Secondment at Tsinghua University

Zhen Deng

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics

Technical Aspects of Multimodal Systems

13. 11. 2018

Outline

1. Learning in-hand manipulation using DRL

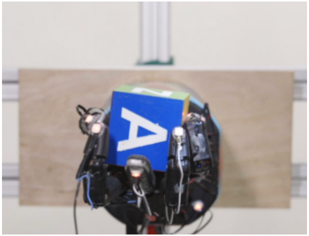
2. Prof.Sun's lab from Tsinghua University

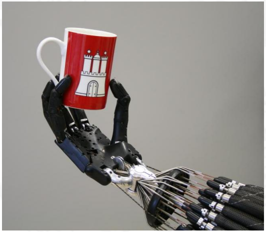
3. Tactile sensor from Prof Pan's lab

Learning in-hand manipulation using DRL

Motivations

1. Programming in-hand manipulation with multi-fingered robotic hands is a challenging problem.
2. Deep reinforcement learning has used successfully to learn complex manipulation skills.
3. Incorporating domain-specific knowledge into DRL in order to reduce its sample complexity
4. Improving the exploration efficiency of DRL





The motivation of this work includes four points

Related works

1. Different control patterns of in-hand manipulation [3]: Rolling, Sliding, Finger gaiting, Finger-privoting/tracking, ...
2. Previous methods: trajectory optimization methods [4] and imitation learning methods [1].
3. Recent works focus on DRL [6, 5], eg., OpenAI's work.

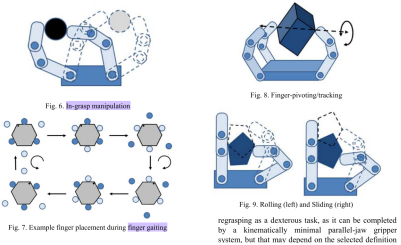


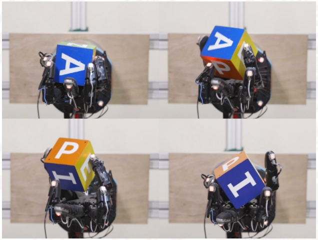
Fig. 6. In-grasp manipulation

Fig. 7. Example finger placement during finger gaiting

Fig. 8. Finger-pivoting/tracking

Fig. 9. Rolling (left) and Sliding (right)

regrasping as a dexterous task, as it can be completed by a kinematically minimal parallel-jaw gripper system, but that may depend on the selected definition

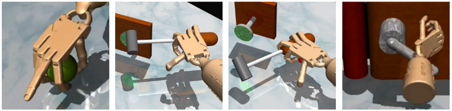


In-hand manipulation has different control patterns. For example

Previous methods include trajectory optimization rely on a

Related works: Two important papers:

1. Learning Complex In-hand Manipulation with Deep Reinforcement Learning and Demonstrations (Rajeswaran et al. [7], 2017)



2. Learning Dexterous In-Hand Manipulation (OpenAI [8], 2018)

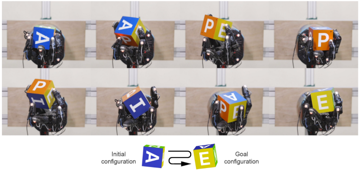


Figure 1: A five-fingered humanoid hand trained with reinforcement learning manipulating a block from an initial configuration to a goal configuration using vision for sensing.

the first one is ..., demonstration data is used to fast the skill

Learning in-hand manipulation using DRL

- Palli et al[6]. In-hand manipulation can be represented as deviations from a reference grasp.

- Odhner et al[5]. utilized *precision* grasp configuration for the planning of dexterous manipulation using under-actuated hand.
- Saut et al[8]. built a probabilistic roadmap in grasp subspaces and searched a trajectory in this roadmap for dexterous manipulation.

Assumptions 1

Precision grasp configuration can be taken as a reference for in-hand manipulation skill learning.

The learning agent is preferred to fully explore a state subspace with high probability success for skill learning.

