Multi-Agent Deep Reinforcement Learning for Variable-Finger Dexterous Grasping through Multi-Stream Embedding Fusion

Mahdi Bonyani [1], Maryam Soleymani[1], and Chao Wang[2]

***Abstract*— Dexterous robotic hands offer unparalleled potential for high-precision, contact-rich manipulation, but their control remains a formidable challenge due to high-dimensional action spaces and diverse object-hand interactions. In this paper, we propose a novel framework for dexterous grasping based on multi-agent deep reinforcement learning (MADRL) and multi-stream embedding fusion. Each component of the robotic hand, fingers, wrist, and arm, is modeled as an independent agent that learns cooperative control strategies guided by multi-stream embedding fusion. By leveraging high-quality static grasp data from the MultiDex dataset as reference targets, our method eliminates the need for human demonstrations or generative sampling during training. Experimental results demonstrate that our method achieves stable, compliant, and generalizable grasps across diverse objects and hand configurations, outperforming traditional single-agent baselines.**

I. INTRODUCTION

Dexterous robotic manipulation is a cornerstone of human-level autonomy in unstructured environments [1]. Unlike simple parallel-jaw grippers, multi-fingered robotic hands can achieve complex and adaptive interactions with objects of varying shapes, sizes, and functionalities [2]. However, the high degrees of freedom (DoF), intricate hand-object dynamics, and multi-modal sensory feedback significantly complicate control and policy learning for such systems [3].

While recent works have explored imitation learning and reinforcement learning for dexterous manipulation, most methods are constrained by reliance on either simplified hands or limited sensory feedback [4]. Visual input alone often proves insufficient when fine-grained force adjustments or occluded contact cues are required [5]. Tactile sensors, though essential, produce sparse and noisy data [6]. To address these challenges, we propose a novel framework that models each joint group of the robotic hand, fingers, wrist, and arm, as a separate agent in a multi-agent deep reinforcement learning (MADRL) setup by fusion of multi-stream embedding. Our approach incorporates high-quality static reference grasps extracted from the MultiDex dataset [7], which provides physically plausible grasp poses across multiple dexterous hand types. These reference configurations serve as supervisory targets during training, avoiding the need for simulation generation or full demonstrations.

A central component of our method is a multi-stream embedding fusion mechanism. Each agent's policy network processes multi-stream embedding data through a dual-attention pipeline, first extracting modality-specific features using self-attention [8], then merging them through cross-attention [9]. This design emulates multi-stream integration in the human approach, enabling robust and precise grasp behavior. We demonstrate that this framework achieves stable and functional grasp poses under various conditions, while maintaining generalization across multiple hand morphologies. Our contributions are threefold:

- We introduce a novel multi-agent reinforcement learning framework tailored for dexterous robotic hands using static reference grasp data.
- We develop a multi-stream attention-based fusion network that effectively integrates multi-stream embedding for precise manipulation.
- We validate our method on diverse object-hand scenarios using MultiDex dataset [7], outperforming baseline and ablated models in grasp success rate and stability.

II. RELATED WORK

Reinforcement learning (RL) has been increasingly adopted for dexterous grasping tasks due to its capacity to learn control policies without explicit modeling of dynamics. Works such as DAPG [10] and PPO-based approaches [11] have demonstrated success in learning high-DoF manipulation strategies. However, these methods often require extensive human demonstrations, collected using VR or motion tracking systems, which are costly and difficult to generalize. To mitigate the complexity of full-hand control, Jia et al. [12] proposed decomposing the hand into finger-level agents, each learning its own subtask. Their Visuo-Tactile Multi-Agent Grasping framework introduced a hierarchical structure for training the wrist, arm, and fingers separately via MADRL. However, their policy relied heavily on demonstration-free end-to-end learning and lacked diverse training grasps. Recent advancements in generative modeling have enabled grasp synthesis across varied hand types. [13], [14] proposed a diffusion-based grasp synthesis pipeline for multiple dexterous hands guided by affordance-aware discriminators. While their approach excels in generating diverse and functional grasp candidates, it focuses primarily on generation, not on closed-loop control or policy learning. In addition, integrating vision and touch is critical for robust grasping, especially under partial observability or environmental uncertainty. Prior methods have explored concatenation [15] or late fusion of sensor modalities. Our multi-stream attention network processes each modality with self-attention and then fuses them using cross-attention,

