NEURAL INFORMATION PROCESSING SYSTEMS

Contact Map Transfer with Conditional Diffusion Model for Generalizable Dexterous Grasp Generation

Yiyao Ma, Kai Chen, Kexin Zheng, Qi Dou

Department of Computer Science and Engineering,
The Chinese University of Hong Kong

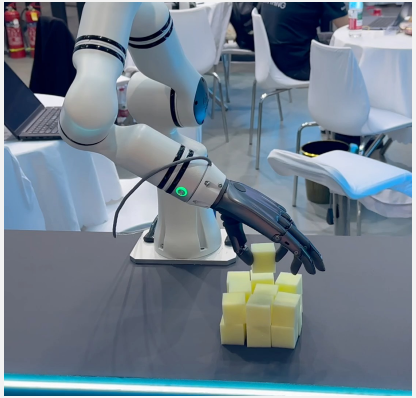
香港中文大學
The Chinese University of Hong Kong

Dexterous Manipulation

Dexterous manipulation plays a crucial role in robotics, enabling robotic hands to manipulate objects with human-like precision and versatility in real-world applications

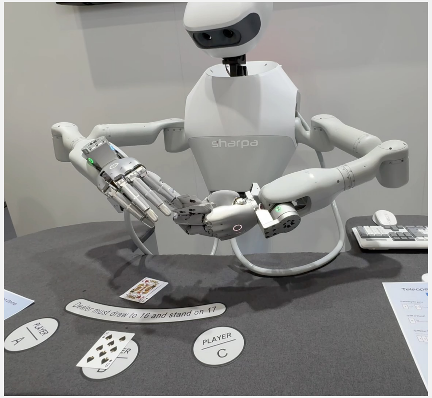
Dexterous Robotic Manipulation

Grasping



Deformable Object Manipulation

Complex Task Completion

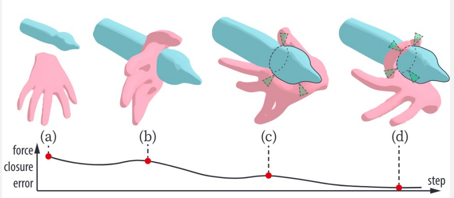


Card Game Playing

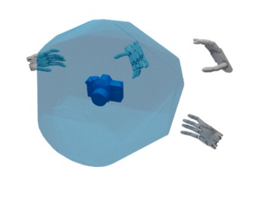
Motivation

Dexterous grasp generation is the foundation for manipulation, requiring both grasp stability and adaptability across diverse objects and tasks.

Analytical Methods

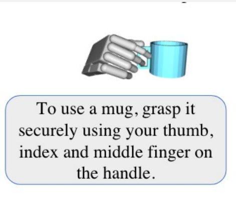


[DFC. RA-L 2021]

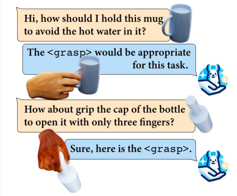


[DexGraspNet. ICRA 2023]

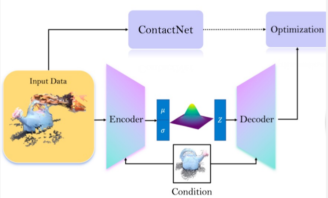
Generative Methods



[DexGYS. NeurIPS 2024]



[SemGrasp. ECCV 2024]

