Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MIN Faculty
Department of Informatics

# Industrial Robotic Assembly

## Oberseminar TAMS


Yunlei Shi

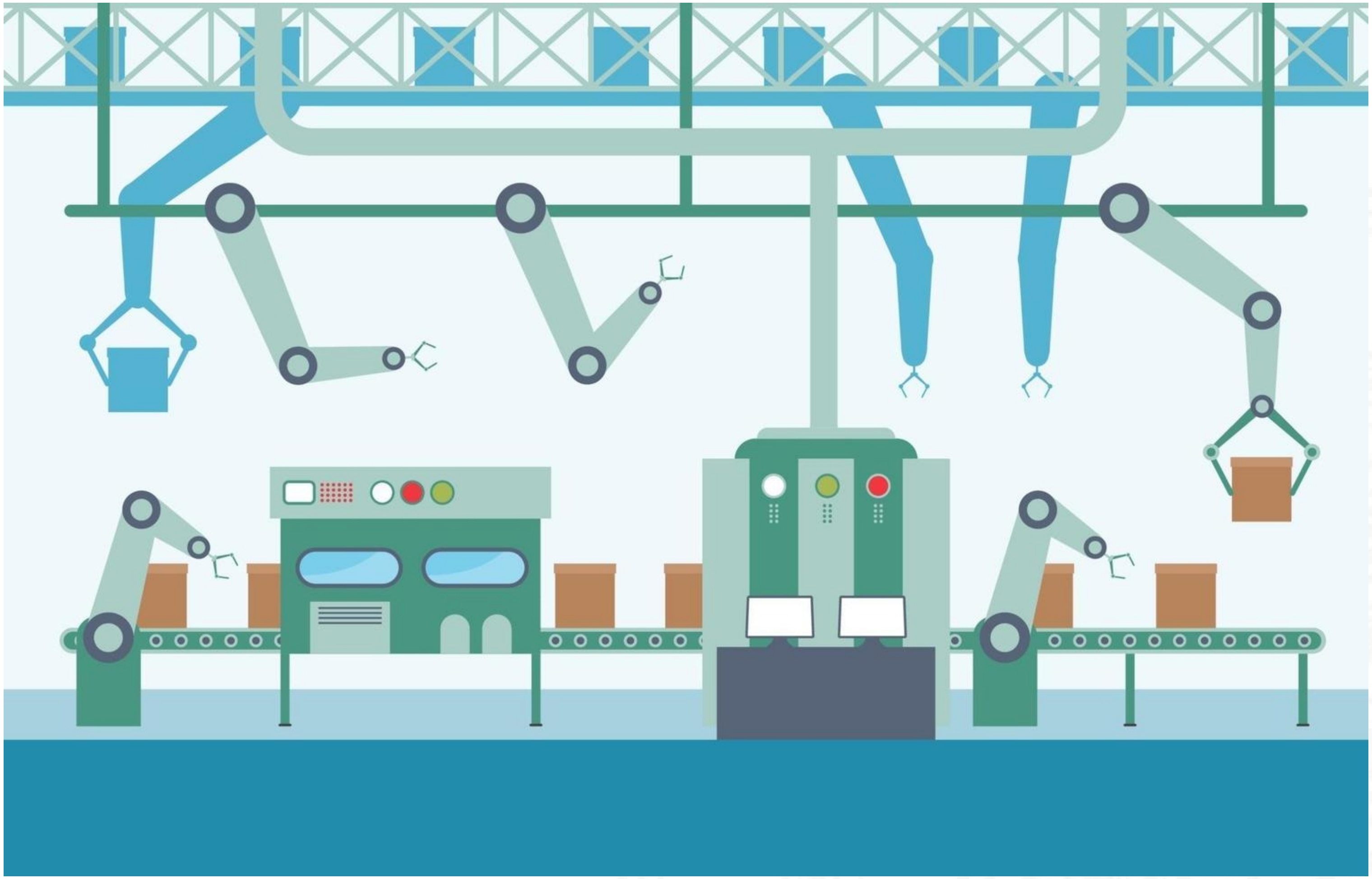
University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
**Technical Aspects of Multimodal Systems**


20.04.2021

# Table of Contents

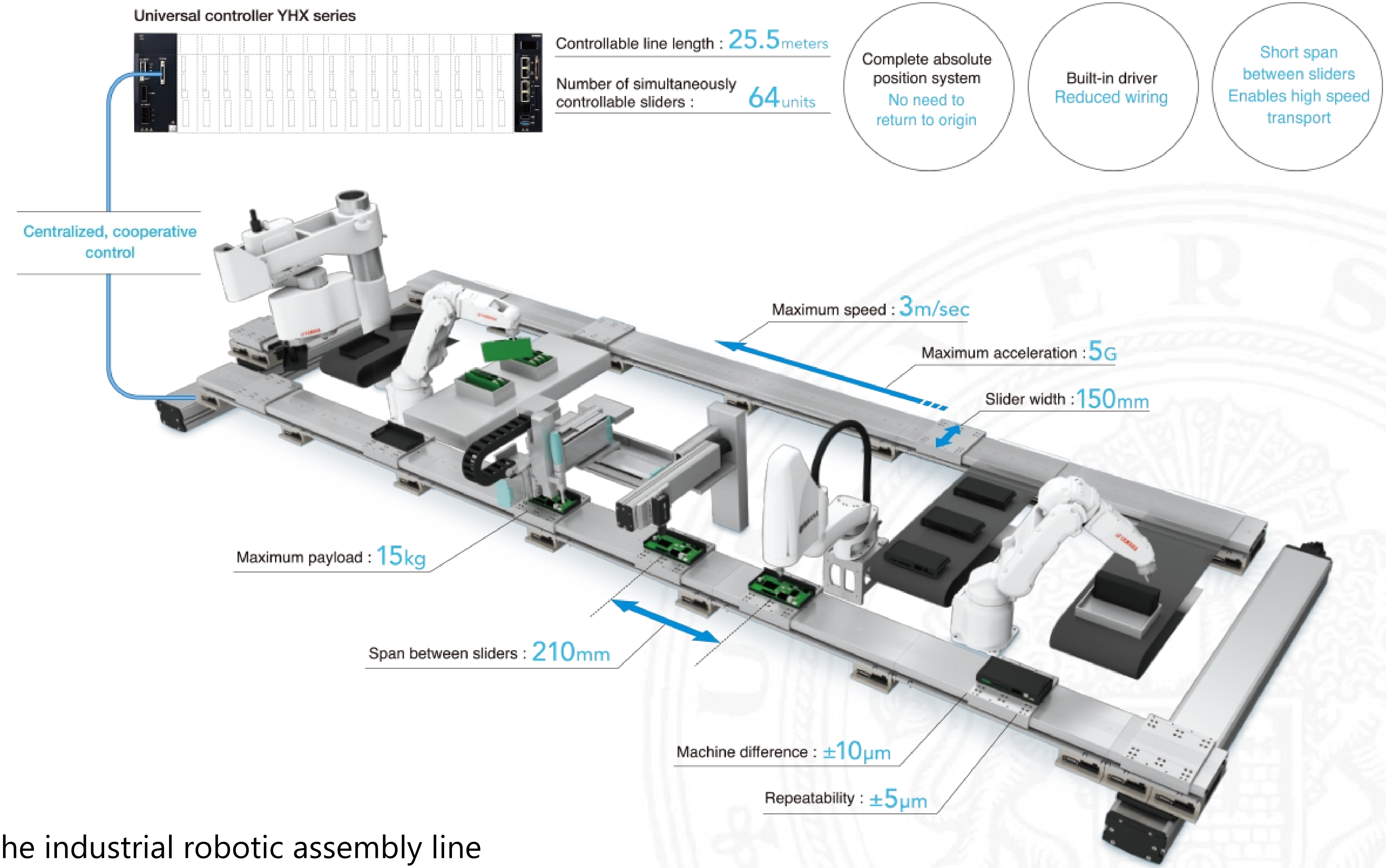
# Table of Contents

# Introduction

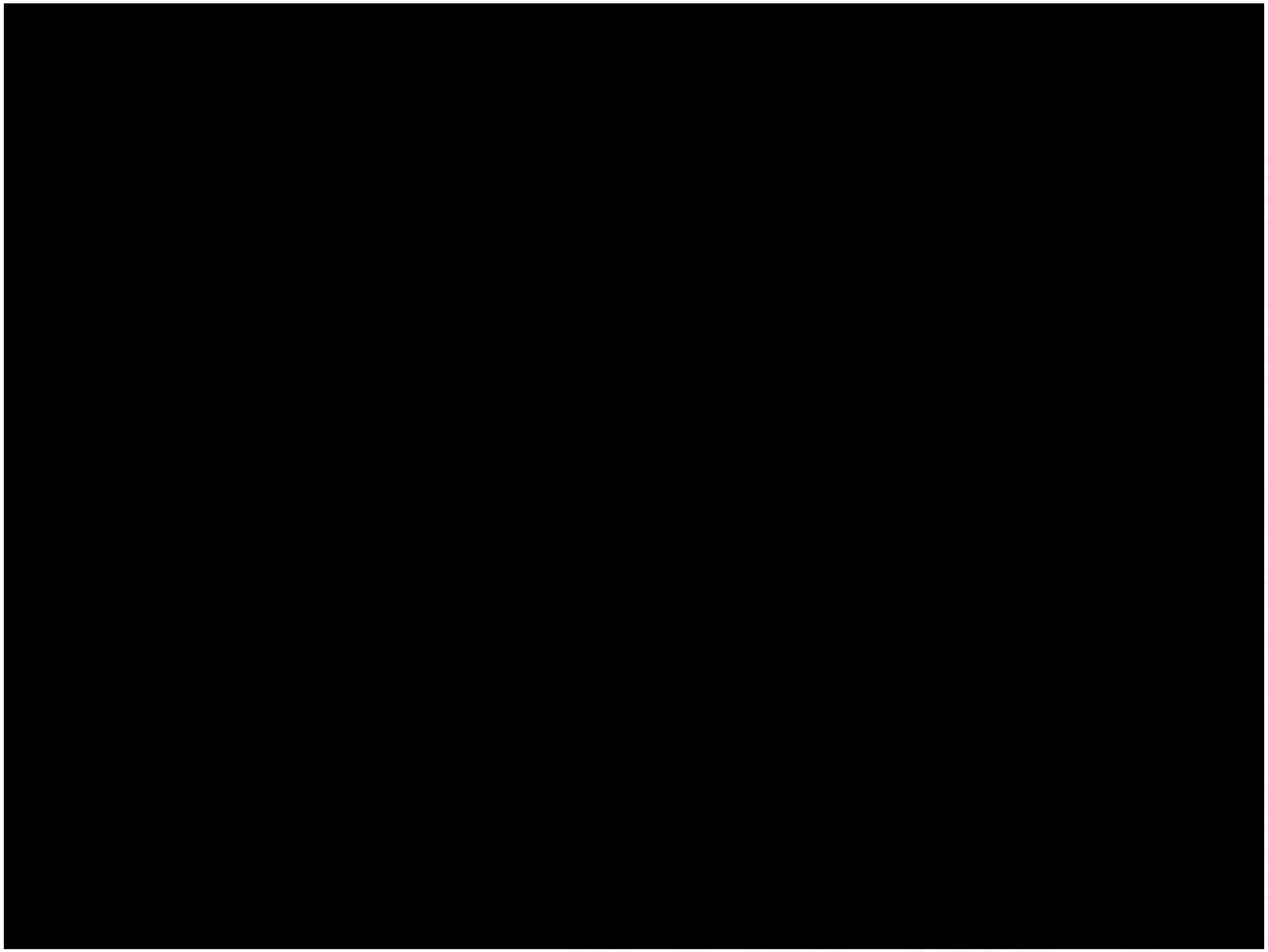


The typical industry production line scenario

# Introduction




The industrial robotic assembly line
(YAMAHA Advanced Robotics Automation Platform)



Example of typical tedious, monotonous tasks

# Introduction

Motivation

- Robotics has largely contributed to increasing industrial productivity and
- to helping factory workers on tedious, monotonous, dangerous tasks

| | Human Operator | Collaborative Systems | Traditional Robot |
|---|---|---|---|
| Assembly | High dexterity and flexibility | Combines human dexterity with robot capabilities [24] | Dexterity/flexibility could be unreachable [24] |
| Placement | High dexterity | Commercial cobots have lower repeatability | High repeatability and payload |
| Handling | Product weight restricted [19] | Typical cobots have low payload | High payload and speed [23] |
| Picking | Product weight restricted [19] | Typical cobots have low payload | High payload and repeatability [23] |

The main industry tasks

## Motivation

Problem Statement:

1. Robotic assembly production lines and tasks are difficult to set up
    - Installation and tuning of robots and devices cost lots of time
    - Ease-of-programming has been identified as an open challenge in robot assembly
2. Assembly task success rates requirement are high (>99%)
3. TAKT time requirements are high
    - Normally, less time than human worker

# Table of Contents

# Related Work

University of California, Berkeley

•Schoettler, Gerrit, et al. "Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards." 2020 International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020.

•Johannink, Tobias, et al. "Residual reinforcement learning for robot control." 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.

•Luo, Jianlan, et al. "Reinforcement learning on variable impedance controller for high-precision robotic assembly." 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.

•Luo, Jianlan, et al. "Deep reinforcement learning for robotic assembly of mixed deformable and rigid objects." 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018.

