Student Lecture:

Staged Approaches for Universal Dexterous Grasping

Team 3

Seungwan Kang, Hajun Park

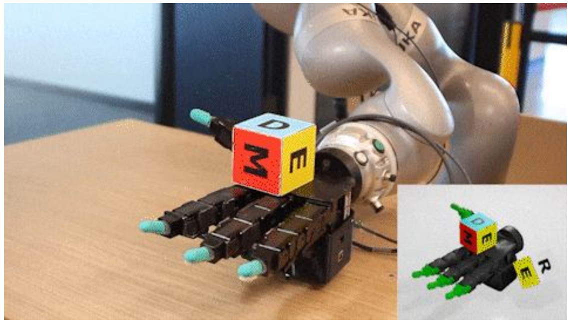
Contents

1. Introduction to Dexterous Grasping

1.1. Grasping in Robotics

1.2. What is Dexterous Grasping?

1.1. Grasping in Robotics

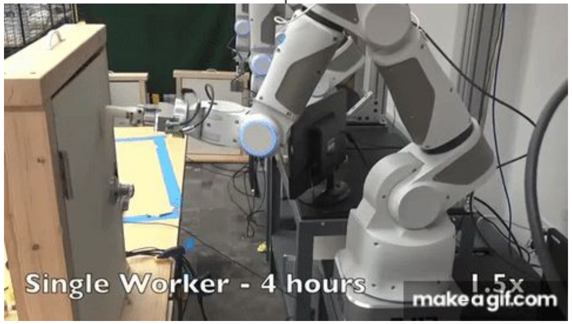


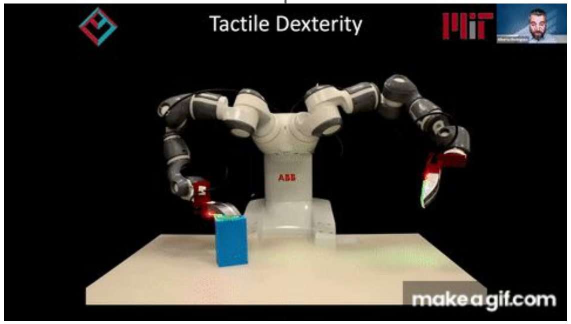
In-hand manipulation

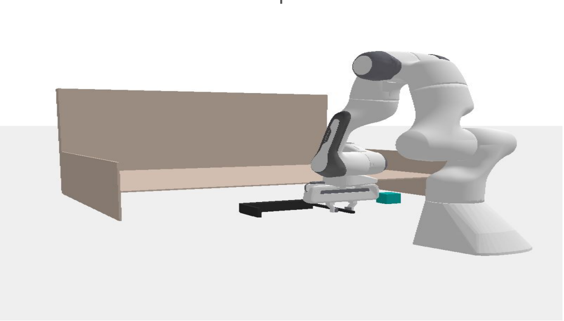
Grasping

Non-prehensile

Manipulation

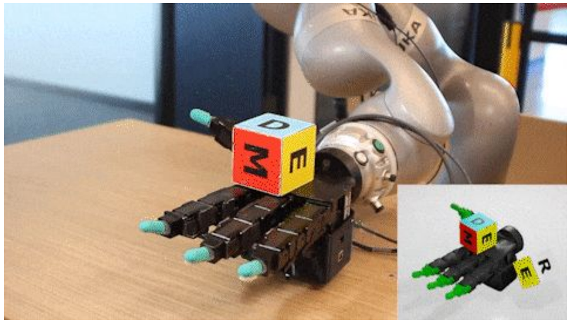






Some high-level tasks

1.1. Grasping in Robotics

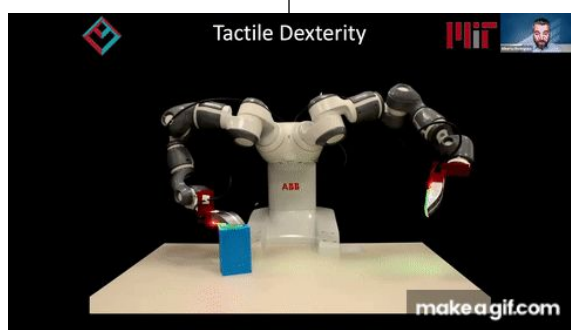


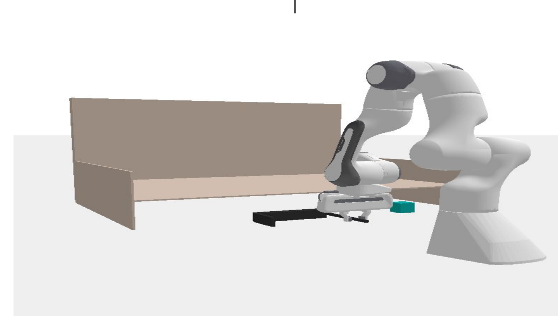
In-hand manipulation

Grasping

Non-prehensile

Manipulation



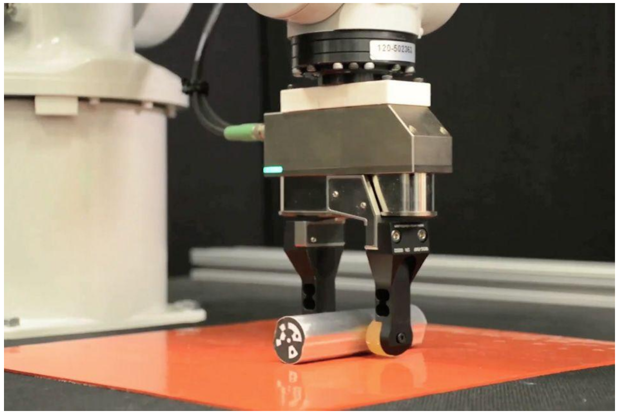


Some high-level tasks

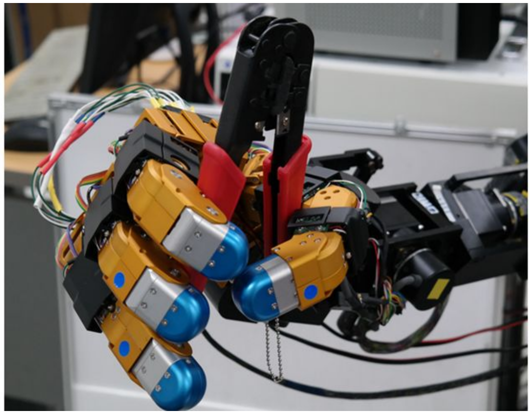
1.1. Grasping in Robotics

- Serving as a foundation for **manipulation tasks** such as picking, holding, and moving objects.

- There are typically two types in grasping mechanisms: **parallel grippers** and **dexterous hands**



- Image from MIT News



- Image from Robotics & Automation

1.1. Grasping in Robotics

- Grasping with **parallel grippers** has achieved great success in universally grasping unknown objects
  (Grasp-Anything, 2024 ICRA)

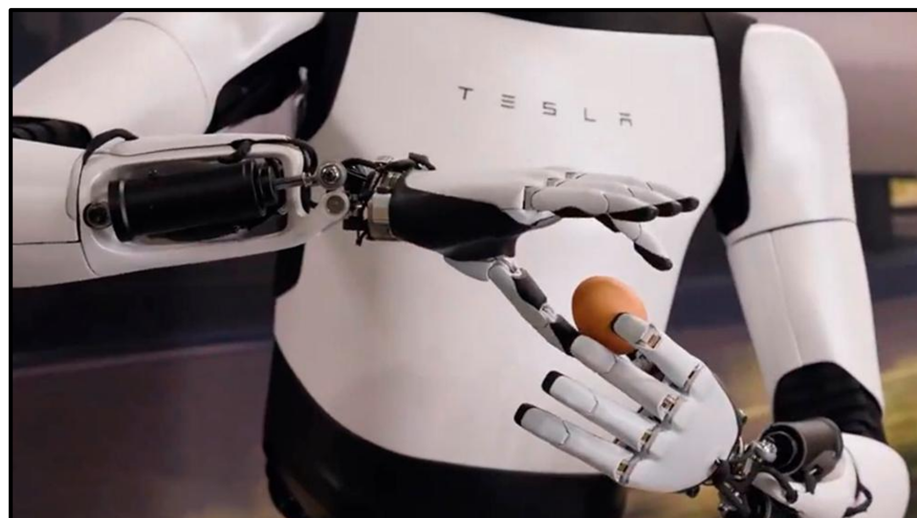
- But, parallel grippers have their **limitation** on **complicated and functional manipulation**

- Image from NVIDIA developer

1.2. What is Dexterous Grasping?

- Use of **robotic hands with high DoFs**
  → **more diverse and intricate interactions**

- Recent humanoid robots are all equipped with advanced dexterous hands
  → dexterous manipulation become more important!



