Geometry-Guided Modeling of Foundation Features Enables Generalizable Object Shape Deformation Learning

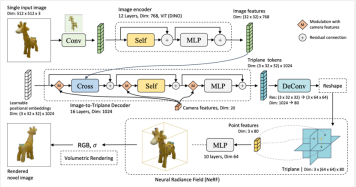
Yiyao Ma[1], Kai Chen[1], Zhongxiang Zhou[2], Zhuheng Song[1], Dongsheng Xie[1], Zelong Tan[1], Rong Xiong[2,3], Qi Dou[1]

[1]The Chinese University of Hong Kong
[2]Zhejiang Humanoid Robot Innovation Center Co., Ltd
[3]Zhejiang University

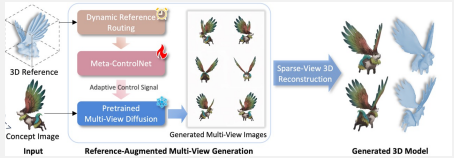
Monocular Shape Reconstruction

Recovering object shape from a monocular image is fundamental to geometric understanding and spatial reasoning.

3D Generative Methods

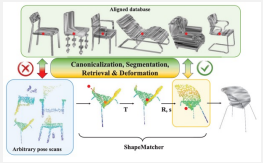


[OpenLRM. ICLR 2024]



[Phidias. ICLR 2025]

Deformation-based Methods



[ShapeMatcher. CVPR 2024]

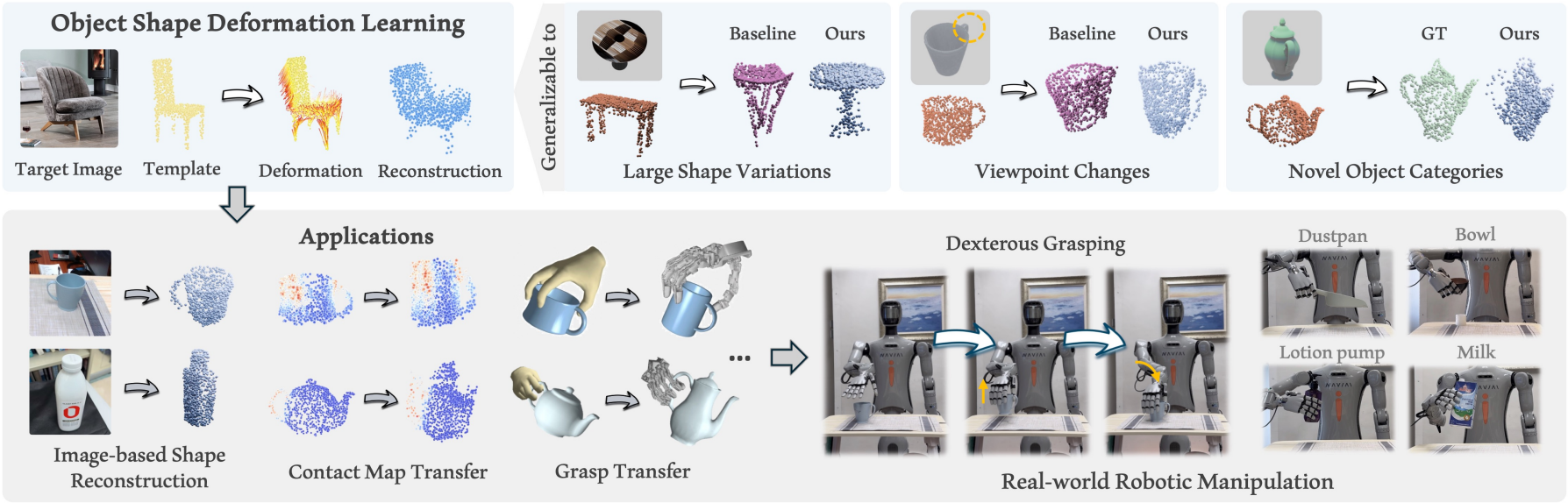
[KP-RED. CVPR 2024]

- ❑ 3D generative methods achieve high-fidelity reconstruction, but remain **sensitive to viewpoint changes** and **partial observability**, often hallucinating inconsistent geometry in occluded regions.
- ❑ Deformation-based methods provide structural priors through template topology, but struggle to generalize across **large shape variations** and **unseen categories** due to limited task-specific features.