# Related Work

University of California, Berkeley

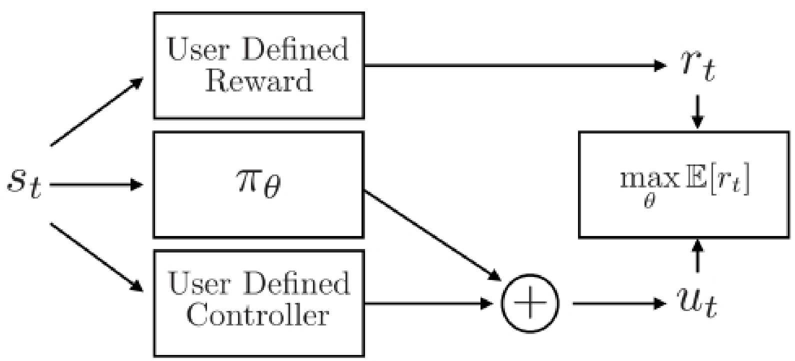
Deep Reinforcement Learning for Industrial Insertion Tasks

Luo, Jianlan, et al. "Reinforcement learning on variable impedance controller for high-precision robotic assembly." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

# Related Work

University of California, Berkeley

## Method overview




**Algorithm 1** Residual reinforcement learning

**Require:** policy $\pi_\theta$, hand-engineered controller $\pi_{\mathrm{H}}$.

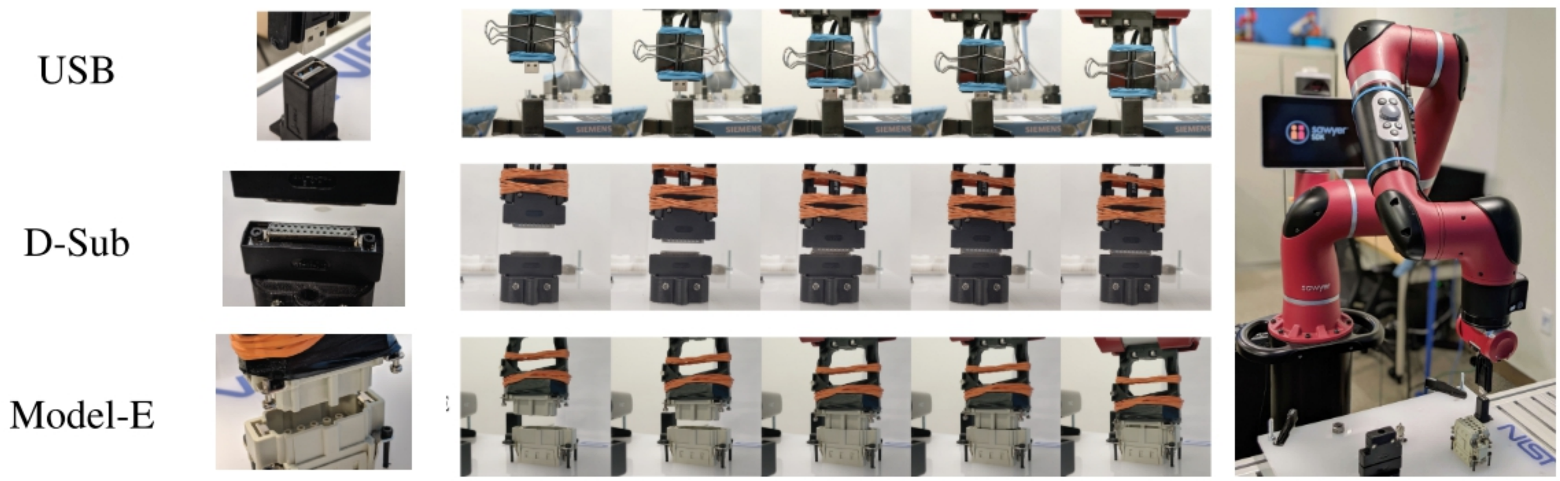
**for** $n = 0, ..., N-1$ episodes **do**
  Sample initial state $s_0 \sim E$.
  **for** $t = 0, ..., H-1$ steps **do**
    Get policy action $u_t \sim \pi_\theta(u_t|s_t)$.
    Get action to execute $u'_t = u_t + \pi_{\mathrm{H}}(s_t)$.
    Get next state $s_{t+1} \sim p(\cdot \mid s_t, u'_t)$.
    Store $(s_t, u_t, s_{t+1})$ into replay buffer $\mathcal{R}$.
    Sample set of transitions $(s, u, s') \sim \mathcal{R}$.
    Optimize $\theta$ using RL with transitions.
  **end for**
**end for**

Luo, Jianlan, et al. "Reinforcement learning on variable impedance controller for high-precision robotic assembly." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

# Related Work

University of California, Berkeley

## Experiment result




Luo, Jianlan, et al. "Reinforcement learning on variable impedance controller for high-precision robotic assembly." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

# Related Work

University of California, Berkeley

## Experiment result

| D-Sub Connector | | Goal: Perfect | Goal: Noisy |
|---|---|---|---|
| Pure RL | Dense | 16% | 0% |
| | Images, SAC | 0% | 0% |
| | Images, TD3 | 12% | 12% |
| RL + LfD | Images | 52% | 52% |
| Residual RL | Dense | **100%** | 60% |
| | Images, SAC | **100%** | **64%** |
| | Images, TD3 | 52% | 52% |
| Human | P-Controller | **100%** | 44% |

| Model-E Connector | | Goal: Perfect | Goal: Noisy |
|---|---|---|---|
| Pure RL | Dense | 0% | 0% |
| | Images, SAC | 0% | 0% |
| | Images, TD3 | 0% | 0% |
| RL + LfD | Images | 20% | 20% |
| Residual RL | Dense | **100%** | **76%** |
| | Images, SAC | **100%** | **76%** |
| | Images, TD3 | 0% | 0% |
| Human | P-Controller | 52% | 24% |

| USB Connector | | Goal: Perfect | Goal: Noisy |
|---|---|---|---|
| Pure RL | Dense | 28% | 20% |
| | Sparse, SAC | 16% | 8% |
| | Sparse, TD3 | 44% | 28% |
| | Images, SAC | 36% | 32% |
| | Images, TD3 | 28% | 28% |
| RL + LfD | Sparse | **100%** | 32% |
| | Images | 88% | 60% |
| Residual RL | Dense | **100%** | **84%** |
| | Sparse, SAC | 88% | **84%** |
| | Sparse, TD3 | **100%** | 36% |
| | Images, SAC | **100%** | 80% |
| | Images, TD3 | 0% | 0% |
| Human | P-Controller | **100%** | 60% |

Luo, Jianlan, et al. "Reinforcement learning on variable impedance controller for high-precision robotic assembly." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

- Lee, Michelle A., et al. "Making sense of vision and touch: Learning multimodal representations for contact-rich tasks." IEEE Transactions on Robotics 36.3 (2020): 582-596.

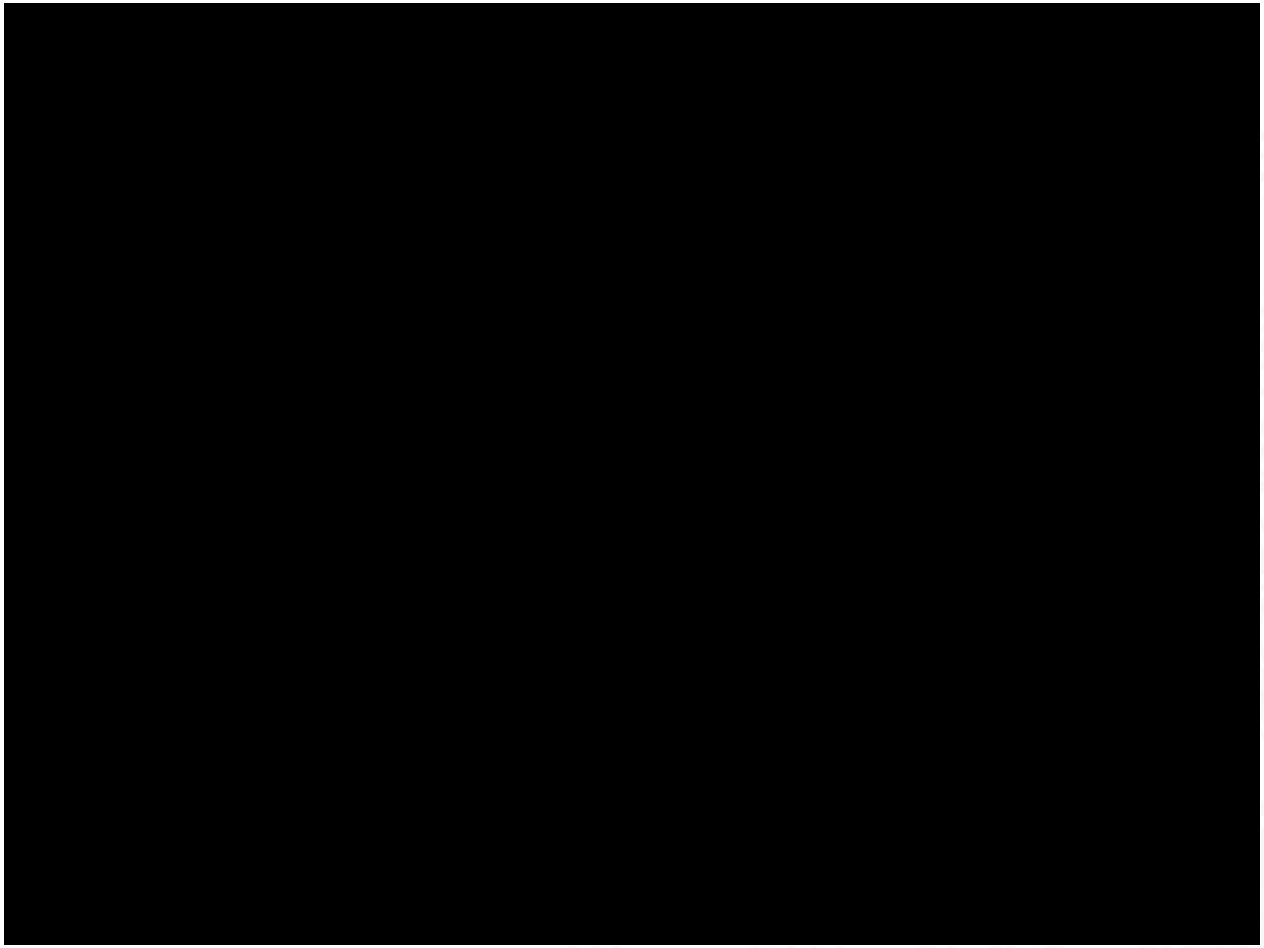
- Lee, Michelle A., et al. "Guided uncertainty-aware policy optimization: Combining learning and model-based strategies for sample-efficient policy learning." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.

- Lee, Michelle A., et al. "Multimodal Sensor Fusion with Differentiable Filters." arXiv preprint arXiv:2010.13021 (2020).

- Martín-Martín, Roberto, et al. "Variable impedance control in end-effector space: An action space for reinforcement learning in contact-rich tasks." arXiv preprint arXiv:1906.08880 (2019).

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

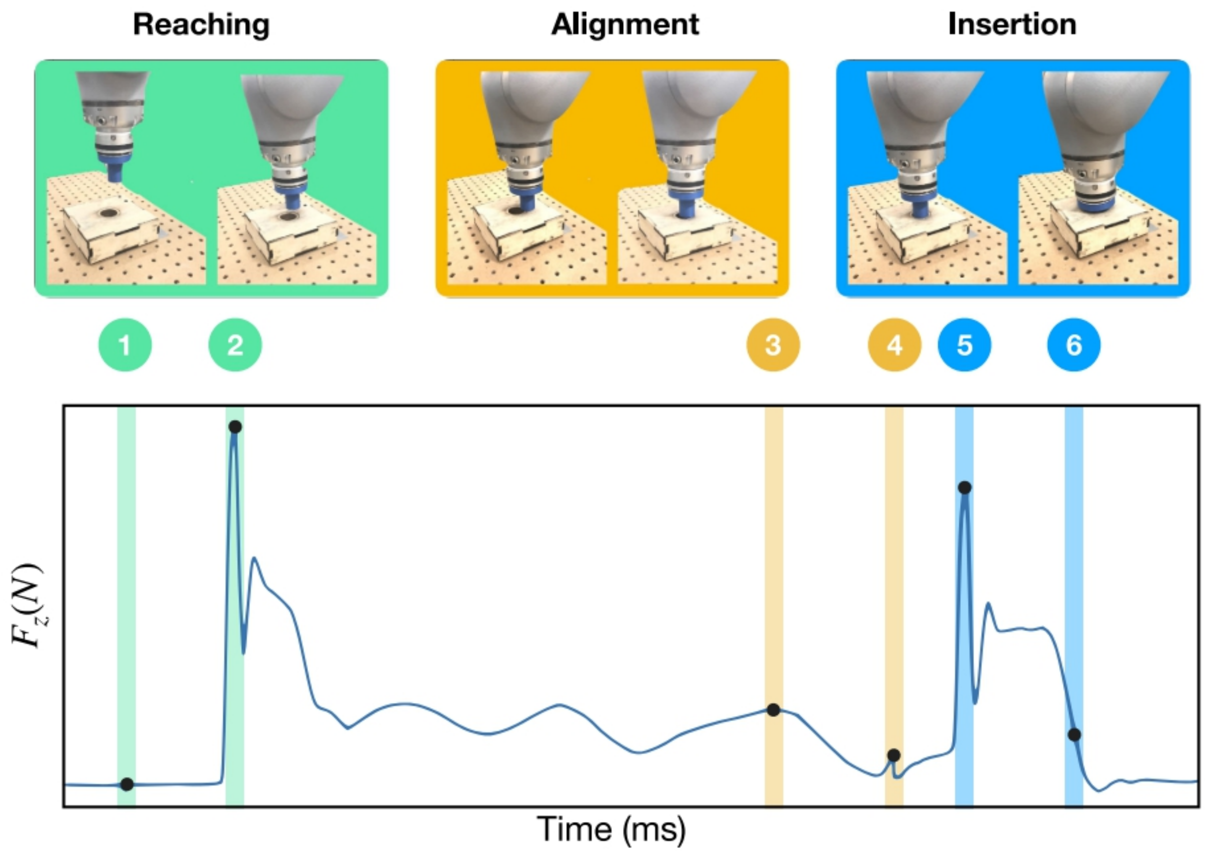


Making Sense of Vision and Touch:
Learning Multimodal Representations for Contact-Rich Tasks

