Physics-informed digital twin from video
for task and motion planning of deformable objects

Introduction:

Achieving proficient robotic manipulation of deformable objects, such as ropes, textiles, and other flexible materials, remains a significant challenge in the robotics community. While recent advances in vision-language-action models [1] demonstrated remarkable capabilities in learning robust, long-horizon behaviors from demonstrations, they are extremely data-hungry and struggle with out-of-distribution tasks. Model-based task and motion planners, on the other hand, can generalize better to unseen tasks because of explicit reasoning at both task and motion levels.

Effective task and motion planning (TAMP) for deformable objects heavily depends on accurate physics simulations to predict how objects respond to robot control commands. However, traditional model-based physics simulations often encounter significant sim-to-real gaps due to unmodeled or difficult-to-simulate physical behaviors and inaccuracies in system identification. Recent advancements in physics-informed digital twins, such as the `PhysTwin` framework proposed by [2], present a promising approach by enabling robots to quickly and autonomously construct realistic and interactive simulations of deformable objects from minimal video data. This method holds great potential to bridge the sim-to-real gap inherent in robotics planning.

Our goal is to evaluate the potential of integrating `PhysTwin` into an optimization-based TAMP framework developed by the mentor. The TAMP framework plans robot control commands to manipulate a rope on a table from any starting configuration to a user-specified target configuration, and co-optimizes a sequence of inputs specifying (1) where on the rope to grasp and (2) how to manipulate after grasp is engaged (velocity of the end-effector motion). The current TAMP framework employs a model-based physics simulation to formulate system evolution rules as constraints. Our goal is to replace this with the `PhysTwin`-learned model (fig. 1).

If successful, this implementation will set the stage for future work involving "plan-to-learn" strategies [3], where robots iteratively refine their "intuitive physics engine" [4] of a target object through autonomously planned, interactive exploration of unknown deformable objects.
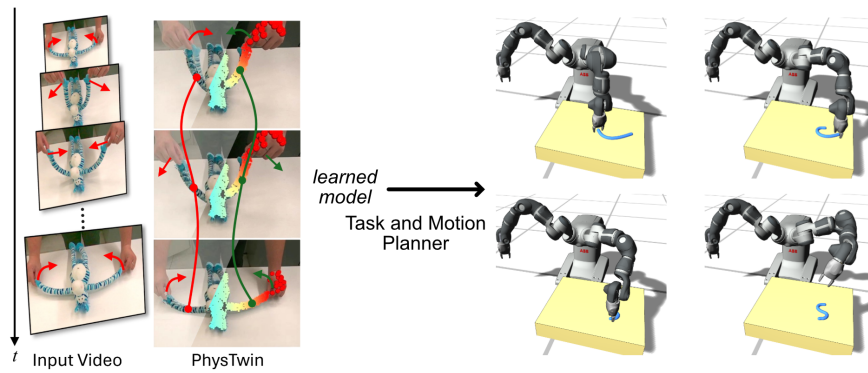

Figure 1: Left: Learning a physics simulation model from interaction videos using PhysTwin [2]. Right: Integration of the learned PhysTwin model into a model-based task and motion planner to enable robotic manipulation of a deformable object (a rope).

Tasks:

- Re-implement `PhysTwin` using their [publicly available codebase](#) and evaluate its performance and limitations.

- Integrate the `PhysTwin`-trained model into a TAMP framework to plan long-horizon manipulations for rope-like deformable objects, validated through simulations and testing on a real robot manipulator system.

- Document guidelines and insights regarding the practical utility and constraints of physics-informed digital twins in robotic manipulation tasks.

- Prepare a thorough written report.

- Present findings and contributions formally.

Requirements:

- Strong background in optimization techniques and numerical methods.

- Proficiency in Python (`PhysTwin` validation) and experience with Julia or C++ (TAMP integration).

- Basic knowledge in physics simulation and trajectory optimization. Students who have taken the *Physically-Based Simulation in Computer Graphics* and/or *Computational Models of Motion* courses at ETH are preferred but not required.

- Self-driven attitude and persistence to read and experiment with existing cutting-edge research code, and the ability to independently explore and solve problems.

Remarks:

This master's thesis project is limited to ETH Zurich students. The project is overseen by Prof. Dr. Stelian Coros and supervised by Dr. Yijiang Huang and Jimmy Envall. For further information or to apply for the thesis project, please send your CV, transcript, and a short motivation letter to: Yijiang (yijiang.huang@inf.ethz.ch), and Jimmy (jimmy.envall@inf.ethz.ch).

References

[1] Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. π_0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

[2] Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, and Yunzhu Li. PhysTwin: Physics-Informed Reconstruction and Simulation of Deformable Objects from Videos, March 2025.

[3] Nishanth Kumar, Tom Silver, Willie McClinton, Linfeng Zhao, Stephen Proulx, Tomás Lozano-Pérez, Leslie Pack Kaelbling, and Jennifer Barry. Practice Makes Perfect: Planning to Learn Skill Parameter Policies, February 2024.

[4] Peter W Battaglia, Jessica B Hamrick, and Joshua B Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332, 2013.