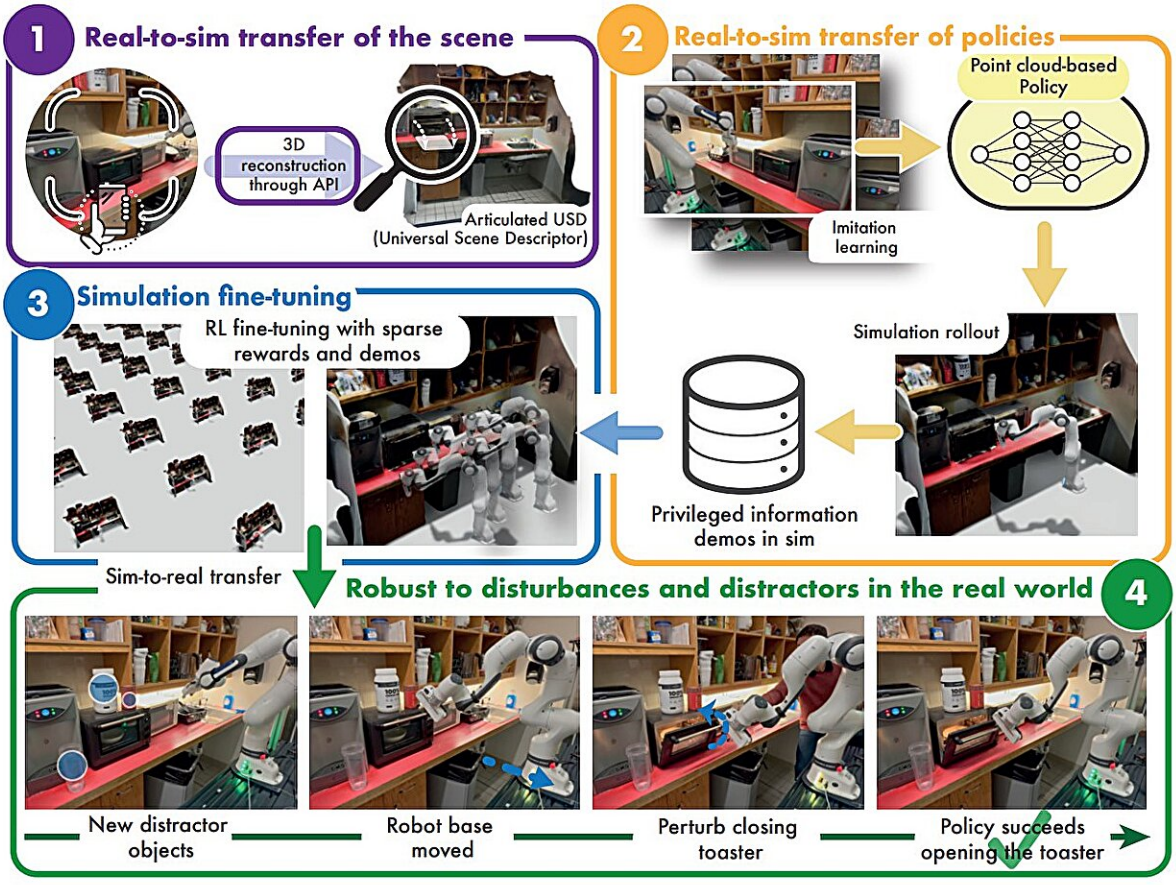
Using photos or videos, these AI systems can conjure simulations that train robots to function in physical spaces

August 7 2024, by Stefan Milne



RialTo system overview. 1) Transfer the real-world scene to the simulator through an easy-to-use API (see Section III-B). 2) Transfer a policy learned from real-world demonstrations to collect a set of demonstrations with privileged information in simulation. We note this step is optional, and RialTo is

compatible with skipping this step and providing demonstrations in simulation (see Section IV-C2) 3) Use the collected set of demonstrations to bias exploration in the RL fine-tuning with sparse rewards of a state-based policy (see Section III-C) 4) Perform teacher-student distillation and deploy the policy in the real world obtaining robust behaviors (see Section III-D). Credit: Torne et al.

Researchers working on large artificial intelligence models like ChatGPT have vast swaths of internet text, photos and videos to train systems. But roboticists training physical machines face barriers: Robot data is expensive, and because there aren't fleets of robots roaming the world at large, there simply isn't enough data easily available to make them perform well in dynamic environments, such as people's homes.

Some researchers have turned to simulations to train robots. Yet even that process, which often involves a graphic designer or engineer, is laborious and costly.