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

## Method overview

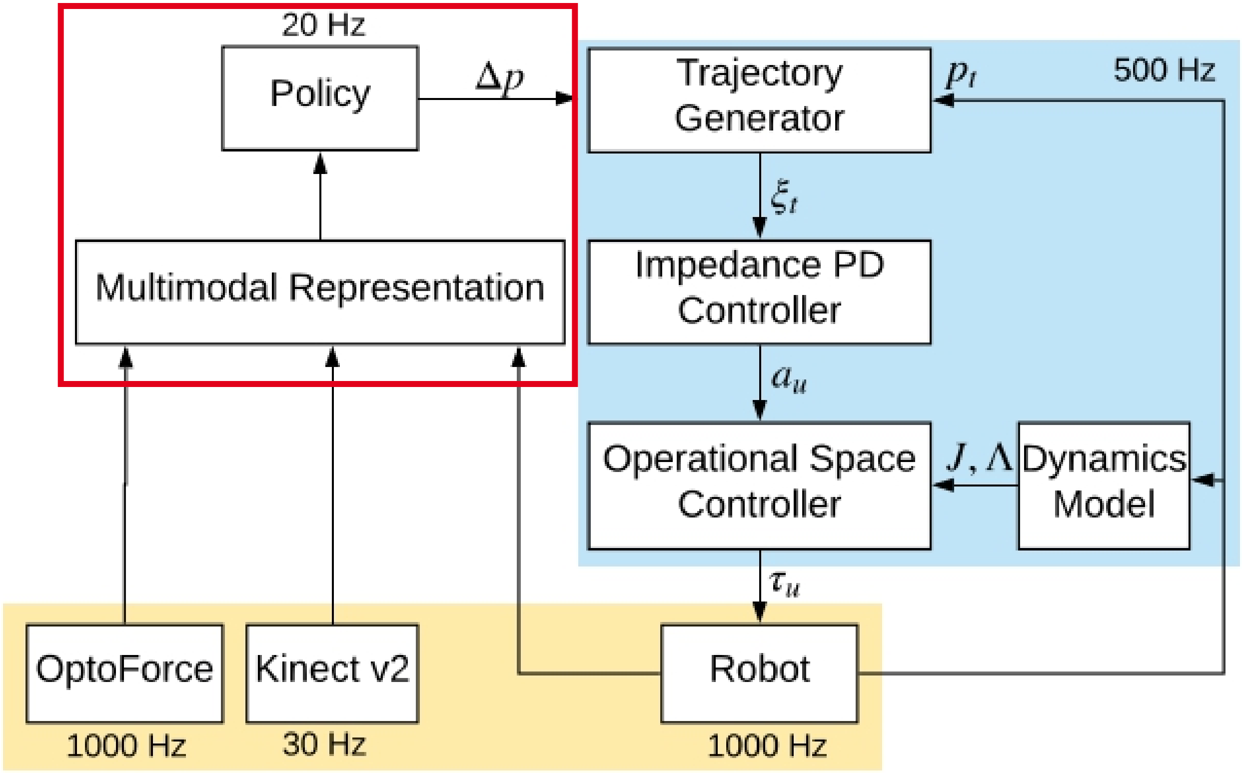



Force sensor readings in the z-axis (height) and visual observations are shown with corresponding stages of a peg insertion task

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

## Method overview

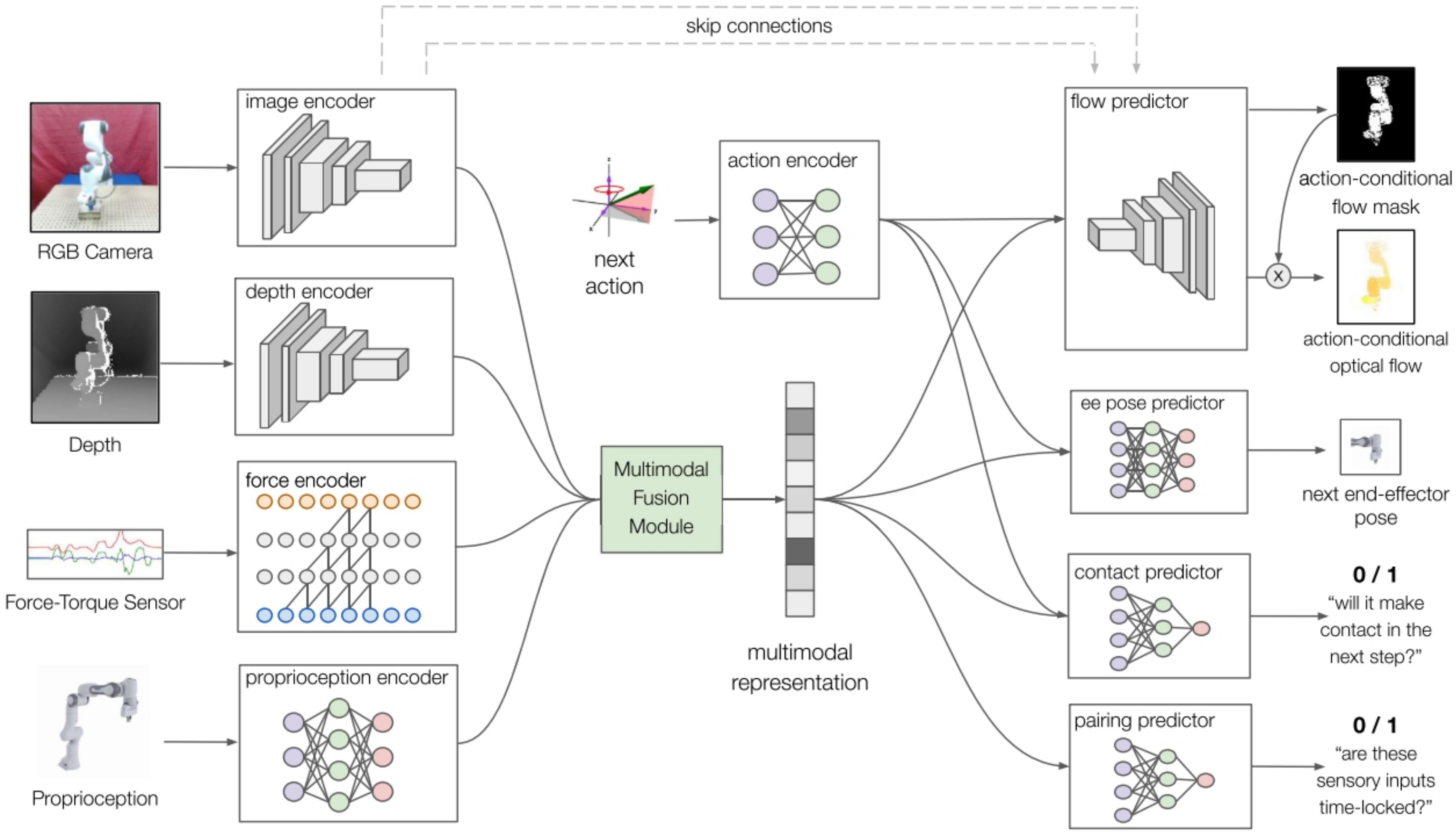



Controller structure

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

## Method overview




Neural network architecture for multimodal representation learning with self-supervision

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

## Method overview
## Encoder's architecture settings:

•For visual feedback, use a six-layer convolutional neural network (CNN) to encode $128 \times 128 \times 3$ RGB images.
•For depth feedback, use an eighteen-layer CNN with $3 \times 3$ convolutional filters of increasing depths to encode $128 \times 128 \times 1$ depth images. A single fully connected layer to the end of both the depth and RGB encoders to transform the final activation maps into a $2 \times$ d-dimensional variational parameter vector.
•For haptic feedback, we take the series and perform five-layer causal convolutions with last 32 readings from the six-axis F/T sensor as a $32 \times 6$ time variational parameter vector.
•For proprioception, we encode the stride 2 to transform the force readings into a $2 \times$ d-dimensional current position, roll, linear velocity, and roll angular velocity of the end-effector with a four-layer multilayer perceptron (MLP) to produce a $2 \times$ d-dimensional variational parameter vector.

For "Self-Supervised Predictions and Decoder Architecture Representations", please check paper

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

## Method overview

RL policy: trust-region policy optimization (TRPO)

$$r(\mathbf{s}) = \begin{cases} c_r(1-(\tanh \lambda \|\mathbf{s}\|_2)(1-s_\psi) & & \text{(r)} \\ 1 + c_a(1 - \frac{\|\mathbf{s}\|_2}{\|\boldsymbol{\varepsilon}_1\|_2})(1 - \frac{s_\psi}{\varepsilon_\psi}) & \text{if } \mathbf{s} \leq \boldsymbol{\varepsilon}_1 \ \&\ s_\psi \leq \boldsymbol{\varepsilon}_\psi & \text{(a)} \\ 2 + c_i(h_d - |s_z|) & \text{if } s_z < 0 & \text{(i)} \\ 5 & \text{if } h_d - |s_z| \leq \varepsilon_2 & \text{(c)} \end{cases}$$

Reward design for: reaching (r), aligning (a), inserting (i), and completed (c).

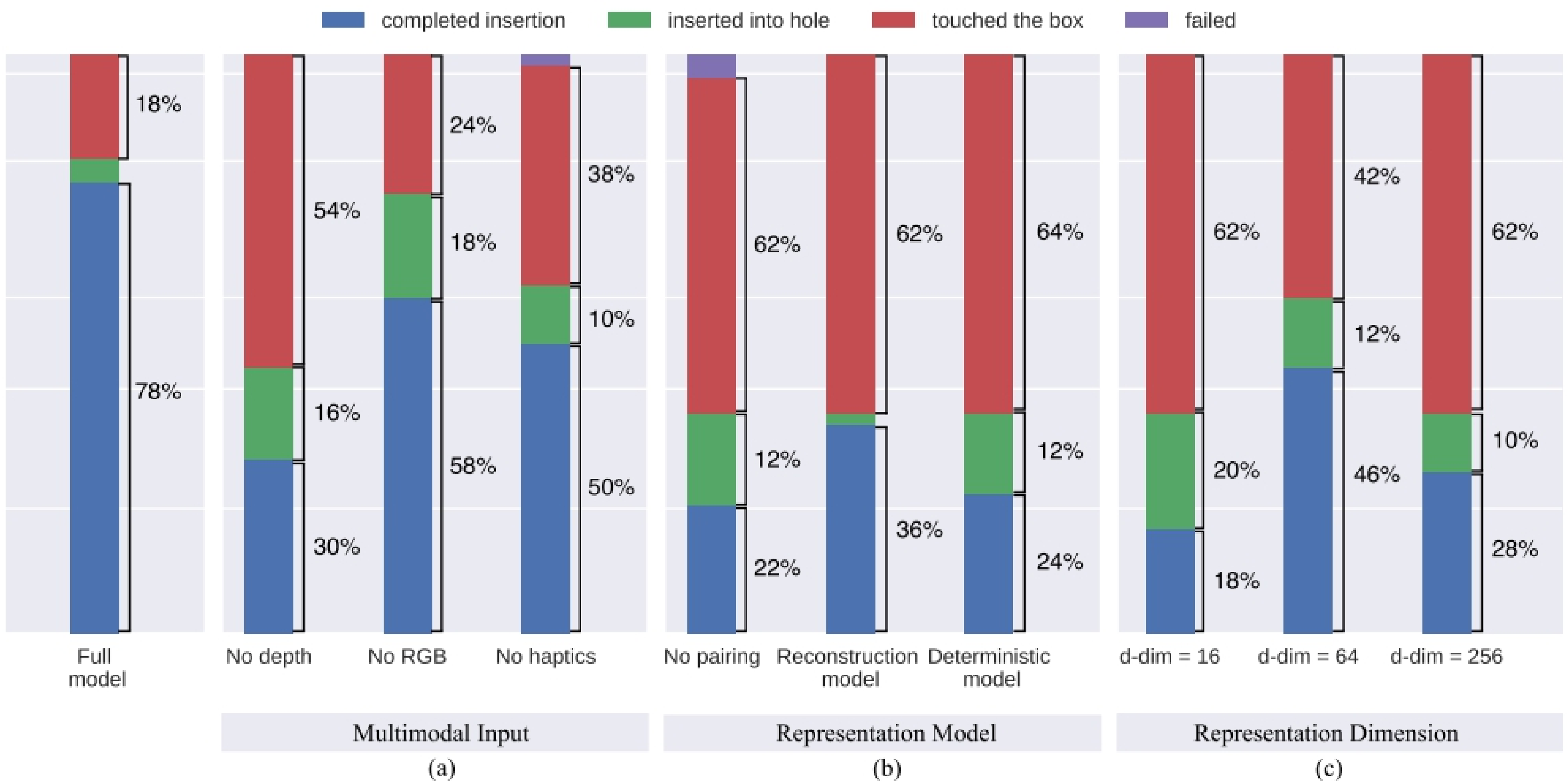
Cartesian end-effector position displacements: $\Delta \mathbf{x}$

