

New framework allows robots to learn via online human demonstration videos

July 19 2024, by Ingrid Fadelli



Our proposed TieBot performs a tie-knotting task. We leverage cloth simulation to recover the tie's motion from human demonstration and learn a goal-condition policy to accomplish the tie-knotting task. Credit: Peng et al.

To be successfully deployed in real-world settings, robots should be capable of reliably completing various everyday tasks, ranging from

household chores to industrial processes. Some of the tasks they could complete entail manipulating fabrics, for instance when folding clothes to put them in a wardrobe or helping older adults with mobility impairments to knot their ties before a social event.

Developing robots that can effectively tackle these tasks has so far proved fairly challenging. Many proposed approaches to train robots on fabric manipulation tasks rely on imitation learning, a technique to train robot control using videos, motion capture footage, and other data of humans completing the tasks of interest.

While some of these techniques achieved encouraging results, to perform well they typically require substantial amounts of human demonstration data. This data can be expensive and difficult to collect, while existing open-source datasets do not always contain as much data as those for training other computational techniques, such as computer vision or generative AI models.

Researchers at National University of Singapore, Shanghai Jiao Tong University, and Nanjing University recently introduced an alternative approach that could enhance and simplify the training of robotics algorithms via human demonstrations. This approach, outlined in a [paper](#) pre-published on *arXiv*, is designed to leverage some of the many videos posted online every day, utilizing them as human demonstrations of everyday tasks.

"This work begins with a simple idea, that of building a system that allows robots to utilize the countless human demonstration videos online to learn complex manipulation skills," Weikun Peng, co-author of the paper, told Tech Xplore. "In other words, given an arbitrary human demonstration video, we wanted the robot to complete the same task shown in the video."

While previous studies also introduced imitation learning techniques that leveraged video footage, they utilized domain-specific videos (i.e., videos of humans completing specific tasks in the same environment in which the robot would be later be tackling the task), as opposed to arbitrary videos collected in any environment or setting.

The framework developed by Peng and his colleagues, on the other hand, is designed to enable robot imitation learning from arbitrary demonstration videos found online.

The team's approach has three primary components, dubbed Real2Sim, Learn@Sim and Sim2Real. The first of these components is the central and most important part of the framework.

"Real2Sim tracks the object's motion in the demonstration video and replicates the same motion on a mesh model in a simulation," Peng explained. "In other words, we try to replicate the human demonstration in the simulation. Finally, we get a sequence of object meshes, representing the ground truth object trajectory."

The researchers' approach utilizes meshes (i.e., accurate digital representations of an object's geometry, shape and dynamics) as intermediate representations. After the Real2Sim component replicates a human demonstration in a simulated environment, the framework's second component, dubbed Learn@Sim, learns the grasping points and placing points that would allow a robot to perform the same actions via reinforcement learning.

"After learning grasping points and placing points in the simulation, we deployed the policy to a real dual-arm robot, which is our pipeline's third step (i.e., Sim2Real)," Peng said. "We trained a residual policy to mitigate the Sim2Real gap."

The researchers evaluated their proposed approach in a series of tests, specifically focusing on the task of knotting a tie. While this [task](#) can be extremely difficult for robots, the team's approach allowed a robotic manipulator to successfully complete it.

"Notably, many previous works require 'in domain' demonstration videos, which means the setting of demonstration videos should be the same as the setting of the robot execution environment," Peng said. "Our method, on the other hand, can learn from 'out of domain' demonstration videos since we extract the object's motion in 3D space from the demonstration video."

In the future, the new approach introduced by Peng and his colleagues could be applied to other complex and challenging [robot](#) manipulation tasks. Ultimately, it could facilitate the training of robots via imitation learning, potentially enabling new advancements in their skills.

"My plan for future work would be to expand the Real-Sim-Real idea to other tasks," Peng added.

"If we can replicate an object's motion in simulation, could we replicate the real world in simulation? The robotics community is facing a data scarcity problem, and in my opinion, if we can replicate the real world in simulation, we can collect data more efficiently and better transfer learned policy to real robots."

More information: Weikun Peng et al, TieBot: Learning to Knot a Tie from Visual Demonstration through a Real-to-Sim-to-Real Approach, *arXiv* (2024). [DOI: 10.48550/arxiv.2407.03245](https://doi.org/10.48550/arxiv.2407.03245)

Citation: New framework allows robots to learn via online human demonstration videos (2024, July 19) retrieved 19 July 2024 from <https://techxplore.com/news/2024-07-framework-robots-online-human-videos.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.