

聂步青

(+86) 132-6272-9360 | niebuqing@sjtu.edu.cn | [Google Scholar](#) | 预备党员

教育经历

上海交通大学 | 计算机科学与技术 | 博士研究生 (硕博连读, 导师: 高岳) 2022.04—2026.03 (预计)
主要研究方向为强化学习算法与具身智能, 对强化学习算法 (深度强化学习, 鲁棒强化学习, 元强化学习等), 强化学习的机器人具身智能应用 (足式机器人、机械臂、人形机器人等) 有一定的理论研究和工程经验。

上海交通大学 | 控制科学与技术 | 硕士研究生 2020.09—2022.04

上海交通大学 | 计算机科学与技术 (IEEE 试点班) | 工学学士 2016.09—2020.06

技术能力

- 深度学习/强化学习基础: Pytorch, Gym, Wandb, DRL (CleanRL, Stable Baselines 3, Rsl-RL, etc.).
- 强化学习算法: Model-free/based RL (PPO, TD3, SAC, etc.); Diffusion Policy, etc.
- 机器人学习: Simulation (Mujoco, Isaac, Coppeliasim); Robots (Legged robot, Robot Manipulator, Humanoid).

主要研究经历

在具身智能 VLA 大模型上实现强化学习后训练, 提升机械臂 VLA 操作成功率 | 进行中

- 在 VLA 具身大模型 OpenVLA-OFT 上, 基于偏好学习/强化学习算法进行后训练, 从而减少对 SFT 微调的依赖, 减少专家数据需求量, 提升 VLA 策略的成功率; 部署到 ALOHA 机械臂平台, 提升策略在实机上的效果。
- 进度: 正在收集真机 SFT 数据, 编写 OpenVLA-OFT 的强化学习仿真环境训练代码。

在强化学习的机器人运控/模仿任务中, 策略对称等变性的效率与鲁棒性研究 | 在投

- 在强化学习实现人形机器人的运控和模仿任务中, 基于机器人结构对称性, 构造策略面对观测的对称等变特性, 提升训练效率, 提升机器人指令跟随精度, 提高机器人动作在时间和空间上的协调性;
- 已实现运控任务的对称等变性; 模仿任务已完成仿真训练, 正在进行性能调优和 Sim2Real。

基于动作空间解耦思路, 实现连续控制的强化学习决策频率自主调整 | ICLR-25

- 传统强化学习受限于 MDP 的固定决策频率, 存在动作序列的时序一致性差、探索效率低下、训练效率低的问题;
- 实现 DRL 动作空间解耦机制, 每个动作维度独立进行闭环动作复用算法 SDAR, 平衡动作的 persistence 和 diversity, 实现了智能体动作决策频率的自主调节, 提高了训练效率、策略效果与动作平滑度。

实现最优动作扰动下的策略优化, 提升强化学习算法的动作噪音鲁棒性 | 在投

- 对强化学习的动作鲁棒性进行问题建模; 提出了动作扰动下的策略迭代优化框架, 并证明理论上的有效性;
- 应用到 DRL 算法 TD3 和 PPO 中, 在常见连续控制任务上开展实验, 证明策略在多种动作攻击下仍具有较高性能, 即提升了动作鲁棒性; 同时保持较高采样效率, 在无扰动情况下仍然保持较高性能。

基于 Lipschitz 连续特性, 构建对观测噪音鲁棒的深度强化学习策略 | AAAI-24

- 将 Lipschitz 连续特性引入到强化学习策略网络结构中, 在理论层面阐述了 Lipschitz 连续性与策略的观测鲁棒性之间的关联; 基于 SortNet 构建了有数学保障的鲁棒强化学习 SortRL, 基于策略蒸馏设计了鲁棒训练流程;
- 在 Atari、ProcGen 等任务与多种攻击上实验, SortRL 相比 SOTA 有显著提升 (小扰动超过 20%, 强扰动超过 28%)。