2. Overview of Fundamental Studies

2.1. Overview of Dexterous Manipulation (ICRA, 2000)

2.2. Popular Tool: GraspIt!

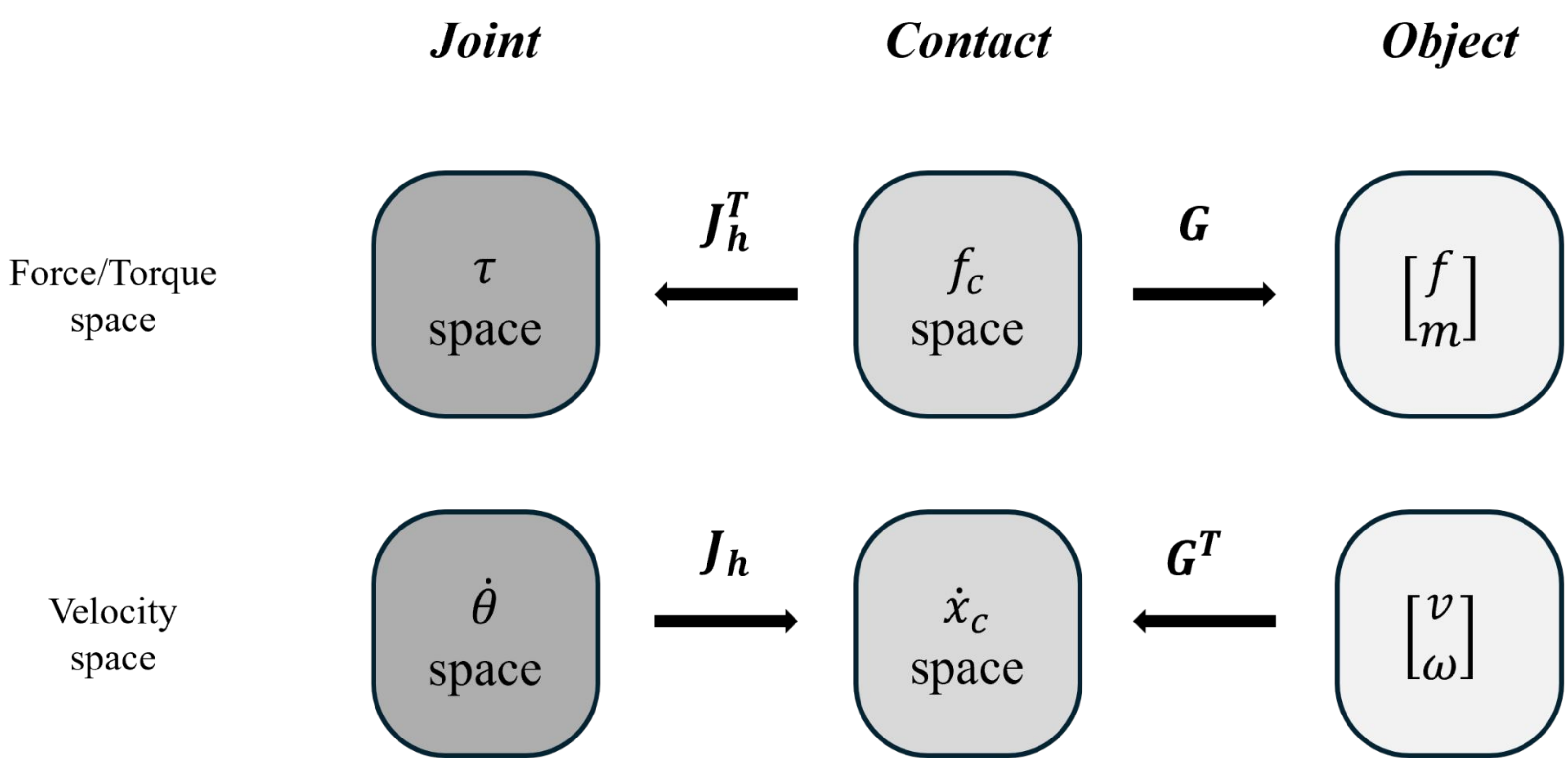
2.3. Some Advances in Analytical Approaches

2.4. Limitation of Non-Learning Methods

2.1. Overview of Dexterous Manipulation

1. Model-Based Approach

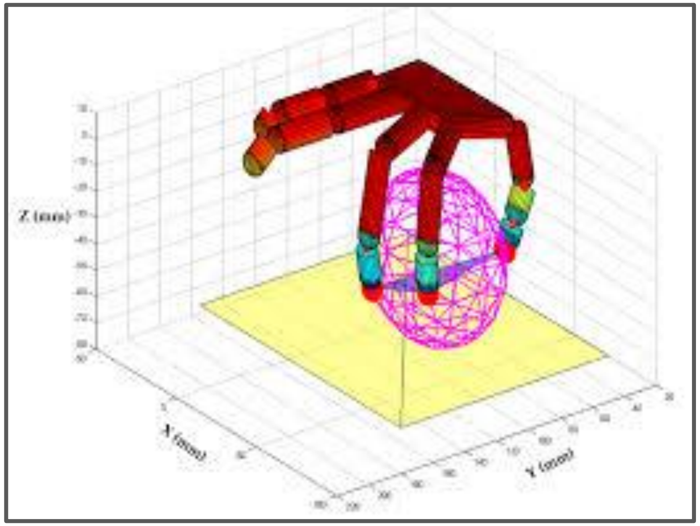
- J_h: hand jacobian, G: grasp jacobian
- Get each joint force



2.1. Overview of Dexterous Manipulation

2. How to Immobilize an Object

- **Form closure**
- **Force closure**



3. Grasp Planning

- **Grasp map**
    - Graphical representation of **all possible stable grasps**
- **Grasp gait**
    - The sequence of **finger motions and regrasping**

2.1. Overview of Dexterous Manipulation

4. Grasp Optimization and Quality Measures

- **What is the best grasp?**
  - Consider contact location, contact force, finger pose, task property

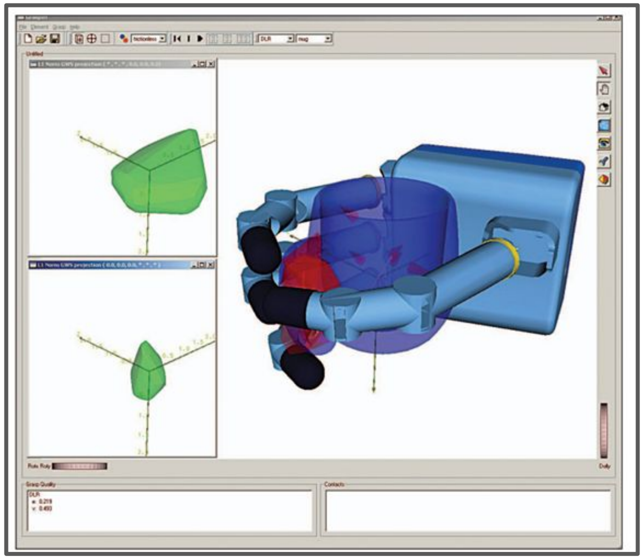
- **Some quality metrics**
  - Task ellipsoid (wrench requirement + twist requirement)
  - Dynamic stability (against external force)
  - Q1
  - object penetration
  - Others…

