THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

RAIDS

Opening Minds • Shaping the Future
啟迪思維 • 成就未來

A Meta-Reinforcement Learning-Based Adaptive Robot Control for Human-Robot Collaboration in Personalized Production

Hin Chi Kwok, Chengxi Li,

Yatming Pang, Pai Zheng

Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong S.A.R., China

Overview

Motivation

Methodology

Experiments

Results & Demonstration

Discussions & Future Work

Motivation

- **Collaborative Robot:**

Industry 4.0 strategy

- **Efficiency:**

Increasing demand for individualised products and shorter product cycles [1].

- **Human-robot Collaboration (HRC):**

Solution to deal with personalized production.

[1] Yoram Koren. *The global manufacturing revolution: product-process-business integration and reconfigurable systems*. Vol. 80. John Wiley & Sons, 2010.

https://sharework-project.eu/project/sharework-ontology-for-human-robot-collaboration-cirp-conference-2020/

https://robot-hub.org/project/human-robot-collaboration-in-industry-4-0/

Motivation

- **Human Labor:**

Processes rely on human operators with low efficiency.

- **Flexibility:**

Requires robot agents with agile learning capabilities.

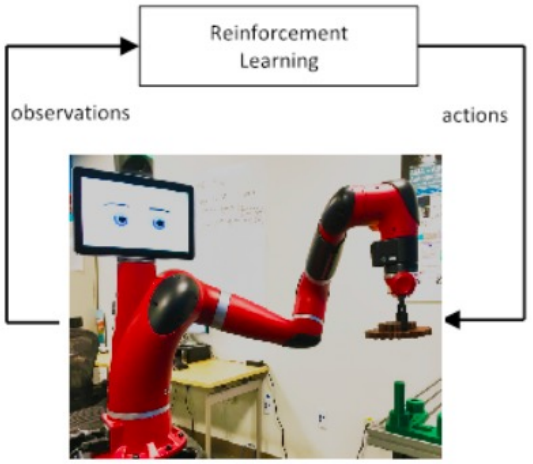
- **Productivity**:

Enhances producing power and advance the development of HRC.

Gervasi, Riccardo, Luca Mastrogiacomo, and Fiorenzo Franceschini. "A conceptual framework to evaluate human-robot collaboration." *The International Journal of Advanced Manufacturing Technology* 108.3 (2020): 841-865.

Motivation

- Deep Reinforcement Learning (DRL)
- Not yet been well-explored in human-in-the-loop manufacturing/production processes.



- Requires **a large amount of** annotated training data

-> impractical [2]

- Relies on **the exploration of the environment**

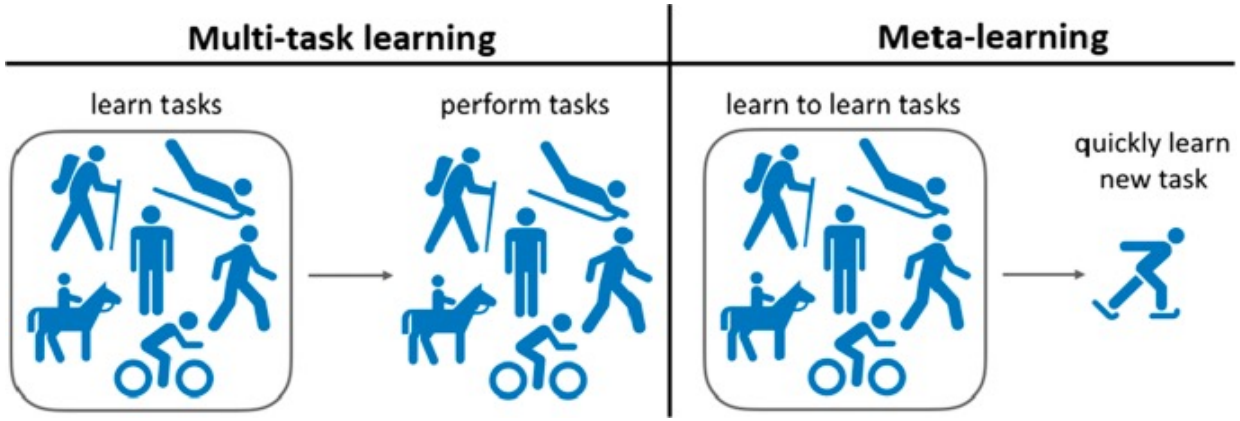
-> safety constraints [3]

[2] Afonso Castro, Filipe Silva, and Vitor Santos. "Trends of Human-Robot Collaboration in Industry Contexts: Handover, Learning, and Metrics". In: *Sensors 21 (June 2021)*, p. 4113. DOI: 10.3390/s21124113.
[3] Mohamed El-Shamouty et al. "Towards Safe Human- Robot Collaboration Using Deep Reinforcement Learning". In: *2020 IEEE International Conference on Robotics and Automation (ICRA).* IEEE. 2020, pp. 4899–4905.