The proposed method:

1. Using a PPO algorithm provided by OpenAI to learn in-hand manipulation skill, like the object rotation.
2. Using reward shaping method to improve the learning efficiency.
   - Guide the exploration of DRL by using the *precision* grasp configuration as a reference.
   - Measuring the uncertainty of explored states for full exploration.
3. Using multi-agent reinforcement learning (MARL) to incorporate multiple reward functions.

So we want the agent to fully explore...., In the proposed method, we use ....., DRL algorithms:

1 Trust region policy optimization (TRPO):

$$\underset{\theta}{\text{maximize}} \quad \hat{\mathbb{E}}_t\left[\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}\hat{A}_t\right]$$
$$\text{subject to} \quad \hat{\mathbb{E}}_t[\mathrm{KL}[\pi_{\theta_{\text{old}}}(\cdot \mid s_t), \pi_\theta(\cdot \mid s_t)]] \leq \delta.$$

2 Proximal Policy Optimization(PPO) Algorithms:

$$L^{CLIP}(\theta) = \hat{E}_t\left[\min(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t)\right]$$

- θ is the policy parameter
- $\hat{E}_t$ denotes the empirical expectation over timesteps
- r_t is the ratio of the probability under the new and old policies, respectively
- $\hat{A}_t$ is the estimated advantage at time t
- ε is a hyperparameter, usually 0.1 or 0.2

PPO algorithm is developed from TRPO algorithm. The different

Three reward functions:

1. An extrinsic reward r^{ext} that specifies the task goal.
2. A hand-based reward r^{hand} that encourage the agent to explore the *precision* hand posture subspace.
3. An uncertainty-based reward r^{unc} that used to balance the trade-off between exploration and exploitation of the MARL algorithm.

$$r := \{r^{ext}, r^{hand}, r^{unc}\}$$

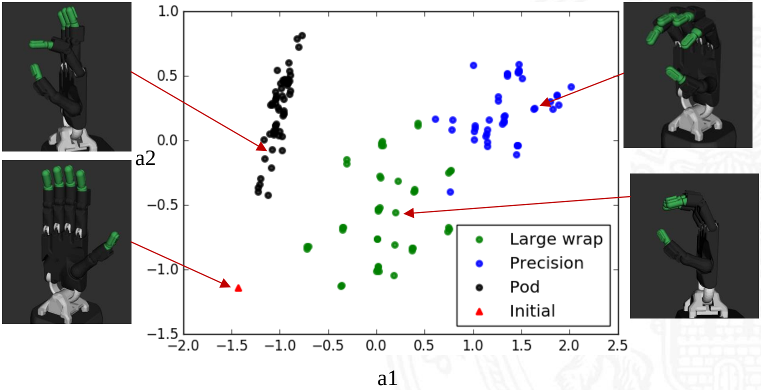
r^{ext} is computed based on the different of two quaternions.

$$r^{ext} = \begin{cases} +2 & \text{if the goal is reached} \\ ||q_{curr} \ominus q_{target}||^2 & \text{otherwise} \end{cases} \tag{1}$$

Learning in-hand manipulation using DRL

r^{hand} is measured based on the similarity between the current explored hand configuration and the reference configuration.

$$r^{hand} = ||A - A^{precision}_{center}||_2^2, \quad A = \{a_1, a_2\} \tag{2}$$



r^{unc} is measured the uncertainty of explored states.
The uncertainty of explored states is approximated by the prediction error of a transitional dynamic model.

- Transition dynamic model: $f_{\theta_f} : s_t \times a_t \to \hat{s}_{t+1} \approx s_{t+1}$.

$$\begin{aligned}\mathscr{L}(\theta_f) &= \frac{1}{|D|} \sum_{\{s_t, a_t, s_{t+1}\} \in D} H(s_{t+1}, \hat{s}_{t+1}) \\ &= -\frac{1}{|D|} \sum_{\{s_t, a_t, s_{t+1}\} \in D} log f_{\theta_f}(s_t, a_t)\end{aligned} \tag{3}$$