Cartesian roll angle displacements: $\Delta \alpha$

# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

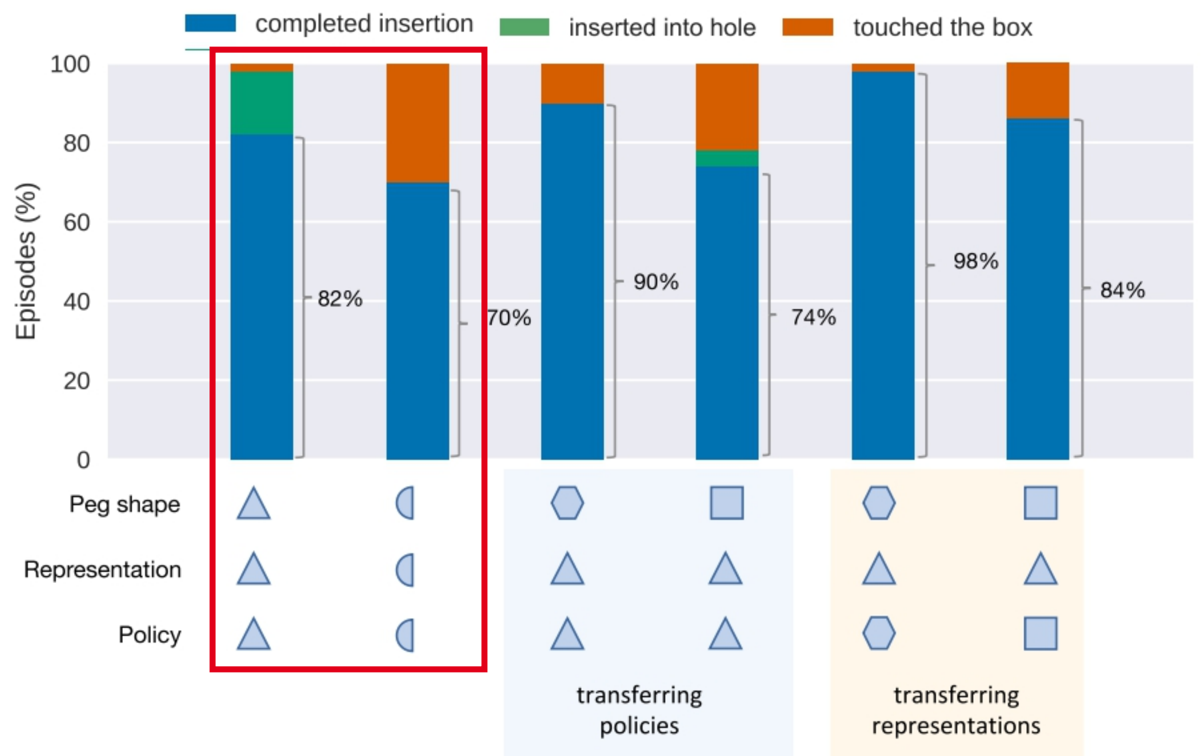
## Experiment result



# Related Work

Stanford University, Stanford Artificial Intelligence Lab (SAIL)

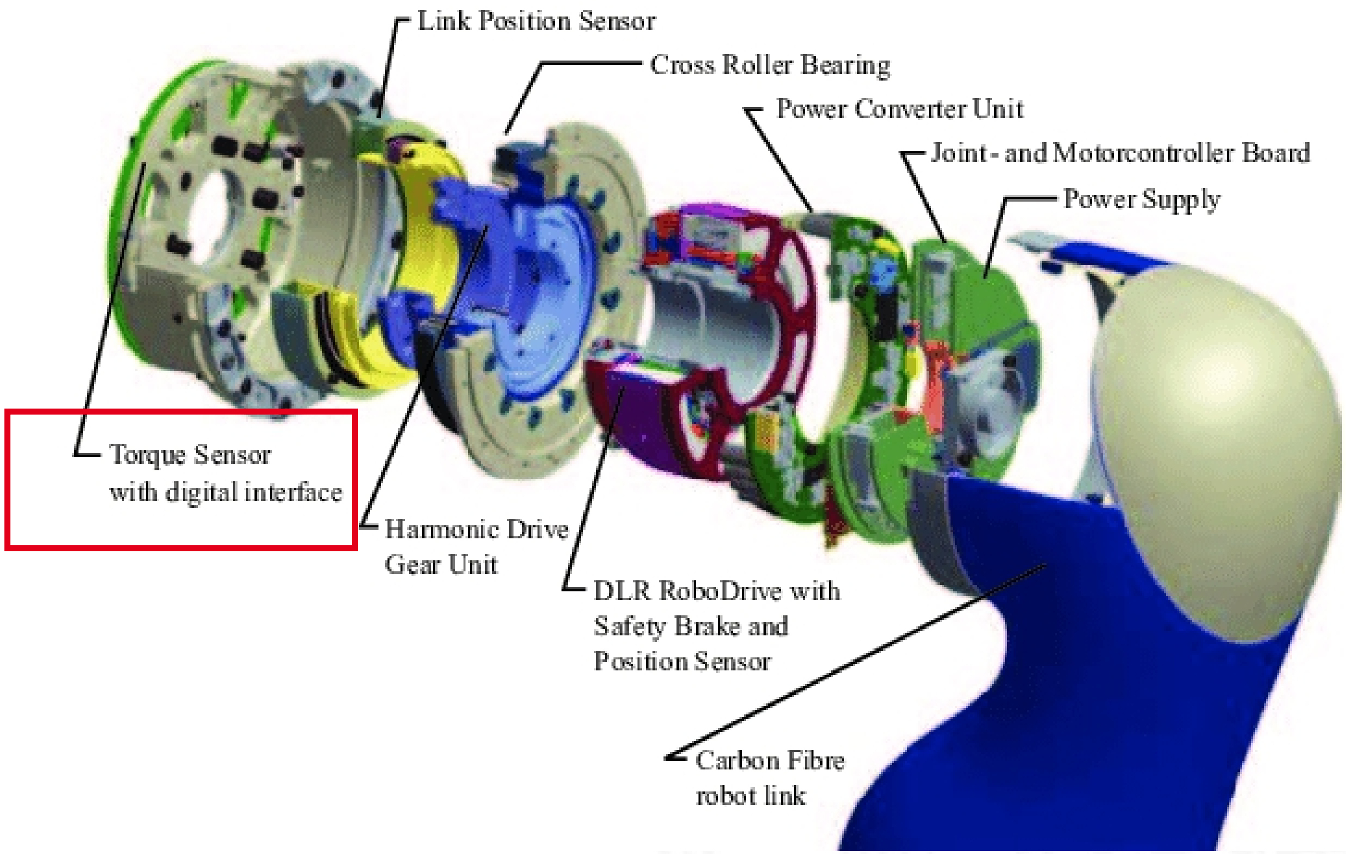
## Experiment result




Peg Insertion and Transfer Learning Results on Different Geometry

# Table of Contents

## Impedance controller, admittance controller

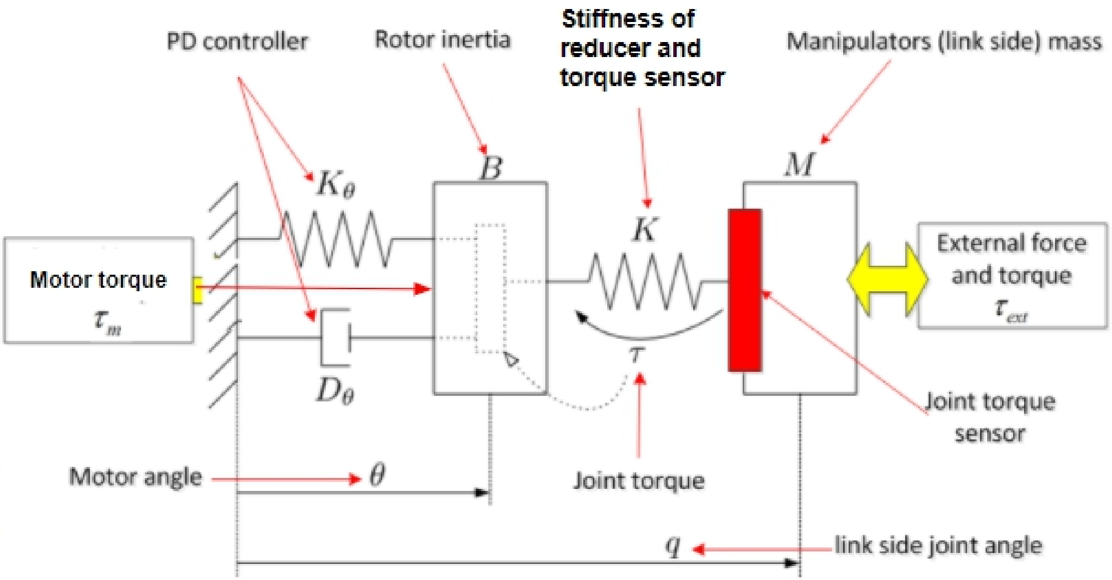



DLR LWR joint structure

## Impedance controller, admittance controller

## DLR LWR joint modeling:

$$\begin{aligned} M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) &= \tau + \tau_{ext} \\ B\ddot{\theta} + \tau &= \tau_m \\ \tau &= K(\theta - q) \end{aligned}$$

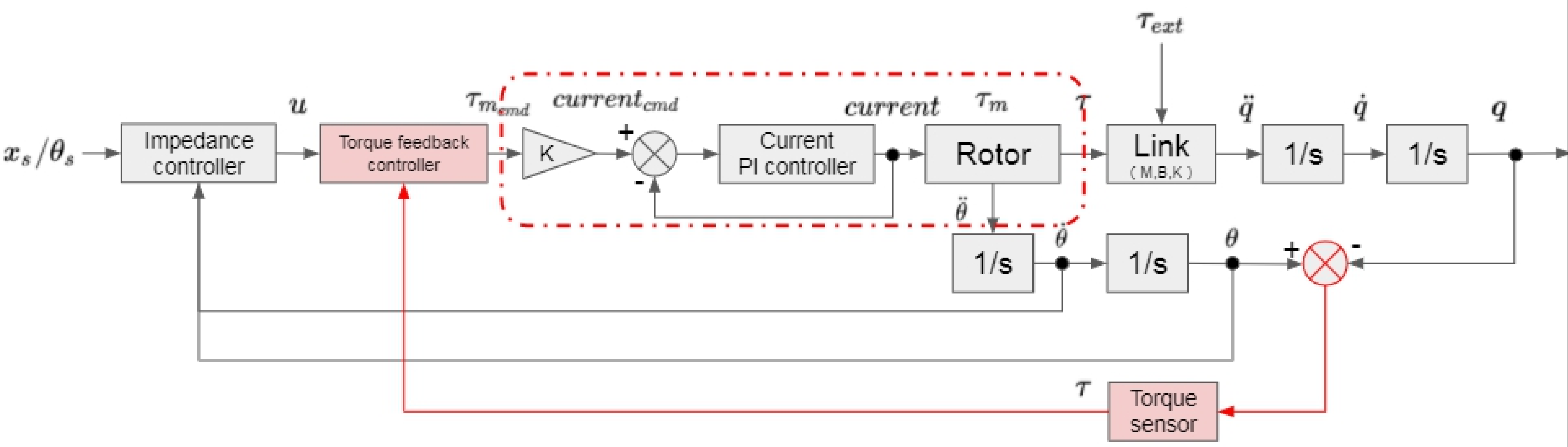



## Force controller:

$$\tau_m = BB_\theta^{-1}u + (I - BB_\theta^{-1})\tau = \boxed{BB_\theta^{-1}(u-\tau) + I\tau}$$

$$u = J(\theta)^T F_{des}$$

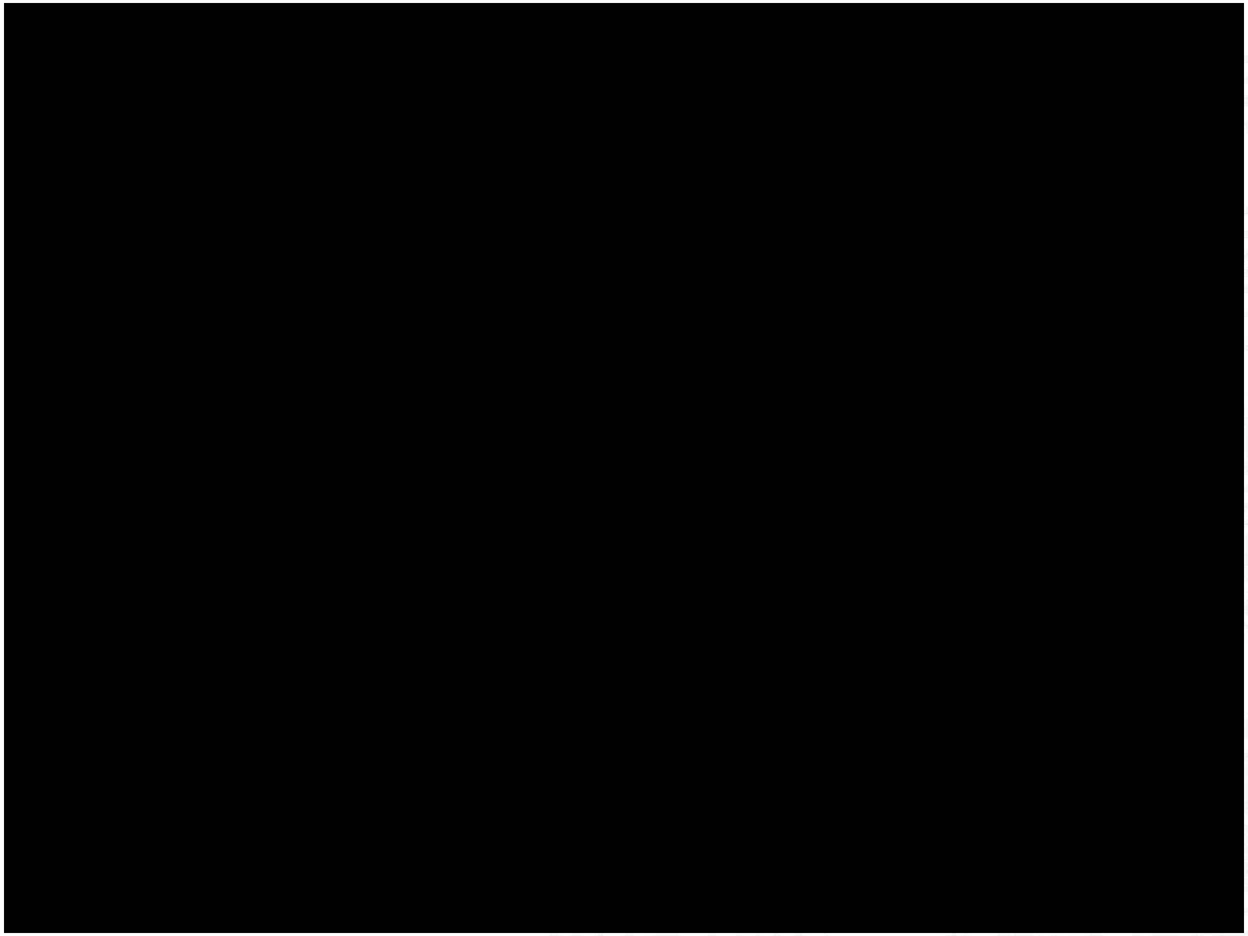
## Impedance controller, admittance controller




| | |
|---|---|
| Joint space impedance controller: | $u = -K_\theta(\theta - \theta_s) - D_\theta \dot{\theta}$ |
| Cartesian space impedance controller: | $u = -J(\theta)^T (K_x \tilde{x} + D_x \dot{x}(\theta)) + \bar{g}(\theta)$ <br> $\tilde{x}(\theta) = f(\theta) - x_s$ |