2.2. Popular Tool: GraspIt!

(Robotics & Automation, 2004)

- **Simulator** for synthesizing stable grasps via collision detection and optimization

- **Limitations**
  - **High computational cost (~10mins for an optimization)**
  - Due to simplified physics, limited in real-world applications
  - Requires full object geometry (oracle inputs)
  - Lack of pose diversity



2.3. Some Advances…

Grasping-Force Optimization for Multifingered Robotic Hands Using a Recurrent Neural Network (T-RO, 2004)

- **Real-time grasp-force optimization**

Hand posture subspaces for dexterous robotic grasping (IJRR, 2009)

- **Using low-dimensional subspace** of the hand DOF space for finding hand postures appropriate for a given task

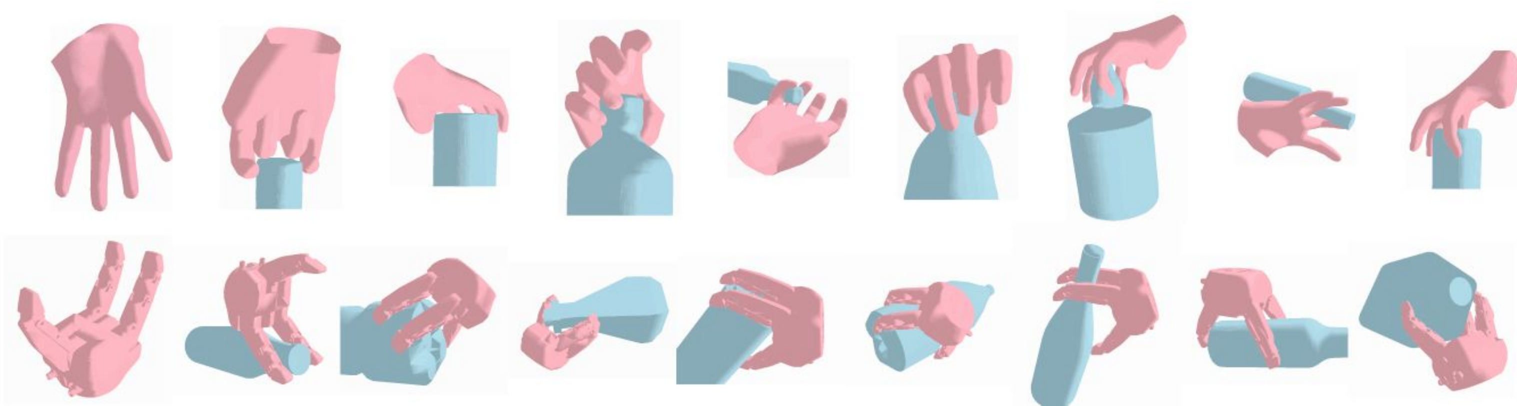
Push-grasping with dexterous hands: Mechanics and a method (IROS, 2010)

- Enabling an agent to grasp more object using **push-grasping**

2.3. Some Advances…

Synthesizing diverse and physically stable grasps with arbitrary hand structures using differentiable force closure estimator (RA-L 2021)

- Overcome flaws of classic force closure evaluation (**speed** ↑)
- **Various grasp poses** without prior data
- Achieved some **generality**
- **Limitations**
  - Unrealistic grasp for concave object
  - Incompleteness in penetration detection
  - Require object’s exact 3D model

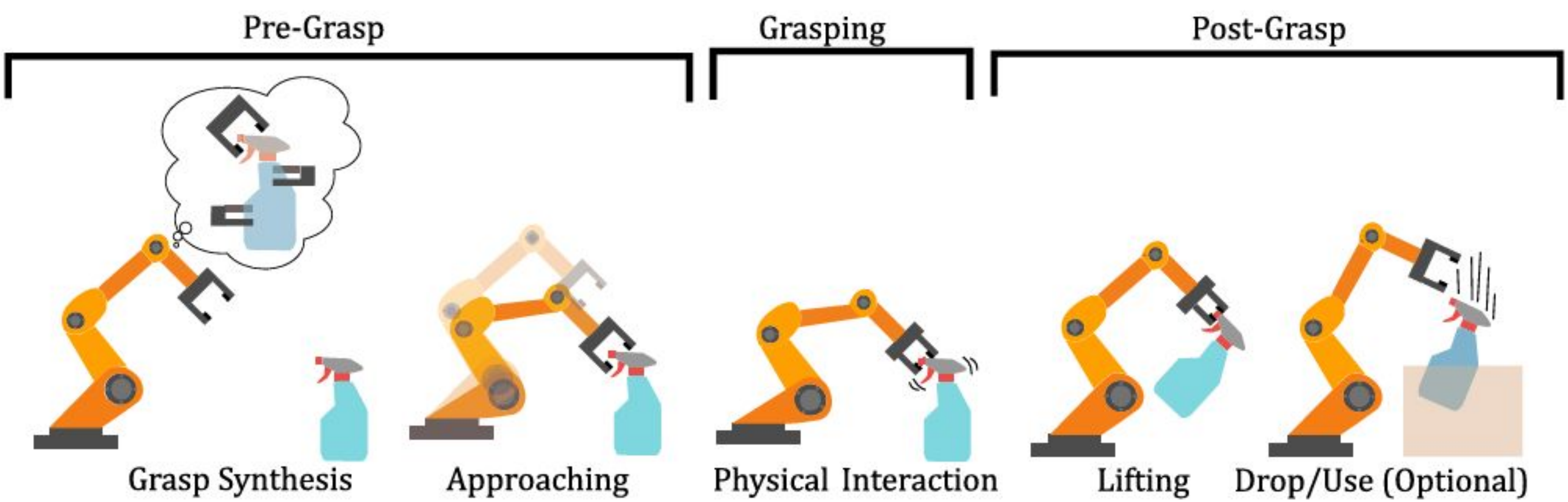


3. Learning-Based Methods

3.1. Learning-Based Methods: Synthesizing

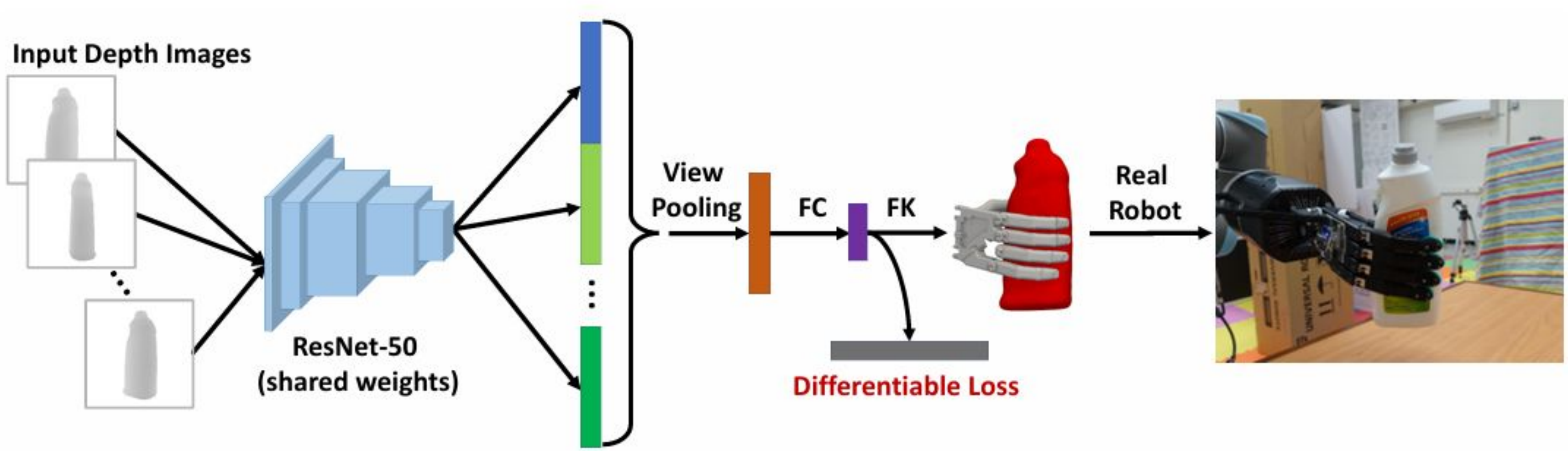
3.2. Learning-Based Methods: Grasping Policy

