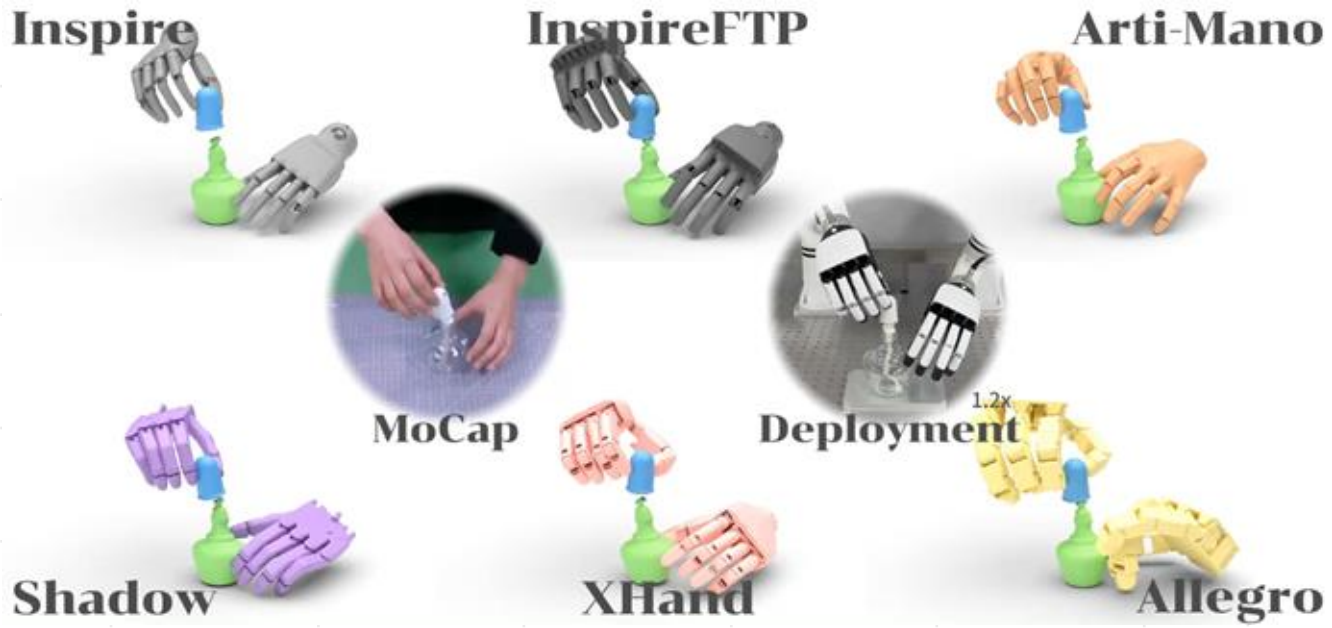
ManipTrans: Efficient Dexterous Bimanual Manipulation Transfer via Residual Learning

Jikai Wang

10/10/2025



Motivation

Rapid acquisition of precise, large-scale, and human-like dexterous manipulation sequences

Two Methods: RL & Teleoperation

- **Method One: Reinforcement Learning (RL)**
  - Use RL to explore and generate dexterous hand actions.
  - Requires carefully designed, task-specific reward functions ➔ Restricting Scalability and and task Complexity
- **Method Two: Teleoperation**
  - Labor-intensive and Costly, ➔ Embodiment-specific Datasets