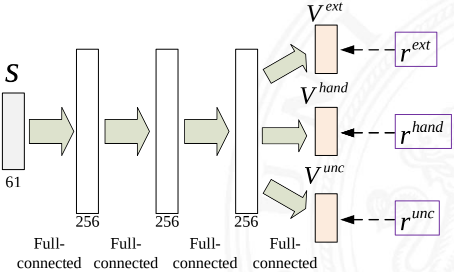
- Use the cross entropy between the distribution of the explored state s_{t+1} and the predicted state $\hat{s}_{t+1}$ to compute the prediction error.

$$r^{unc} = -log f_{\theta_f}(s, a) \tag{4}$$

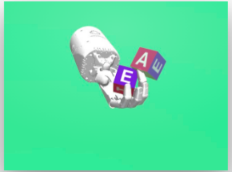
Learning in-hand manipulation using DRL

MARL: multiple agents represented by value functions respectively are trained separately with respect to each own reward and then cooperate to optimize the skill policy.

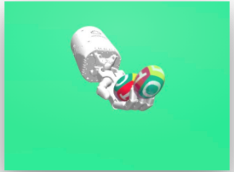
$$V_{comp}(s) = \beta_1 V^{ext}(s) + \beta_2 V^{hand}(s) + \beta_2 V^{unc}(s) \qquad (5)$$



Object rotation tasks in the Openai gym robotic environments:



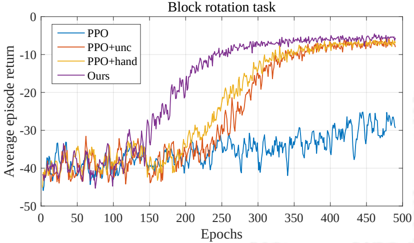
HandManipulateBlock-v0
Orient a block using a robot hand.

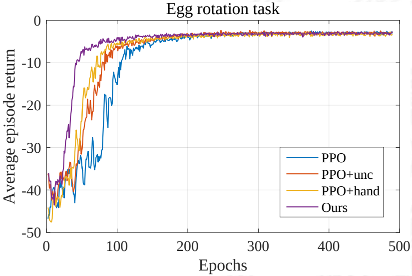


HandManipulateEgg-v0
Orient an egg using a robot hand.

Experiments

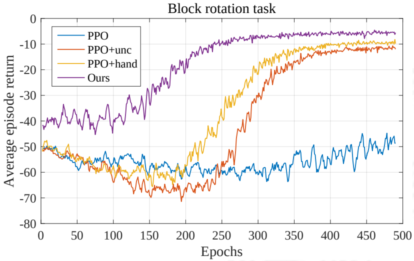
the average episode return with respect to the extrinsic reward r^{ext}:

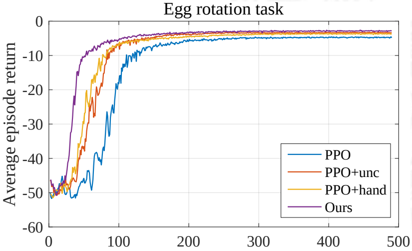




Experiments

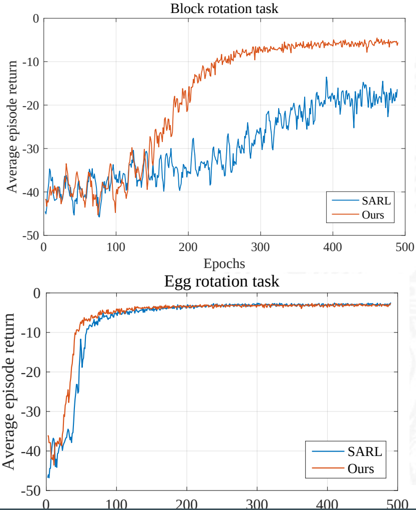
The average episode return with respect to the hand-based reward r^{hand}:





Experiments

The MARL is compared with the single-agent RL which train a value function on a composition reward function.

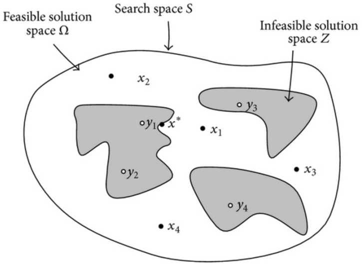


Conclusion:

- Using DRL to learn in-hand manipulation skills.
- Incorporating domain-specific knowledge into DRL algorithm by designing additional reward.