Monocular Shape Reconstruction

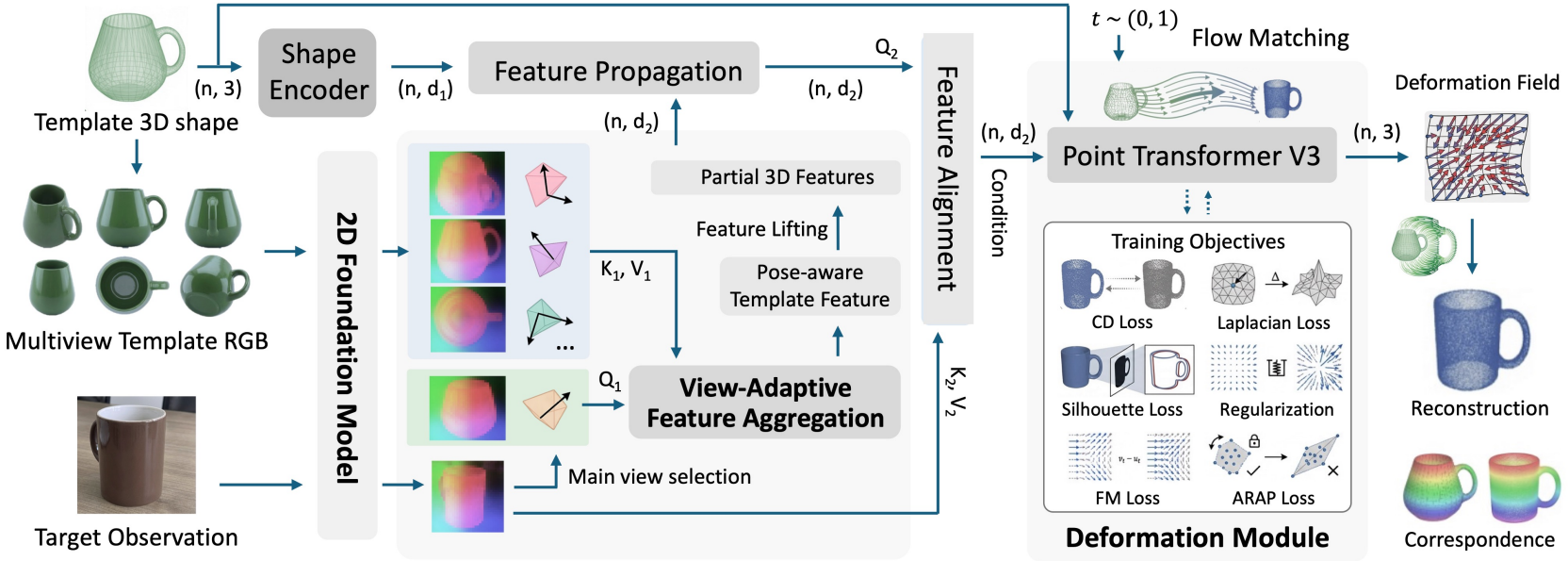
A Generalizable Template Deformation Framework

We present a generalizable deformation learning framework that leverages 2D foundation features to explicitly deform a category-level shape template for robust monocular 3D reconstruction.



Method Overview

- **Geometry-guided foundation feature modeling** for spatially aligned 2D-to-3D conditioning.
- **View-adaptive aggregation** for pose-aware and occlusion-robust feature representation.
- **Flow-matching deformation** for smooth and topology-preserving shape recovery.



Results

Comparison with 3D Generative Methods

- Our method is more robust to viewpoint variations and produces more accurate and efficient reconstructions than existing 3D generative methods.

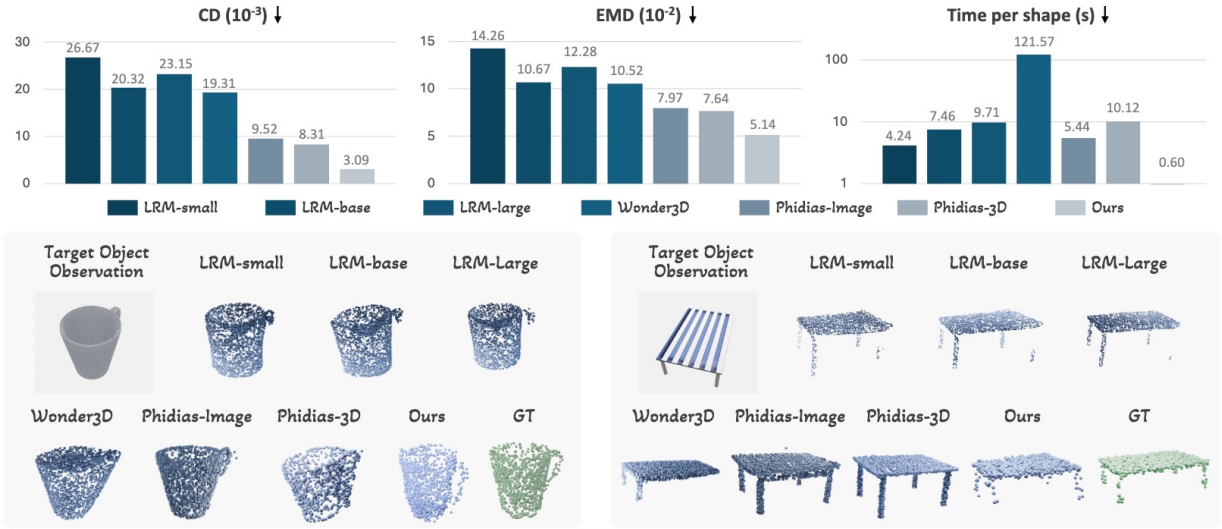
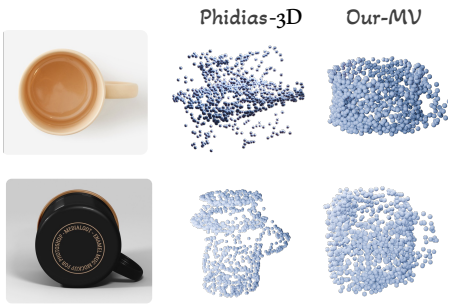


