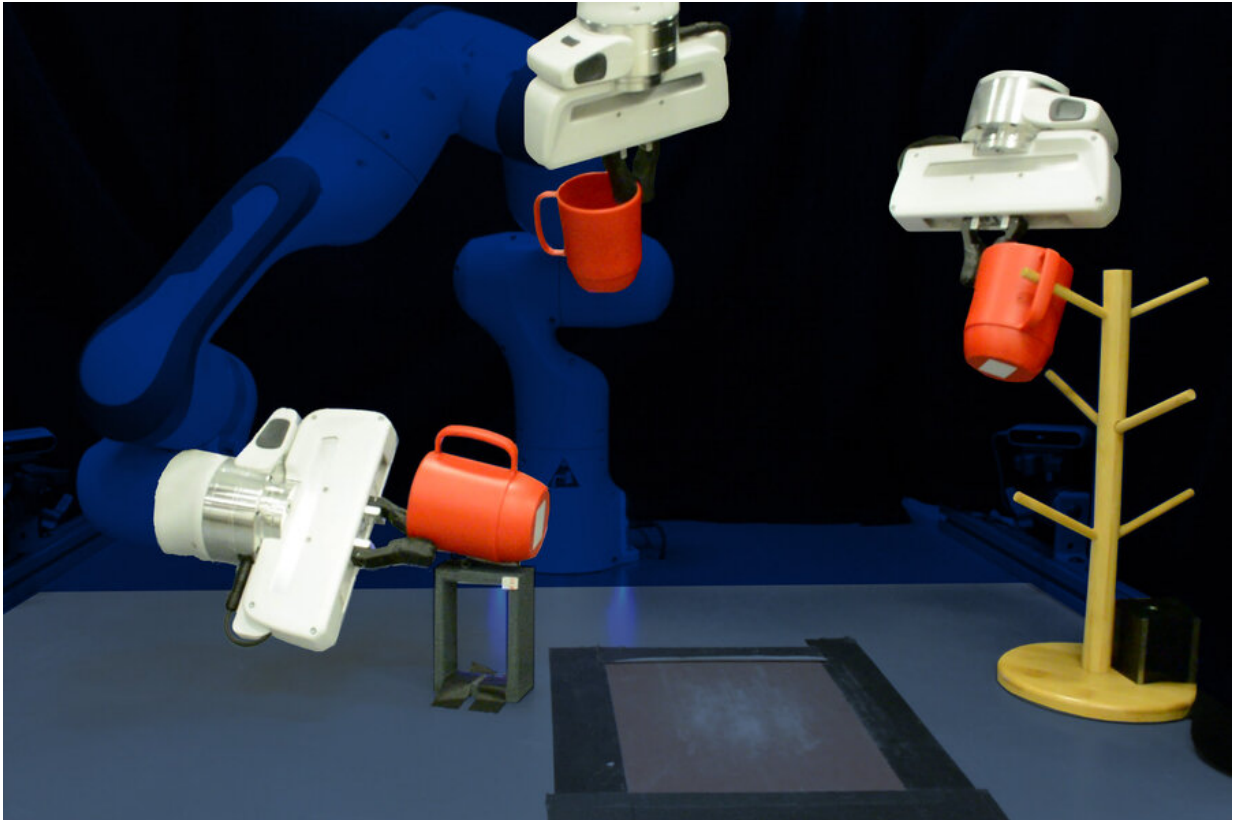


An easier way to teach robots new skills

April 25 2022, by Adam Zewe



MIT researchers have developed a system that enables a robot to learn a new pick-and-place task based on only a handful of human examples. This could allow a human to reprogram a robot to grasp never-before-seen objects, presented in random poses, in about 15 minutes. Credit: Massachusetts Institute of Technology

With e-commerce orders pouring in, a warehouse robot picks mugs off a

shelf and places them into boxes for shipping. Everything is humming along, until the warehouse processes a change and the robot must now grasp taller, narrower mugs that are stored upside down.

Reprogramming that [robot](#) involves hand-labeling thousands of images that show it how to grasp these new mugs, then training the system all over again.

But a new technique developed by MIT researchers would require only a handful of human demonstrations to reprogram the robot. This machine-learning method enables a robot to pick up and place never-before-seen objects that are in random poses it has never encountered. Within 10 to 15 minutes, the robot would be ready to perform a new pick-and-place task.

