Low-Rank Modular Reinforcement Learning via Muscle Synergy

Heng Dong[1*], Tonghan Wang[2*], Jiayuan Liu[1], Chongjie Zhang[1]
[1]IIIS, Tsinghua University, [2]Harvard University

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

- **Background of Modular RL:** To control robots with different number of actuators, previous work proposed an elegant solution: Modular RL. In this learning paradigm, the control policy is decentralized, and each actuator is controlled by a shared local policy.
- **Limitations of Modular RL**: Despite the significant progress, Modular RL is still struggled with robots with many joints like Humanoids. The large degree of control freedom presents a major challenge for learning control policies.
- **Inspirations from Muscle Synergy**: A question is why humans can control hundreds of muscles without effort. In fact, human central nervous systems decrease the control complexity by activating muscles in groups, which is known as *muscle synergy*.
- **Incorporating Synergy into Modular RL:** We thus proposed our **Synergy-Oriented LeARning framework (SOLAR).**
- **Challenges**: We mainly faced two challenges (1) How to discover the synergy structures of different robots. (2) How to exploit the learned synergy structures.

Method

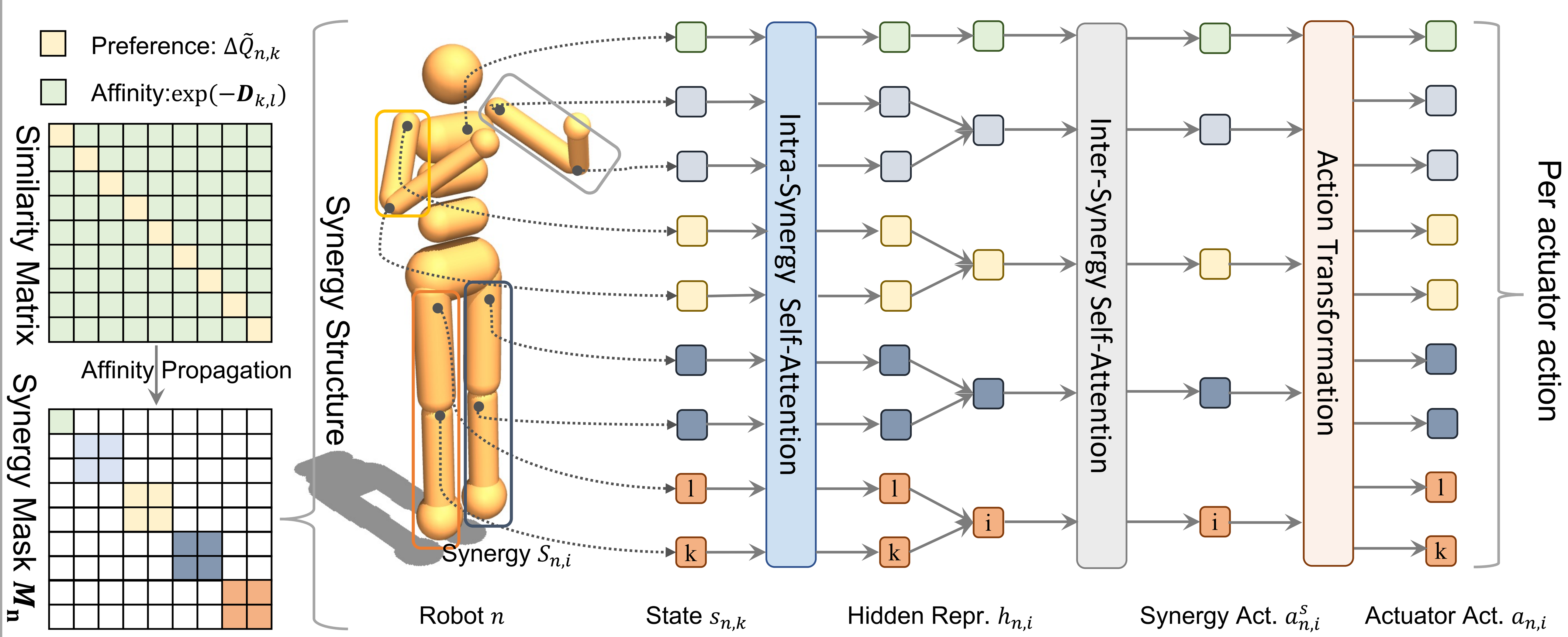
Our method has two major components. The first one is an unsupervised learning module that utilizes the morphological structure and value information to discover the synergy hierarchy. The second is a novel attention-based policy architecture that supports synergy-aware learning

Discovering Synergy Structures

1. **Input:** Similarity Matrix = non-diagonal elements (Affinity: morphological information) + diagonal elements (Preference: learning information).
2. **Process:** Affinity Propagation. A clustering algorithm that does not need to specify the number of clusters.
3. **Output:** Actuators in the same cluster are regarded as a synergy.