Figure 1. Quantitative and qualitative comparisons with existing 3D generative methods on single-view shape reconstruction.



[1]. LRM: Large Reconstruction Model for Single Image to 3D. ICLR, 2024.

[2]. Wonder3d: Single image to 3d using cross-domain diffusion. CVPR, 2024.

[3]. Phidias: A Generative Model for Creating 3D Content from Text, Image, and 3D Conditions with Reference-Augmented Diffusion. ICLR, 2025.

Results

Comparison with Deformation Learning Methods

- Our method demonstrates stronger deformation capability under large shape variations and better generalization to novel object categories.

Methods	Retrieved Template			Random Template		
	CD $(10^{-3})\downarrow$	EMD $(10^{-2})\downarrow$	S-IoU $(\%)\uparrow$	CD $(10^{-3})\downarrow$	EMD $(10^{-2})\downarrow$	S-IoU $(\%)\uparrow$
ShapeMatcher (Di et al., 2024)	5.92	6.43	40.47	13.02	8.82	34.36
KP-RED (Zhang et al., 2024)	3.05	5.23	46.73	5.10	6.35	42.05
Our-SV	2.45	4.76	48.45	2.61	4.94	46.78
Our-MV	**2.38**	**4.69**	**48.79**	**2.46**	**4.86**	**47.31**

Table 1. Comparison with state-of-the-art deformation learning methods.

Figure 1. Qualitative comparison with shape deformation methods on novel target observation.

[1]. Kp-red: Exploiting semantic keypoints for joint 3d shape retrieval and deformation. CVPR, 2024.
[2]. Shapematcher: Self-supervised joint shape canonicalization segmentation retrieval and deformation. CVPR, 2024.

Results

Ablation Study

- Each proposed module contributes to robust deformation learning, with the full model achieving the best performance on novel objects.

Methods	Retrieved Template			Random Template		
	CD $(10^{-3})\downarrow$	EMD $(10^{-2})\downarrow$	S-IoU (%) $\uparrow$	CD $(10^{-3})\downarrow$	EMD $(10^{-2})\downarrow$	S-IoU (%) $\uparrow$
w/o FM	2.66	4.87	45.78	2.74	4.95	43.57
w/o Prop.	2.74	4.96	44.19	2.95	5.18	41.10
w/o Rel.	2.56	4.85	45.36	2.70	5.06	44.67
w/o PrimSel.	2.64	4.87	44.50	2.84	5.08	44.40
w/o PoseAware.	2.60	4.87	44.98	2.79	5.06	44.47
Our-MV	**2.38**	**4.69**	**48.79**	**2.46**	**4.86**	**47.31**

Table 1. Quantitative ablation study results on novel objects.