3.1. Learning-based Methods: Synthesizing



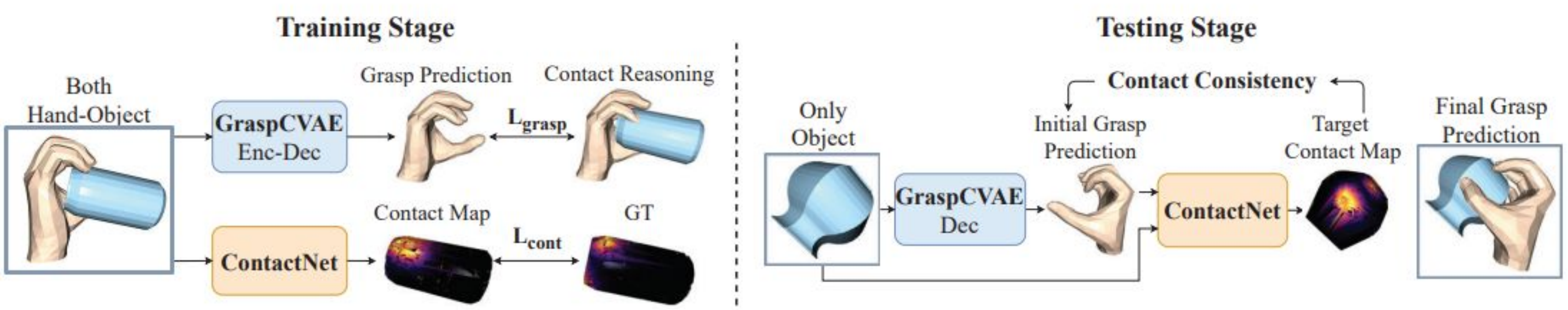
3.1. Learning-based Methods: Synthesizing

- **Deep differentiable grasp planner for high-DoF grippers** (RSS 2020)
    - Framework: ResNet-50 based neural network
    - Small scale dataset + simple network
    - Slow learning… (fine-tuning)
    - **Low diversity**



3.1. Learning-based Methods: Synthesizing

- **Hand-Object Contact Consistency Reasoning for Human Grasps Generation** (ICCV 2021)
  - Framework: Conditional VAE + ContactNet
  - More stable, natural, generalizable, fast
  - Suffer from severe mode collapse leading to **limited diversity**
  - Require large-scale, high-quality dataset



3.1. Learning-based Methods: Synthesizing

- **Benchmark Datasets** (based on simulation)
  - **DexGraspNet** (ICRA 2023)
    - 1.32 million ShadowHand grasps on 5,355 objects
    - >133 object categories
    - >200 diverse grasps for each object instance
  - **DexGraspNet 2.0** (CoRL 2024)
    - For cluttered scene
    - 1319 objects, 8270 scenes, and 427 million grasps

3.2. Learning-based Methods: Grasping Policy

- **DAPG** (RSS 2018)
  - Imitation learning + RL

- **HGA-Dex** (CoRL 2021)
  - Affordance map → RL conditioned

- **UniDexGrasp** (CVPR 2023)
  - Pose generation → Teacher-Student Framework (RL + distillation)

4. Recent Works

4.1. UniDexGrasp (2023, CVPR)

4.2. UniDexGrasp++ (2023, ICCV)

4.3. DexGrasp Anything (2025 CVPR)

4.1. Recent Works: UniDexGrasp (2023, CVPR)

Problem:

- In **Dexterous grasping**, previous researches had problem of **Low generalization quality, Relying on oracle state**

Solution:

- Use **Two task division** (Synthesizing and Grasping Policy)
- “**Rotation Generation**” and” **Translation & Articulation generation**” **separately** to avoid mode collapse
- **Teacher-Student Framework** for grasp with realistic input

4.1. Recent Works: UniDexGrasp (2023, CVPR)

Dexterous Grasp Proposal Generation

Grasp Orientation Generation

Input: $X_0 \in \mathbb{R}^{N\times 3}$ (input point cloud)
Output: $R \in \mathrm{SO}(3)$ (rotation matrix)

Grasp Translation and Articulation Generation

Input: $\widetilde{X}_0 = R^{-1}X_0$ (rotated point cloud)
Output: $t \in \mathbb{R}^3, \quad q \in \mathbb{R}^K$ (translation and joint angles)

Goal-Conditioned Dexterous Grasping Policy

Teacher Policy $\pi^{\mathcal{E}}$

Input: $\tilde{s}_t^o$ (object state) $\in \mathbb{R}^{D_s}$
Output: c (category label) $\in \{1, \ldots, C\}$,
a_t (action) $\in \mathbb{R}^{26}$
[using reinforcement learning]

Student Policy π^S

Input: $\widetilde{X}_t$ (raw scene point cloud) $\in \mathbb{R}^{N\times 3}$
Output: a_t (action) $\in \mathbb{R}^{26}$
[using MLP]

4.1. Recent Works: UniDexGrasp (2023, CVPR)

Details:

- Correct pose using **ControlNet** by using contact map
- **Object Curriculum Learning** by increasing training data set from one object to categories

Result:

- High performance in both “Grasping quality” and “Object penetration”
- Language-guided Dexterous Grasping by combining with CLIP