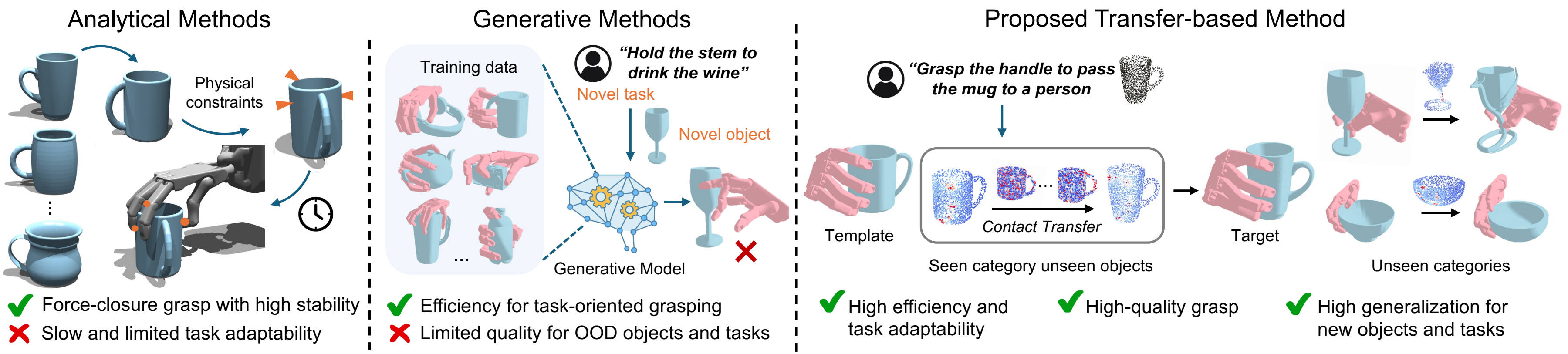


[RealDex. IJCAI 2024]

- ❑ Analytical methods primarily focus on low-level spatial and mechanical constraints to ensure stable grasps; however, they are **computationally expensive** and **lack task adaptability**.
- ❑ Generative methods improve efficiency and task integration but **generalize poorly to unseen objects and tasks** due to data limitations.

Motivation

Dexterous grasp generation is the foundation for manipulation, requiring both grasp stability and adaptability across diverse objects and tasks.



- ❑ We propose a **transfer-based framework** for dexterous grasp generation.
- ❑ It enables the generation of **high-quality**, **task-oriented** grasps while enhancing **generalization** to new object instances, unseen tasks, and even novel object categories

Motivation

Challenges for Transfer-based Method

We propose a novel transfer-based framework which first sample grasps around shape templates and then transfer the grasp to various novel objects using generative models.

- ❑ The substantial shape variations across objects.
- ❑ The complex contact interactions between robot hands and diverse object geometries.
- ❑ The need to accommodate varying task specifications

Shape variation

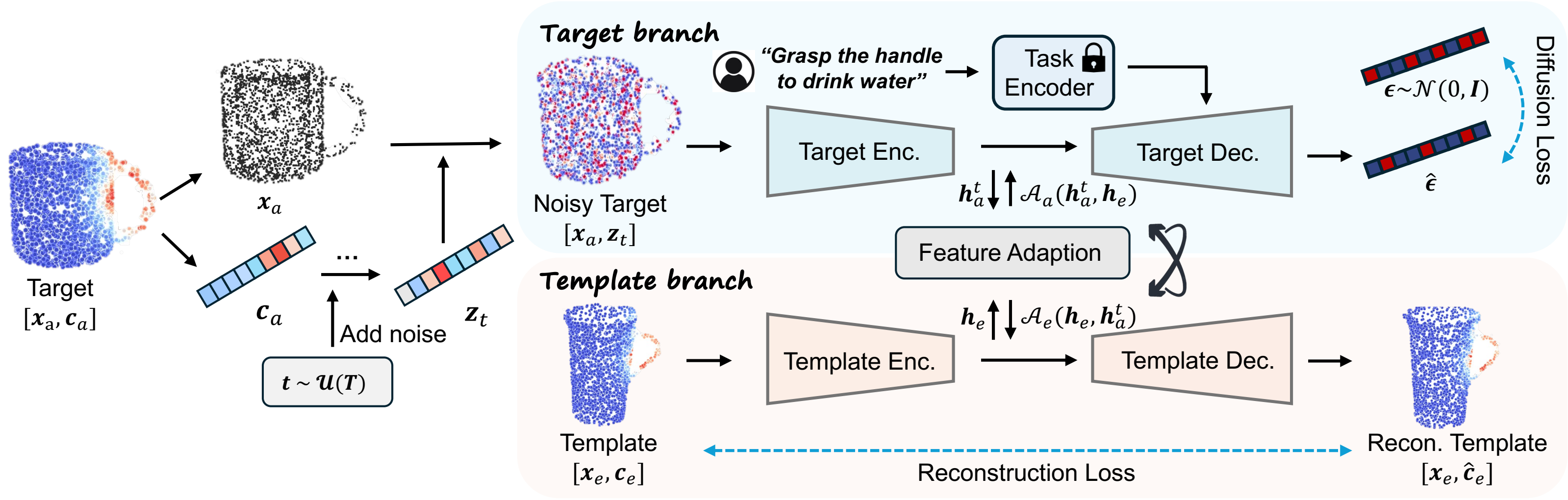
Complex contact interactions

Varying task specifications

Our Solution

Conditional Diffusion Model for Contact Map Transfer

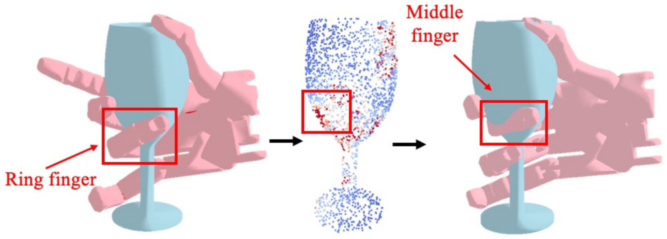
- A conditional diffusion model to jointly capture object geometric similarity and textual task embeddings for a more generalizable dexterous grasp generation.



