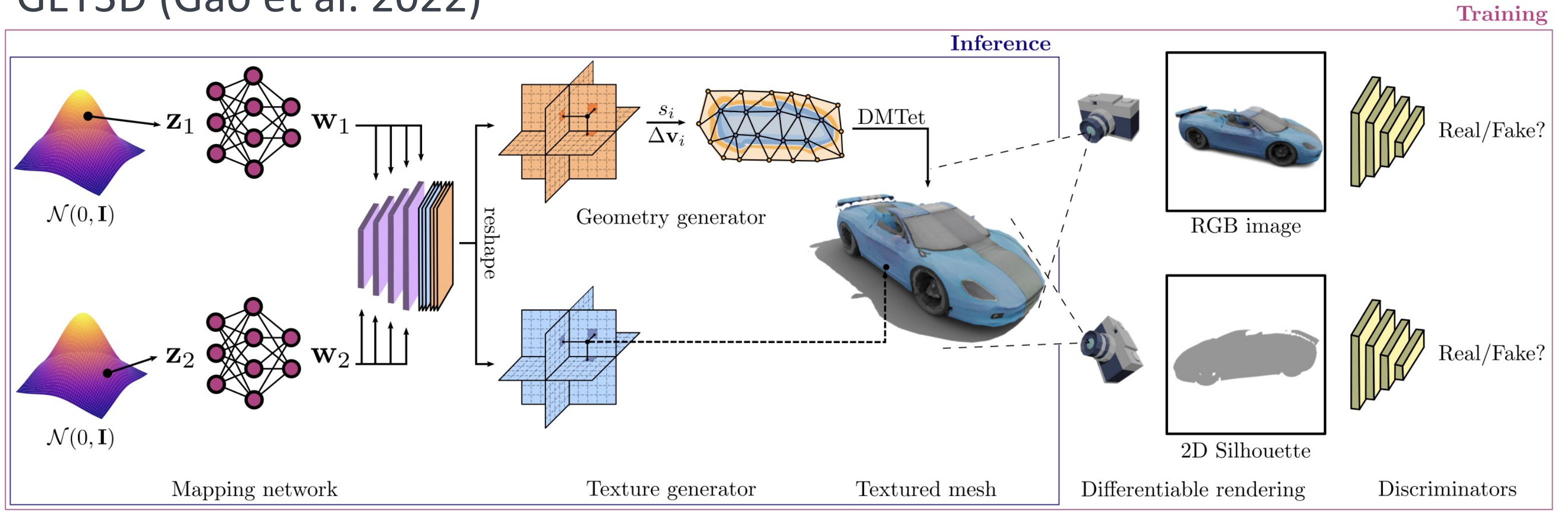


Figure S12. **Distributions of the four subsets.**

GET3D (Gao et al. 2022)



