Research Statement

Pingcheng Jian

My long-term research goal is to enable robots to assist humans in various daily tasks intelligently and reliably. To achieve this goal, I conduct research on novel robot control approaches and use them to enable skillful and robust robot actions, especially in **novel environments** unseen in the training data and **adversarial environments** with disturbances.

Adaptation to novel environments. Robots are typically trained to excel at specific tasks such as relocating objects or navigating to predetermined locations. However, such robots need to be retrained from scratch when faced with new tasks, body changes, or visual configuration changes. This retraining process is usually time-consuming and computationally expensive, hindering the deployment of robots in the real world where the environment is changing continuously and sometimes drastically. In contrast, humans demonstrate remarkable capabilities to continuously acquire new skills by drawing on past experiences. Even under physical constraints imposed by injuries, we can still rapidly adapt to perform new tasks. Despite significant advancements in robotics and machine learning, robots still cannot generalize their experiences across a wide range of tasks, body configurations, or visual settings. It is crucial to empower robots with an adaptation capability similar to that of humans if we want to deploy them in the real world.

Previous works have attempted to enable fast adaptation to novel environments by regularizing a large policy network for multi-task learning. They achieve this by learning routing connections to reuse part of the network weights [1] or by assigning task-specific sub-networks [2]. However, the capacity of the large policy network grows exponentially as the number of tasks increases, leading to enormous memory consumption and computational expense.

Motivated by the need for compact yet highly adaptable robot policies, I adopted a **modular structure design** of the policy networks and developed the Policy Stitching algorithm [3] and the Perception Stitching algorithm [4]. The key idea behind these two methods is to disentangle the knowledge of different objectives such as tasks, sensors, robot kinematics, and then learn them in different neural network modules. This allows us to reuse modules trained in different environments by assembling them into novel combinations and deploying them in previously unseen environments. For instance, given one policy trained for a 3-DoF manipulator to pick up a cube and another policy trained for a 5-DoF manipulator to pick up a stick, if we would like to have the 3-DoF manipulator pick up a stick now, Policy Stitching algorithm [3] can directly take the robot module in the first policy and stitch it with the task module in the second policy.

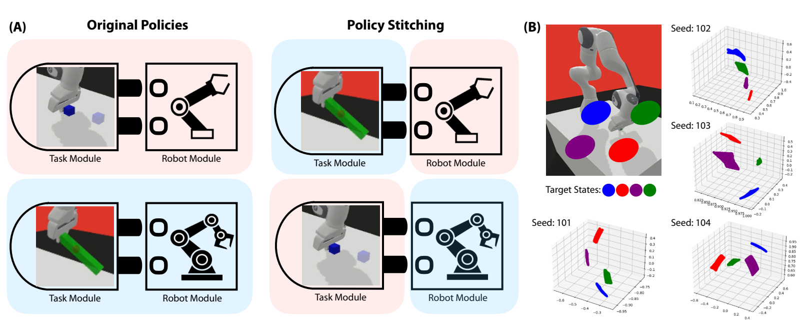

Figure 1: Policy Stitching [3]. (A) Our framework facilitates robot transfer learning among novel combinations of robots and tasks by decoupling and stitching robot and task modules. (B) Motivation example: A robot arm is trained to reach goals in four different target regions using the modular policy. Results from separate training runs with different random seeds (101-104) show misaligned latent representations.

Through my work on **training neural network modules to be reusable** in other environments, I have found that the latent representation misalignment issue [5] between the encoders and decoders trained in different environments is the major problem that diminishes the performance of the reassembled policy in a new environment. This observation drives my research into developing techniques inspired by the relative representation algorithm [5] and the covariance regularization method [6, 7, 8, 9] to enforce the invariance at the latent space of different encoders. Our experience has shown that training the latent states to be invariant to isometric transformations and enforcing the latent features to be independent can significantly alleviate the latent space misalignment issue and improve the performance of the reassembled policy. Although proven to be successful in many real-world experiments, my current approaches cannot achieve very satisfying performance on long-horizon dexterous tasks yet, leading to my future research to further understand the latent space of neural networks for standardizing the interfaces between different network modules.

Robustness against disturbances. Many robot learning approaches show satisfactory performance in the ideal disturbance-free environments. However, in a real-world environment, noise and disturbances are inevitable, which often lead to a drastic drop in robot performance. If we want to deploy robots in the real world, we need control policies that are not only robust to external disturbances caused by outside perturbations, but also robust to internal disturbances caused by hardware noise and errors. To this end, previous works [10, 11] have proposed various data augmentation methods that inject random noise into the trajectory data. However, these methods generally cannot be effective against more malicious disturbances beyond random noise. For example, suppose a protagonist robot arm wants to pick up a cube on the desk, while there is another adversarial robot arm deliberately blocks in the way of the protagonist or moves the cube away from the grasp of the protagonist. In this case, a robot trained with data augmentation techniques can no longer be robust enough to accomplish the task. How can we teach a robot to avoid attacks and recover from failures in an adversarial environment?

(a) (b)

Figure 2: Adversarial skill learning for robust manipulation [12]. The protagonist tries to maximize the reward while the adversary tries to minimize it. (a) An external disturbance, the protagonist is disturbed by another adversary robot. (b) An internal disturbance, the protagonist joint action is under the disturbance of adversarial action on each joint of the robot.