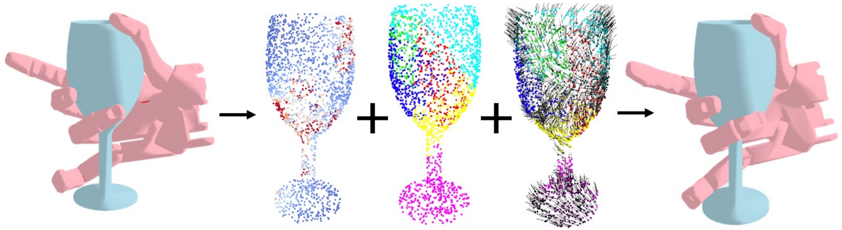
Method

Cascaded Conditional Diffusion for Joint Contact, Part, and Direction Transfer

- Contact maps alone are insufficient to fully capture the complex hand-object interaction.



Only use contact map for grasp generation

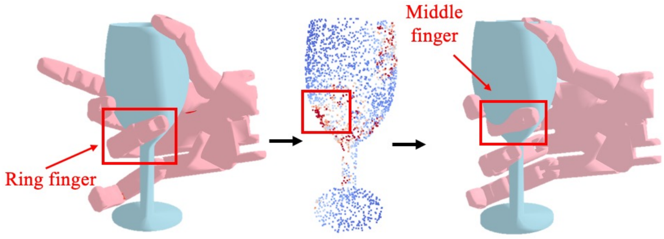


Apply contact + part + direction maps for grasp generation

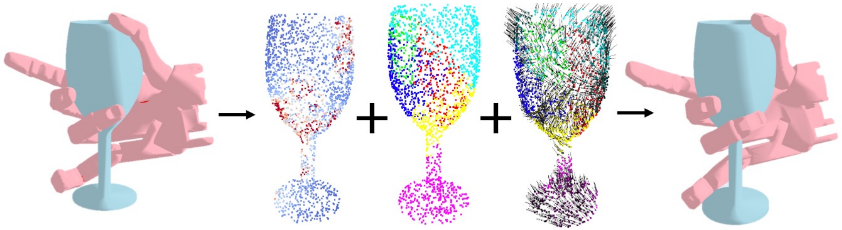
Method

Cascaded Conditional Diffusion for Joint Contact, Part, and Direction Transfer

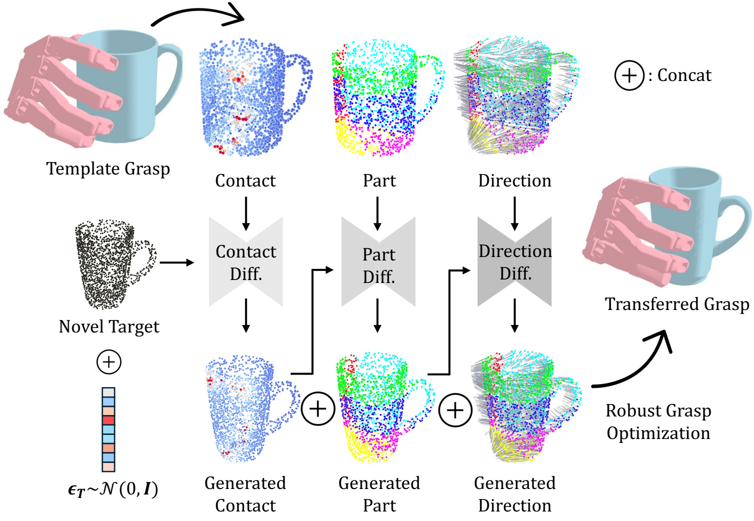
- Therefore, we jointly use object-centric contact map, part map, and direction map to model complex hand-object interaction for a more comprehensive representation.