3D Object Generation

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

3D Object Generation

Interpolation across different categories

3D Object Generation

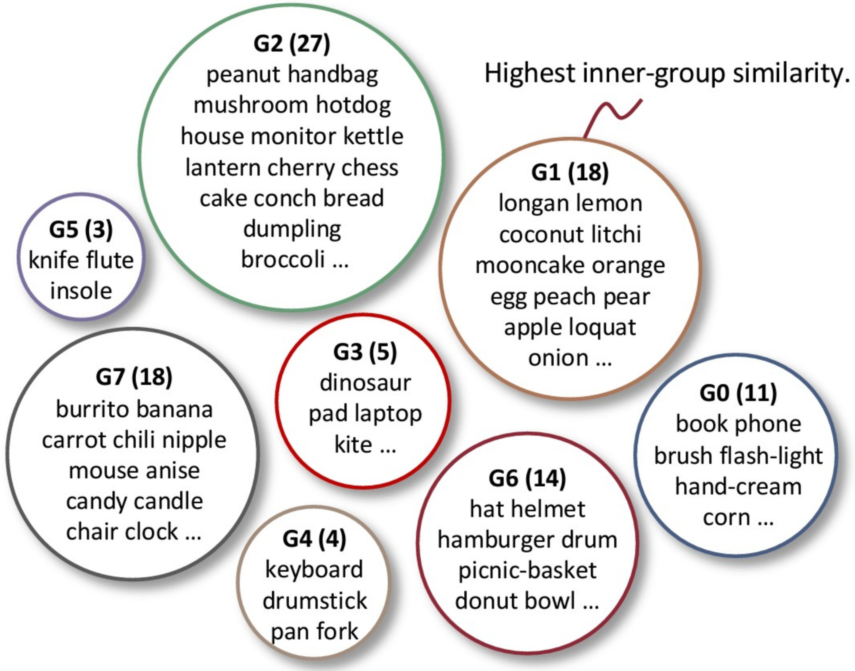
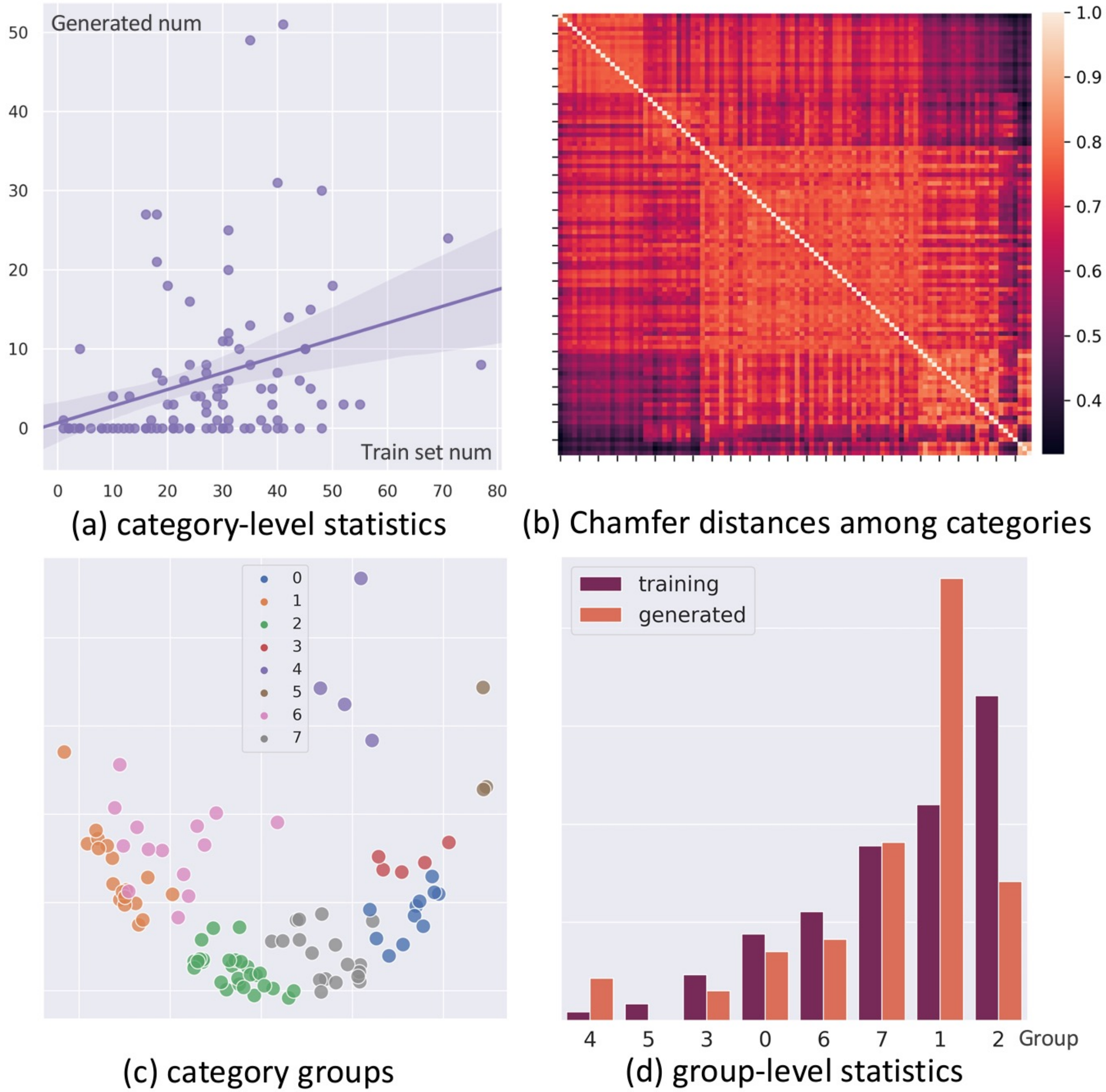


Figure S11. **Categories in each group after the KMeans clustering.** Categories in Group 1 are highly similar to each other, while those in Group 2 bear a high inner-group divergence.



Limitations and future works

❖ Data

- ***More data**: to support more extensive task requirements. Our data is still growing.*
- ***Broader distribution**: both domestically and internationally.*
- ***More modalities**: including language and various sensor types.*
- ***Higher complexity**: pushing beyond the limitations of 3D scanning technology.*

❖ Tasks

- *2D/3D detection; 6D pose estimation*
- *Human-object interaction*
- *Object in scene*

Limitations and future works

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thank you!

Project page