The technique uses a [neural network](#) specifically designed to reconstruct the shapes of 3D objects. With just a few demonstrations, the system uses what the neural network has learned about 3D geometry to grasp new objects that are similar to those in the demos.

In simulations and using a real robotic arm, the researchers show that their system can effectively manipulate never-before-seen mugs, bowls, and bottles, arranged in random poses, using only 10 demonstrations to teach the robot.

"Our major contribution is the general ability to much more efficiently provide new skills to robots that need to operate in more unstructured environments where there could be a lot of variability. The concept of generalization by construction is a fascinating capability because this problem is typically so much harder," says Anthony Simeonov, a graduate student in [electrical engineering](#) and computer science (EECS) and co-lead author of the paper.

Simeonov wrote the paper with co-lead author Yilun Du, an EECS graduate student; Andrea Tagliasacchi, a staff research scientist at Google Brain; Joshua B. Tenenbaum, the Paul E. Newton Career Development Professor of Cognitive Science and Computation in the Department of Brain and Cognitive Sciences and a member of the Computer Science and Artificial Intelligence Laboratory (CSAIL); Alberto Rodriguez, the Class of 1957 Associate Professor in the Department of Mechanical Engineering; and senior authors Pulkit Agrawal, a professor in CSAIL, and Vincent Sitzmann, an incoming assistant professor in EECS. The research will be presented at the International Conference on Robotics and Automation.

Grasping geometry

A robot may be trained to pick up a specific item, but if that object is lying on its side (perhaps it fell over), the robot sees this as a completely new scenario. This is one reason it is so hard for machine-learning systems to generalize to new object orientations.

To overcome this challenge, the researchers created a new type of neural network model, a Neural Descriptor Field (NDF), that learns the 3D geometry of a class of items. The model computes the geometric representation for a specific item using a 3D point cloud, which is a set of data points or coordinates in three dimensions. The data points can be obtained from a depth camera that provides information on the distance between the object and a viewpoint. While the network was trained in simulation on a large dataset of synthetic 3D shapes, it can be directly applied to objects in the real world.

The team designed the NDF with a property known as equivariance. With this property, if the model is shown an image of an upright mug, and then shown an image of the same mug on its side, it understands that the second mug is the same object, just rotated.

"This equivariance is what allows us to much more effectively handle cases where the object you observe is in some arbitrary orientation," Simeonov says.

As the NDF learns to reconstruct shapes of similar objects, it also learns to associate related parts of those objects. For instance, it learns that the handles of mugs are similar, even if some mugs are taller or wider than others, or have smaller or longer handles.

"If you wanted to do this with another approach, you'd have to hand-label all the parts. Instead, our approach automatically discovers these parts from the shape reconstruction," Du says.

The researchers use this trained NDF model to teach a robot a new skill with only a few physical examples. They move the hand of the robot onto the part of an object they want it to grip, like the rim of a bowl or the handle of a mug, and record the locations of the fingertips.

Because the NDF has learned so much about 3D geometry and how to reconstruct shapes, it can infer the structure of a new shape, which enables the system to transfer the demonstrations to new objects in arbitrary poses, Du explains.

Picking a winner

They tested their model in simulations and on a real robotic arm using mugs, bowls, and bottles as objects. Their method had a [success rate](#) of 85 percent on pick-and-place tasks with new objects in new orientations, while the best baseline was only able to achieve a success rate of 45 percent. Success means grasping a new object and placing it on a target location, like hanging mugs on a rack.

Many baselines use 2D image information rather than 3D geometry,

which makes it more difficult for these methods to integrate equivariance. This is one reason the NDF technique performed so much better.

While the researchers were happy with its performance, their method only works for the particular object category on which it is trained. A robot taught to pick up mugs won't be able to pick up boxes or headphones, since these objects have geometric features that are too different than what the network was trained on.

"In the future, scaling it up to many categories or completely letting go of the notion of category altogether would be ideal," Simeonov says.

They also plan to adapt the system for nonrigid objects and, in the longer term, enable the system to perform pick-and-place tasks when the target area changes.

More information: Anthony Simeonov et al, Neural Descriptor Fields: SE(3)-Equivariant Object Representations for Manipulation. arXiv:2112.05124v1 [cs.RO], doi.org/10.48550/arXiv.2112.05124

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: An easier way to teach robots new skills (2022, April 25) retrieved 24 April 2024 from <https://techxplore.com/news/2022-04-easier-robots-skills.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.