Only use contact map for grasp generation



Apply contact + part + direction maps for grasp generation



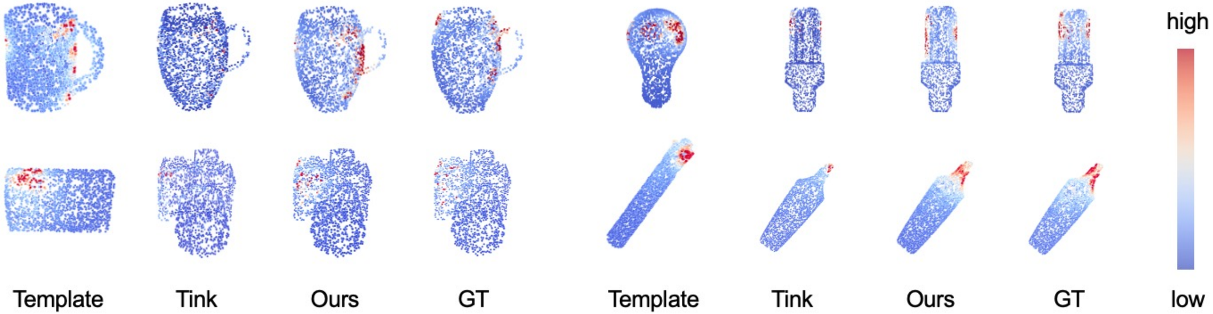
A cascaded conditional diffusion model to transfer three maps with consistency.

Results

Grasp Quality Evaluation

- Comparison with different types of dexterous grasp generation methods:

Methods	Analytical		Generative		Transfer		
	DFC [20]	DexGraspNet [36]	ContactGen [19]	GenDexGrasp [15]	Tink [44]	Ours-Contact	Ours
SR (%) ↑	78.98	83.63	73.00	68.31	60.21	78.46	**84.65**
Pen. (mm) ↓	3.15	4.52	4.11	12.89	**1.18**	1.87	1.47
Cov. (%) ↑	32.28	31.87	34.78	**44.93**	23.49	36.28	38.16



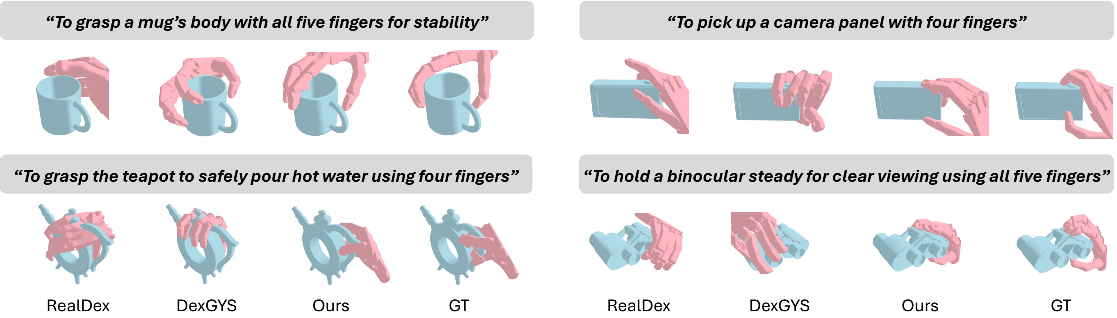
[1] Liu, et al. "Synthesizing diverse and physically stable grasps with arbitrary hand structures using differentiable force closure estimator" RA-L, 2021
[2] Wang, et al. "Dexgraspnet: A large-scale robotic dexterous grasp dataset for general objects based on simulation." ICRA, 2023.
[3] Liu, et al. "Contactgen: Generative contact modeling for grasp generation." ICCV, 2023.
[4] Li, Puhao, et al. "Gendexgrasp: Generalizable dexterous grasping." ICRA, 2023.
[5] Yang, et al. "Oakink: A large-scale knowledge repository for understanding hand-object interaction" CVPR, 2022.