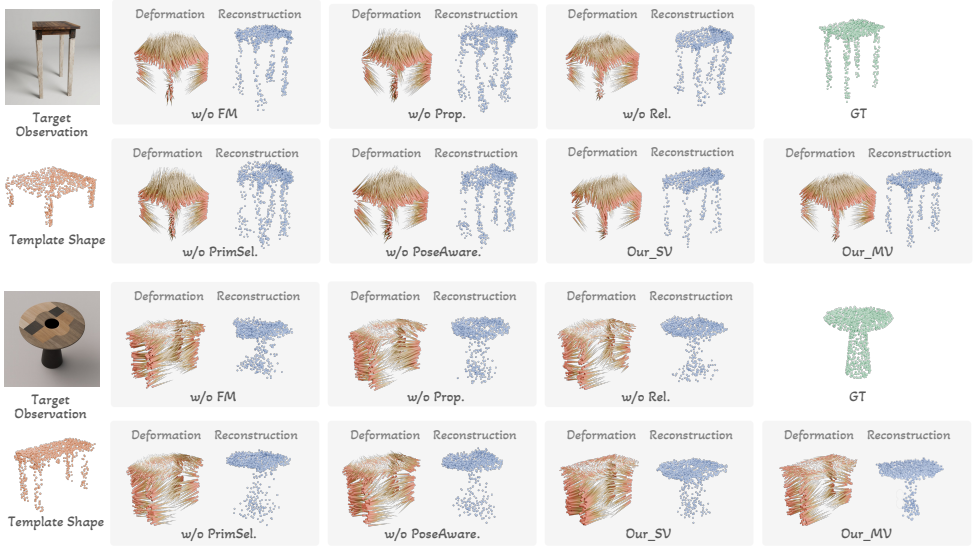


Figure 1. Qualitative comparison of deformation and reconstruction results under different ablation settings.

Results

Downstream Applications

- Our deformation fields preserve category-level dense correspondences, enabling efficient and generalizable transfer of contact maps and dexterous grasps to novel objects.

Methods	SR (%) ↑	Pen. (mm) ↓	Cov. (%) ↑	Time (s) ↓
Analytical Method				
DFC (Liu et al., 2021)	65.00%	8.01	30.07%	>1000
Generative Methods				
ConGen (Liu et al., 2023c)	47.59%	3.05	23.19%	17.1
UGG (Lu et al., 2024)	58.00%	8.37	**36.65%**	76.0
Transfer-Based Methods				
Tink (Yang et al., 2022)	61.96%	4.60	27.91%	87.6
cmtDiff (Ma et al., 2025)	69.59%	**2.70**	31.17%	62.2
Our-MV	**76.92%**	2.86	28.39%	**15.7**

Table 1. Object-centric deformation field enables transfer-based dexterous grasp generation.

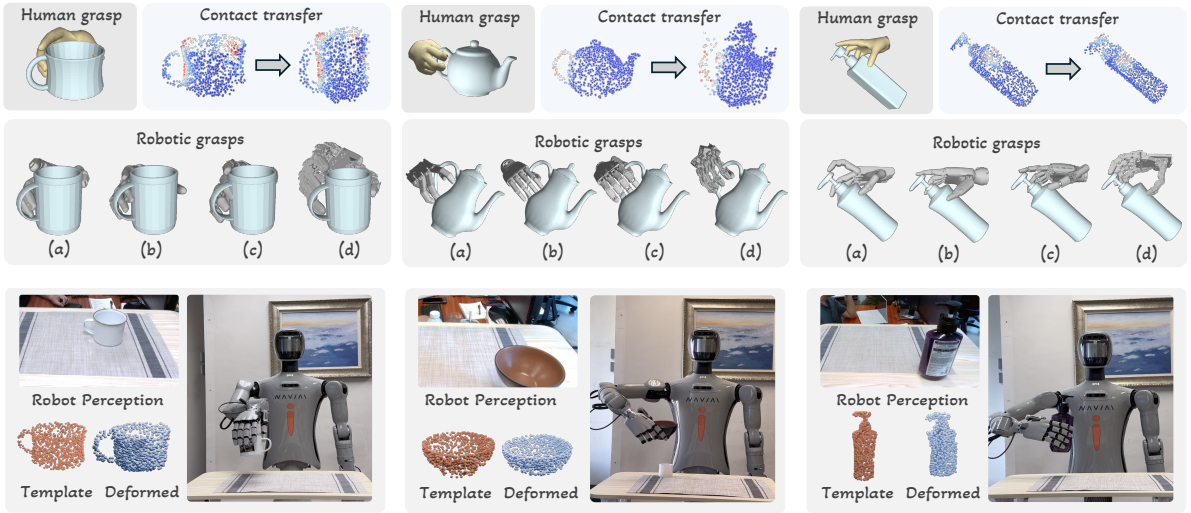


Figure 1. Our method facilitates generalizable dexterous manipulation for humanoid robots.

Conclusion

Conclusion

- A geometry-guided flow matching framework bridging **2D semantics and 3D priors** for single-view shape recovery.
- Robust across **arbitrary viewpoints**, **large shape variations**, and **unseen categories**.
- Facilitates generalizable **dexterous manipulation** for humanoid robots.

Future Works

- Extend the framework to multi-view target inputs for richer geometric cues and more accurate deformation learning.
- Incorporate semantic priors from vision-language models to better resolve ambiguous object structures.

浙江人形机器人创新中心
ZHEJIANG INNOVATION CENTER FOR HUMANOID ROBOTICS

ICML
International Conference
On Machine Learning

Thank you for your attention!

Geometry-Guided Modeling of Foundation Features Enables Generalizable Object Shape Deformation Learning

Project Page