Motivation

Meta-learning, a method learning to gain prior knowledge of new tasks.

- **Training phase:** quickly learn
- **Testing phase:** simply observes some trials in a new task to conform to a new task that is similar to the training task



Huisman, Mike, Jan N. Van Rijn, and Aske Plaat. "A survey of deep meta-learning." *Artificial Intelligence Review* 54.6 (2021): 4483-4541.

Methodology

- **Model diagnostic meta-learning (MAML)**, is used in combination with the **proximal policy optimization (PPO)** DRL algorithm [4,5].
- **Meta-reinforcement learning (meta-RL)** is an algorithm that trains an agent on a set of different tasks and extracts high-level knowledge that is shared across all training tasks.

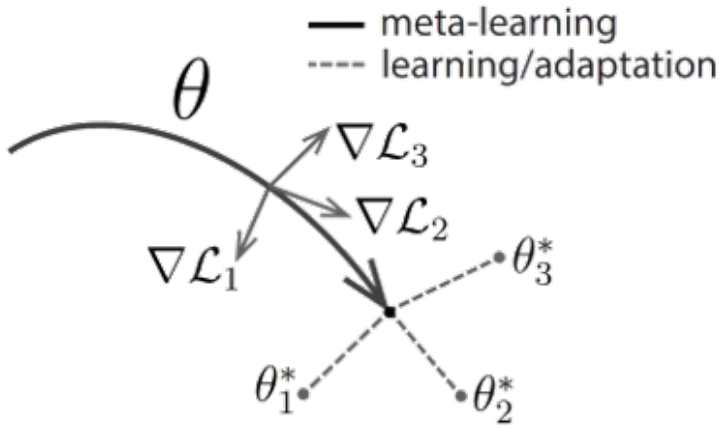


Diagram of our model-agnostic meta-learning algorithm (MAML), which optimizes for a representation ϑ that can quickly adapt to new tasks

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks". In: *Proceedings of the 34th Interna- tional Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 1126–1135.

[5] John Schulman et al. "Proximal policy optimiza- tion algorithms". In: *arXiv preprint arXiv:1707.06347* (2017).

Methodology

- Could the DRL-based solution improve the **learning capability** of the robotic system in Self-learning Robotic assisted Systems (SLRAS) and enable the robots to perform robustly to unseen HRC robot tasks?
- Compared to train-from-scratch DRL, does the proposed meta-RL solution perform better in terms of **training time and reward**?

https://www.epfl.ch/labs/lasa/human-robot-interaction/

Methodology

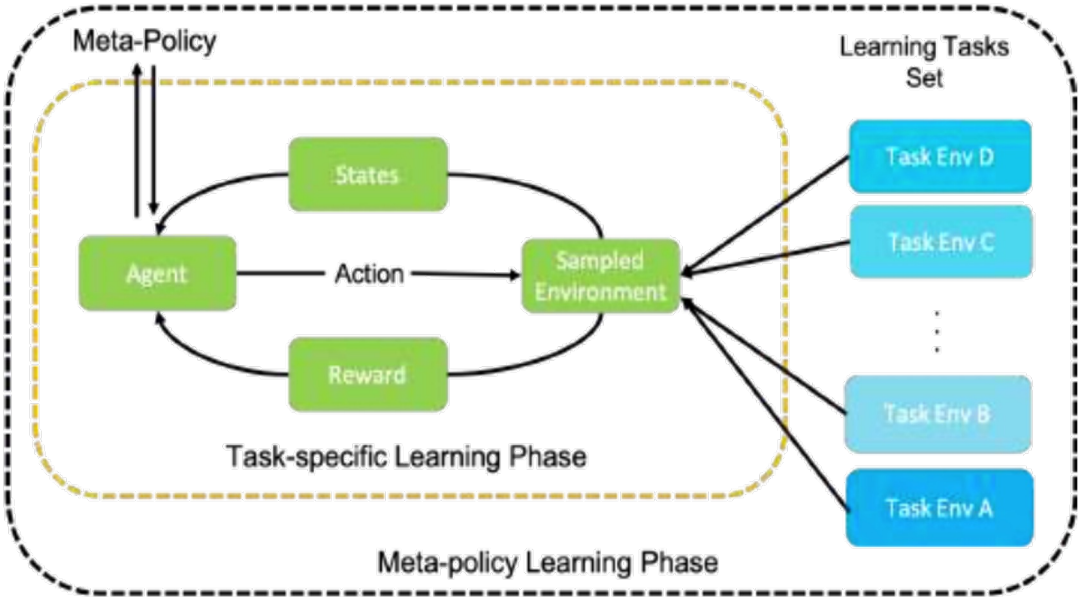
With the SLRAS proposed, which could self-adapt DRL algorithms to assist human operators by assisting with task goal setting through **meta-DRL approach**