[1] are Ph.D. Student, Bert S. Turner Department of Construction Management, Louisiana State University, USA mbonya1@lsu.edu msoley1@lsu.edu

[2]Associate Professor and Graduate Program Advisor, Bert S. Turner Department of Construction Management, Louisiana State University, USA chaowang@lsu.edu

enhancing responsiveness to contact events and visual cues simultaneously.

we treat the high-quality grasp poses as static supervision for RL-based learning. This allows our system to benefit from the grasp diversity and quality without inheriting the computational complexity of generative models. Compared to earlier tactile-aware methods limited to two-finger grippers [16], our model supports variable dexterous hands and leverages agent-level fusion for more localized decision-making. This structure improves compliance, stability, and coordination across all DoF during grasp execution.

III. Method

Our objective is to achieve stable and generalizable dexterous manipulation using multi-agent deep reinforcement learning (MADRL), leveraging a shared dataset [7] of functional grasp poses. We propose a modular learning framework that models each component of a robotic hand (fingers, wrist, arm) as an independent agent, trained via MADRL, with a focus on multi-stream sensor fusion for enhanced robustness and adaptability.

A. Problem Formulation

Given an object represented by a 3D point cloud $\mathcal{O} \in R^{N\times 3}$, and a reference grasp pose $h^* = (t^*, \theta^*)$ sampled from a MultiDex dataset [7], where $t^* \in R^3$ is the target position and $\theta^* \in R^k$ is the joint configuration for a hand with k degrees of freedom, the goal is to train an agent-based control policy that achieves this grasp in a physically plausible, compliant, and robust manner. We define the system as a team of M agents, each controlling a subset of joints in the hand-arm system. The joint action at time t is $\mathbf{a}_t = [\mathbf{a}_t^1, \ldots, \mathbf{a}_t^M]$, with each $\mathbf{a}_t^i$ representing the torque control signal for agent i. Each agent $\mathcal{A}_i$ receives a local observation $\mathbf{s}_t^i$, and the overall state $\mathbf{s}_t = [\mathbf{s}_t^1, \ldots, \mathbf{s}_t^M]$ is used for centralized training.

B. Multi-Stream Embedding Fusion

To robustly perceive and react to condition of objects, each agent's policy network uses a multi-stream feature embedding. Let $\mathbf{v}_t^{i,j}$ denotes the features for agent i at time t in stream j. These are passed through separate self-attention modules to obtain unimodal features:

$$\phi_v^{i,j} = \text{SelfAttn}_v(\mathbf{v}_t^{i,j}), \tag{1}$$

$$\tag{2}$$

The fused representation ψ^i is produced via a cross-attention mechanism:

$$\psi^i = \text{CrossAttn}(\phi_v^i, \phi_u^i), \tag{3}$$

which is then passed through a fully connected network to output the action:

$$\mathbf{a}_t^i = \pi_{\theta_i}(\psi^i). \tag{4}$$

C. Reinforcement Learning with Centralized Critic

Training is performed using Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm [17]. Each agent's actor π_{θ_i} is updated to maximize the Q-value estimated by a shared critic $Q_i(\mathbf{s}_t, \mathbf{a}_t)$:

$$\nabla_{\theta_i} J(\theta_i) = E\left[\nabla_{\theta_i}\pi_{\theta_i}(\mathbf{s}_t^i)\nabla_{\mathbf{a}_t^i} Q_i(\mathbf{s}_t, \mathbf{a}_t)\right]. \tag{5}$$