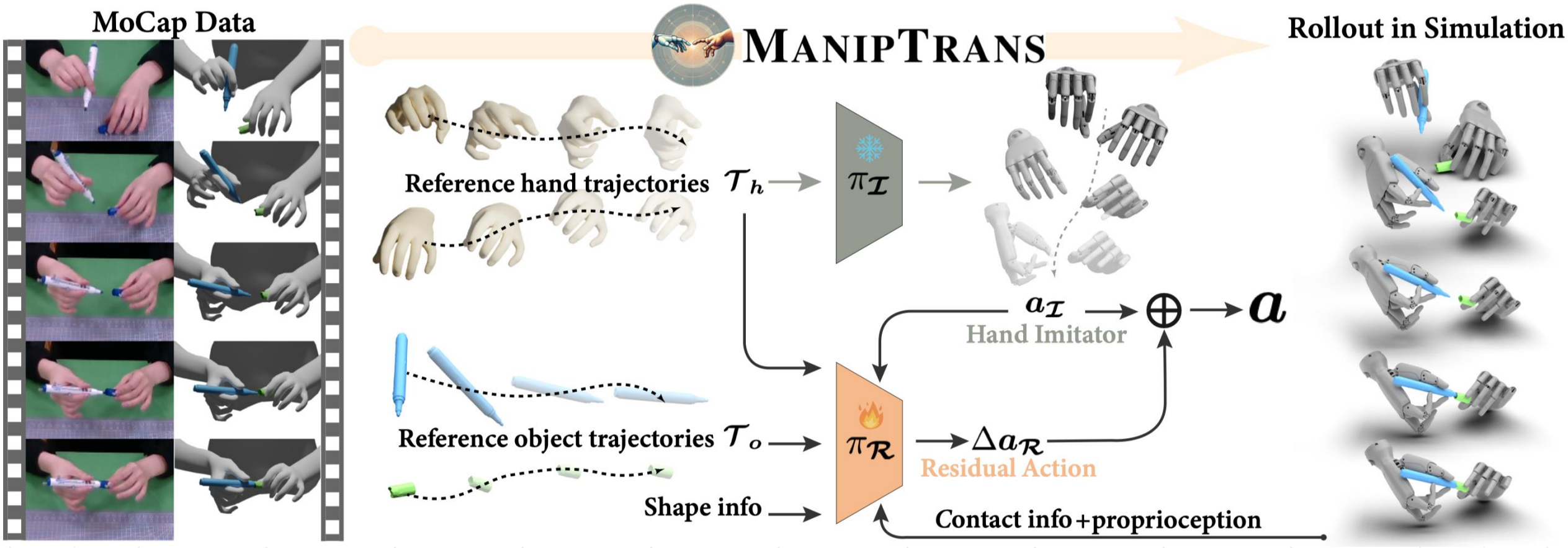
Solution: Imitation Learning

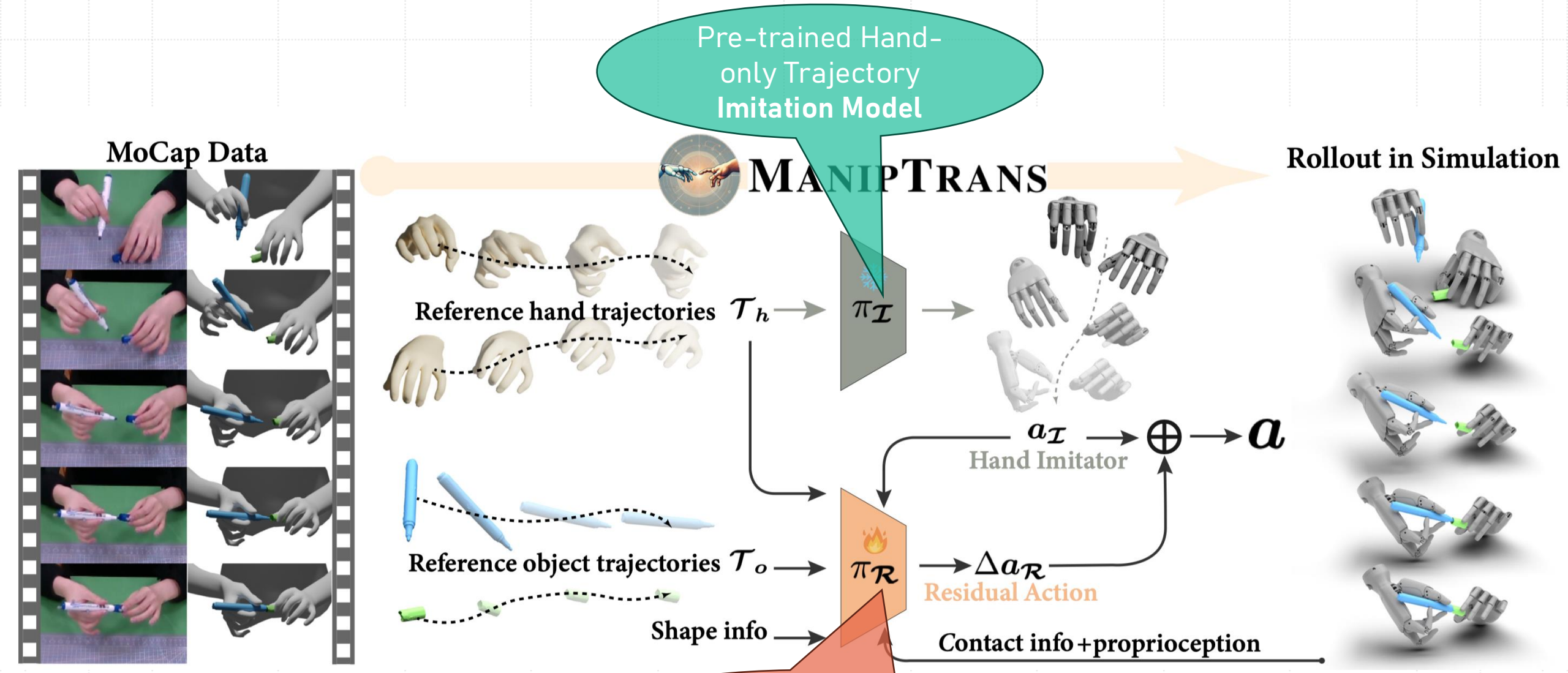
- **Imitation Learning in Simulated Environments**
  - **Advantages**
    - **Naturalistic** Hand-Object Interactions
    - Human Demonstrations are **Easily Accessible**: MoCap datasets and hand pose estimation techniques
    - **Cost-effective Validation** in Simulations
  - **Difficulties**
    - **Pose Retargeting**: morphological differences between human and robotic
    - **Critical Failures** during High-precision Tasks : error accumulation of Dataset
    - **High-dimensional Action Space**: significantly increasing the difficulty of efficient policy learning

Contributions

- **ManipTrans**: two-stage transfer framework for transfer of human bimanual manipulation to dexterous robotic hands in simulation.
- **DexManipNet**: a large-scale dataset of bimanual manipulation tasks.

Pipeline


MoCap Data
ManipTrans
Rollout in Simulation
Reference hand trajectories $\mathcal{T}_h$
$\pi_{\mathcal{I}}$
$a_{\mathcal{I}}$
Hand Imitator
$\oplus$
a
Reference object trajectories $\mathcal{T}_o$
$\pi_{\mathcal{R}}$
$\Delta a_{\mathcal{R}}$
Residual Action
Shape info
Contact info + proprioception


Pre-trained Hand-only Trajectory Imitation Model
MoCap Data
ManipTrans
Rollout in Simulation
Reference hand trajectories $\mathcal{T}_h$
$\pi_{\mathcal{I}}$
$a_{\mathcal{I}}$
Hand Imitator
$\oplus$
a
Reference object trajectories $\mathcal{T}_o$
$\pi_{\mathcal{R}}$
$\Delta a_{\mathcal{R}}$
Residual Action
Shape info
Contact info+proprioception
Residual Module that fine-tunes the coarse actions to ensure task compliance

Problem Formulation

- Dexterous Hands: $\boldsymbol{d} = \{d_l, d_r\}$
- Human Hands: $\boldsymbol{h} = \{h_l, h_r\}$
- Objects: $\boldsymbol{o} = \{o_l, o_r\}$
- Reference Trajectories:
  - Hand: $\boldsymbol{\mathcal{T}_h} = \{\boldsymbol{\tau_h^t}\}_{t=1}^{T}$
    $\boldsymbol{w_h} \in \mathbb{SE}(3)$ $\dot{\boldsymbol{w}}_{\boldsymbol{h}} = \{\boldsymbol{v_h}, \boldsymbol{u_h}\}$ $\boldsymbol{j_h} \in \mathbb{R}^{F\times 3}$
  - Object: $\boldsymbol{\mathcal{T}_o} = \{\boldsymbol{\tau_o^t}\}_{t=1}^{T}$
    $\boldsymbol{p_o} \in \mathbb{SE}(3)$ $\dot{\boldsymbol{p}}_{\boldsymbol{o}} = \{\boldsymbol{v_o}, \boldsymbol{u_o}\}$

- **Translations** are Relative to the dexterous hand's Wrist Position.
- **Original Rotations** are Preserved: for the correct gravity direction.