To evaluate if,

- DRL algorithm: improve the **generality** of the robot in HRC
- meta-RL: increase the **learning efficiency** and improve the **speed of adaptation**

Methodology

- Task-specific learning phase;
- Meta-policy learning phase;
- Meta-policy generalizing phase



The diagram of meta-RL learning process [6].

[6] Mathew Botvinick et al. “Reinforcement learning, fast and slow”. In: *Trends in cognitive sciences* (2019).

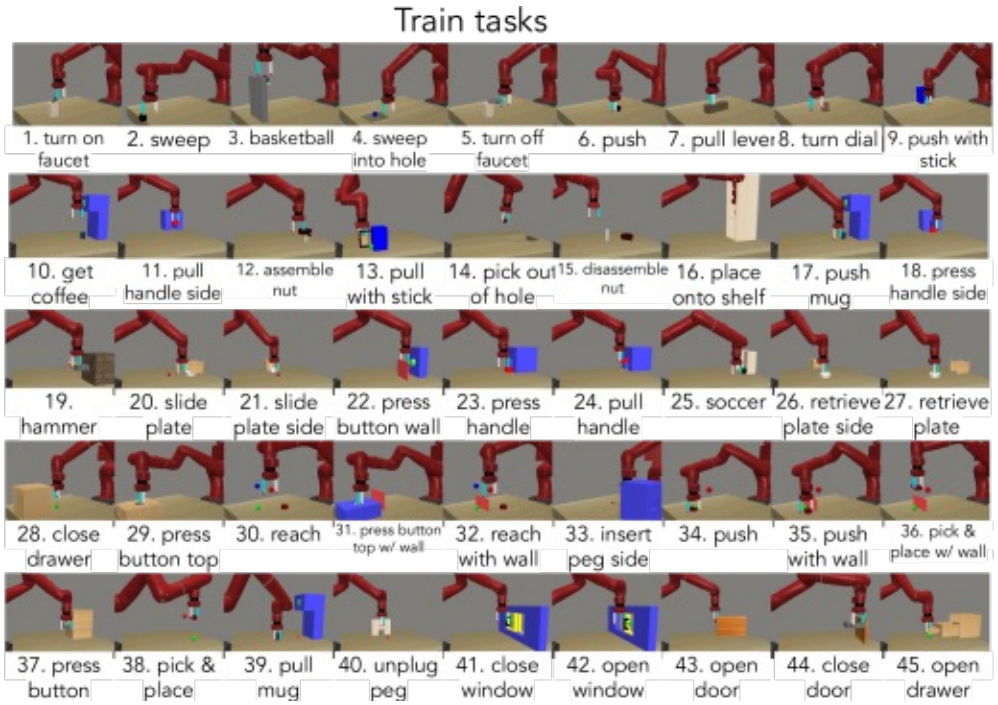
Experiments

- **Pre-training set:**

Manipulation of different objects

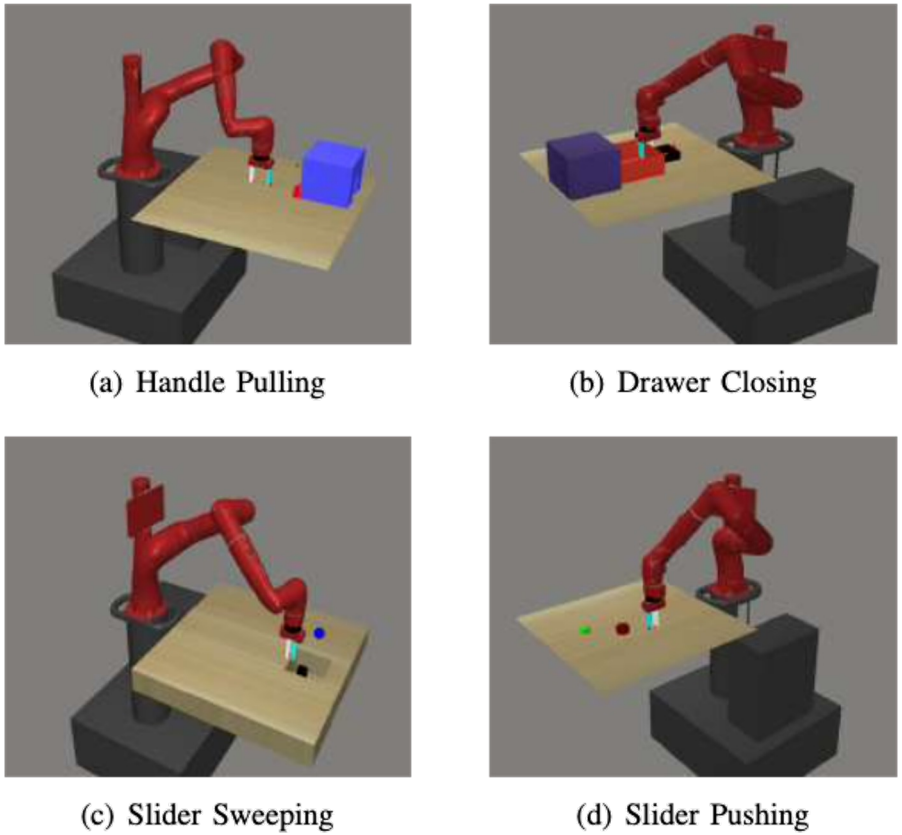
- **Unseen sampled tasks:**

Complete a specific practical use task



Experiments

- 12 robot control tasks from similar HRC tasks in meta-world [7].
- Phase (1)(2): 10 **pre-training tasks**.
- Phase (3): 2 tasks from training set + 2 **unseen tasks** were set as evaluation tasks.