Learning Synergy-Aware Policies

1. **Intra-Synergy Self-Attention:** aggregates information of actuators in the same synergy. This module takes actuators' states as input and outputs a hidden representation for each synergy. In practice, we use a self-attention layer where the attentions between different synergies are masked out. We then use a mean pooling to aggregate information of actuators in the same synergy.
2. **Inter-Synergy Self-Attention**: aggregates information from different synergies and outputs the synergy actions.
3. **Action Transformation:** the synergy actions are combined linearly to generate physical actions for actuators. As the number of synergies is much fewer than actuators, the learning complexity is reduced. In this way, we learned a low-rank control policy.

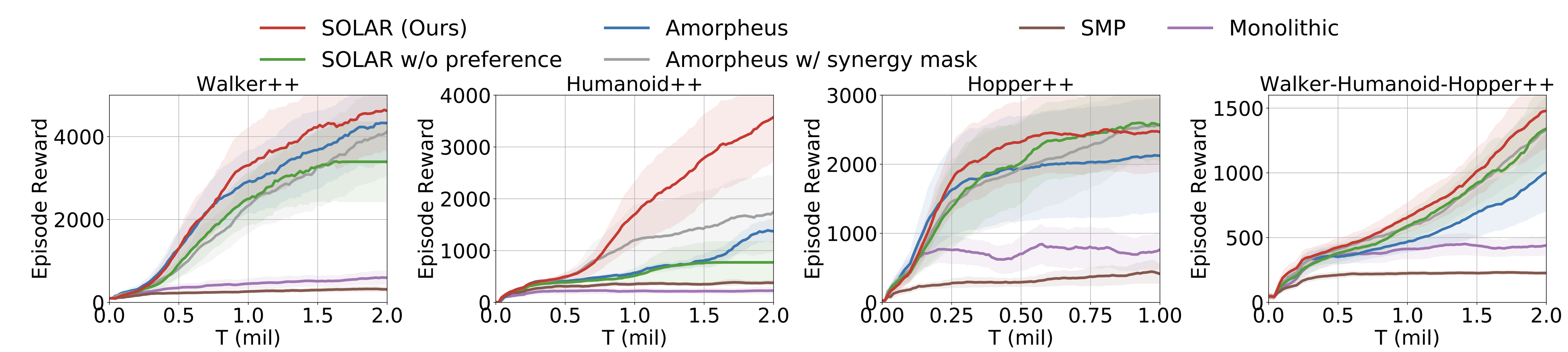


Synergy-Oriented Learning Framework

Results

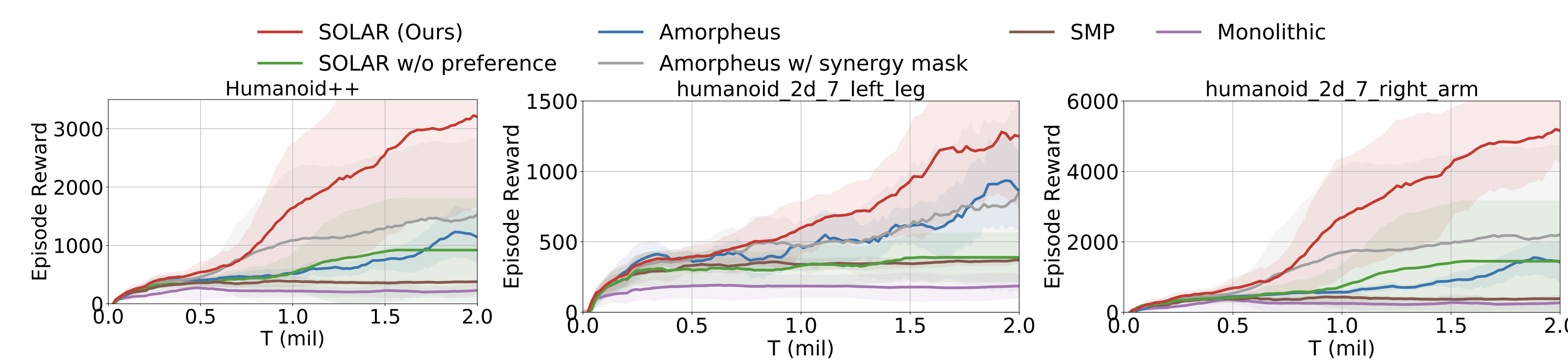
We benchmark our method SOLAR on various MuJoCo locomotion tasks.