Problem Formulation

- **Modeled as an Implicit Markov Decision Process (MDP)**

$$\mathcal{M} = \langle \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{A}}, \mathbf{T}, \mathbf{R}, \gamma \rangle$$

- Action for each dexterous hand at time t: $\boldsymbol{a}^t \in \boldsymbol{\mathcal{A}}$
  - **Target positions** of each dexterous hand's joint: $\boldsymbol{a}_{\boldsymbol{q}}^t \in \mathbb{R}^K$
  - **6-DoF force** applied to the robotic wrist: $\boldsymbol{a}_{\boldsymbol{w}}^t \in \mathbb{R}^6$
- The State and Reward are defined separately for each stage

- State Space
- Action Space
- Transform Dynamics
- Reward Functions
- Discount Factor

Hand Trajectory Imitating

Stage One: Imitating Learning

- **Objective:** learn a general hand trajectory imitation model
- **Method:** RL, PPO
- **Policy:** $\pi_{\mathcal{I}}(\boldsymbol{a}^t | \boldsymbol{s}_{\mathcal{I}}^t, \boldsymbol{a}^{t-1})$
- **State Space:** for each dexterous hand at $\boldsymbol{s}_{\mathcal{I}}^t = \{\boldsymbol{\tau}_{\boldsymbol{h}}^t, \boldsymbol{s}_{\text{prop}}^t\}$
  - target hand trajectory $\boldsymbol{\tau}_{\boldsymbol{h}}^t$ current pro-prioception $\boldsymbol{s}_{\text{prop}}^t = \{\boldsymbol{q}_{\boldsymbol{d}}^t, \dot{\boldsymbol{q}}_{\boldsymbol{d}}^t, \boldsymbol{w}_{\boldsymbol{d}}^t, \dot{\boldsymbol{w}}_{\boldsymbol{d}}^t\}$
- **Reward:**

$$r_{\mathcal{I}}^t = w_{\text{wrist}} \cdot r_{\text{wrist}}^t + w_{\text{finger}} \cdot r_{\text{finger}}^t + w_{\text{smooth}} \cdot r_{\text{smooth}}^t$$

Wrist Tracking Reward

$$\boldsymbol{w}_{\boldsymbol{d}}^t \ominus \boldsymbol{w}_{\boldsymbol{h}}^t \quad \dot{\boldsymbol{w}}_{\boldsymbol{d}}^t - \dot{\boldsymbol{w}}_{\boldsymbol{h}}^t$$

Stage One: Imitating Learning

- **Objective:** learn a general hand trajectory imitation model
- **Method:** RL, PPO
- **Policy:** $\pi_{\mathcal{I}}(\boldsymbol{a}^t | \boldsymbol{s}_{\mathcal{I}}^t, \boldsymbol{a}^{t-1})$
- **State Space:** for each dexterous hand at $\boldsymbol{s}_{\mathcal{I}}^t = \{\boldsymbol{\tau}_{\boldsymbol{h}}^t, \boldsymbol{s}_{\text{prop}}^t\}$
  - target hand trajectory $\boldsymbol{\tau}_{\boldsymbol{h}}^t$ current pro-prioception $\boldsymbol{s}_{\text{prop}}^t = \{\boldsymbol{q}_{\boldsymbol{d}}^t, \dot{\boldsymbol{q}}_{\boldsymbol{d}}^t, \boldsymbol{w}_{\boldsymbol{d}}^t, \dot{\boldsymbol{w}}_{\boldsymbol{d}}^t\}$
- **Reward:**

$$r_{\mathcal{I}}^t = w_{\text{wrist}} \cdot r_{\text{wrist}}^t + w_{\text{finger}} \cdot r_{\text{finger}}^t + w_{\text{smooth}} \cdot r_{\text{smooth}}^t$$

Wrist Tracking Reward

Finger Imitation Reward

$$r_{\text{finger}}^t = \sum_{f=1}^{F} w_f \cdot \exp\left(-\lambda_f \| \boldsymbol{j}_{\boldsymbol{d}_f}^t - \boldsymbol{j}_{\boldsymbol{h}_f}^t \|_2^2\right)$$

Stage One: Imitating Learning

- **Objective:** learn a general hand trajectory imitation model
- **Method:** RL, PPO
- **Policy:** $\pi_{\mathcal{I}}(\boldsymbol{a}^t | \boldsymbol{s}_{\mathcal{I}}^t, \boldsymbol{a}^{t-1})$
- **State Space:** for each dexterous hand at $\boldsymbol{s}_{\mathcal{I}}^t = \{\boldsymbol{\tau}_{\boldsymbol{h}}^t, \boldsymbol{s}_{\text{prop}}^t\}$
  - target hand trajectory $\boldsymbol{\tau}_{\boldsymbol{h}}^t$ current pro-prioception $\boldsymbol{s}_{\text{prop}}^t = \{\boldsymbol{q}_{\boldsymbol{d}}^t, \dot{\boldsymbol{q}}_{\boldsymbol{d}}^t, \boldsymbol{w}_{\boldsymbol{d}}^t, \dot{\boldsymbol{w}}_{\boldsymbol{d}}^t\}$
- **Reward:**

$$r_{\mathcal{I}}^t = w_{\text{wrist}} \cdot r_{\text{wrist}}^t + w_{\text{finger}} \cdot r_{\text{finger}}^t + w_{\text{smooth}} \cdot r_{\text{smooth}}^t$$

Wrist Tracking Reward

Finger Imitation Reward

Smoothness Reward

element-wise product of joint velocities and torques

Stage One: Imitating Learning

- **Training Strategy:**
  - **Hand-only Datasets:** mirrored to balance the left and right hands
  - **Early Termination:**
    - If the dexterous hand keypoints j_d deviate beyond a threshold ϵ_{finger}, the episode terminates early
  - **Reference State Initialization (RSI):**
    - Reset to a randomly sampled MoCap state
  - **Curriculum Learning:**
    - Initially, gradually reducing ϵ_{finger} to encourage broad exploration
    - Then, focusing on fine-grained finger control