# Studies

## Impedance controller, admittance controller



Impedance controller, admittance controller

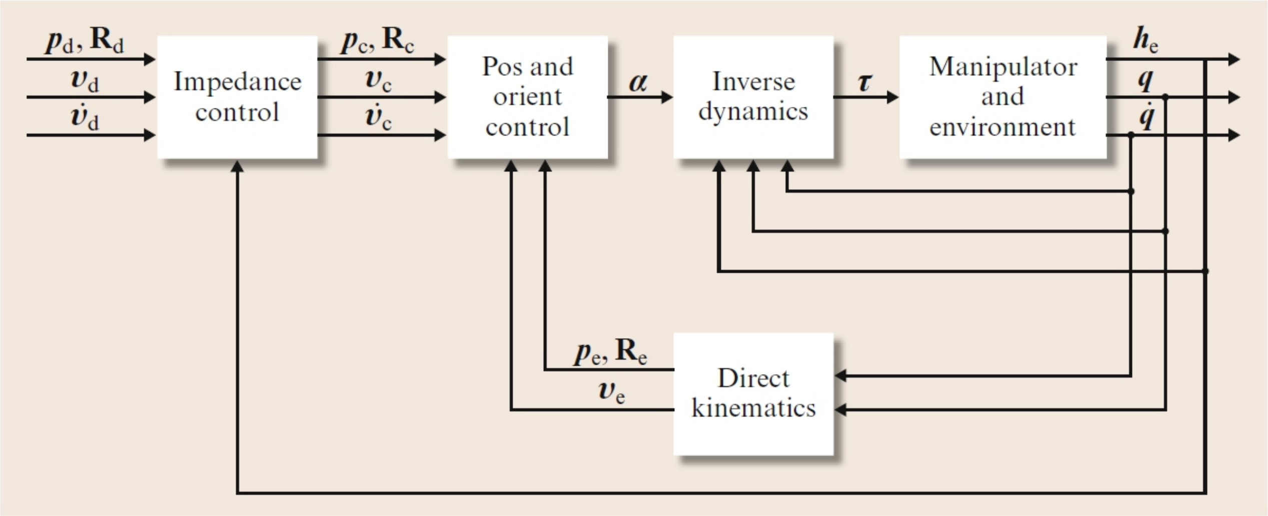



**Fig. 9.2** Impedance control with inner motion control loop (admittance control)

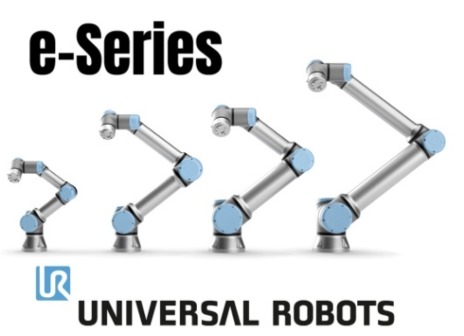
Force controller:

$$M\ddot{x}_e + B\dot{x}_e = F_e$$

$$\ddot{x}_e = M^{-1}(F_e - B\dot{x}_e)$$

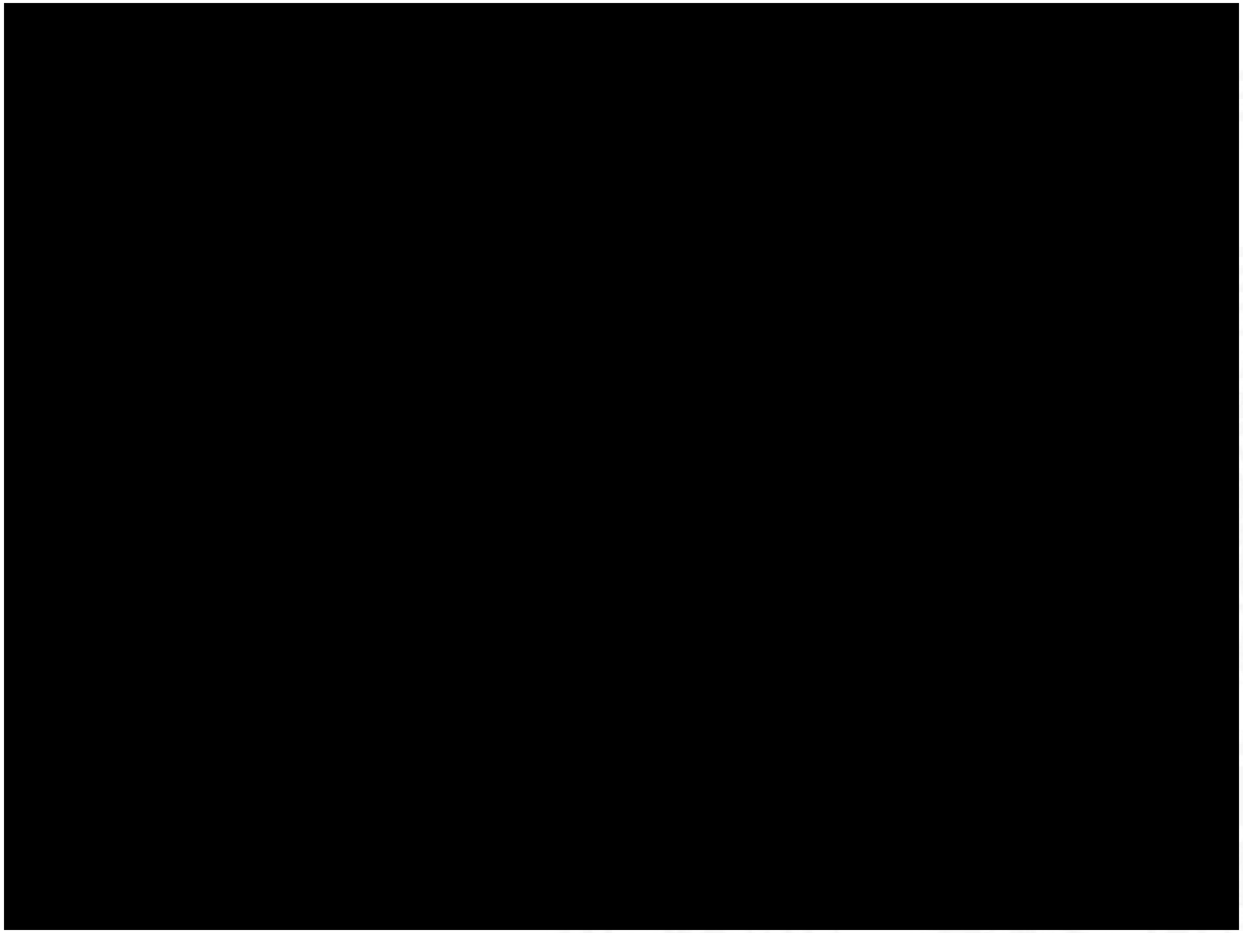
$$\dot{x}_e{}^{t+1} = \dot{x}_e{}^{t} + \ddot{x}_e{}^{t+1}T$$

$$x_e{}^{t+1} = x_e{}^{t} + \dot{x}_e{}^{t+1}T$$



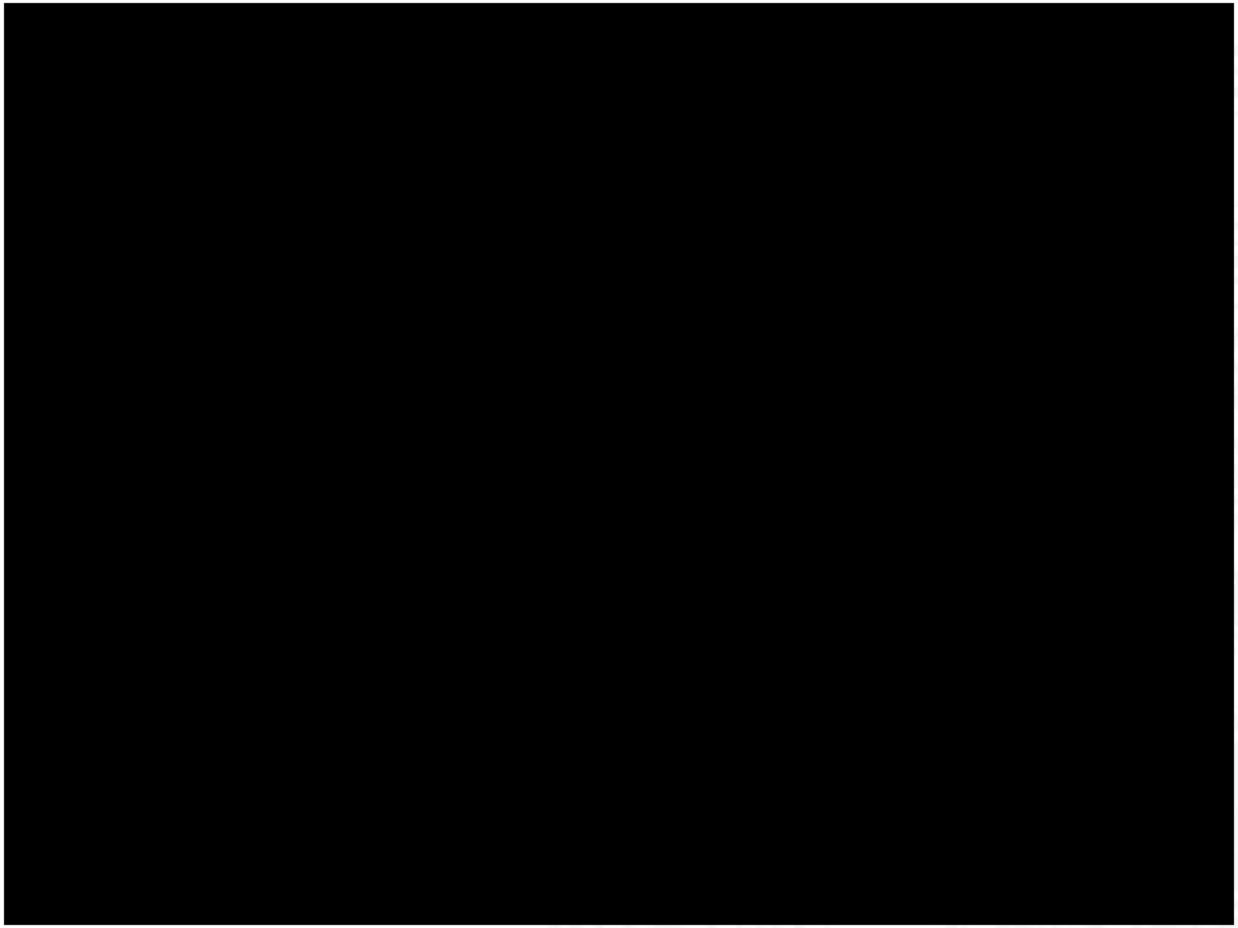
# Studies

Impedance controller, admittance controller



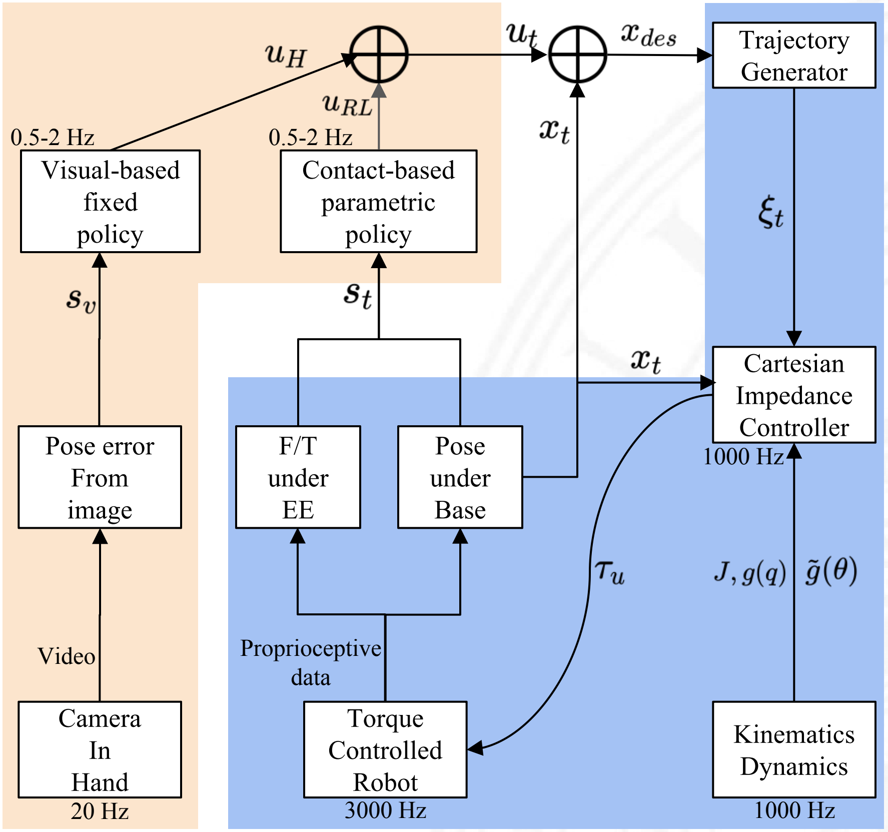
# Studies

1. Proactive Action Visual Residual Reinforcement Learning for Contact-Rich Tasks Using a Torque-Controlled Robot (ICRA2021)