Results

Generalization Evaluation

❑ Comparison with generative methods and transfer-based method:

Methods	Seen Categories SR ↑	Pen. ↓	Cov. ↑	Cont. Err. ↓	Consis. ↑	Unseen Categories SR ↑	Pen. ↓	Cov. ↑	Cont. Err. ↓	Consis. ↑
RealDex [21]	42.16	3.26	23.61	0.1002	80.62	29.90	3.35	17.15	0.0975	70.85
DexGYS [38]	41.56	22.25	30.50	0.0834	74.85	39.16	23.63	32.37	0.1332	68.08
Tink [44]	62.60	**1.29**	24.86	0.0327	68.75	—	—	—	—	—
Ours-Contact	76.14	1.68	33.55	0.0322	75.26	70.34	1.59	**35.66**	0.0410	75.00
Ours	**79.32**	1.74	**36.77**	**0.0287**	**83.60**	**74.14**	**1.36**	30.05	**0.0363**	**79.28**



[1] Liu, et al. “RealDex: towards human-like grasping for robotic dexterous hand.” IJCAI, 2024.
[2] Wei, et al. “Grasp as you say: Language-guided dexterous grasp generation.” NeurIPS, 2024.
[3] Yang, et al. “Oakink: A large-scale knowledge repository for understanding hand-object interaction” CVPR, 2022.

Results

Real world Experiments

- We conducted real-world dexterous grasping experiments to evaluate the effectiveness of our method in practical scenarios.

"To pass the bottle to another person" | *"To grab the headband of the headphone"* | *"To carry a cup full of liquid"* | *"To grasp a bowl for stacking with other bowls"*

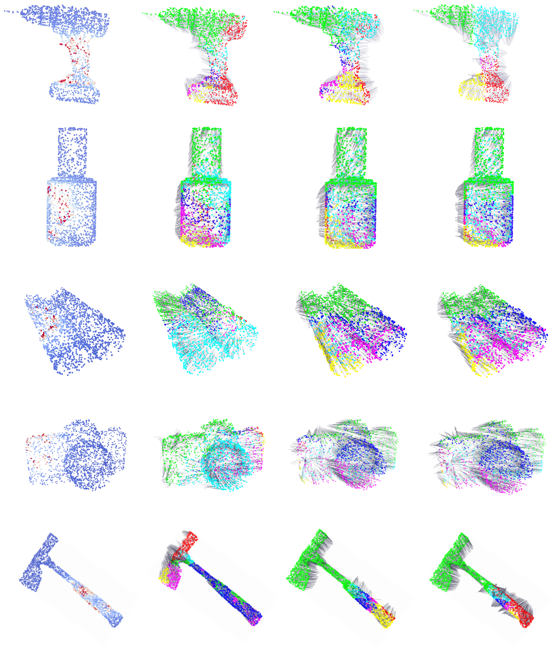
For each task, the left, middle, and right images show the template contact map, the transferred contact map, and the generated dexterous grasp, respectively.

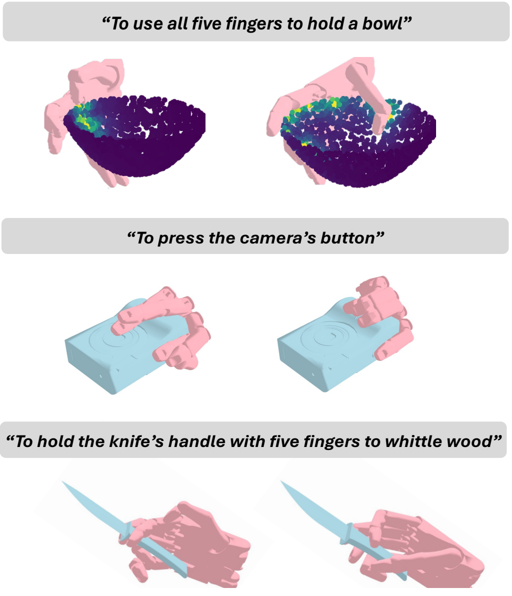
Results

Ablation Study Results

Methods	SR ↑	Pen. ↓	Cov. ↑	Cont. Err. ↓
w/o $\mathcal{A}_d$	37.57	1.95	21.11	0.0522
w/o $\mathcal{A}_s$	38.03	1.99	21.91	0.0512
w/o cascaded diff.	60.71	1.87	25.08	0.0489
w/o task desc.	73.85	1.83	34.03	0.0322
w/o robust opt.	55.04	**1.74**	27.95	0.0448
Ours	**79.32**	**1.74**	**36.77**	**0.0287**

Ours-Contact w/o Robust. Ours GT





w/o textural embeddings Ours

Conclusion

In summary, our work highlights the following contributions:

- We propose a novel **transfer-based framework** for dexterous grasp generation.
- We introduce a **conditional diffusion model** that leverages task embeddings to learn geometric similarities for contact map transfer.
- We further develop **a cascaded diffusion framework** to jointly transfer contact, part, and direction maps while maintaining their consistency.

Thanks for your attention!

Project
Homepage