Residual Learning for Interaction

Stage Two: Residual Learning

- **Objective:** refine coarse actions and satisfy task-specific constraints by residual actions $\Delta \boldsymbol{a}_{\mathcal{R}}^{t}$
- **Object's Collider Mesh:** convex hull of the object's mesh
- **State Space Expansion:** $\boldsymbol{s}_{\mathcal{R}}^{t} = \boldsymbol{s}_{\mathcal{I}}^{t} \cup \boldsymbol{s}_{\text{interact}}^{t}$

$$\boldsymbol{s}_{\mathcal{I}}^{t} = \{\boldsymbol{\tau}_{\boldsymbol{h}}^{t}, \boldsymbol{s}_{\text{prop}}^{t}\}$$

$$\boldsymbol{s}_{\text{interact}}^{t} = \{\boldsymbol{\tau}_{\boldsymbol{o}}^{t}, \boldsymbol{p}_{\hat{\boldsymbol{o}}}^{t}, \dot{\boldsymbol{p}}_{\hat{\boldsymbol{o}}}^{t}, \boldsymbol{m}_{\hat{\boldsymbol{o}}}^{t}, \boldsymbol{G}_{\hat{\boldsymbol{o}}}^{t}, \text{BPS}(\hat{\boldsymbol{o}}), \boldsymbol{D}(\boldsymbol{j}_{\boldsymbol{d}}^{t}, \boldsymbol{p}_{\hat{\boldsymbol{o}}}^{t}), \boldsymbol{C}^{t}\}$$

Stage Two: Residual Learning

- **Objective:** refine coarse actions and satisfy task-specific constraints
- **Object's Collider Mesh:** convex hull of the object's mesh
- **State Space Expansion:** $s_{\mathcal{R}}^t = s_{\mathcal{I}}^t \cup s_{\text{interact}}^t$

$$s_{\text{interact}}^t = \{\tau_o^t, p_{\hat{o}}^t, \dot{p}_{\hat{o}}^t, m_{\hat{o}}^t, G_{\hat{o}}^t, \text{BPS}(\hat{o}), D(j_d^t, p_{\hat{o}}^t), C^t\}$$

Object's Position

Object's Velocities

Target Object Trajectory

Object's Center of Mass

Stage Two: Residual Learning

- **Objective:** refine coarse actions and satisfy task-specific constraints
- **Object's Collider Mesh:** convex hull of the object's mesh
- **State Space Expansion:** $s_{\mathcal{R}}^t = s_{\mathcal{I}}^t \cup s_{\text{interact}}^t$

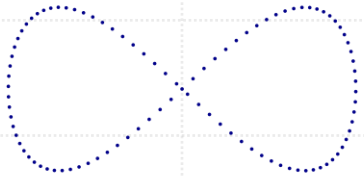
$$D(j_d^t, p_{\hat{o}}^t) = \|j_d^t - p_{\hat{o}}^t\|_2^2$$

Gravitational Force Vector

H-O Spatial Relationship

$$s_{\text{interact}}^t = \{\tau_o^t, p_{\hat{o}}^t, \dot{p}_{\hat{o}}^t, m_{\hat{o}}^t, G_{\hat{o}}^t, \text{BPS}(\hat{o}), D(j_d^t, p_{\hat{o}}^t), C^t\}$$

BPS Representation of Object Shape



Contact Force

- point cloud $X = \{x_1, \ldots, x_n\}$

Stage Two: Residual Learning

- **Objective:** refine coarse actions and satisfy task-specific constraints
- **Object's Collider Mesh:** convex hull of the object's mesh
- **State Space Expansion:** $s_{\mathcal{R}}^t = s_{\mathcal{I}}^t \cup s_{\text{interact}}^t$
- **Residual Actions Combining Strategy:**
  - First, sample the imitation action
  - Then, sample the residual correction
  - a^t is then clipped to adhere to the dexterous hand's joint limits

$$a_{\mathcal{I}}^t \sim \pi_{\mathcal{I}}(a^t | s_{\mathcal{I}}^t, a^{t-1})$$
$$\Delta a_{\mathcal{R}}^t \sim \pi_{\mathcal{R}}(\Delta a^t | s_{\mathcal{R}}^t, a_{\mathcal{I}}^t, a^{t-1})$$

$$a^t = a_{\mathcal{I}}^t + \Delta a_{\mathcal{R}}^t$$

Stage Two: Residual Learning

- **Objective:** refine coarse actions and satisfy task-specific constraints
- **Object's Collider Mesh:** convex hull of the object's mesh
- **State Space Expansion:** $s_{\mathcal{R}}^t = s_{\mathcal{I}}^t \cup s_{\text{interact}}^t$
- **Residual Actions Combining Strategy:** $a^t = a_{\mathcal{I}}^t + \Delta a_{\mathcal{R}}^t$
- **Reward:** (avoid task-specific reward engineering)

$$r_{\mathcal{R}}^t = r_{\mathcal{I}}^t + w_{\text{object}} \cdot r_{\text{object}}^t + w_{\text{contact}} \cdot r_{\text{contact}}^t$$

Object Following Reward

Contact Force Reward

Stage Two: Residual Learning

- **Objective:** refine coarse actions and satisfy task-specific constraints
- **Object's Collider Mesh:** convex hull of the object's mesh
- **State Space Expansion:** $s_{\mathcal{R}}^t = s_{\mathcal{I}}^t \cup s_{\text{interact}}^t$
- **Residual Actions Combining Strategy:** $a^t = a_{\mathcal{I}}^t + \Delta a_{\mathcal{R}}^t$
- **Reward:** (avoid task-specific reward engineering)

$$r_{\mathcal{R}}^t = r_{\mathcal{I}}^t + w_{\text{object}} \cdot r_{\text{object}}^t + w_{\text{contact}} \cdot r_{\text{contact}}^t$$