The critic is updated using the Bellman target [18]:

$$y_t^i = r_t^i + \gamma Q_i(\mathbf{s}_{t+1}, \mathbf{a}_{t+1}), \tag{6}$$

and minimizing the loss:

$$\mathcal{L}_i = E\left[\left(Q_i(\mathbf{s}_t, \mathbf{a}_t) - y_t^i\right)^2\right]. \tag{7}$$

D. Reward Design

Each agent receives a dense reward tailored to its role in achieving a stable and functional grasp. For finger agents ($i = 1, \ldots, 5$), the reward encourages convergence to the reference fingertip positions:

$$r_t^i = -\|\mathbf{f}_i(t) - \mathbf{f}_i^*\|^2. \tag{8}$$

For arm and wrist agents, the reward promotes object lift and pose alignment:

$$r_t^j = -\lambda\|\mathbf{p}_{\text{palm}}(t) - \mathbf{p}_{\text{ref}}\|^2 + \mu h(t), \tag{9}$$

where $h(t)$ denotes the object's elevation and λ, μ are weighting factors.

E. Dataset

We utilize reference grasps from the MultiDex dataset [7], which provides diverse object-grasp pairs for multiple dexterous hands. These reference poses serve as static supervision signals for target configuration learning. Our framework integrates rich sensory fusion, agent-level specialization, and centralized training into a coherent architecture that enables robust dexterous grasping. By leveraging high-quality real-world-inspired grasp data, we demonstrate that our model can achieve generalizable, compliant, and functionally appropriate grasps across multiple hand configurations.

IV. Results and Discussion

We evaluate our proposed multi-agent reinforcement learning framework with multi-stream fusion on four dexterous robotic hands and multiple object categories. The primary objective is to assess generalization and control fidelity under variations in hand morphology and grasp complexity. Performance is measured by success rate (%), averaged over 10 test objects, using the static grasp pose from the MultiDex dataset as the goal configuration.

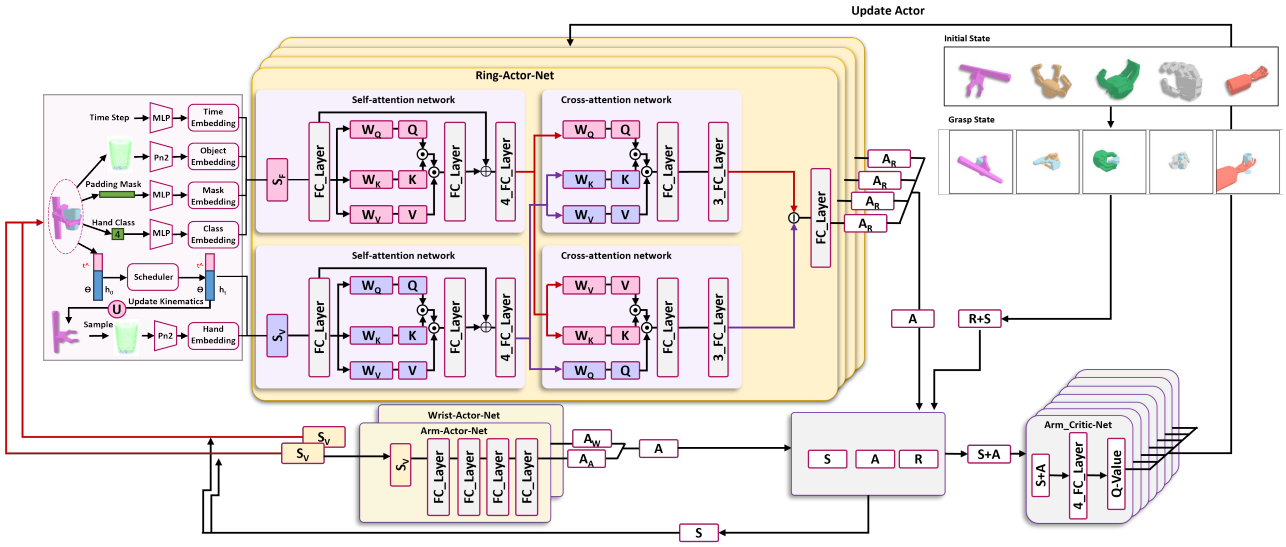

Fig. 1. An overview of the proposed method.