4.2. Recent Works: UniDexGrasp++ (2023, ICCV)

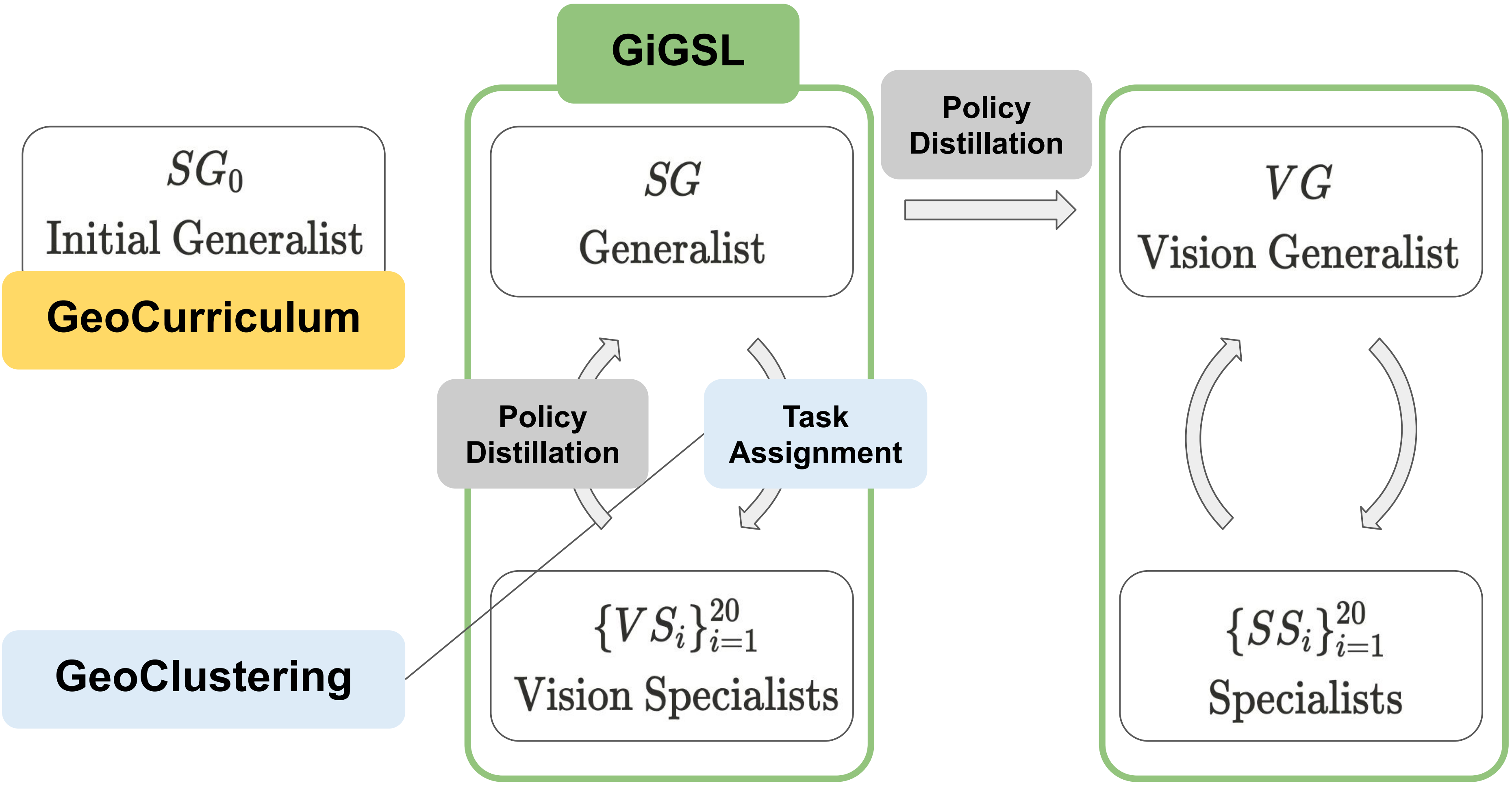
Problem:

- Previous research had **limit in Grasping Policy**
  - Even Teacher policy had poor performance
  - Previously, category did not consider grasping pose

Solution:

- **GeoCurriculum**: Train in **order of similar geometry features**
- **GeoClustering**: **Cluster tasks** and distribute to specific models
- **GiGSL**: **Repeat** “Task assignment” and “Policy distillation”

4.2. Recent Works: UniDexGrasp++ (2023, ICCV)



4.2. Recent Works:
UniDexGrasp++ (2023, ICCV)

Result:

- 10~12% of performance improvement
- Also worked well in Meta-World benchmark

4.3. Recent Works:
DexGrasp Anything (2025, CVPR)

Problem:

- Previous researches had **limit in pose generation**
  - Lack of constraints about physical rules
  - Small dataset and are based on simulation

Solution:

- **DDPM based diffusion pose generation**:
  - Add loss about **physical constraints**
  - Give additional **guidance** using physical loss

Result:

- Performance improved in all aspect (SOTA)
  - Success Rate, Penetration and Diversity

4.4. Limitations of recent works

- Rigid objects only. Not considering **elasticity, texture, fluidity**

- **Task-oriented manipulation** can be different from grasping

  → **synthesis of reward for arbitrary dexterous task**

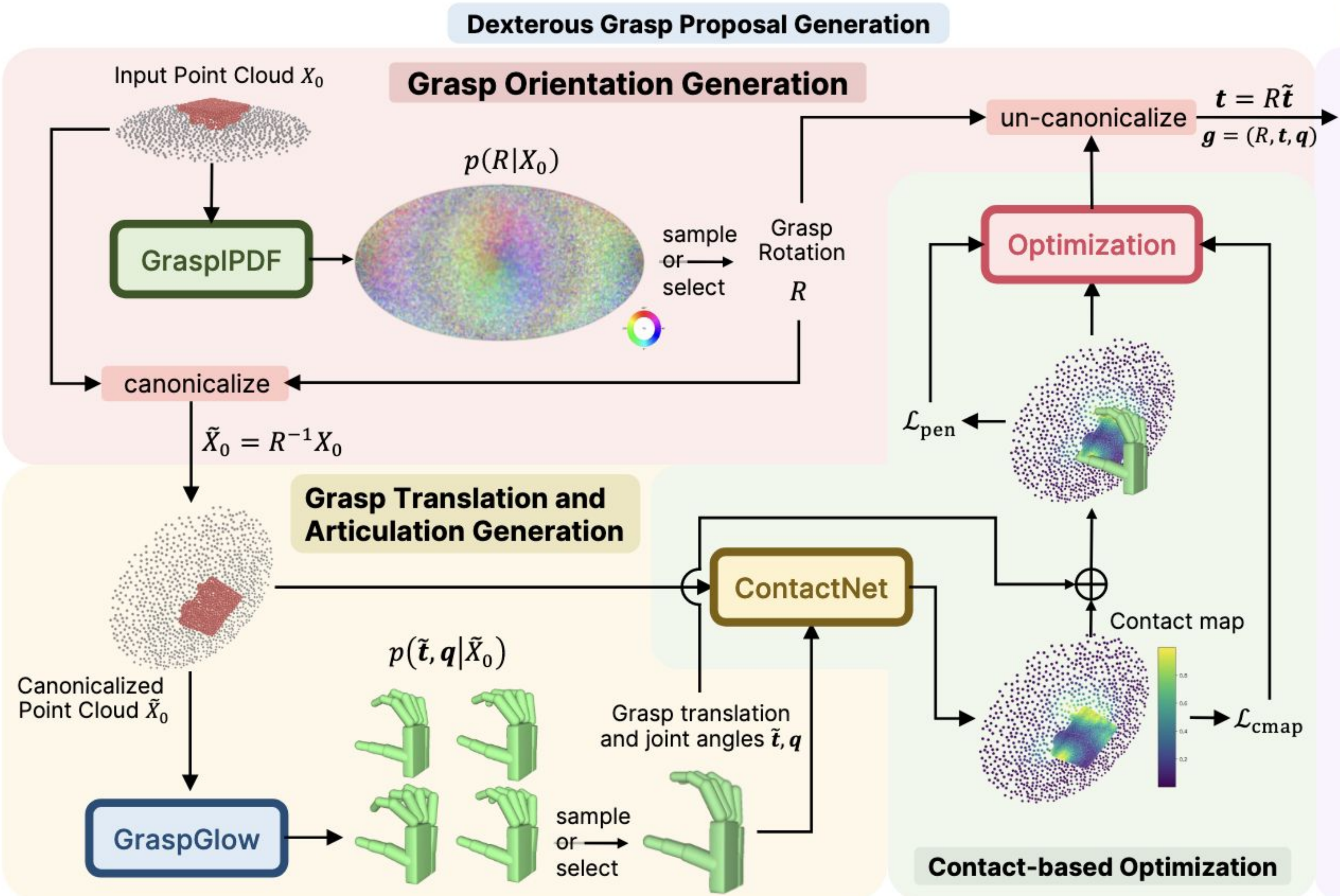
- **Integration** with other dexterous manipulation is needed for both **dataset and policy**