基于对抗学习框架, 实现机器人模仿人类行为时的自身稳定性 | ICRA-22

- 实现了音乐驱动的人形机器人模仿学习算法, 模仿人类舞蹈动作实现动作相似性, 同时满足自身稳定性;
- 基于对抗训练思路, 学习机器人稳定性预测器与动作生成器, 在 Coppeliasim、Mujoco 仿真环境中对 NAO、ROBOTIS-OP2 两款人形机器人进行训练, 均达到了预定的稳定性的效果。

主要研究成果

- Buqing Nie**, Yangqing Fu, Yue Gao, “Select before Act: Spatially Decoupled Action Repetition for Continuous Control.” International Conference on Learning Representations (ICLR) 2025.
- Buqing Nie**, Jingtian Ji, Yangqing Fu, Yue Gao. “Improve robustness of reinforcement learning against observation perturbations via l_∞ lipschitz policy networks.” AAAI Conference on Artificial Intelligence (AAAI) 2024.
- Buqing Nie**, Yangqing Fu, Yue Gao. “Action Robust Reinforcement Learning via Optimal Adversary Aware Policy Optimization.” (Ready to submit for TNNLS)
- Buqing Nie**, and Yue Gao. “DanceHAT: Generate Stable Dances for Humanoid Robots with Adversarial Training.” International Conference on Robotics and Automation (ICRA) 2022.

- [5] **Buqing Nie**, Yidong Mei, Yue Gao, Feng Gao. “Capability Iteration Network for Robot Path Planning.” International Journal of Robotics and Automation (IJRA) 2022.
- [6] **Buqing Nie**, Yang Zhang, Rongjun Jin, Zhanxiang Cao, Huangxuan Lin, Yue Gao. “Learning Symmetry Equivariant Deep Reinforcement Learning Policy for Humanoid Robots.” (Submit to AAAI 2026)
- [7] Yangqing Fu, Ming Sun, **Buqing Nie**, Yue Gao. “Accelerating Monte Carlo Tree Search with Probability Tree State Abstraction.” Advances in Neural Information Processing Systems (NeurIPS) 2024.
- [8] Yang Zhang, **Buqing Nie**, and Yue Gao. “Robust Locomotion Policy with Adaptive Lipschitz Constraint for Legged Robots.” IEEE Robotics and Automation Letters (RAL) 2024.
- [9] Jingtian Ji, **Buqing Nie**, and Yue Gao. “DAGA: Dynamics Aware Reinforcement Learning With Graph-Based Rapid Adaptation.” IEEE Robotics and Automation Letters (RAL) 2023.
- [10] Zhanxiang Cao, **Buqing Nie**, Yang Zhang, Yue Gao. “Minimizing Acoustic Noise: Enhancing Quiet Locomotion for Quadruped Robots in Indoor Applications.”, IROS 2025.
- [11] Yang Zhang, **Buqing Nie**, Zhanxiang Cao, Yangqing Fu, Yue Gao. “Disturbance-Aware Adaptive Compensation in Hybrid Force-Position Locomotion Policy for Legged Robots.” (RAL Under Review)
- [12] Zhanxiang Cao, Yang Zhang, **Buqing Nie**, Huangxuan Lin, Haoyang Li, Yue Gao. “Learning Motion Skills with Adaptive Assistive Curriculum Force in Humanoid Robots.” (RAL Under Review)
- [13] Yangqing Fu, Ming Sun, **Buqing Nie**, Yue Gao. “General Tree State Abstraction for Efficient Monte Carlo Tree Search with Automatic Optimization.” (T-PAMI Under Review)
- [14] Wei Xu, Yue Gao, **Buqing Nie**. “Structure-Aware Policy to Improve Generalization among Various Robots and Environments.” International Conference on Robotics and Biomimetics (ROBIO) 2022.

主要奖项

- 华为奖学金 (2021 年 11 月)
- 上海交通大学博士生学业奖学金 (2022 年-2024 年)
- 上海交通大学研究生一等学业奖学金 (2020 年-2021 年)
- 美国大学生数学建模竞赛 M 奖 (<7%, 组长) (2019 年 5 月)
- 上海交通大学学业进步奖学金 (2018 年 9 月)