Inspired by the training pipeline of the Generative Adversarial Network (GAN) [13], I proposed the Adversarial Skill Learning method [12]. The key contribution of this approach lies in the **adversarial pipeline design** of the training process. Specifically, we introduce an adversary agent to disturb the protagonist agent. During the reinforcement learning process, the protagonist agent is optimized to accomplish the task, while the adversary agent is optimized to obstruct the task. These two phases of optimization occur iteratively. After this competitive GAN-style training process, the adversary agent learns dexterous skills to attack the protagonist agent, while the protagonist agent is trained to be robust against malicious adversarial attacks as well as various random perturbations.

While my proposed method shows a pronounced advantage over vanilla data augmentation, it still requires manually tuning the optimal amplitude range of the disturbance for the training. In my future research, I aim to enhance the Adversarial Skill Learning framework with the recent advances in the diffusion policy [14] and the foundation models [15]. Along this path, I believe I can push the limits of robot control robustness, enabling reliable deployment of robots in the real world.

Recently, I began to explore the application of foundation models in my research agenda. Previous research has shown that large language models (LLMs) can effectively design and improve the reward functions of some robotic tasks trained with reinforcement learning (RL) [16, 17]. We tested this idea on some difficult quadruped robot locomotion tasks and found that the training process consumes very expensive computation resources for parallel reward functions searching, yet the trained robots sometimes cannot achieve satisfying performance, showing unnatural gait patterns or struggling to accomplish the tasks. In the short term, my research objective is to apply vision-language models (VLMs) on model-based reinforcement learning for more efficient training computation and better performance that satisfies the human preferences [18]. In the long term, I would like to explore leveraging the LLMs and VLMs for state and action abstraction learning, which I believe is a promising path towards general robotic intelligence.

Looking further, my very long-term goal is to develop general embodied intelligence that assists people in every aspect of their work and daily lives. To achieve this goal, we need to have a deeper understanding of the mechanisms behind both natural and artificial intelligence. Based on the advances in understanding the source of "intelligence", we can build robotic systems that push the limits of robustness and adaptation capabilities that I previously focused on, as well as many other performance facets that I am excited to explore.

References

[1] R. Yang, H. Xu, Y. Wu, and X. Wang. Multi-task reinforcement learning with soft modularization. *Advances in Neural Information Processing Systems*, 33:4767–4777, 2020.

[2] J. A. Mendez, H. van Seijen, and E. Eaton. Modular lifelong reinforcement learning via neural composition. *arXiv preprint arXiv:2207.00429*, 2022.

[3] P. Jian, E. Lee, Z. Bell, M. M. Zavlanos, and B. Chen. Policy stitching: Learning transferable robot policies. *arXiv preprint arXiv:2309.13753*, 2023.

[4] P. Jian, E. Lee, Z. Bell, M. M. Zavlanos, and B. Chen. Perception stitching: Zero-shot perception encoder transfer for visuomotor robot policies. *arXiv preprint arXiv:2406.19971*, 2024.

[5] L. Moschella, V. Maiorca, M. Fumero, A. Norelli, F. Locatello, and E. Rodolà. Relative representations enable zero-shot latent space communication. *arXiv preprint arXiv:2209.15430*, 2022.

[6] A. Bardes, J. Ponce, and Y. LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv preprint arXiv:2105.04906*, 2021.

[7] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pages 12310–12320. PMLR, 2021.

[8] I. Higgins, L. Matthey, A. Pal, C. P. Burgess, X. Glorot, M. M. Botvinick, S. Mohamed, and A. Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. *ICLR (Poster)*, 3, 2017.

[9] I. Higgins, D. Amos, D. Pfau, S. Racaniere, L. Matthey, D. Rezende, and A. Lerchner. Towards a definition of disentangled representations. *arXiv preprint arXiv:1812.02230*, 2018.

[10] N. Hansen and X. Wang. Generalization in reinforcement learning by soft data augmentation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13611–13617. IEEE, 2021.

[11] A. D'Innocente, F. M. Carlucci, M. Colosi, and B. Caputo. Bridging between computer and robot vision through data augmentation: a case study on object recognition. In *Computer Vision Systems: 11th International Conference, ICVS 2017, Shenzhen, China, July 10-13, 2017, Revised Selected Papers 11*, pages 384–393. Springer, 2017.

[12] P. Jian, C. Yang, D. Guo, H. Liu, and F. Sun. Adversarial skill learning for robust manipulation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2555–2561. IEEE, 2021.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[14] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song. Diffusion policy: Visuomotor policy learning via action diffusion. *arXiv preprint arXiv:2303.04137*, 2023.

[15] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.

[16] Y. J. Ma, W. Liang, G. Wang, D.-A. Huang, O. Bastani, D. Jayaraman, Y. Zhu, L. Fan, and A. Anandkumar. Eureka: Human-level reward design via coding large language models. *arXiv preprint arXiv:2310.12931*, 2023.

[17] Y. J. Ma, W. Liang, H.-J. Wang, S. Wang, Y. Zhu, L. Fan, O. Bastani, and D. Jayaraman. Dreureka: Language model guided sim-to-real transfer. *arXiv preprint arXiv:2406.01967*, 2024.

[18] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.