(a) Handle Pulling (b) Drawer Closing

(c) Slider Sweeping (d) Slider Pushing

The four scenarios in evaluating experiments.

[7] Tianhe Yu et al. “Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning”. In: *Conference on Robot Learning*. PMLR. 2020, pp. 1094–1100.

Experiments

- **DRL algorithm setting:** A multi-layer neural network consisting of two fully connected layers with 100 hidden units in each layer.
- **Learning settings:**
  - Activation function: Tanh
  - Optimizer: Adam
- **Hardware:** CPU Intel Core i5- 5200U 2-Core Processor 2.20GHz and each iteration took an average of 280s.
- **Software Environment:** Pytorch+Mujuco+Metaworld

I. Experiment Settings

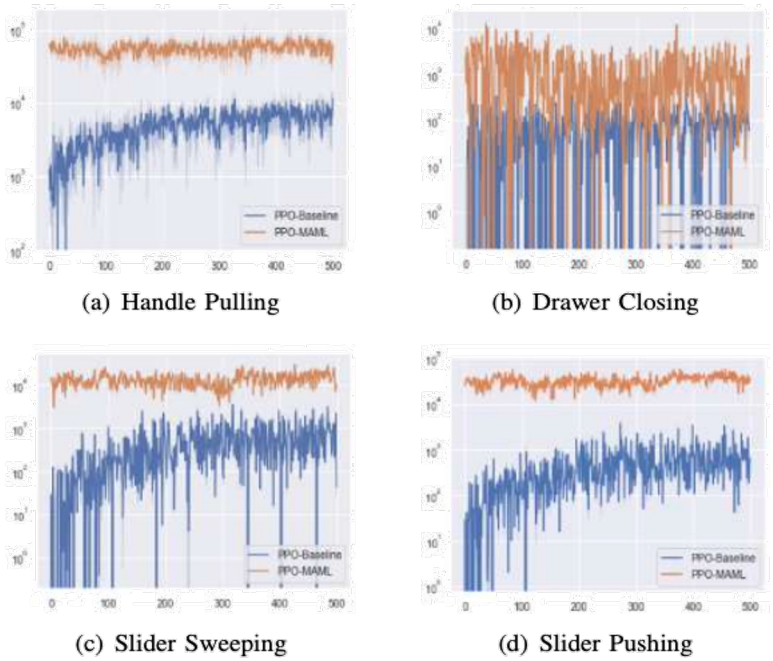
Parameter	Value
PPO epochs	3
PPO clip ratio	0.1
Inner learning rate	0.01
Outer learning rate	0.01
Training adapt_steps	1
Validation adapt_steps	5