Two new studies from University of Washington researchers introduce AI systems that use either video or photos to create simulations that can train robots to function in real settings. This could significantly lower the costs of training robots to function in complex settings.

In the first study, a user quickly scans a space with a smartphone to record its geometry. The system, called RialTo, can then create a "digital twin" simulation of the space, where the user can enter how different things function (opening a drawer, for instance).

A robot can then virtually repeat motions in the simulation with slight variations to learn to do them effectively. In the second study, the team built a system called URDFormer, which takes images of real environments from the internet and quickly creates physically realistic

simulation environments where robots can train.

The teams presented their studies—the first on July 16 and the second on July 19—at the Robotics Science and Systems conference in Delft, Netherlands.

"We're trying to enable systems that cheaply go from the real world to simulation," said Abhishek Gupta, a UW assistant professor in the Paul G. Allen School of Computer Science & Engineering and co-senior author on both papers.

"The systems can then train robots in those simulation scenes, so the robot can function more effectively in a physical space. That's useful for safety—you can't have poorly trained robots breaking things and hurting people—and it potentially widens access. If you can get a robot to work in your house just by scanning it with your phone, that democratizes the technology."

While many robots are currently well suited to working in environments like assembly lines, teaching them to interact with people and in less structured environments remains a challenge.

"In a factory, for example, there's a ton of repetition," said lead author of the URDFormer study Zoey Chen, a UW doctoral student in the Allen School. "The tasks might be hard to do, but once you program a robot, it can keep doing the task over and over and over. Whereas homes are unique and constantly changing. There's a diversity of objects, of tasks, of floorplans and of people moving through them. This is where AI becomes really useful to roboticists."

The two systems approach these challenges in different ways.

RialTo—which Gupta created with a team at the Massachusetts Institute

of Technology—has someone pass through an environment and take video of its geometry and moving parts. For instance, in a kitchen, they'll open cabinets and the toaster and the fridge.

The system then uses existing AI models—and a human does some quick work through a graphic user interface to show how things move—to create a simulated version of the kitchen shown in the video. A virtual robot trains itself through trial and error in the simulated environment by repeatedly attempting tasks such as opening that toaster oven—a method called reinforcement learning.

By going through this process in the simulation, the robot improves at that task and works around disturbances or changes in the environment, such as a mug placed beside the toaster. The robot can then transfer that learning to the physical environment, where it's nearly as accurate as a robot trained in the real kitchen.

The other system, URDFormer, is focused less on relatively high accuracy in a single kitchen; instead, it quickly and cheaply conjures hundreds of generic kitchen simulations. URDFormer scans images from the internet and pairs them with existing models of how, for instance, those kitchen drawers and cabinets will likely move.

It then predicts a simulation from the initial real-world image, allowing researchers to quickly and inexpensively train robots in a huge range of environments. The trade-off is that these simulations are significantly less accurate than those that RialTo generates.

"The two approaches can complement each other," Gupta said. "URDFormer is really useful for pre-training on hundreds of scenarios. RialTo is particularly useful if you've already pre-trained a robot, and now you want to deploy it in someone's home and have it be maybe 95% successful."

Moving forward, the RialTo team wants to deploy its system in peoples' homes (it's largely been tested in a lab), and Gupta said he wants to incorporate small amounts of real-world training data with the systems to improve their success rates.

"Hopefully, just a tiny amount of real-world data can fix the failures," Gupta said. "But we still have to figure out how best to combine data collected directly in the real world, which is expensive, with data collected in simulations, which is cheap, but slightly wrong."

On the URDFormer paper additional co-authors include the UW's Aaron Walsman, Marius Memmel, Alex Fang—all doctoral students in the Allen School; Karthikeya Vemuri, an undergraduate in the Allen School; Alan Wu, a masters student in the Allen School; and Kaichun Mo, a research scientist at NVIDIA. Dieter Fox, a professor in the Allen School, was a co-senior author.

On the URDFormer paper additional co-authors include MIT's Marcel Torne, Anthony Simeonov, Tao Chen—all doctoral students; Zechu Li, a research assistant; and April Chan, an undergraduate. Pulkit Agrawal, an assistant professor at MIT, was a co-senior author. The URDFormer research was partially funded by Amazon Science Hub.

More information: Torne et al. Reconciling Reality through Simulation: A Real-to-Sim-to-Real Approach for Robust Manipulation, enriquecoronadozu.github.io/rs … s2024/rss20/p015.pdf

Chen et al. URDFormer: A Pipeline for Constructing Articulated Simulation Environments from Real-World Images, enriquecoronadozu.github.io/rs … s2024/rss20/p124.pdf

Provided by University of Washington