Thank you

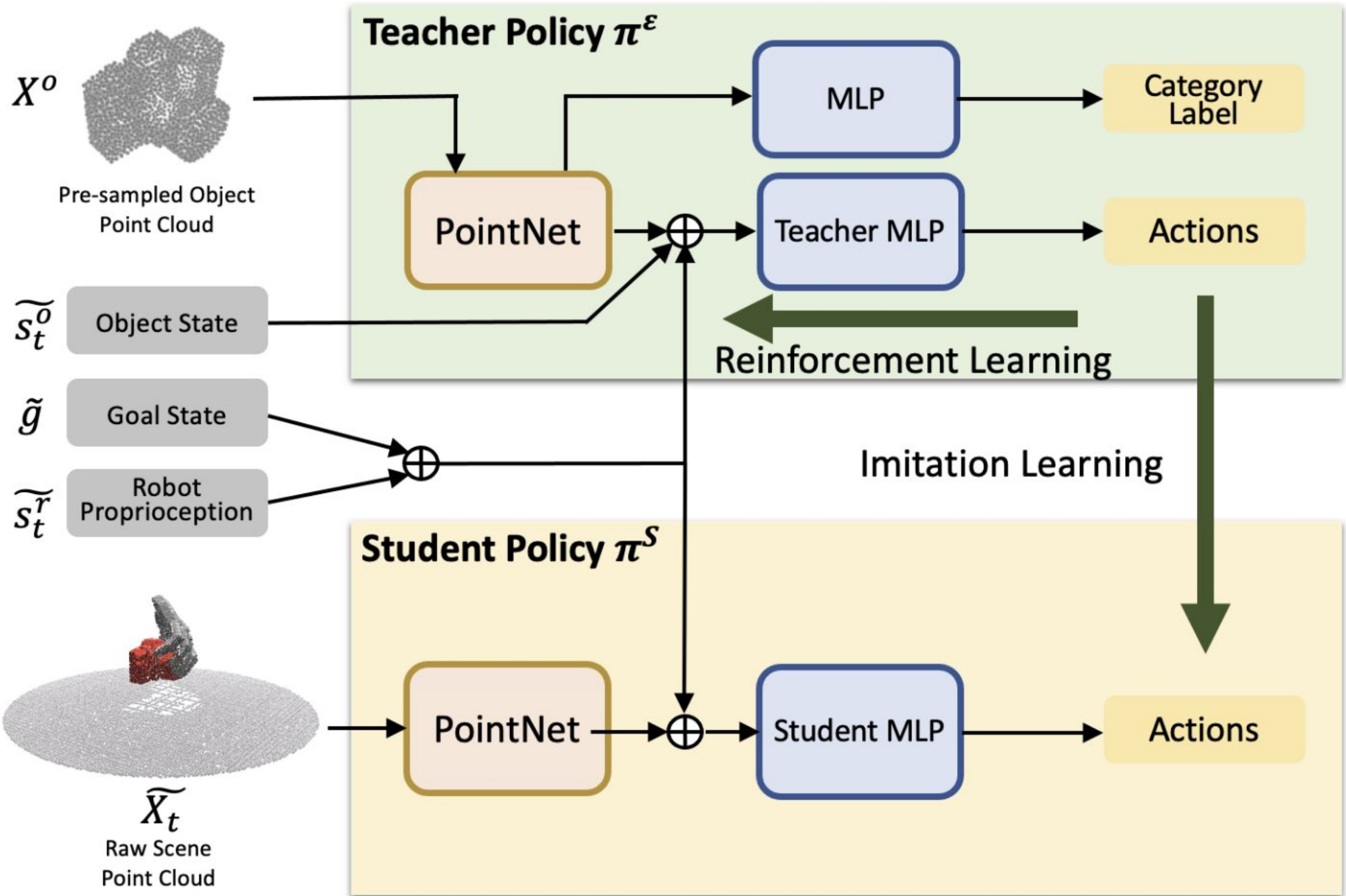
Prepared Appendix

- Detailed figure for UniDexGrasp
- Detailed figure for UniDexGrasp++
- Detailed figure for DexGrasp Anything

Appendix UniDexGrasp



Appendix UniDexGrasp



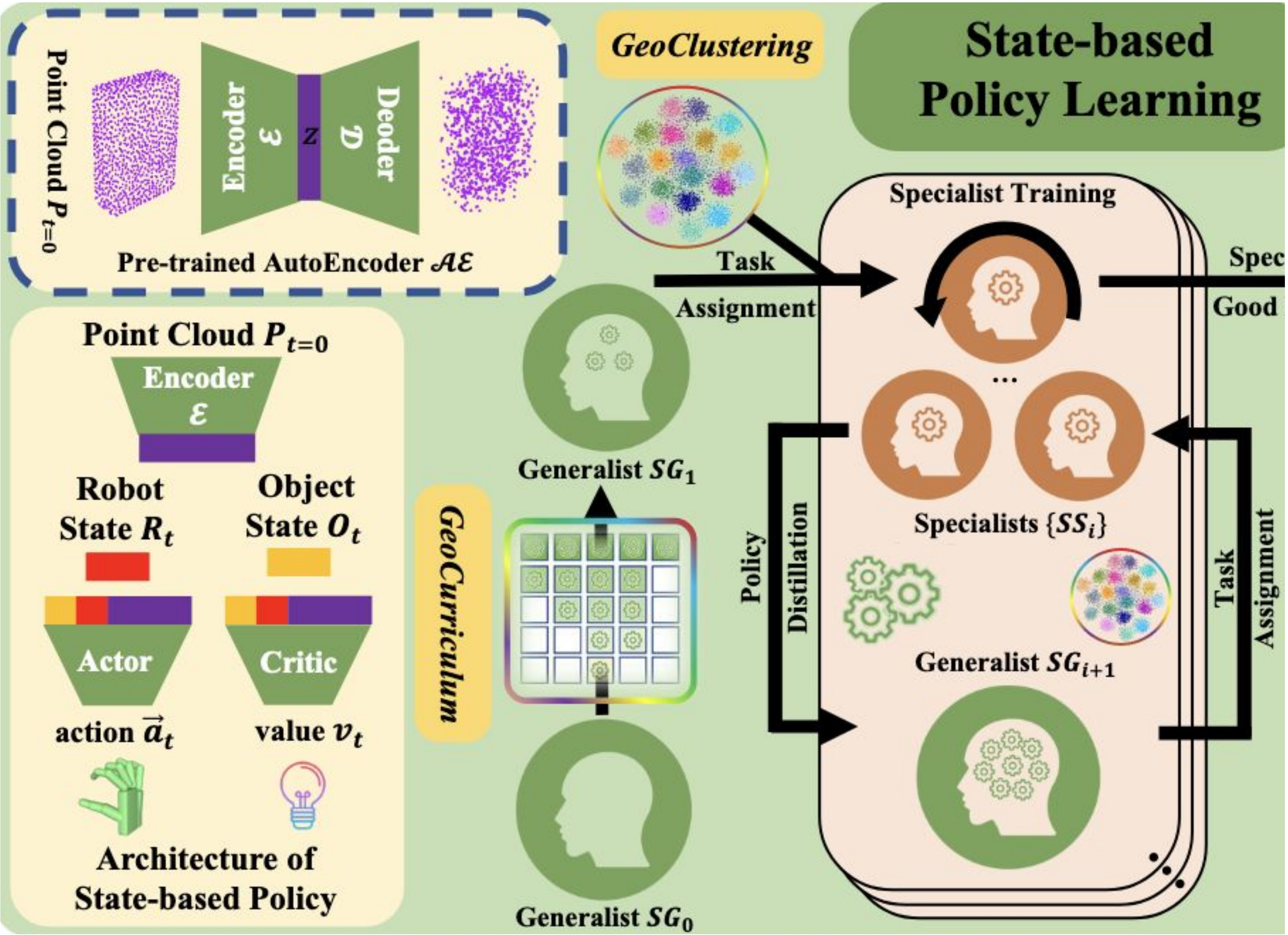
Appendix UniDexGrasp

Method	seen cat		unseen cat		$\sigma_R \uparrow$	$\sigma_{T\|R} \uparrow$	$\sigma_{\theta\|R} \uparrow$	$\sigma_{\text{keypoints}} \uparrow$
	$Q_1 \uparrow$	obj. pen.$\downarrow$	$Q_1 \uparrow$	obj. pen.$\downarrow$	(degree)	(cm)	(degree)	(cm)
GraspTTA [24] (C + T)	0.0269	0.354	0.0239	0.363	4.9	/	/	2.909
DDG [28]	0.0357	0.319	0.0223	0.338	0.0	/	/	0.000
R + C + T	<u>0.0362</u>	0.251	**0.0336**	0.235	**128.0**	<u>0.095</u>	<u>0.227</u>	5.982
ReLie [16] + T	0.0190	0.219	0.0191	0.225	109.9	/	/	**6.698**
ProHMR [26] + T	0.0210	**0.202**	0.0221	**0.192**	88.4	/	/	5.837
ours (R + GL + T)	**0.0423**	<u>0.205</u>	<u>0.0322</u>	<u>0.220</u>	<u>127.6</u>	**1.143**	**5.806**	<u>6.389</u>