Object Following Reward

Contact Force Reward

$$p_{\hat{o}}^t \ominus p_o^t$$

$$\dot{p}_{\hat{o}}^t - \dot{p}_o^t$$

Stage Two: Residual Learning

- **Objective:** refine coarse actions and satisfy task-specific constraints
- **Object's Collider Mesh:** convex hull of the object's mesh
- **State Space Expansion:** $\boldsymbol{s}_{\mathcal{R}}^{t} = \boldsymbol{s}_{\mathcal{I}}^{t} \cup \boldsymbol{s}_{\text{interact}}^{t}$
- **Residual Actions Combining Strategy:** $\boldsymbol{a}^{t} = \boldsymbol{a}_{\mathcal{I}}^{t} + \Delta \boldsymbol{a}_{\mathcal{R}}^{t}$
- **Reward:** (avoid task-specific reward engineering)

$$r_{\mathcal{R}}^{t} = r_{\mathcal{I}}^{t} + w_{\text{object}} \cdot r_{\text{object}}^{t} + w_{\text{contact}} \cdot r_{\text{contact}}^{t}$$

Object Following Reward

Contact Force Reward

$$r_{\text{contact}}^{t} = w_{\text{c}} \cdot \exp\left(\frac{-\lambda_{\text{c}}}{\sum_{f=1}^{F} \boldsymbol{C}_{\boldsymbol{d}_f}^{t} \cdot \mathbb{1}\left(\boldsymbol{D}(\boldsymbol{j}_{\boldsymbol{h}_f}^{t}, \boldsymbol{p}_{\boldsymbol{o}}^{t} \cdot \boldsymbol{o}) < \xi_{\text{c}}\right)}\right)$$

Stage Two: Residual Learning

- **Training Strategy:**
  - **Physical Constraints Adjust: Gravity $\mathcal{G}$ and friction coefficient $\mathcal{F}$**
    - Initially, set $\mathcal{G}$ to zero and $\mathcal{F}$ to a high value
    - Then, gradually restore $\mathcal{G}$ to its true value and reduce $\mathcal{F}$ to a suitable value to approximate real interactions
  - **Early Termination:**
    - If the object's pose $p_{\hat{o}}^{t}$ deviates beyond a predefined threshold ϵ_{object}
    - If MoCap data indicates a firm grasp by the human hands, the contact force must be nonzero
  - **Reference State Initialization (RSI):**
    - Reset the robotic hands by randomly selecting a non-colliding near-object state
  - **Curriculum Learning:**
    - Progressively reducing ϵ_{object} to encourage more precise object manipulation
    - Then, focusing on fine-grained finger control

DexManipNet Dataset

- Derived from **FAVOR** and **OakInk-V2**
  - 61 diverse and challenging tasks
  - 3.3K episodes of robotic hand manipulation over 1.2K objects
  - totaling 1.34 million frames, including ~ 600 sequences involving complex bimanual tasks
- **Robotic Hand:** Inspire Hand with 12-DoF configuration
- **Simulation :** Isaac Gym

Experiments

Datasets and Metrics

- **Datasets**:
  - OakInk-V2, GRAB, FAOVR, ARCTIC
- **Metrics**:
  - Per-frame Average Object Rotation and Translation Error

$$E_r = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{p}_{\text{rot}}{}_{\hat{\boldsymbol{o}}}^{t} \cdot (\boldsymbol{p}_{\text{rot}}{}_{\boldsymbol{o}}^{t})^{-1}) \qquad E_t = \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{p}_{\text{tsl}}{}_{\hat{\boldsymbol{o}}}^{t} - \boldsymbol{p}_{\text{tsl}}{}_{\boldsymbol{o}}^{t}\|_2^2$$

  - Mean Per-Joint Position Error

$$E_j = \frac{1}{T \cdot F}\sum_{t=1}^{T}\sum_{f=1}^{F} \|\boldsymbol{j}_{\boldsymbol{d}_f}^{t} - \boldsymbol{j}_{\boldsymbol{h}_f}^{t}\|_2^2.$$

  - Mean Per-Fingertip Position Error

$$E_{ft} = \frac{1}{T \cdot M}\sum_{t=1}^{T}\sum_{ft=1}^{M} \|\boldsymbol{t}_{\boldsymbol{d}_{ft}}^{t} - \boldsymbol{t}_{\boldsymbol{h}_{ft}}^{t}\|_2^2$$

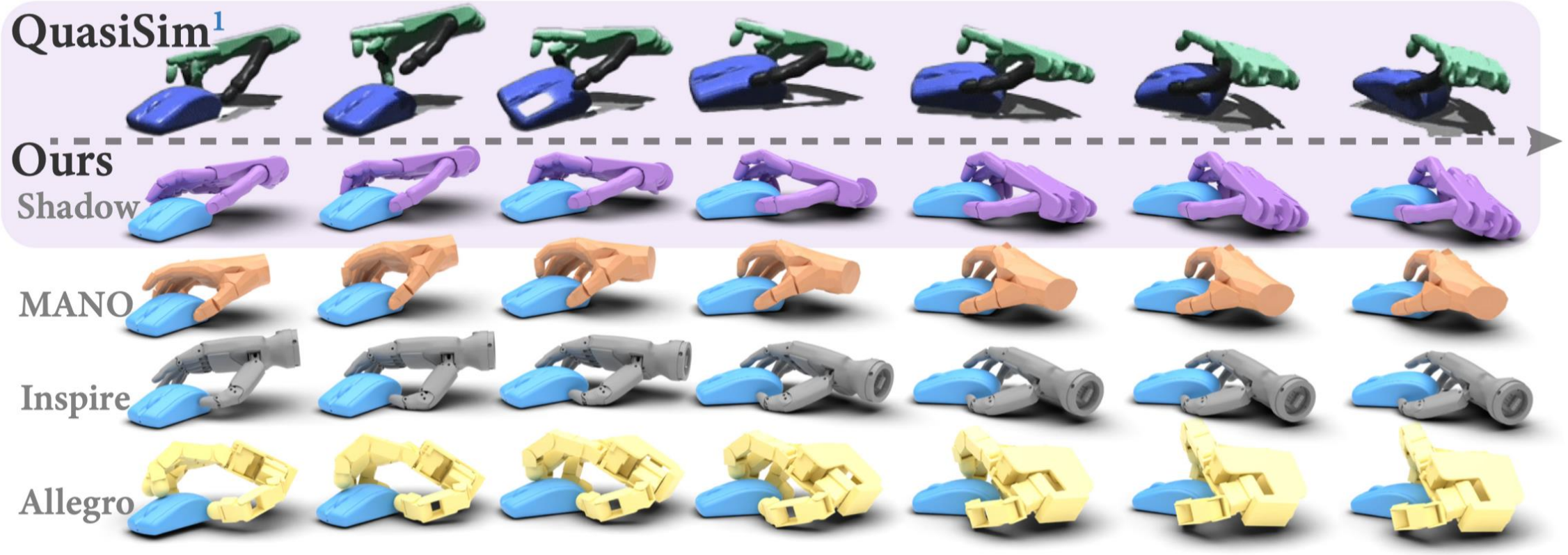
  - Success Rate (SR):
    - A tracking attempt is deemed successful if all Errors are all below the specified thresholds