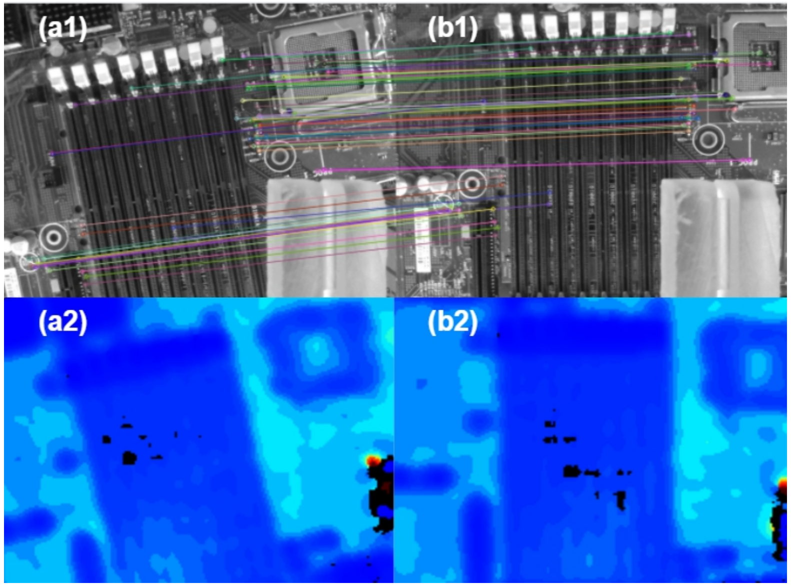


Contribution 1: Representation of policies and controller scheme

$$u_t = (1-\alpha)\pi_H(s_v) + \alpha * \pi_\theta(s_t)$$



Visual-based fixed policy




$$s_v = ({}^{c*}t_c, \theta u)$$

$$\pi_H(s_v) = -k_p \cdot s_v$$

Contact-based parametric policy

We use a simple Q-learning algorithm:

$$Q^\pi(s_t, u_t) = \mathbb{E}_{r_t, s_{t+1} \sim E}[r_t + \gamma \mathbb{E}_{u_{t+1} \sim \pi}[Q^\pi(s_{t+1}, u_{t+1})]]$$

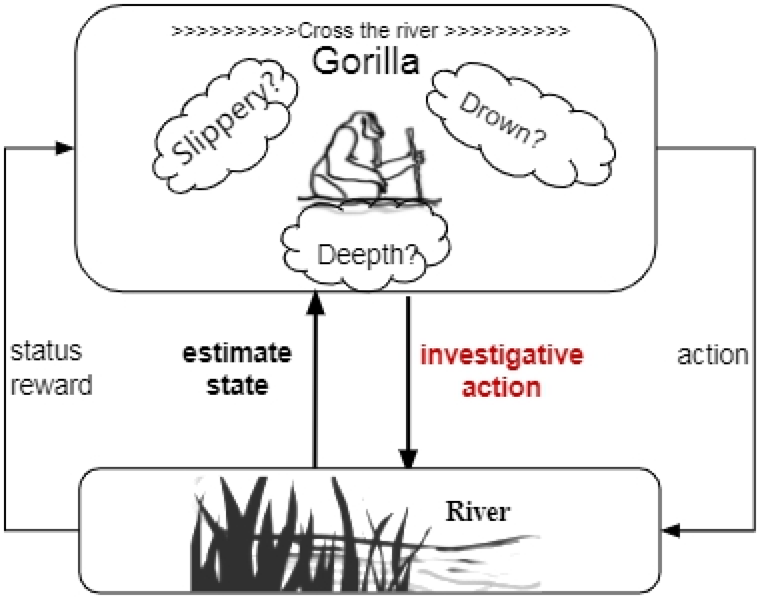
$$a = \lambda[P^d_{\sigma x}, P^d_{\sigma y}, P^d_{\sigma z}, R^d_{\sigma x}, R^d_{\sigma y}, R^d_{\sigma z}]$$

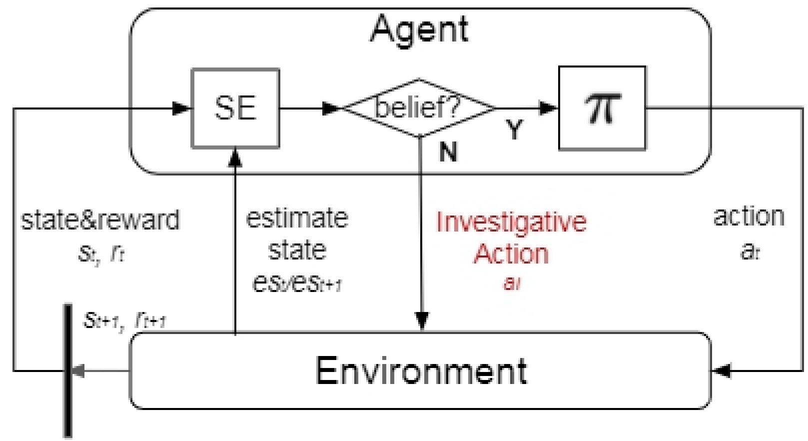
$$s = [F_x, F_y, F_z, M_x, M_y, M_z]$$

$$r = \begin{cases} 1, & \text{(success)} \\ -2, & \text{(failed).} \\ 1 - 150\|s_{xy}\|_2 - s/s_{max}, & \text{(otherwise).} \end{cases}$$

# Studies

Contribution 2: Proactive  Action



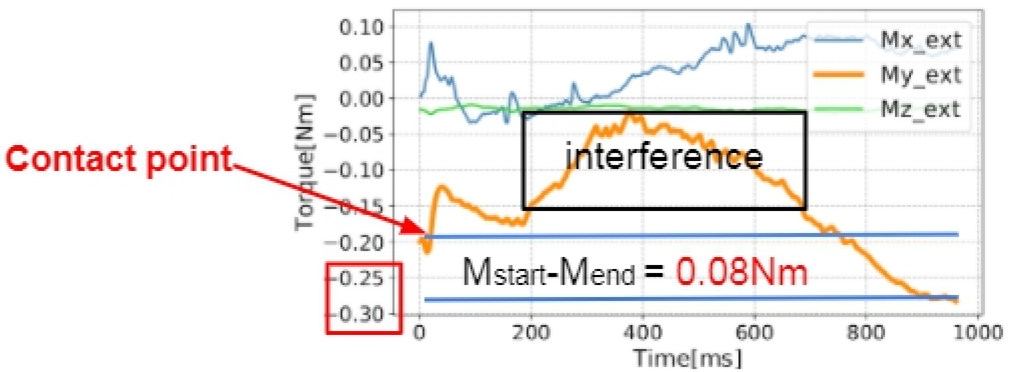

**The gorilla uses a stick to investigate water depth state**

**The agent uses a robot to investigate contact state**

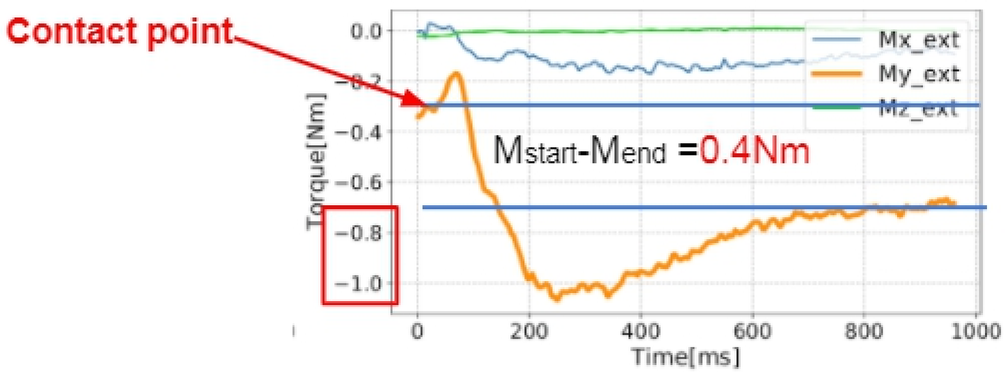
Note: RL action is **an impedance controller**, while investigative action use **a force controller**