A. Baseline Comparisons

We compare our method with four learning-based baselines:

- **SAPG (Single-Agent Policy Gradient)**: A modified DAPG [10] trained using only static grasp references.
- **Single-Agent PPO**: An end-to-end control baseline using Proximal Policy Optimization with global observation [19].
- **SAC**: A single-policy variant using the Soft Actor-Critic algorithm, known for sample efficiency [20].
- **A2C**: Advantage Actor-Critic with discrete-time actor updates and centralized reward signals [21].

All models are trained with identical object-hand combinations, using the same reward shaping and static grasp target configuration as supervision.

TABLE I

SUCCESS RATE (%) COMPARISON ACROSS ROBOTIC HANDS AND CONTROL STRATEGIES.

Method	EZGripper	Barrett	Allegro	ShadowHand
A2C	21.7	15.2	22.4	33.6
SAC	26.9	18.5	27.1	42.8
Single-Agent PPO	29.8	20.6	31.4	50.3
SAPG (DAPG-style)	33.5	22.9	33.2	56.4
Ours (MADRL + Fusion)	**49.2**	**25.8**	**35.9**	**67.9**

B. Analysis by Hand Morphology

EZGripper (Low DoF): Our method achieves 49.2% success, outperforming the SAPG baseline by over 15%. This is significant for underactuated hands where fine control is limited. Multi-agent coordination allows individual fingers to adaptively adjust force distribution, while tactile fusion informs the agent about object displacement and resistance, a key advantage over model-free single-agent controllers.

Barrett (Symmetric Tri-finger): The gain here is more modest (25.8% vs. 22.9%), as the symmetric topology simplifies control. However, our method still excels on asymmetric objects like hammers or flashlights, where task-relevant force redirection is needed. Notably, SAC and A2C failed to generalize grasp patterns when object poses varied.

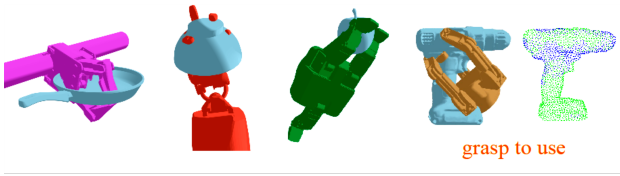

Fig. 2. An overview of failure cases wherein the model endeavors to grasp an object with two fingers while extending another finger toward the opposing end.

Allegro (Moderate DoF): Our MADRL system yields a 35.9% success rate. The gain over SAPG (+2.7%) and PPO (+4.5%) highlights the value of agent-level specialization. In several cases, the thumb and index fingers coordinated in power-wraps while other fingers stabilized the base. This behavior was rarely seen in flat policy baselines.

ShadowHand (High DoF): Here, the largest absolute gain is observed. Our method achieves 67.9% success versus 56.4% for SAPG and 50.3% for PPO. This shows that agent-level policy modularity and sensory fusion are essential when joint control complexity increases. With fusion, local tactile signals drive rapid reconfiguration after partial object contact, an ability not learned in vanilla policy gradients.

Also, for our qualitative results, Fig. 3 demonstrates the successful grasping sequences across various objects using different finger configurations, while Fig. 2 illustrates common failure cases where the system struggled with highly reflective surfaces and complex geometric features.

