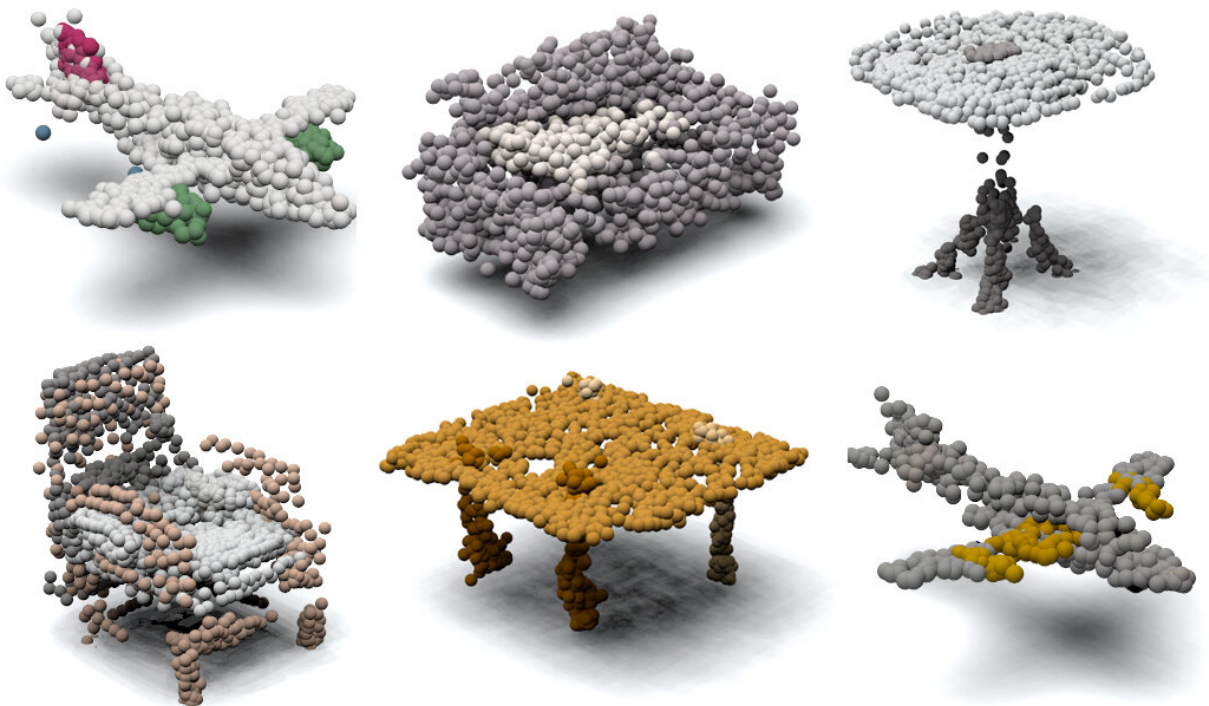# How to train a robot (using AI and supercomputers)

January 20 2021, by Aaron Dubrow

Examples of 3D point clouds synthesized by the progressive conditional generative adversarial network (PCGAN) for an assortment of object classes. PCGAN generates both geometry and color for point clouds, without supervision, through a coarse to fine training process. Credit: William Beksi, Mohammad Samiul Arshad, UT Arlington

Before he joined the University of Texas at Arlington as an Assistant Professor in the Department of Computer Science and Engineering and

founded the Robotic Vision Laboratory there, William Beksi interned at iRobot, the world's largest producer of consumer robots (mainly through its Roomba robotic vacuum).

To navigate built environments, robots must be able to sense and make decisions about how to interact with their locale. Researchers at the company were interested in using machine and deep learning to train their robots to learn about objects, but doing so requires a large dataset of images. While there are millions of photos and videos of rooms, none were shot from the vantage point of a robotic vacuum. Efforts to train using images with human-centric perspectives failed.

Beksi's research focuses on robotics, computer vision, and cyber-physical systems. "In particular, I'm interested in developing algorithms that enable machines to learn from their interactions with the physical world and autonomously acquire skills necessary to execute high-level tasks," he said.

Years later, now with a research group including six Ph.D. computer science students, Beksi recalled the Roomba training problem and begin exploring solutions. A manual approach, used by some, involves using an expensive 360 degree camera to capture environments (including rented Airbnb houses) and custom software to stitch the images back into a whole. But Beksi believed the manual capture method would be too slow to succeed.