Extension:

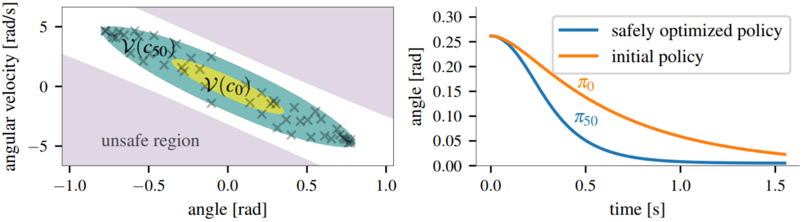
1. Can we explicitly represent the feasible state subspace with high probability success.



Conclusion

Extension:

2 The safety constraint should be considered in DRL which restricts the agent to explore the reasonable states
3 Combine model-based optimal control and model-free RL for stable skill learning.



(a) Estimated safe set.   (b) State trajectory (lower is better).

Safe Model-based Reinforcement Learning with Stability Guarantees [2]

Prof.Sun’s lab from Tsinghua University

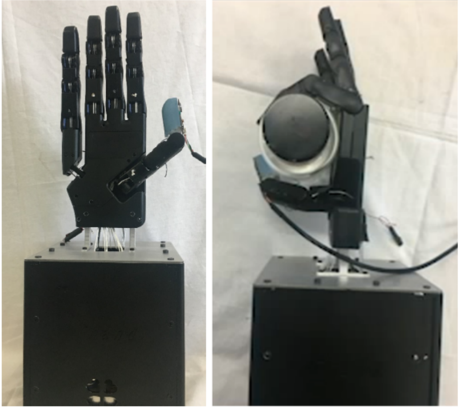
Two main research directions of the lab: Active perception and Dexterous manipulation.

PhD student’s work:

- Yikai Wang: "Visual-tactile fusion" for CML B5 sub-project.
- Chao Yang: "Imitation learning considering structure inconsistent"
- Mingxuan Jing: "preference-based reinforcement learning" and "pose estimation"

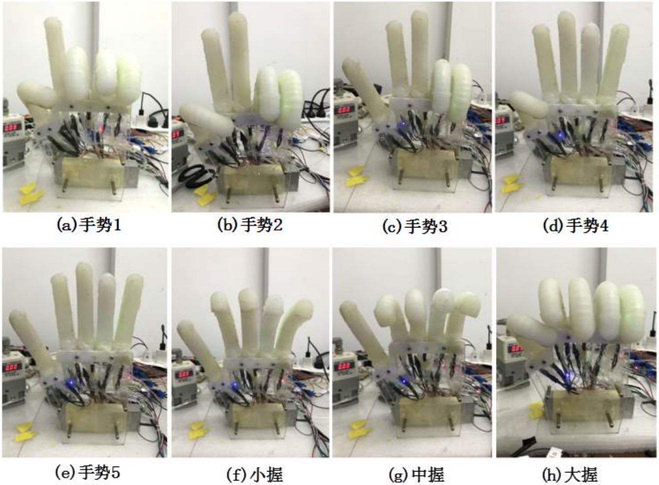
Prof.Sun's lab from Tsinghua University

Five-fingered robotic hand:



Prof.Sun’s lab from Tsinghua University

Five-fingered soft hand:



(a)手势1 (b)手势2 (c)手势3 (d)手势4

(e)手势5 (f)小握 (g)中握 (h)大握

Tactile sensor from Prof Pan's lab

https://bme.ucdavis.edu/people/departmental-faculty/tingrui-pan/

Tingrui Pan

Associate Professor and Director

Micro-Nano Innovations (MiNI) Laboratory

Department of Biomedical Engineering

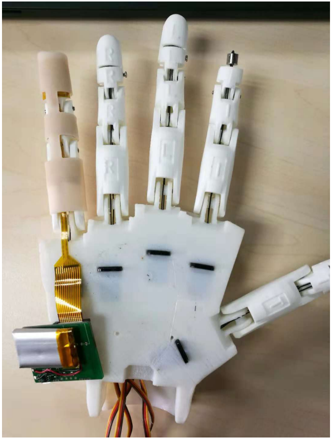
University of California Davis, CA 95616