Comparison with RL-Combined Methods

- **RL-Only:** using only trajectory-following rewards, employing the PPO algorithm to train the robotic hand from scratch
- **Retarget + Residual:** applying residual action to retargeted robotic hand poses obtained via alignment between human and robot keypoints
- **Retarget-Only:** retargeting without any learning

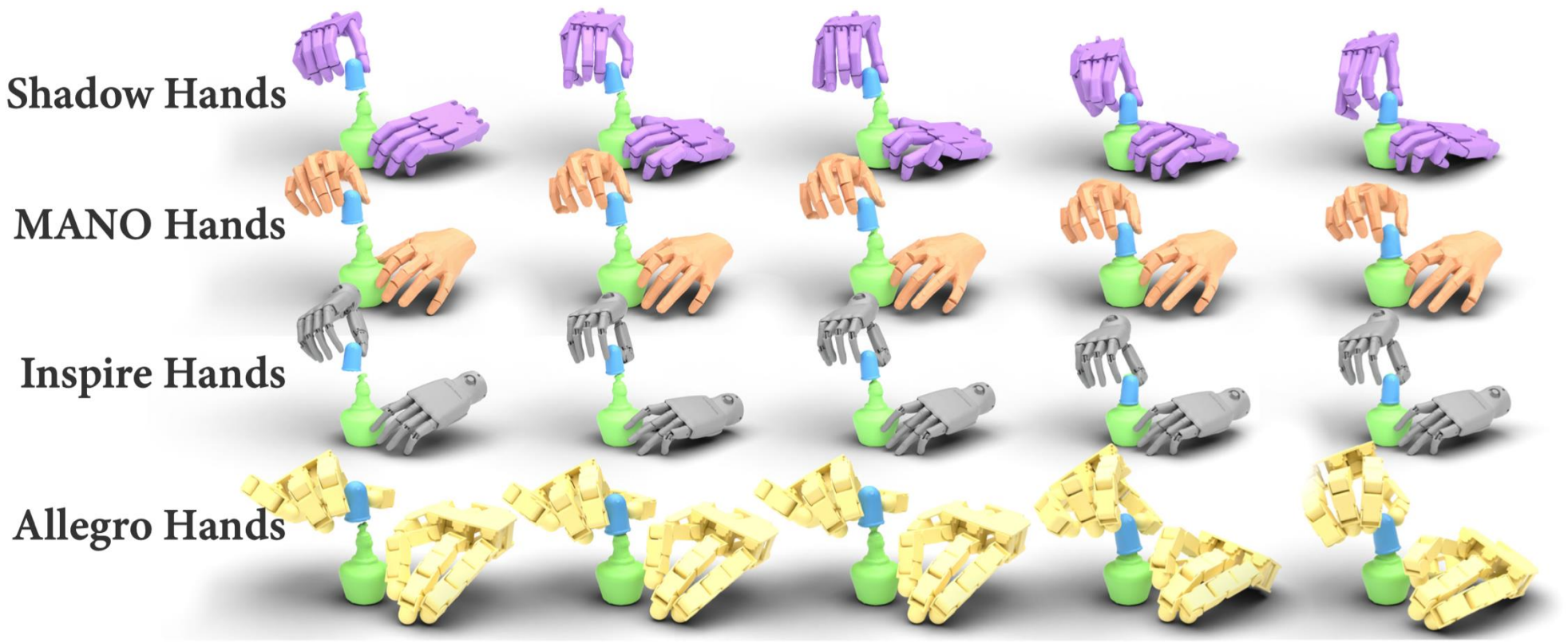
Methods	$E_r \downarrow$	$E_t \downarrow$	$E_j \downarrow$	$E_{ft} \downarrow$	$SR \uparrow$
Retarget-Only	N/A	N/A	N/A	N/A	4.6 / 0.0
RL-Only	9.72	1.23	2.96	2.38	34.3 / 12.1
Retarget + Residual	11.58	0.79	2.54	1.74	47.8 / 13.9
ManipTrans	**8.60**	**0.49**	**2.15**	**1.36**	**58.1 / 39.5**

Comparison with Optimization-Based Method (qualitative comparison)



Cross-Embodiments Validation

- Robotic Hands: Shadow Hand, MANO hand, Inspire Hand, and Allegro Hand



Real-World Deployment

- **Platform:** two 7-DoF Realman arms & two upgraded Inspire Hands
- **To bridge the gap** between the simulated 12-DoF robotic hands and the 6-DoF real hardware:
  - Fingertip Alignment Loss:

