ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

ManiMorph: Object Representations in Robot Manipulators Morphology for Improving Multi-Task Manipulation Performance

Ali Midhat[1], Michael Przystupa[2], Xinrui Zu[2], Kevin Sebastian Luck[2], Glen Berseth[3]

***Abstract*— Robot manipulation tasks involve direct interactions with objects, which can be viewed as dynamic changes to the robot's kinematic chain. Morphology-aware learning frameworks, in which robot embodiment is explicitly modeled, do not account for these object-induced changes in their architectures. We address this gap by proposing ManiMorph, a multi-task, morphology-aware manipulation-learning framework in which object features are integrated into the robot's morphological graph. We demonstrate that this node-centric representation, combined with a Feature-wise Linear Modulation (FiLM) task component, enhances the performance of the morphology-aware frameworks for robotic manipulation and generalizes effectively to new object variations.**

I. Introduction

A robot's morphology plays a significant role in the development of machine-learning control models for manipulation tasks. Morphological variations can drastically affect the reusability of control systems; even benign variations can reduce model performance [1]. Robot manipulators often overlap in topological structure, but their morphologies dynamically change due to object interactions [2], [3]; the target object often behaves as an unactuated additional node in the kinematic chain [4]. We study the consequences of treating these object interactions as dynamically changing aspects of a robot's morphology.

This paper addresses this question through *morphology-aware* policies, where the robot's embodiment is explicitly modeled as a directed acyclic graph (DAG). Such policies can be robust under minor embodiment variations and reduce additional data collection when transferring the model to a new robot [5]. Most prior morphology-aware work focuses on locomotion [6], [7] and ignores the effects that holding an object has on the morphological structure, which is central to our research.

Our primary contribution is the **ManiMorph framework**, a *multi-task morphology-aware manipulation learning framework*, supported by three key claims:

(1) Representing objects as nodes in the robot-morphology graph helps policies achieve better manipulation task performance compared to treating objects as an external sensor input [5], [8].

(2) FiLM layers [9] can modulate a morphology-aware policy's features to switch between tasks more easily than providing task information as direct model input.

(3) ManiMorph surpasses alternative frameworks in cumulative reward and generalizes well to variations in target objects, demonstrated on the Robosuite benchmark [10].

[1]African Institute for Mathematical Sciences. [2]Vrije Universiteit Amsterdam. [3]University of Montréal & Mila. Contact: m.v.przystupa@vu.nl

II. Morphology-Aware Manipulation Model

a) Modeling manipulators as controllable graphs: Each robot arm is represented as a DAG with three node types: base, joint(s), and gripper. Node features are extracted by depth-first traversal and fed to a Transformer as a sequence. This node-centric representation enables switching between joint-space (JNT) and operational-space (OSC) control through a per-node action masking scheme, without architectural changes.

b) Integrating object representations into the morphology graph: Manipulation tasks involve direct contact with an object, which can be viewed as adding a new link to the serial chain. We model the object as an additional non-actuated node appended to the morphology graph: $\mathbf{s} = [\mathbf{s}^1, \ldots, \mathbf{s}^{l(c^m)}; \mathbf{o}]$, where $\mathbf{o} \in \mathcal{O}$ is projected into the same embedding space as limb features. Each robot node can then attend directly to the target object via the Transformer attention mechanism, enabling bi-directional interaction between robot morphology and the manipulation target.

c) Multi-task learning via hypernetworks and FiLM: ManiMorph decouples morphology and task interactions through a global FiLM layer that conditions on the average morphology graph embedding $\bar{\mathbf{c}}^{c_m}$ and a task embedding $\mathbf{e}_t$:

$$\gamma, \beta = HN_{\text{task}}(\bar{\mathbf{c}}, \mathbf{e}_t), \qquad \hat{h} = \gamma \odot h_{\text{transformer}} + \beta. \quad (1)$$

This uniformly adapts the entire morphological representation to the current task while preserving morphology-specific features learned by the Transformer, providing a framework for multi-robot multi-task learning.

III. Experiments

We evaluate ManiMorph on Robosuite [10] using four robot morphologies (Jaco, Kuka, Gen3, Sawyer) across three tasks: Lift, Door, and Wipe, trained with PPO. We compare against MetaMorph [5] and ModuMorph [8], both augmented with our object-as-node representation to isolate the contribution of the FiLM task adapter.

a) Node-centric representation improves performance.: The object-as-node representation surpasses the specialized single-robot upper bound on both Lift and Door tasks. In contrast, the late-fusion baseline — where object information is concatenated to the policy decoder's MLP inputs — fails

TABLE I

SUCCESS RATE (%) IN THE MULTI-ROBOT MULTI-TASK SETTING ACROSS OPERATIONAL-SPACE (OSC) AND JOINT-SPACE (JNT) CONTROL.