Instead, he looked to a form of deep learning known as generative adversarial networks, or GANs, where two neural networks contest with each other in a game until the 'generator' of new data can fool a 'discriminator.' Once trained, such a network would enable the creation of an infinite number of possible rooms or outdoor environments, with different kinds of chairs or tables or vehicles with slightly different forms, but still—to a person and a robot—identifiable objects with

recognizable dimensions and characteristics.

"You can perturb these objects, move them into new positions, use different lights, color and texture, and then render them into a training image that could be used in dataset," he explained. "This approach would potentially provide limitless data to train a robot on."

"Manually designing these objects would take a huge amount of resources and hours of human labor while, if trained properly, the generative networks can make them in seconds," said Mohammad Samiul Arshad, a graduate student in Beksi's group involved in the research.

Generating objects for synthetic scenes

After some initial attempts, Beksi realized his dream of creating photorealistic full scenes was presently out of reach. "We took a step back and looked at current research to determine how to start at a smaller scale—generating simple objects in environments."

Beksi and Arshad presented PCGAN, the first conditional generative adversarial network to generate dense colored [point clouds](point-clouds) in an unsupervised mode, at the International Conference on 3-D Vision (3DV) in Nov. 2020. Their paper, "[A Progressive Conditional Generative Adversarial Network for Generating Dense and Colored 3-D Point Clouds,](link)" shows their network is capable of learning from a training set (derived from ShapeNetCore, a CAD model database) and mimicking a 3-D data distribution to produce colored point clouds with fine details at multiple resolutions.

"There was some work that could generate synthetic objects from these CAD model datasets," he said. "But no one could yet handle color."

	Real	PCGAN			
Resolution	8192 x 6	1024 x 6	2048 x 6	4096 x 6	8192 x 6
Airplane					
Chair					
Motorcycle					
Sofa					
Table					

Examples of 3D point clouds synthesized by a progressive conditional generative adversarial network (PCGAN). Credit: William Beksi, Mohammad Samiul Arshad, UT Arlington

In order to test their method on a diversity of shapes, Beksi's team chose chairs, tables, sofas, airplanes, and motorcycles for their experiment. The tool allows the researchers to access the near-infinite number of possible versions of the set of objects the deep learning system generates.

"Our model first learns the basic structure of an object at low resolutions and gradually builds up towards high-level details," he explained. "The relationship between the object parts and their colors—for examples, the legs of the chair/table are the same color while seat/top are contrasting—is also learned by the network. We're starting small, working with objects, and building to a hierachy to do full synthetic scene generation that would be extremely useful for robotics."

They generated 5,000 random samples for each class and performed an evaluation using a number of different methods. They evaluated both point cloud geometry and color using a variety of common metrics in the field. Their results showed that PCGAN is capable of synthesizing high-quality point clouds for a disparate array of object classes.

SIM2REAL

Another issue that Beksi is working on is known colloquially as 'sim2real.' "You have real training data, and synthetic training data, and there can be subtle differences in how an AI system or robot learns from them," he said. "'Sim2real' looks at how to quantify those differences

and make simulations more realistic by capturing the physics of that scene—friction, collisions, gravity—and by using ray or photon tracing."

The next step for Beksi's team is to deploy the software on a robot, and see how it works in relationship to the sim-to-real domain gap.

The training of the PCGAN model was made possible by TACC's Maverick 2 deep learning resource, which Beksi and his students were able to access through the University of Texas Cyberinfrastructure Research (UTRC) program, which provides computing resources to researchers at any of the UT System's 14 institutions.

"If you want to increase resolution to include more points and more detail, that increase comes with an increase in computational cost," he noted. "We don't have those hardware resources in my lab, so it was essential to make use of TACC to do that."

In addition to computation needs, Beksi required extensive storage for the research. "These datasets are huge, especially the 3-D point clouds," he said. "We generate hundreds of megabytes of data per second; each point cloud is around 1 million points. You need an enormous amount of storage for that."

While Beksi says the field is still a long way from having really good robust robots that can be autonomous for long periods of time, doing so would benefit multiple domains, including health care, manufacturing, and agriculture.

"The publication is just one small step toward the ultimate goal of generating synthetic scenes of indoor environments for advancing robotic perception capabilities," he said.

 More information: International Conference on 3D Vision (3DV),

Nov. 2020, "A Progressive Conditional Generative Adversarial Network for Generating Dense and Colored 3D Point Clouds".
arxiv.org/pdf/2010.05391.pdf

Provided by University of Texas at Austin

Citation: How to train a robot (using AI and supercomputers) (2021, January 20) retrieved 18 April 2024 from https://techxplore.com/news/2021-01-robot-ai-supercomputers.html