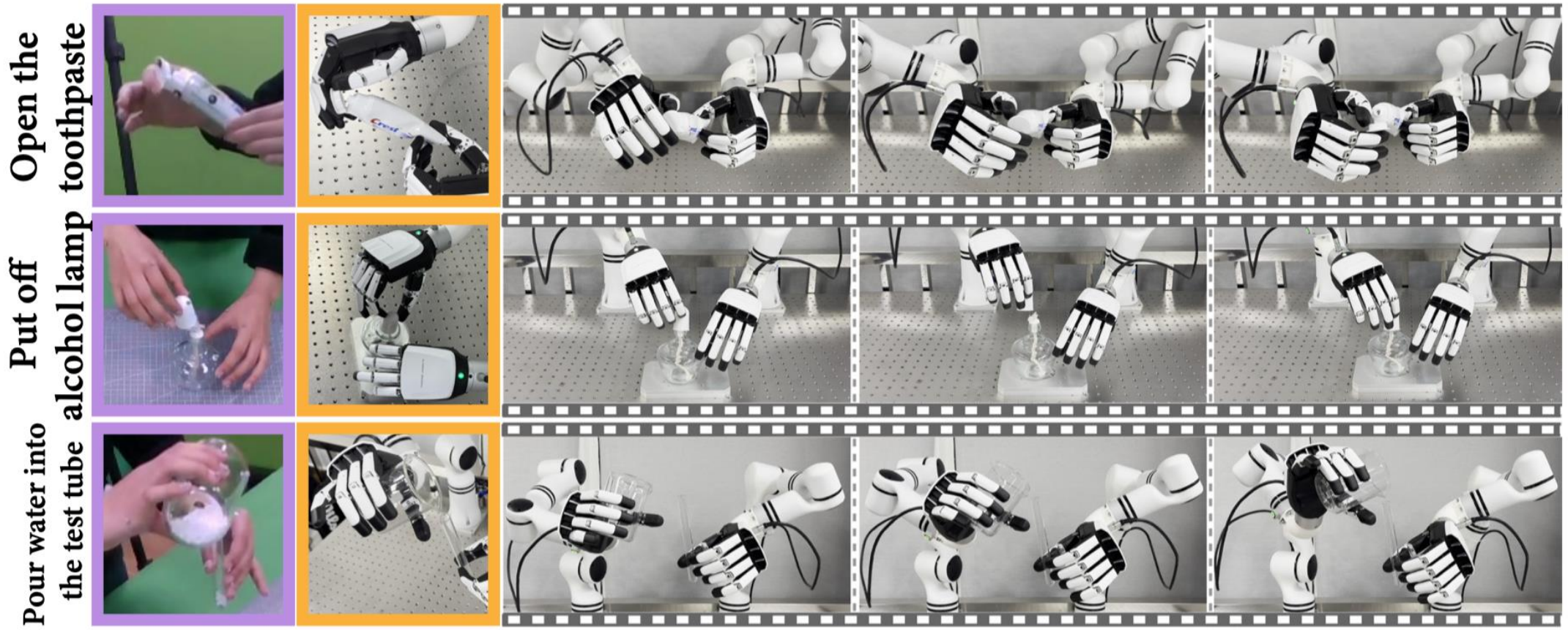
$$\text{argmin}_{\boldsymbol{q}_{\tilde{d}}} \frac{1}{T \cdot M} \sum_{t=1}^{T} \sum_{ft=1}^{M} \| \boldsymbol{t}_{\boldsymbol{d}_{ft}}^{t} - \boldsymbol{t}_{\tilde{\boldsymbol{d}}_{ft}}^{t} \|_2^2$$

  - Temporal Smoothness Loss

$$L_{\text{smooth}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \| \boldsymbol{q}_{\tilde{\boldsymbol{d}}}^{t+1} - \boldsymbol{q}_{\tilde{\boldsymbol{d}}}^{t} \|_2^2$$

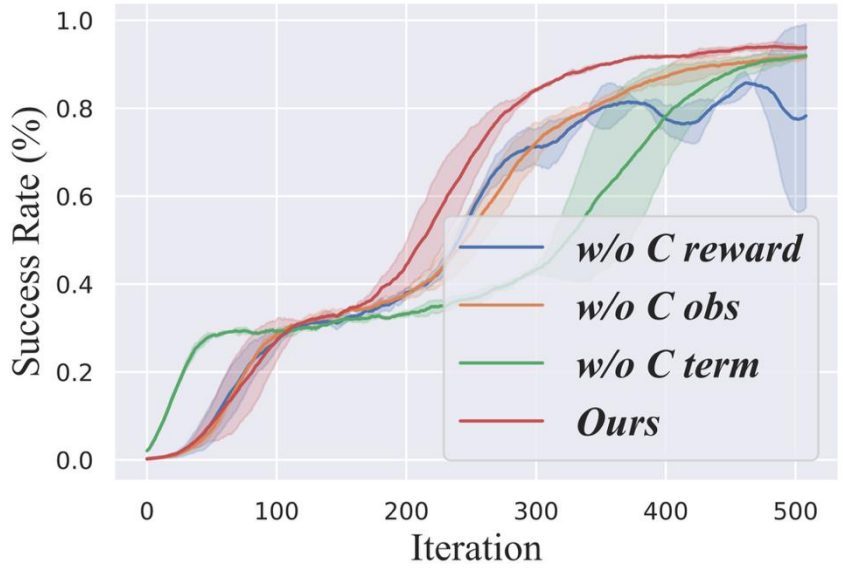
- **Arm Control**: solving inverse kinematics to align the arms' flanges with the dexterous hands' wrists

Real-World Deployment



Abalation Studies

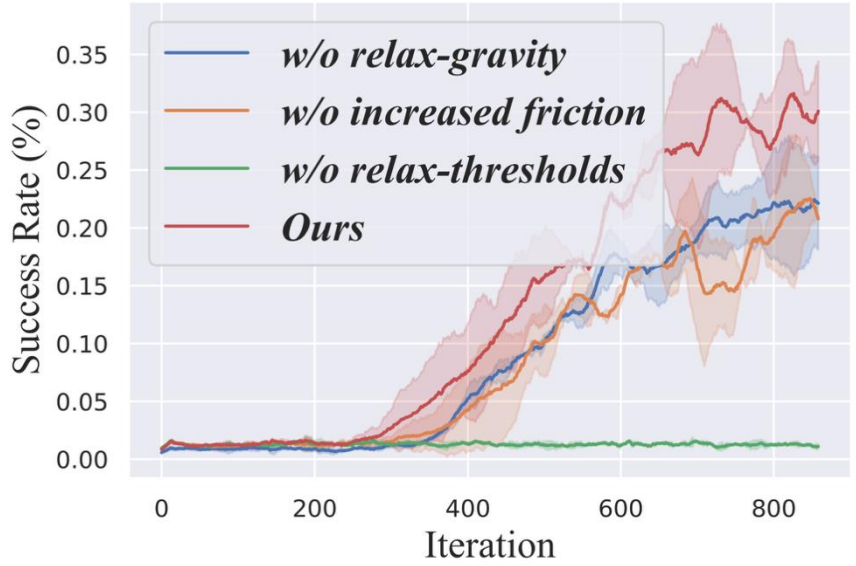
- Tactile Information
  - as an observation input
  - as a reward component to encourage contact
  - as a condition for early termination



(a) Tactile ablations training curve.

Abalation Studies

- Training Strategy
  - relaxing gravity effects
  - increasing friction influence
  - relaxing thresholds



(b) Curve on training strategies.