Multi-task with different morphologies



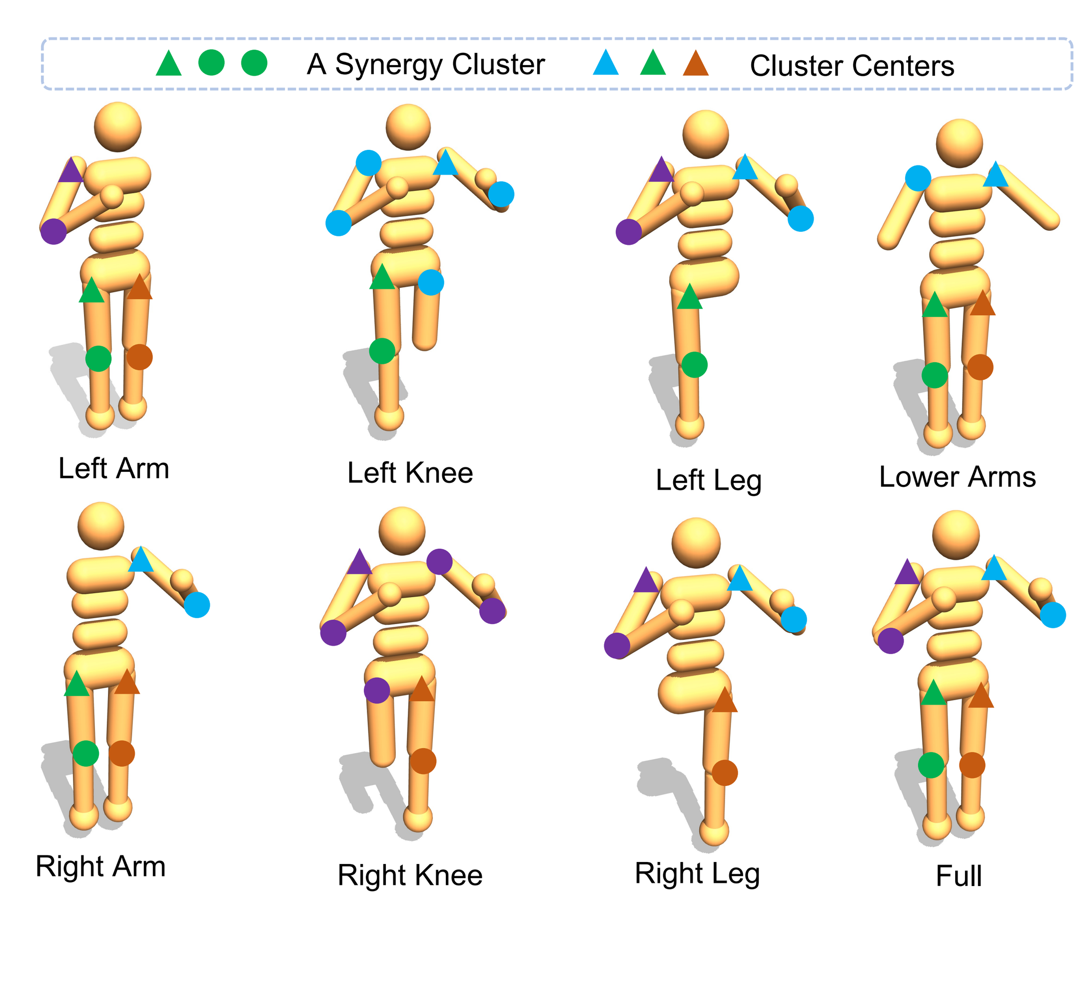
Multi-task performance of our method SOLAR compared to baselines and ablations