	OSC			**JNT**		
Method	Lift	Door	Wipe	Lift	Door	Wipe
ManiMorph (ours)	56.1±28.0	**93.9±2.4**	**70.2±1.3**	55.6±27.6	95.0±1.4	**48.4±8.0**
All Nodes	40.1±27.9	92.1±1.0	69.6±4.2	55.9±28.1	86.2±15.6	47.5±1.2
MetaMorph	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
ModuMorph	**72.1±0.2**	78.0±29.5	41.9±5.5	**71.3±2.1**	**95.9±0.6**	25.8±3.1

to reach this benchmark on Lift and only matches it on Door after significantly more training. These results suggest that allowing controllable limbs to attend to the object directly in the Transformer improves sample efficiency for manipulation tasks.

b) Action space affects morphology-aware learning.: Table I reports success rates across both control spaces. In OSC, ManiMorph achieves the best Door (93.9%) and Wipe (70.2%) results, while ModuMorph leads on Lift. MetaMorph largely fails in the multi-robot multi-task setting across both control spaces. In JNT, ModuMorph is competitive on Lift and Door, but ManiMorph outperforms it substantially on Wipe (48.4% vs. 25.8%), suggesting the FiLM task adapter is particularly beneficial for tasks requiring sustained contact control.

c) Generalization to new object types: To evaluate zero-shot generalization, we test on unseen object geometries (Ball and Cylinder) with varied density, friction, and scale in the Lift task. ManiMorph and its All Nodes variant successfully complete the task with novel objects; MetaMorph and ModuMorph fail across all object variations in both control spaces. ManiMorph maintains performance even when object dimensions are scaled to 200% of the original size, suggesting HN_{task} modulates the policy flexibly rather than memorizing specific object dimensions. Notably, ModuMorph achieves stronger training performance yet fails to generalize, indicating susceptibility to overfitting.

IV. CONCLUSIONS AND FUTURE WORK

We demonstrate that integrating objects as nodes in the manipulator morphology graph, combined with FiLM-based task modulation, yields a more robust morphology-aware manipulation policy. Results are particularly strong for out-of-distribution object modifications, a generalisation capability not previously demonstrated with morphology-aware models. Future work will explore: (1) how to adapt pre-trained ManiMorph models to new morphologies, and (2) the role of the Transformer architecture relative to other components of the framework.

V. POSTER SUMMARY

The accompanying poster presents the ManiMorph framework across four sections. The **Introduction** motivates the treatment of the object interactions as dynamic changes to robot morphology, extending cross-embodiment learning, previously explored in locomotion, to manipulation. The **Methodology** section illustrates how each robot and its target object are jointly represented as a graph, and describe the Transformer-based architecture with hypernet encoder/decoder, limb cross-attention, and FiLM task modulation. The **Experiments** section compares ManiMorph against MetaMorph and ModuMorph on four robots and four tasks, showing learning curves for Lift and Door and success-rate breakdowns by action space. The **Object Generalisation** panel demonstrates zero-shot transfer to unseen Ball and Cylinder geometries with varied physical properties, where ManiMorph and its All Nodes variant succeed while both baselines fail.

REFERENCES

[1] T. Chen, A. Murali, and A. Gupta, "Hardware conditioned policies for multi-robot transfer learning," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[2] Z. Wu, W. Lian, V. Unhelkar, M. Tomizuka, and S. Schaal, "Learning dense rewards for contact-rich manipulation tasks," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6214–6221.

[3] Í. Elguea-Aguinaco, A. Serrano-Muñoz, D. Chrysostomou, I. Inziarte-Hidalgo, S. Bøgh, and N. Arana-Arexolaleiba, "A review on reinforcement learning for contact-rich robotic manipulation tasks," *Robotics and Computer-Integrated Manufacturing*, vol. 81, p. 102517, 2023.

[4] Z. Liu, S. Tian, M. Guo, K. Liu, and J. Wu, "Learning to design and use tools for robotic manipulation," in *Conference on Robot Learning*. PMLR, 2023, pp. 887–905.

[5] A. Gupta, L. Fan, S. Ganguli, and L. Fei-Fei, "Metamorph: learning universal controllers with transformers," in *International Conference on Learning Representations*. ICLR, 2022.

[6] N. Bohlinger, G. Czechmanowski, M. P. Krupka, P. Kicki, K. Walas, J. Peters, and D. Tateo, "One policy to run them all: an end-to-end learning approach to multi-embodiment locomotion," in *Conference on Robot Learning*. PMLR, 2025, pp. 3356–3378.

[7] M. Shafiee, G. Bellegarda, and A. Ijspeert, "Manyquadrupeds: Learning a single locomotion policy for diverse quadruped robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 3471–3477.

[8] Z. Xiong, J. Beck, and S. Whiteson, "Universal morphology control via contextual modulation," in *International Conference on Machine Learning*. PMLR, 2023, pp. 38 286–38 300.

[9] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[10] Y. Zhu, J. Wong, A. Mandlekar, and R. Martín-Martín, "robosuite: A modular simulation framework and benchmark for robot learning," *CoRR*, vol. abs/2009.12293, 2020. [Online]. Available: https://arxiv.org/abs/2009.12293