DexManipNet for Policy Learning

- **Representative Imitation Learning Methods**:
  - two regression-based behavior cloning approaches: IBC and BET
  - two diffusion policy methods: with UNet and Transformer backbones
- **Policy Learning Task: m**oving a bottle to a goal position
  - Each policy is trained on 85% of the 140 sequences involving the bottle rearrangement task and evaluated on the remaining 15%
  - 20 rollouts per sequence
  - A rollout is considered successful if the object's final position is within 10 cm of the goal

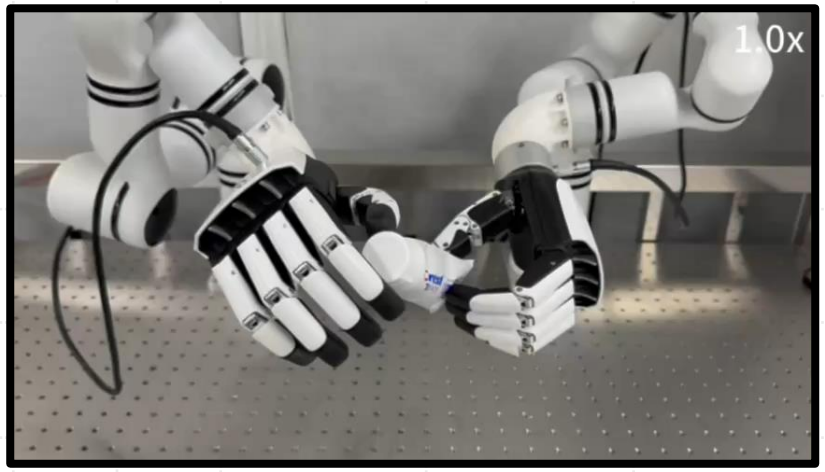
Methods	IBC [33]	BET [101]	DP-UNet [25]	DP-Trans [25]
SR	4.69%	9.69%	**18.44%**	14.69%

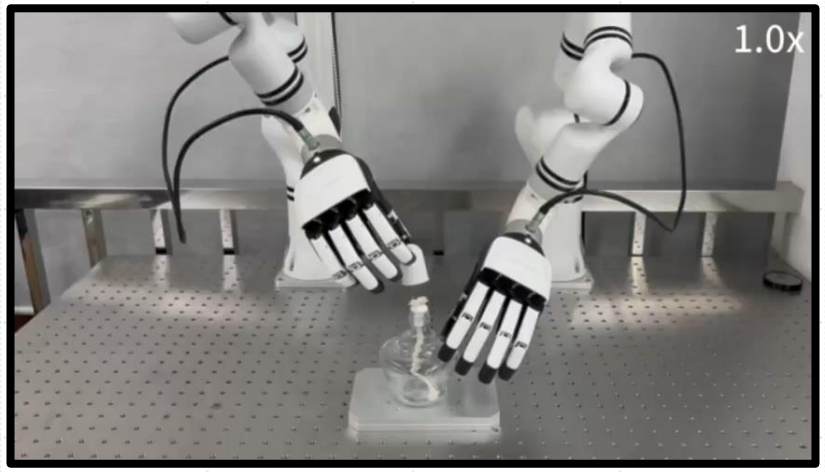
Table 2. **Imitating Learning on Bottle Rearrangement Task.**

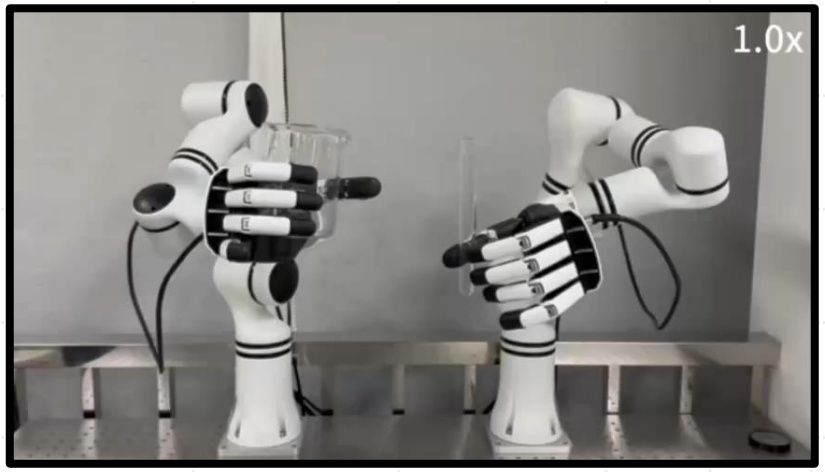
Simulation Results

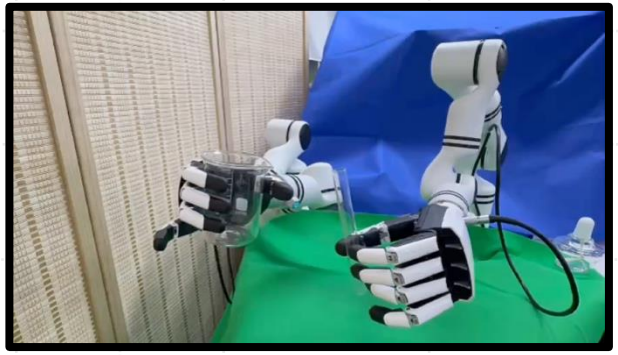
Scoop something	Stir	Put on lid
Write on a whiteboard	Write with a pencil	Cap the pen

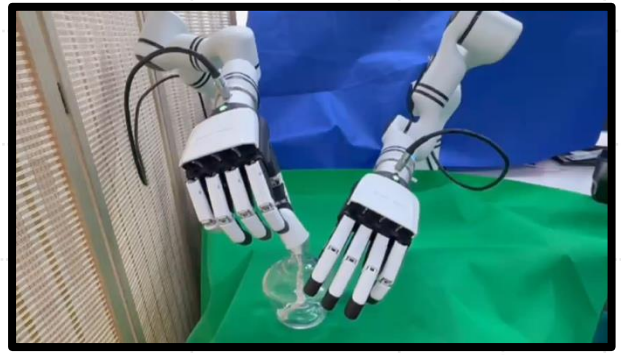
Real-world Results











Thanks!