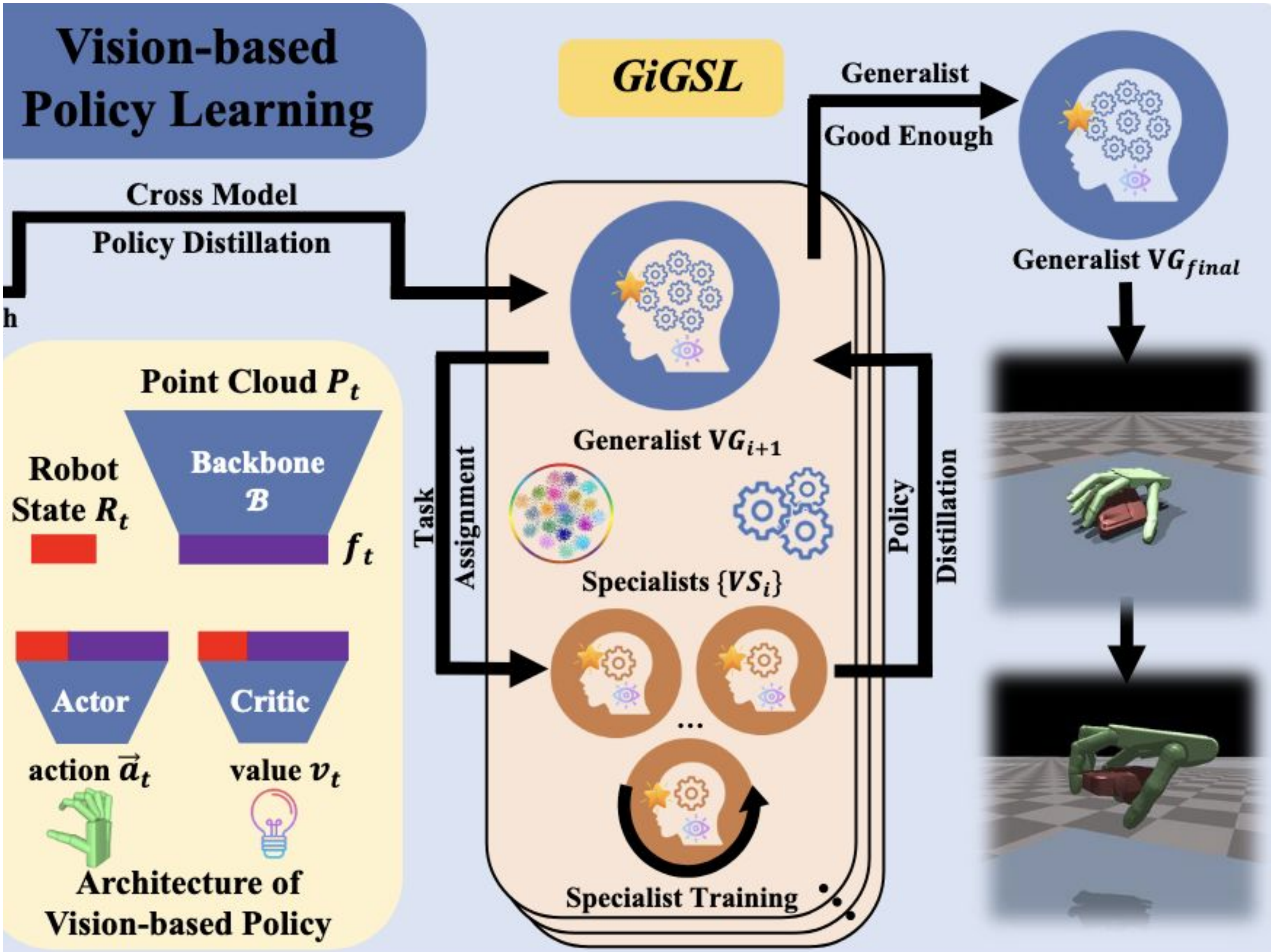
Table 1. **Results on grasp goal generation.** R: GraspIPDF, C: CVAE, T: test-time adaptation, GL: GraspGlow, and obj. pen. is the penetration between the hand and the object.

Model	Train	Test	
		unseen obj seen cat	unseen cat
MP	0.12±0.01	0.02±0.00	0.02±0.01
PPO [52]	0.14±0.06	0.11±0.04	0.09±0.06
DAPG [45]	0.13±0.05	0.13±0.08	0.11±0.05
ILAD [59]	0.25±0.03	0.22±0.04	0.20±0.05
Ours	**0.74±0.07**	**0.71±0.05**	**0.66±0.06**
Ours(w/o SC)	0.59±0.06	0.54±0.07	0.51±0.04
Ours(w/o cls)	0.65±0.05	0.64±0.06	0.60±0.07
Ours(w/o OCL)	0.31±0.07	0.23±0.06	0.21±0.04
Ours(1-stage OCL)	0.58±0.07	0.55±0.03	0.55±0.05
Ours(2-stage OCL)	0.68±0.06	0.67±0.07	0.62±0.05

Appendix UniDexGrasp++



Appendix UniDexGrasp++



Appendix UniDexGrasp++

Model	Train(%)	Test(%)	
		Uns. Obj. Seen Cat.	Uns. Cat.
PPO[55]	24.3	20.9	17.2
DAPG[49]	20.8	15.3	11.1
ILAD[69]	31.9	26.4	23.1
GSL[29]	57.3	54.1	50.9
UniDexGrasp[70]	79.4	74.3	70.8
Ours (state-based)	**87.9**	**84.3**	**83.1**
PPO[55]+DAgger[51]	20.6	17.2	15.0
DAPG[49]+DAgger	17.9	15.2	13.9
ILAD[69]+DAgger	27.6	23.2	20.0
GSL[29]+DAgger	54.1	50.2	44.8
UniDexGrasp[70]	73.7	68.6	65.1
Ours (state)+DAgger	77.4	72.6	68.8
Ours (vision-based)	**85.4**	**79.6**	**76.7**

Table 1: **The Average Success Rate of the Evaluated Objects on Both Training and Test Set**. For better clarity, we use green for the state-based policy and blue for the vision-based policy.