Zero-shot generalization

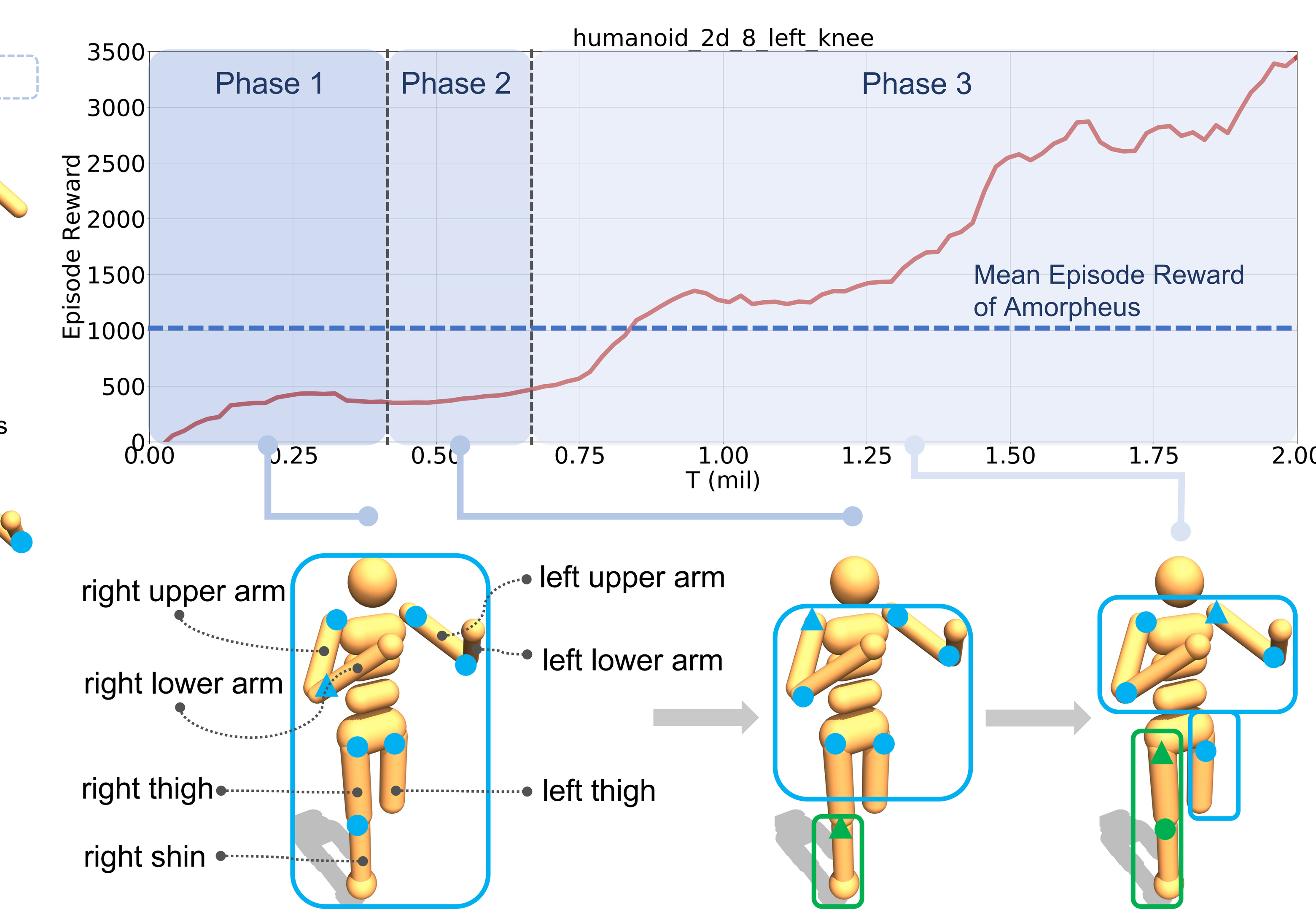


Zero-shot performance of our method SOLAR compared against baseline and ablations.

Analysis of synergies



Synergy clustering results of SOLAR in Humanoid++. Different colors represent different synergy clusters, and joints marked with the same color are in the same cluster. Joints marked with triangles are the centers of their corresponding clusters.



Synergy structure evolution of SOLAR in Humanoid. Phase are divided according to the change of synergy clusters and synergy clusters are masked with colored shapes.