V. Ablation Analysis

To evaluate the contribution of each architectural component in our proposed framework, we conduct a detailed ablation study on the MultiDex dataset, focusing on the

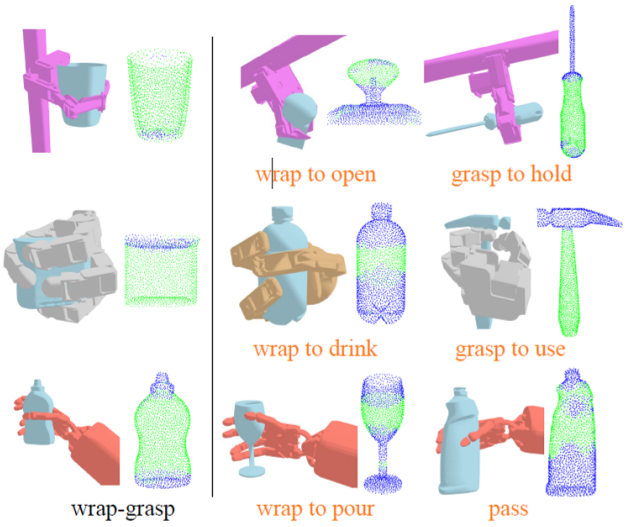

Fig. 3. An overview of qualitative result that unseen object is demonstrated with orange text.

ShadowHand due to its complex kinematic structure and high number of degrees of freedom (DoF). The following components are ablated individually while keeping all other parts fixed: (1) multi-agent policy design, (2) tactile and visual sensory inputs, and (3) attention-based fusion mechanisms. We report success rate, grasp diversity (as the standard deviation across successful joint configurations), and collision depth (mm) as our evaluation metrics.

TABLE II

ABLATION STUDY RESULTS USING THE SHADOWHAND.

Configuration	Success (%)	Diversity (rad)	Collision (mm)
Full model (Ours)	**67.9**	**0.228**	**15.8**
w/o Multi-Agent (single actor)	53.1	0.182	18.4
w/o Cross-Attention (early fusion)	57.5	0.191	18.0
w/o Self-Attention (MLP only)	50.4	0.176	19.5

A. *Effect of Multi-Agent Policy Decomposition*

Disabling the multi-agent structure and reverting to a single shared policy across all joints led to a **14.8% reduction** in grasp success and a noticeable drop in diversity. This performance degradation reflects the inability of a monolithic policy to coordinate localized control actions effectively. The decentralized design enables finer motion primitives at the joint level and allows specialization for different roles, wrist orientation control versus fingertip positioning, which are especially critical in high-DoF hands like the ShadowHand.

B. *Impact of Attention-Based Fusion Architecture*

When we replaced the cross-attention fusion module with early fusion (simple concatenation of visual and tactile features), performance dropped to 57.5% success. This demonstrates the limitations of naive integration strategies. Cross-attention allows the network to model interactions between modalities contextually, learning dependencies between visual cues (e.g., object geometry) and tactile feedback (e.g., contact force).

The most significant degradation occurred when both self-attention and cross-attention were removed, replaced with standard MLP layers. Success dropped to 50.4%, with the worst grasp diversity and the highest average collision depth. This highlights the role of attention in modeling spatial locality, contact semantics, and coordinated motion across fingers. Without attention, the network failed to assign appropriate importance to contact-rich regions, leading to aggressive or unbalanced grasps.

VI. CONCLUSION

In this work, we introduced a novel framework for dexterous robotic grasping that integrates multi-agent deep reinforcement learning with multi-stream fusion. Unlike prior approaches that rely on hand-specific generative models or demonstration-driven policy learning, our method uses static reference grasp data as supervision to train decentralized agents, each specialized for a subset of joints in the robotic hand. This modular design enables precise, compliant, and generalizable grasp execution across a diverse set of high-DoF robotic hands.

We demonstrated that our architecture significantly outperforms conventional single-agent reinforcement learning algorithms, including PPO, SAC, A2C, and a demonstration-free adaptation of DAPG,on both success rate and grasp diversity metrics. Furthermore, ablation analysis revealed the essential roles of multi-agent decomposition, dual-modality fusion, and attention-based encoders in achieving robust grasp performance and contact-safe behavior. Our method achieved high grasp success rates across object types and hand morphologies, with particularly strong results on anthropomorphic hands such as ShadowHand and Allegro. The combination of localized control and context-aware sensory fusion allowed the system to adapt to complex object geometries, unexpected contacts, and asymmetrical affordance regions.