# Studies

## Contribution 2: Proactive  Action Experiment



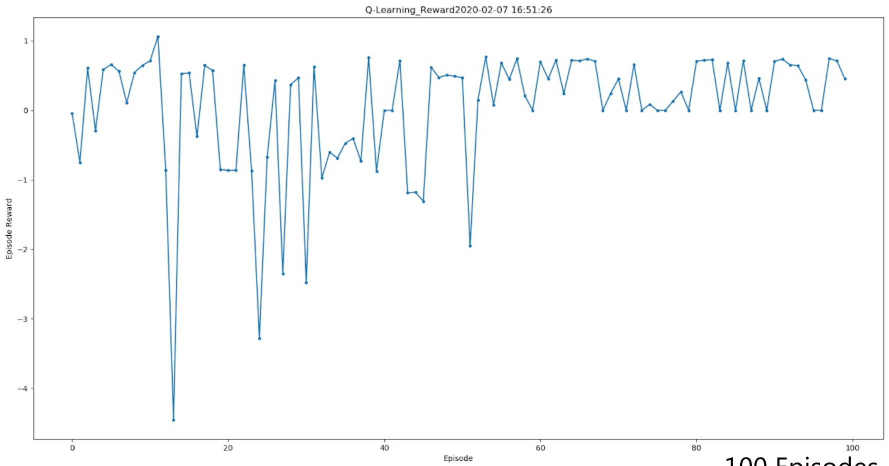

**Without**
**an investigative action,**
**the moment y is unclear**
**with lots of interference**




**With**
**an investigative action,**
**the moment y is detectable**
**5 times than before**

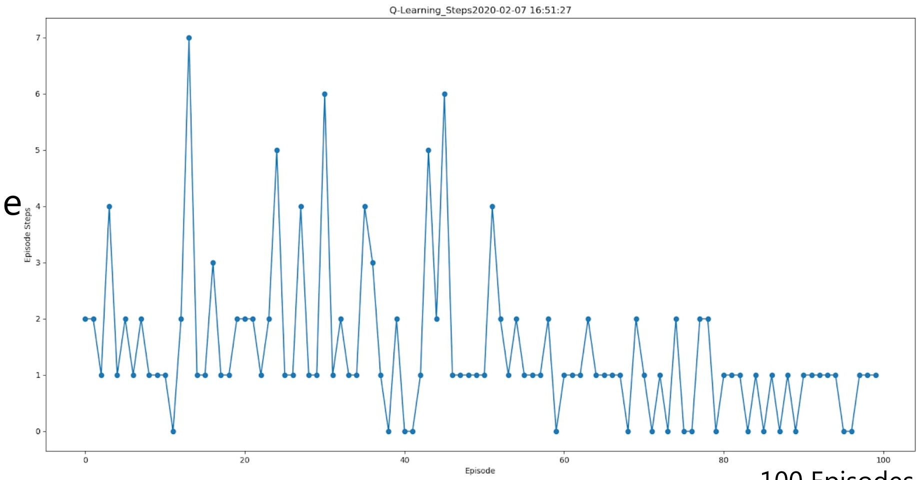
## Training Experiment Example

Episode reward




100 Episodes

Episode steps




100 Episodes

# Studies

## Experiment Results

TABLE I
ABLATION STUDY OF POLICY EVALUATION STATISTICS

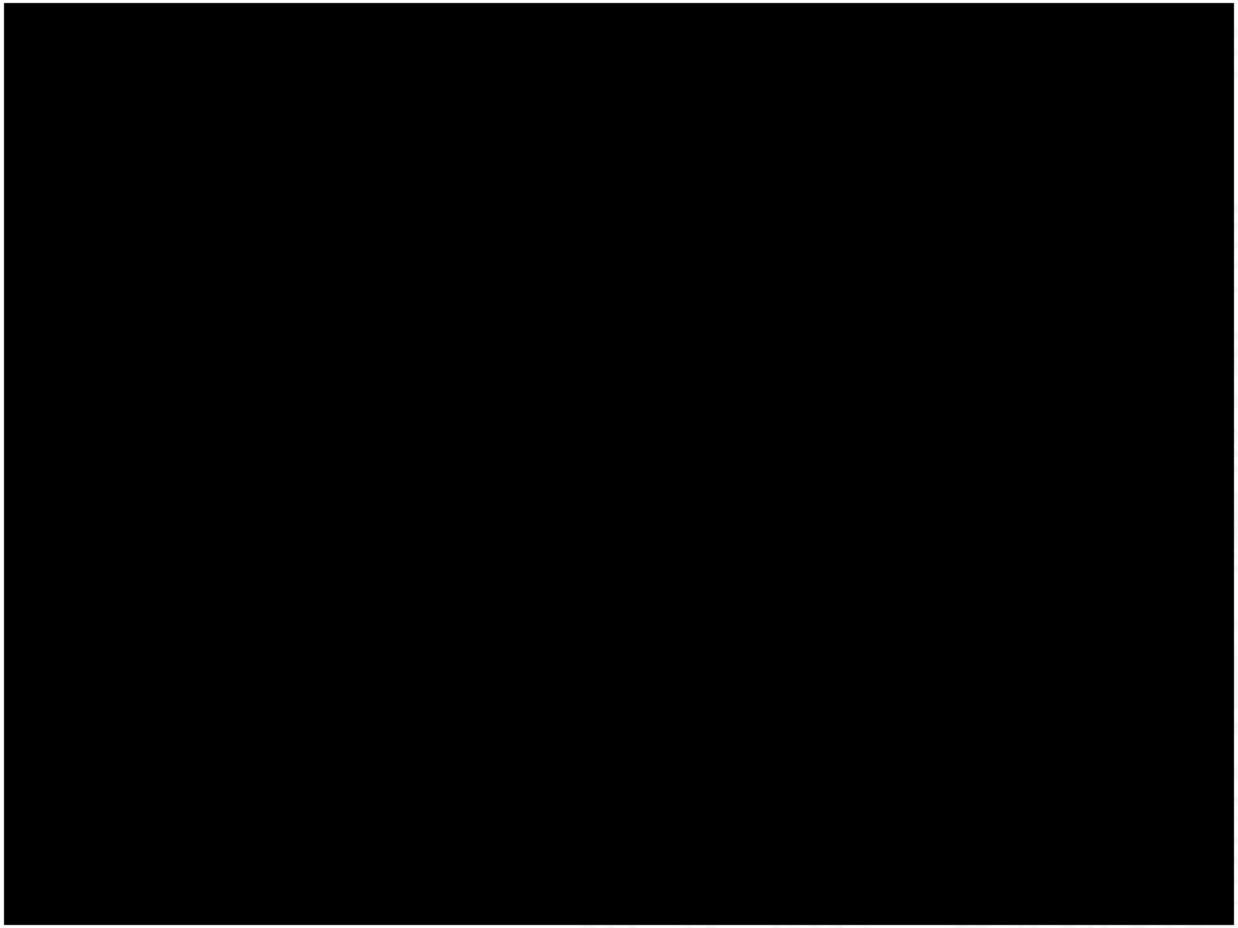
| Baselines | Result(success/total) | Total Time Cost |
|---|---|---|
| No vision | 92/200 | 1.09 h |
| No RL policy | 112/200 | 0.65 h |
| Random RL policy | 77/200 | 2.59 h |
| No investigative action | 66/200 | 0.85 h |
| **Our method** | **179/200** | 1.18 h |

TABLE II
COMPARISON OF SUCCESS RATES FOR DIFFERENT BASELINES

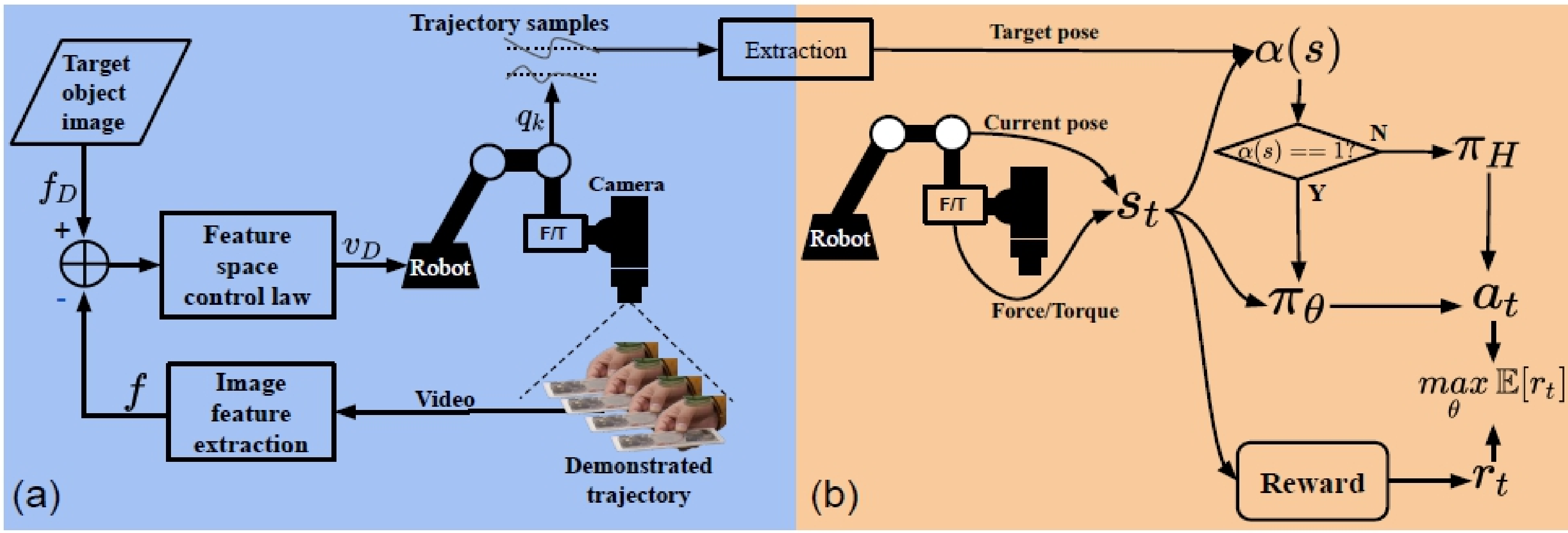
| Baselines | Fix motherboard | Move motherboard |
|---|---|---|
| Baseline 1 | 97/100 | 0/20 |
| Baseline 2 | 100/100 | 0/20 |
| Baseline 3 | 98/100 | 81/100 |
| Baseline 4 | 100/100 | 88/100 |
| **Our method** | **100/100** | **100/100** |

# Studies

2. Combining Learning from Demonstration with Learning by Exploration to Facilitate Contact-Rich Tasks (IROS2021 submission)

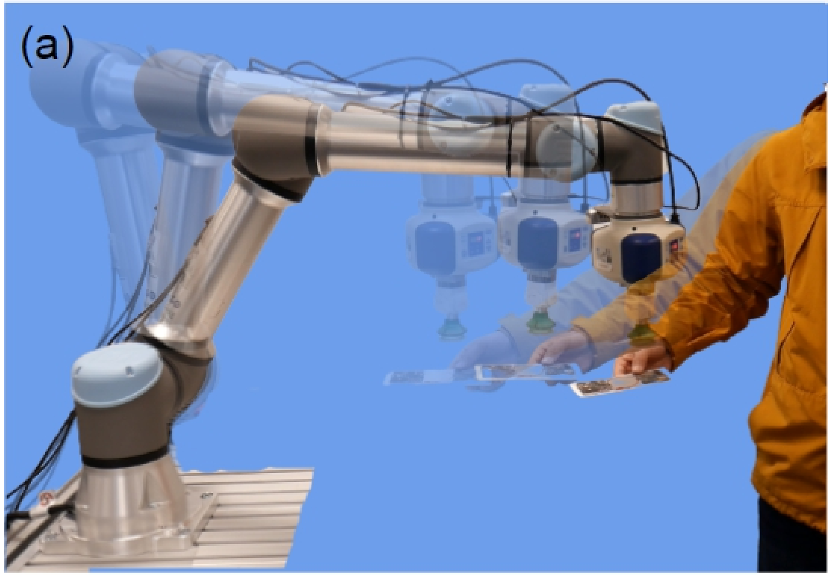


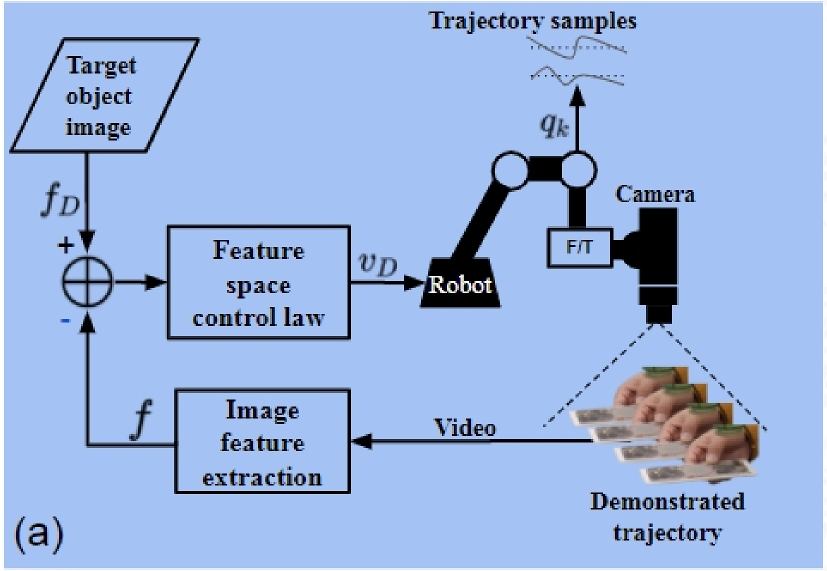
## Representation of policies and controller scheme



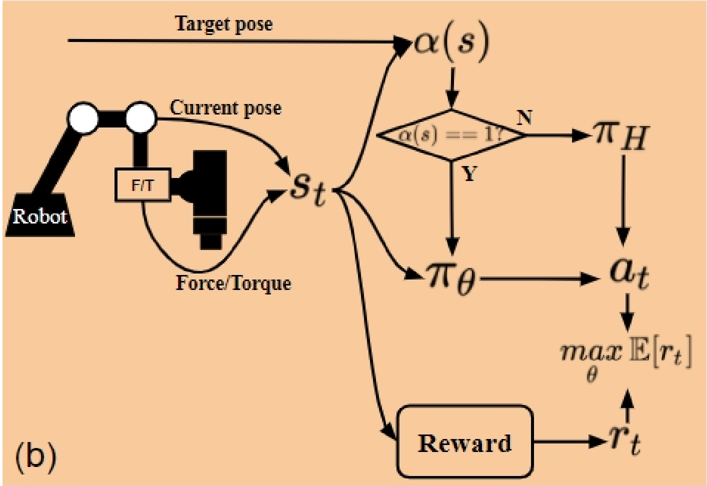
# Studies

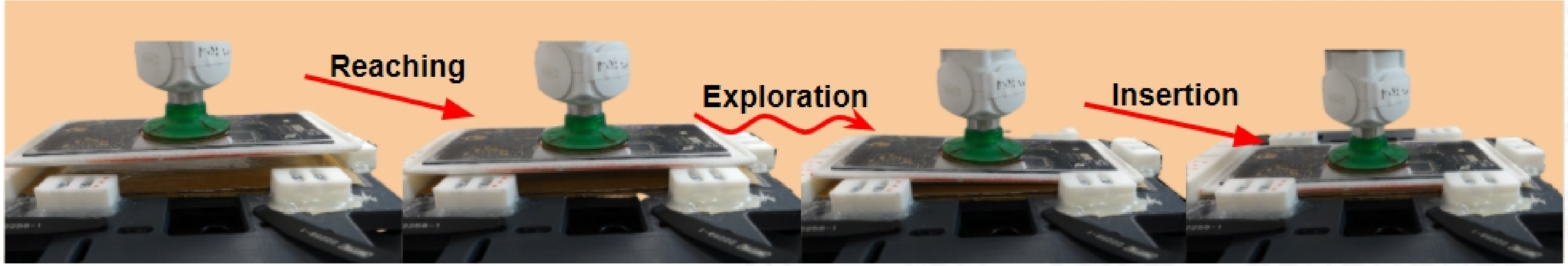
Contribution 1: Learning from demonstration based on visual servoing



Contribution 2: A region-limited residual reinforcement learning(RRRL) policy based on force-torque information



## Contribution 2: A region-limited residual reinforcement learning(RRRL) policy based on force-torque information

**Algorithm 1 RRRL**

**Require:** Model based policy $\pi_H$, learning frequency $C$ target action-value update frequency $C_2$.

Initialize replay memory $\mathcal{H}$ to capacity $N$
Initialize action-value function $Q$ with random weigh $\theta$
Initialize target action-value function $Q_{target}$ wi weights $\theta^- = \theta$
**for** episode = 1 to $M$ **do**
    Sample state $s_0$
    **while** NOT EpisodeEnd **do**
        Calculate $\alpha(s)$ with Equation (8)
        Choose action $a_H$ from $\pi_H(s_t)$
        With probability $\epsilon$ choose a random action $a_{R}$
        Otherwise select $a_{RL} \sim \pi_\theta(s_t)$
        Obtain action $a_t = (1 - \alpha) * a_H + \alpha * a_{RL}$
        Execute $a_t$, observe reward $r_t$ and state $s_{t+1}$
        Store transition $(s_t, a_t, r_t, s_{t+1})$ in $\mathcal{H}$ with pri rity $p_t = max_{i<t} p_i$
        **for** $j$ = 1 to $C_1$ **do**
            Sample minibatch of transitions with priori from $\mathcal{H}$
            Update transition priority
            Update $\theta$ with the method proposed in [40
        **end for**
        Every $C_2$ steps reset $Q_{target} = Q$
    **end while**
**end for**

$$\pi(a|s) = (1 - \alpha(s)) \cdot \pi_H(a|s) + \alpha(s) \cdot \pi_\theta(a|s). \quad (1)$$