[8] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

Experiment Results

II. Success Rate of Evaluating Tasks

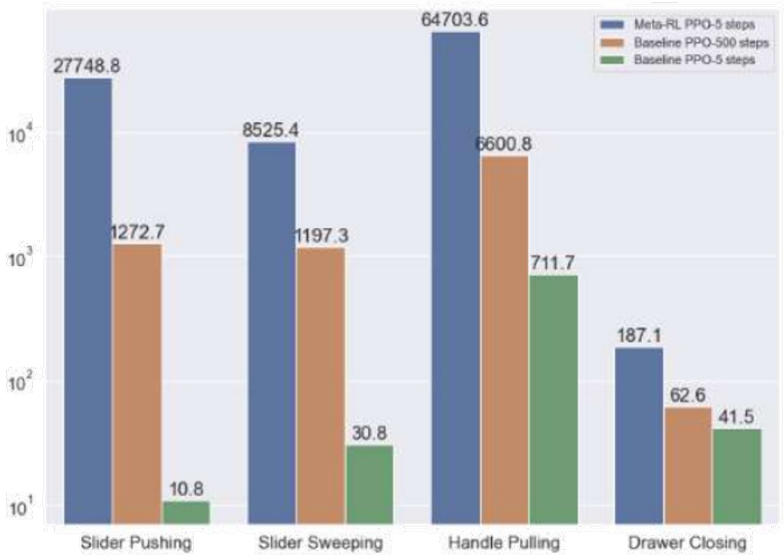
Policy	Slider Pushing	Slider Sweeping	Handle Pulling	Drawer Closing
Baseline-PPO 5 steps	0%	0%	0%	0%
Baseline-PPO 500 steps	30%	66%	0%	0%
MAML-PPO 5 steps	100%	100%	100%	86.7%



(a) Handle Pulling (b) Drawer Closing

(c) Slider Sweeping (d) Slider Pushing

The reward gained with MAML-based PPO and baseline PPO in evaluating scenarios with different learning periods

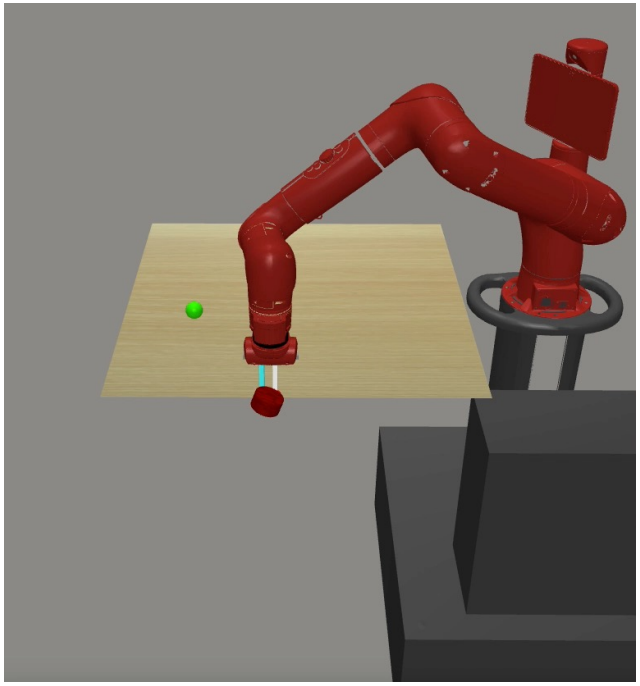


The reward curves of evaluating experiments. Higher reward value means better performance

Experiment Results

Task Demonstration of cobot with MAML-PPO in 5 steps (40sec training)