For future work, we aim to extend the framework to in-the-wild robotic grasping tasks involving real sensor inputs and actuation noise. Additionally, integrating open-vocabulary affordance reasoning with online policy adaptation may further enhance functionality in task-oriented scenarios, such as tool use or human-object handovers. We believe our work takes an important step toward scalable, interpretable, and generalizable dexterous manipulation in unstructured environments.

REFERENCES

[1] N. Hudson, J. Ma, P. Hebert, A. Jain, M. Bajracharya, T. Allen, R. Sharan, M. Horowitz, C. Kuo, T. Howard, *et al.*, "Model-based autonomous system for performing dexterous, human-level manipulation tasks," *Autonomous Robots*, vol. 36, pp. 31–49, 2014.

[2] N. Elangovan, L. Gerez, G. Gao, and M. Liarokapis, "Improving robotic manipulation without sacrificing grasping efficiency: a multi-modal, adaptive gripper with reconfigurable finger bases," *IEEE Access*, vol. 9, pp. 83 298–83 308, 2021.

[3] W. Guo, W. Xu, Y. Zhao, X. Shi, X. Sheng, and X. Zhu, "Toward human-in-the-loop shared control for upper-limb prostheses: a systematic analysis of state-of-the-art technologies," *IEEE transactions on Medical Robotics and Bionics*, vol. 5, no. 3, pp. 563–579, 2023.

[4] S. An, Z. Meng, C. Tang, Y. Zhou, T. Liu, F. Ding, S. Zhang, Y. Mu, R. Song, W. Zhang, *et al.*, "Dexterous manipulation through imitation learning: A survey," *arXiv preprint arXiv:2504.03515*, 2025.

[5] D. Pathak, R. Girshick, P. Dollár, T. Darrell, and B. Hariharan, "Learning features by watching objects move," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2701–2710.

[6] H. Liu, D. Guo, and F. Sun, "Object recognition using tactile measurements: Kernel sparse coding methods," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 656–665, 2016.

[7] P. Li, T. Liu, Y. Li, Y. Geng, Y. Zhu, Y. Yang, and S. Huang, "Gendexgrasp: Generalizable dexterous grasping," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 8068–8074.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[9] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 603–612.

[10] 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.

[11] P. Mandikal and K. Grauman, "Dexvip: Learning dexterous grasping with human hand pose priors from video," in *Conference on Robot Learning*. PMLR, 2022, pp. 651–661.

[12] P. Jia, X. Li, T. Zhu, R. Wu, X. Lin, and Y. Sun, "Multi-fingered hand grasps with visuo-tactile fusion via multi-agent deep reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 14, 2025, pp. 14 594–14 601.

[13] X. Wu, T. Liu, C. Li, Y. Ma, Y. Shi, and X. He, "Fastgrasp: Efficient grasp synthesis with diffusion," *arXiv preprint arXiv:2411.14786*, 2024.

[14] Y. Zhang, Q. He, Y. Wan, Y. Zhang, X. Deng, C. Ma, and H. Wang, "Diffgrasp: Whole-body grasping synthesis guided by object motion using a diffusion model," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 10, 2025, pp. 10 320–10 328.

[15] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, E. H. Adelson, and S. Levine, "More than a feeling: Learning to grasp and regrasp using vision and touch," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3300–3307, 2018.

[16] J. Gao, Z. Huang, Z. Tang, H. Song, and W. Liang, "Visuo-tactile-based slip detection using a multi-scale temporal convolution network," *arXiv preprint arXiv:2302.13564*, 2023.

[17] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in neural information processing systems*, vol. 30, 2017.

[18] Y. Feng, L. Li, and Q. Liu, "A kernel loss for solving the bellman equation," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[19] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[20] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, *et al.*, "Soft actor-critic algorithms and applications," *arXiv preprint arXiv:1812.05905*, 2018.

[21] B. Kiumarsi and F. L. Lewis, "Actor–critic-based optimal tracking for partially unknown nonlinear discrete-time systems," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 1, pp. 140–151, 2014.