Parametric Policy:
Double DQN with proportional prioritization

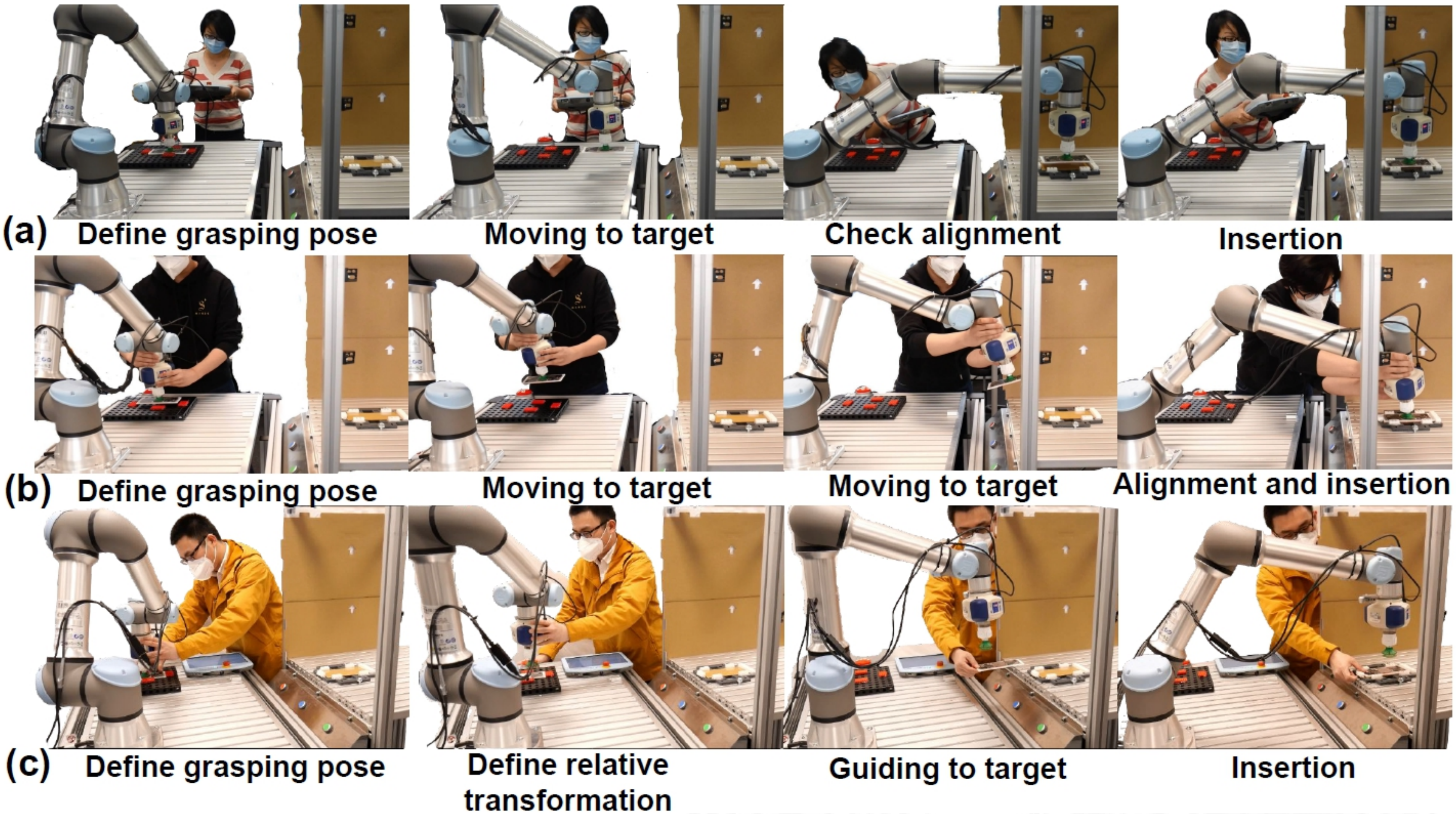
# Studies

Contribution 2: A region-limited residual reinforcement learning(RRRL) policy based on force-torque information

Training Experiment Example

# Studies

Teaching Experiments



# Studies

## Experiments Results

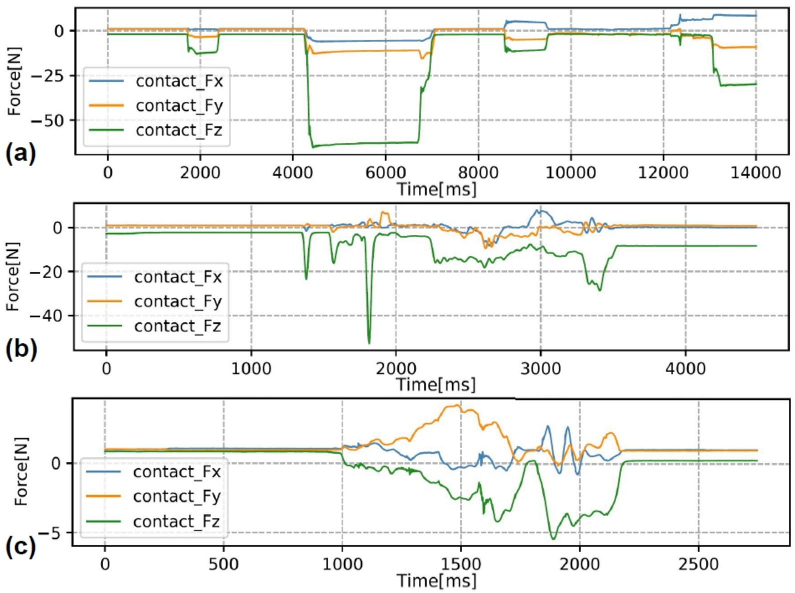



TABLE II

EVALUATION IN THE TEACHING PHASE

| **Teaching phase** | Time cost | Maximum contact force |
|---|---|---|
| Teach-pendant | 60–120 s | 15–50 N |
| Hand-guiding | **15–42 s** | 30–60 N |
| Our method | **23–30 s** | **3–10 N** |

# Studies

## Experiments Results

TABLE III

EVALUATION IN THE EXECUTION PHASE

| Execution phase | Success rate | | Maximum contact force |
|---|---|---|---|
| | Perfect | Uncertainty | |
| Only teach-pendant | 55/100 | 17/100 | 15 N |
| Only hand-guiding | 33/100 | 5/100 | 15 N |
| Teach-pendant + spiral searching | 69/100 | 47/100 | 35 N |
| Hand-guiding + spiral searching | 51/100 | 33/100 | 35 N |
| Our method | **95/100** | **91/100** | **15 N** |

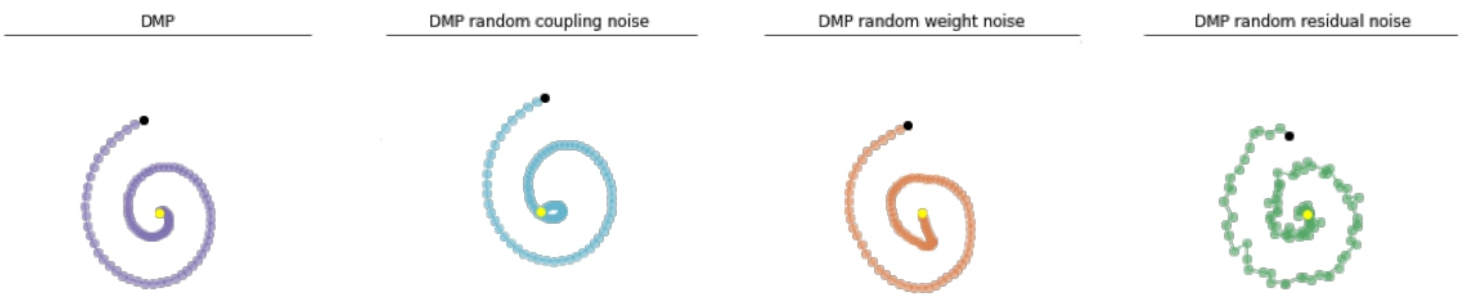
# Table of Contents

1. Introduction
2. Related Work
3. Studies
    - Impedance controller, admittance controller
    - RL based assembly using impedance controller
    - DRL based assembly using admittance controller
4. Future Work

# Future Work

Dynamic Movement Primitives for More Complex Contact-rich Tasks

- Complex trajectory assembly
- Massage Robot

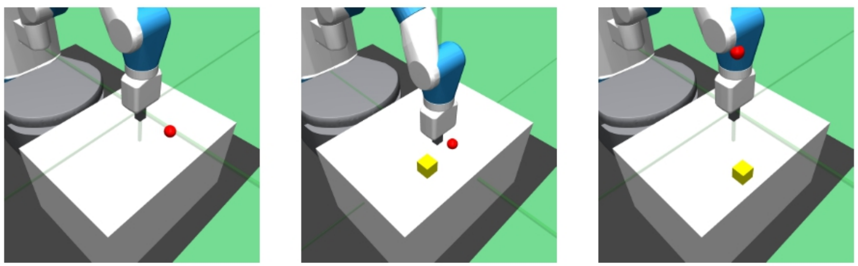



$$\dot{y} = \frac{1}{\tau^2}(\alpha_v(\beta_v(g-x)-\tau y)+zf_{\omega+\boxed{\eta}}+C_t(\boxed{\eta}))+\boxed{\dot{\eta}}$$

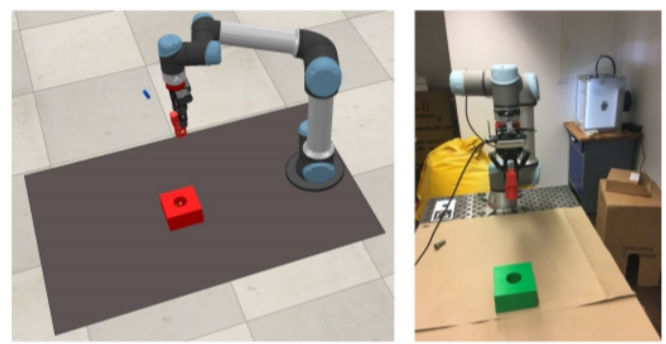
Davchev, T., Luck, K. S., Burke, M., Meier, F., Schaal, S., & Ramamoorthy, S. (2020). Residual Learning from Demonstration: Adapting Dynamic Movement Primitives for Contact-rich Insertion Tasks. *arXiv e-prints*, arXiv-2008.

# Future Work

Sim2Real RL for insertion tasks



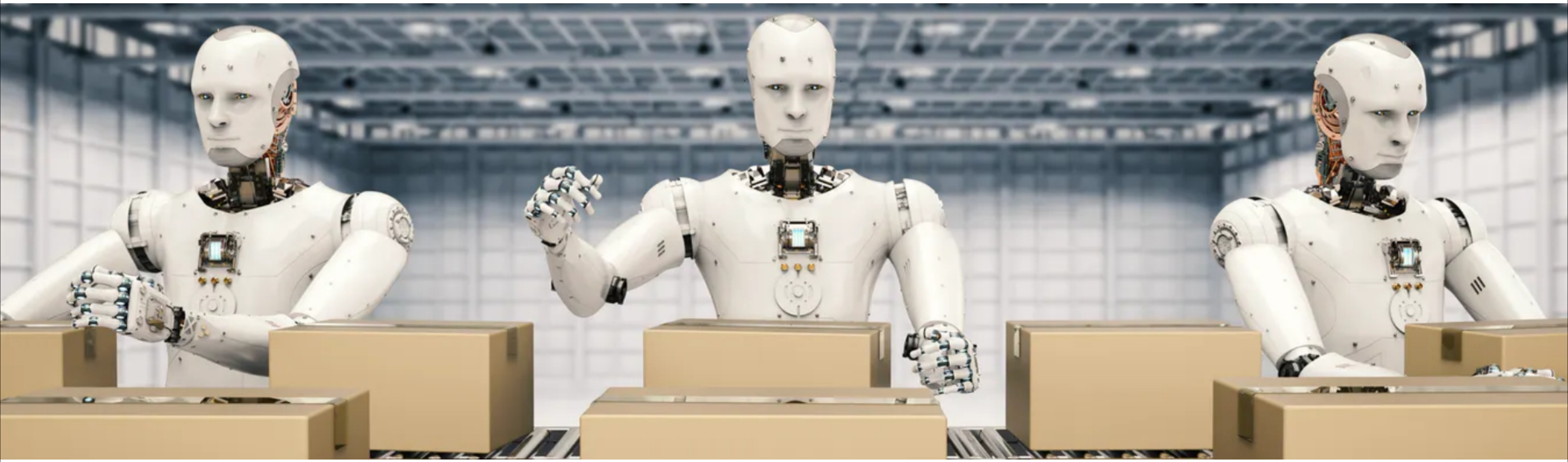
(a) Reach (b) Push (c) Pick-and-place



•Zhan, A., Zhao, P., Pinto, L., Abbeel, P., & Laskin, M. (2020). A Framework for Efficient Robotic Manipulation. *arXiv preprint arXiv:2012.07975*.
•Bogunowicz, Damian, Aleksandr Rybnikov, Komal Vendidandi, and Fedor Chervinskii. "Sim2Real for Peg-Hole Insertion with Eye-in-Hand Camera." *arXiv preprint arXiv:2005.14401* (2020).

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MIN Faculty
Department of Informatics



# Thank you for your attention!