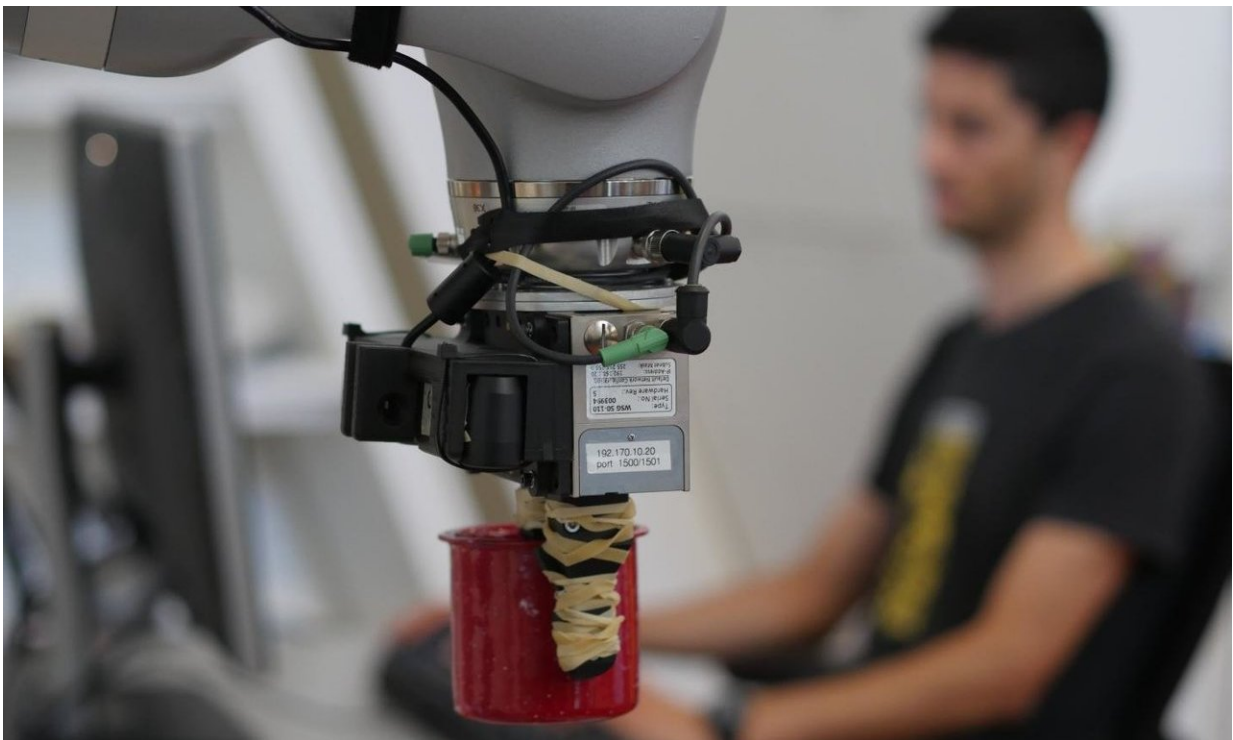# Robot can pick up any object after inspecting it

September 10 2018

Manuelli uses the DON system and Kuka robot to grasp a cup. Credit: Tom Buehler

Humans have long been masters of dexterity, a skill that can largely be credited to the help of our eyes. Robots, meanwhile, are still catching up. Certainly there's been some progress: for decades robots in controlled environments like assembly lines have been able to pick up the same

object over and over again.

More recently, breakthroughs in computer vision have enabled robots to make basic distinctions between objects, but even then, they don't truly understand objects' shapes, so there's little they can do after a quick pick-up.

In a new paper, researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL), say that they've made a key development in this area of work: a system that lets robots inspect random objects, and visually understand them enough to accomplish specific tasks without ever having seen them before.

The system, dubbed "Dense Object Nets" (DON), looks at objects as collections of points that serve as "visual roadmaps" of sorts. This approach lets robots better understand and manipulate items, and, most importantly, allows them to even pick up a specific object among a clutter of similar objects—a valuable skill for the kinds of machines that companies like Amazon and Walmart use in their warehouses.

For example, someone might use DON to get a robot to grab onto a specific spot on an object—say, the tongue of a shoe. From that, it can look at a shoe it has never seen before, and successfully grab its tongue.

"Many approaches to manipulation can't identify specific parts of an object across the many orientations that object may encounter," says Ph.D. student Lucas Manuelli, who wrote a new paper about the system with lead author and fellow Ph.D. student Pete Florence, alongside MIT professor Russ Tedrake. "For example, existing algorithms would be unable to grasp a mug by its handle, especially if the mug could be in multiple orientations, like upright, or on its side."

The team views potential applications not just in manufacturing settings,

but also in homes. Imagine giving the system an image of a tidy house, and letting it clean while you're at work, or using an image of dishes so that the system puts your plates away while you're on vacation.

What's also noteworthy is that none of the data was actually labeled by humans; rather, the system is "self-supervised," so it doesn't require any human annotations.

Making it easy to grasp

Two common approaches to robot grasping involve either task-specific learning, or creating a general grasping algorithm. These techniques both have obstacles: task-specific methods are difficult to generalize to other tasks, and general grasping doesn't get specific enough to deal with the nuances of particular tasks, like putting objects in specific spots.

The DON system, however, essentially creates a series of coordinates on a given object, which serve as a kind of "visual roadmap" of the objects, to give the robot a better understanding of what it needs to grasp, and where.

The team trained the system to look at objects as a series of points that make up a larger coordinate system. It can then map different points together to visualize an object's 3-D shape, similar to how panoramic photos are stitched together from multiple photos. After training, if a person specifies a point on a object, the robot can take a photo of that object, and identify and match points to be able to then pick up the object at that specified point.

This is different from systems like UC-Berkeley's DexNet, which can grasp many different items, but can't satisfy a specific request. Imagine an infant at 18-months old, who doesn't understand which toy you want it to play with but can still grab lots of items, versus a four-year old who

can respond to "go grab your truck by the red end of it."

In one set of tests done on a soft caterpillar toy, a Kuka robotic arm powered by DON could grasp the toy's right ear from a range of different configurations. This showed that, among other things, the system has the ability to distinguish left from right on symmetrical objects.

When testing on a bin of different baseball hats, DON could pick out a specific target hat despite all of the hats having very similar designs—and having never seen pictures of the hats in training data before.

"In factories robots often need complex part feeders to work reliably," says Manuelli. "But a system like this that can understand objects' orientations could just take a picture and be able to grasp and adjust the object accordingly."

In the future, the team hopes to improve the system to a place where it can perform specific tasks with a deeper understanding of the corresponding objects, like learning how to grasp an object and move it with the ultimate goal of say, cleaning a desk.

The team will present their paper on the system next month at the Conference on Robot Learning in Zürich, Switzerland.

Provided by Massachusetts Institute of Technology