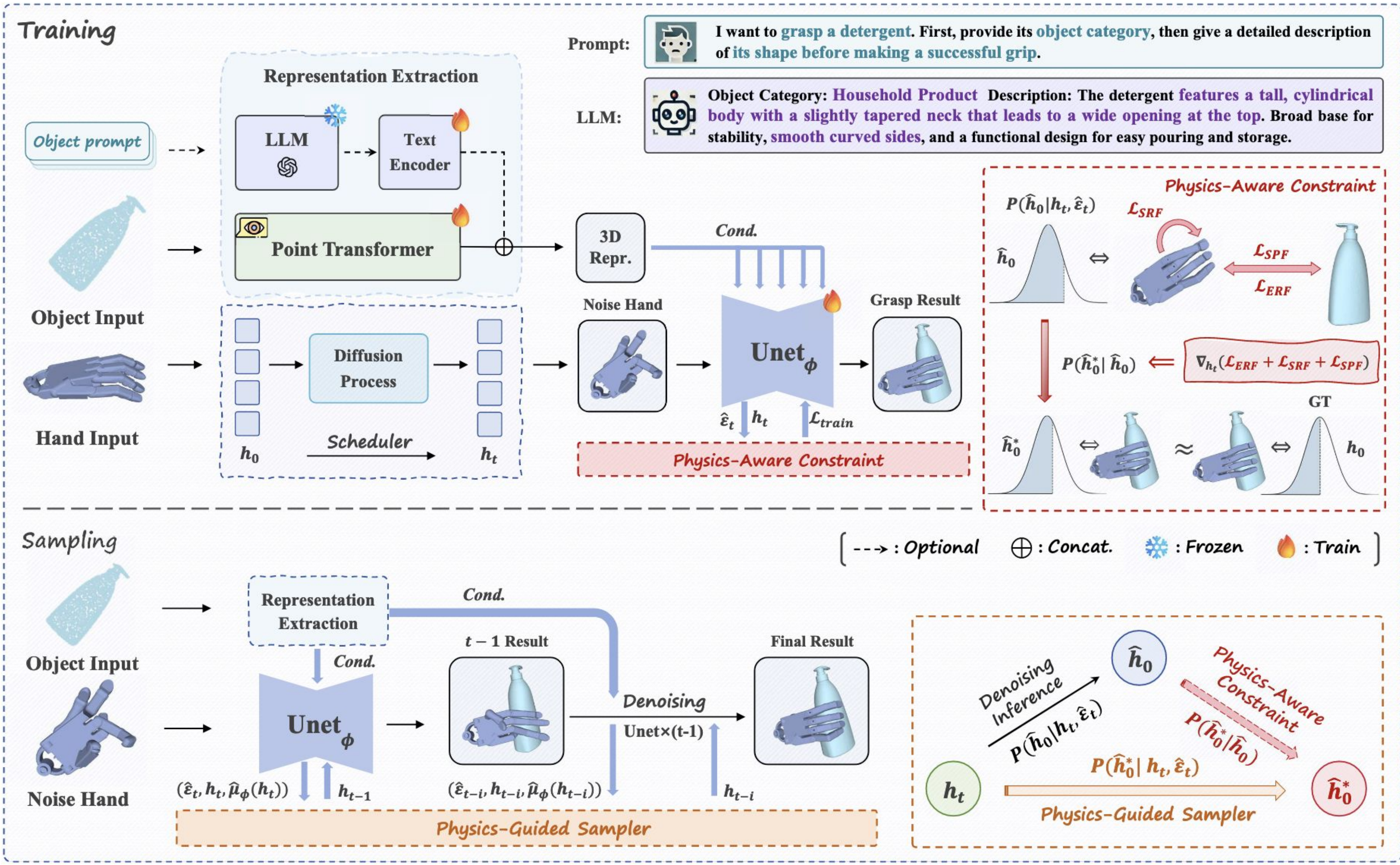
Model	Train(%)	Test(%)	
		Uns. Obj. Seen Cat.	Uns. Cat.
No Curriculum	30.5	23.4	20.6
OCL[70]	79.4	74.3	70.8
GeoCurriculum (3)	81.3	75.6	73.3
GeoCurriculum (4)	82.7	**76.8**	**74.2**
GeoCurriculum (5)	**82.9**	76.4	74.0

Table 6: **Ablation study on *GeoCurriculum*.** OCL refers to the Object Curriculum Learning proposed in [70]. The numbers in brackets represent the number of stages for curriculum learning.

Model	Train(%)	Test(%)	
		Uns. Obj. Seen Cat.	Uns. Cat.
Random	77.0	71.9	68.2
Category Label.	79.7	73.9	74.1
Ours	**85.4**	**79.6**	**76.7**

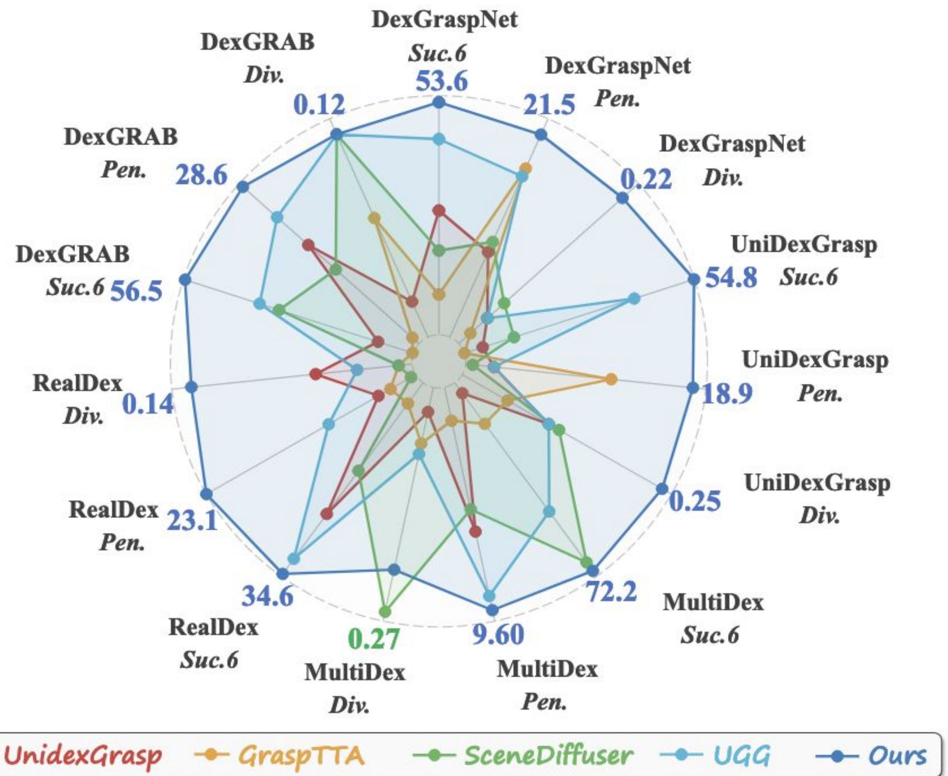
Table 9: **Ablation study on the pre-trained autoencoder.** The features from the encoder are used in *GeoClustering* in the state-based setting.

Appendix DexGrasp Anything



Appendix DexGrasp Anything

Method \ Dataset	DexGraspNet				UniDexGrasp				MultiDex				RealDex				DexGRAB			
	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑
UniDexGrasp [47]	33.9	70.1	31.9	0.14	23.7	65.5	24.5	0.14	21.6	47.5	13.5	0.08	27.1	59.4	39.0	0.11	20.8	55.8	37.4	0.08
GraspTTA [11]	18.6	67.8	24.5	0.13	21.0	65.3	21.2	0.10	30.3	62.8	19.0	0.11	13.3	46.4	40.1	0.09	14.4	51.0	51.4	0.10
SceneDiffuser [10]	26.6	66.9	31.0	0.15	28.3	74.8	25.1	0.15	69.8	85.6	14.6	**0.27**	21.7	56.1	42.0	0.09	39.1	85.0	41.1	0.12
UGG [21]	46.9	79.0	25.2	0.14	46.0	83.2	24.5	0.14	55.3	93.4	10.3	0.12	32.7	63.4	34.4	0.10	42.7	90.6	33.2	0.12
Ours	53.6	90.4	21.5	0.22	**54.8**	90.8	18.9	**0.25**	72.2	96.3	**9.6**	0.23	34.6	71.2	**23.1**	**0.14**	56.5	91.8	**28.6**	0.12
Ours(w/ LLM)	**57.5**	**90.6**	**17.8**	**0.23**	53.1	**91.2**	**18.8**	0.23	**79.1**	**98.1**	11.4	0.22	**44.8**	**73.7**	27.7	0.13	**57.9**	**92.7**	30.4	**0.13**



Suc.6 → Success rate (6 directions) *Pen.* → Penetration *Div.* → Diversity