Slider Sweeping



Slider Pushing

Case Study: HRC disassembly of a retired battery

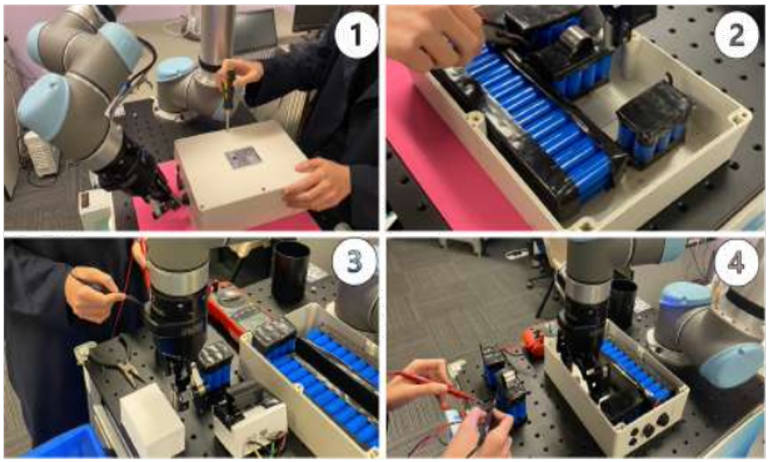
HRC in small batches and customised dismantling tasks

Human: unscrews the battery box

Cobot: press a button to observe the state of the battery

Human: diagnoses the recycling value of the parts in the battery

Cobot: sweeps the disassembled parts into containers by classification



Case Study of HRC for aging battery disassembly

Human: cuts the cable to the battery

Cobot: remove the battery from the box

Human: receives the items

Cobot: pushes new items to the operator

Discussions and Future Work

- Simplify the model and simulation environment
- Handle the task diversity
- Transfer control strategies obtained in simulators to physical robots without any performance degradation

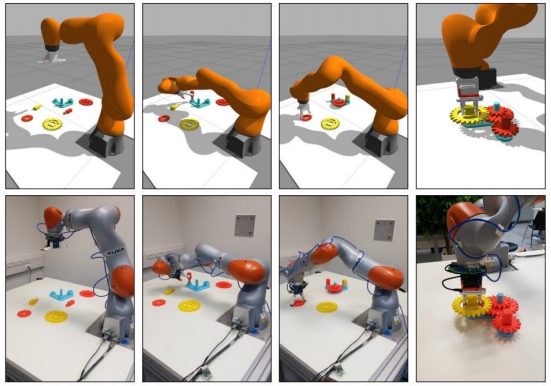


Figure 1: The KUKA LBR IIWA robot performs the *Siemens Innovation Challenge* successfully in simulation (top row) and in reality (bottom row).

Kalashnikov, Dmitry, et al. "Scalable deep reinforcement learning for vision-based robotic manipulation." *Conference on Robot Learning*. PMLR, 2018.

Lee, Robert, et al. "Zero-shot sim-to-real transfer with modular priors." *CoRR* (2018).

Makoviychuk, Viktor, et al. "Isaac gym: High performance gpu-based physics simulation for robot learning." *arXiv preprint arXiv:2108.10470* (2021).

Reference

[1] Yoram Koren. *The global manufacturing revolution: product-process-business integration and reconfigurable systems*. Vol. 80. John Wiley & Sons, 2010.

[2] Afonso Castro, Filipe Silva, and Vitor Santos. “Trends of Human-Robot Collaboration in Industry Contexts: Handover, Learning, and Metrics”. In: *Sensors 21 (June 2021*), p. 4113. DOI: 10.3390/s21124113.

[3] Mohamed El-Shamouty et al. “Towards Safe Human- Robot Collaboration Using Deep Reinforcement Learning”. In: *2020 IEEE International Conference on Robotics and Automation (ICRA).* IEEE. 2020, pp. 4899–4905.

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks”. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 1126–1135.

[5] John Schulman et al. “Proximal policy optimization algorithms”. In: *arXiv preprint arXiv:1707.06347* (2017).

[6] Mathew Botvinick et al. “Reinforcement learning, fast and slow”. In: *Trends in cognitive sciences* (2019).

[7] Tianhe Yu et al. “Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning”. In: *Conference on Robot Learning*. PMLR. 2020, pp. 1094–1100.

[8] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

Thank you

Q & A