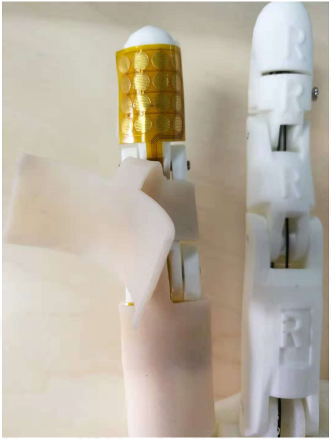
Phone: (530) 754-9508

Email: tingrui@ucdavis.edu

- B. Nie, R. Li, J. D. Brandt, and T. Pan, "Iontronic Microdroplet Array for Flexible Ultrasensitive Tactile Sensing", Lab Chip, vol. 14, pp. 1107-1116.
- R. Li, Y. Si, Z. Zhu, Y. Guo, Y. Zhang, N. Pan, G. Sun, and T. Pan, "Supercapacitive Iontronic Nanofabric Sensing," Adv Mater, vol. 29, 1700253, pp. 1-8.
- Z. Zhu, R. Li, and T. Pan, "Imperceptible Epidermal-Iontronic Interface for Wearable Sensing," Adv Mater, vol. 30, 1705122, pp. 1-9, Jan 2018.

Tactile sensor:





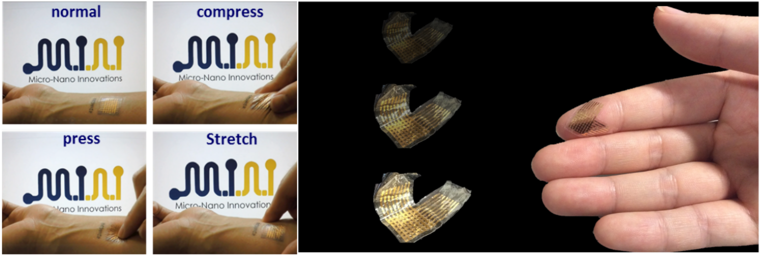
Tactile sensor from Prof Pan's lab

32

Artificial Tactile Sensing

Ultrathin Stick-on Pressure Sensing Array

- Ultrathin, Imperceptible and Invisible
- Conformable Attachment to Skin



This slide is provided by Prof Pan's lab.

D. Antotsiou, G. Garcia-Hernando, and T.-K. Kim. Task-oriented hand motion retargeting for dexterous manipulation imitation. *arXiv preprint arXiv:1810.01845*, 2018.

F. Berkenkamp, M. Turchetta, A. Schoellig, and A. Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in Neural Information Processing Systems*, pages 908–918, 2017.

R. R. Ma and A. M. Dollar. On dexterity and dexterous manipulation. In *Advanced Robotics (ICAR), 2011 15th International Conference on*, pages 1–7. IEEE, 2011.

I. Mordatch, Z. Popović, and E. Todorov. Contact-invariant optimization for hand manipulation. In *Proceedings of the ACM SIGGRAPH/Eurographics symposium on computer animation*, pages 137–144. Eurographics Association, 2012.

L. U. Odhner and A. M. Dollar. Stable, open-loop precision manipulation with underactuated hands. *The International Journal of Robotics Research*, 34(11):1347–1360, 2015.

G. Palli, F. Ficuciello, U. Scarcia, C. Melchiorri, and B. Siciliano. Experimental evaluation of synergy-based in-hand manipulation. *IFAC Proceedings Volumes*, 47(3):299–304, 2014.

A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine. Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. *arXiv preprint arXiv:1709.10087*, 2017.

J.-P. Saut, A. Sahbani, S. El-Khoury, and V. Perdereau. Dexterous manipulation planning using probabilistic roadmaps in continuous grasp subspaces. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 2907–2912. IEEE, 2007.

Thanks for your attention! Any questions?