

Generating a realistic 3D world

December 6 2021, by Lauren Hinkel



The 3DWorld simulates real-world physics and visualizations in a virtual world.
Credit: Chuang Gan et al

While standing in a kitchen, you push some metal bowls across the counter into the sink with a clang, and drape a towel over the back of a chair. In another room, it sounds like some precariously stacked wooden blocks fell over, and there's an epic toy car crash. These interactions with our environment are just some of what humans experience on a daily basis at home, but while this world may seem real, it isn't.

A new study from researchers at MIT, the MIT-IBM Watson AI Lab,

Harvard University, and Stanford University is enabling a rich virtual world, very much like stepping into "The Matrix." Their [platform](#), called ThreeDWorld (TDW), simulates high-fidelity audio and visual environments, both indoor and outdoor, and allows users, objects, and mobile agents to interact like they would in real life and according to the laws of physics. Object orientations, [physical characteristics](#), and [velocities](#) are calculated and executed for fluids, soft bodies, and rigid objects as interactions occur, producing accurate collisions and impact sounds.

TDW is unique in that it is designed to be flexible and generalizable, generating synthetic photo-realistic scenes and audio rendering in real time, which can be compiled into audio-visual datasets, modified through interactions within the scene, and adapted for human and neural network learning and prediction tests. Different types of robotic agents and avatars can also be spawned within the controlled simulation to perform, say, task planning and execution. And using virtual reality (VR), human attention and play behavior within the space can provide real-world data, for example.

"We are trying to build a general-purpose simulation platform that mimics the interactive richness of the real world for a variety of AI applications," says study lead author Chuang Gan, MIT-IBM Watson AI Lab research scientist.

Creating realistic virtual worlds with which to investigate human behaviors and train robots has been a dream of AI and cognitive science researchers. "Most of AI right now is based on supervised learning, which relies on huge datasets of human-annotated images or sounds," says Josh McDermott, associate professor in the Department of Brain and Cognitive Sciences (BCS) and an MIT-IBM Watson AI Lab project lead. These descriptions are expensive to compile, creating a bottleneck for research. And for physical properties of objects, like mass, which

isn't always readily apparent to human observers, labels may not be available at all. A simulator like TDW skirts this problem by generating scenes where all the parameters and annotations are known. Many competing simulations were motivated by this concern but were designed for specific applications; through its flexibility, TDW is intended to enable many applications that are poorly suited to other platforms.

Another advantage of TDW, McDermott notes, is that it provides a controlled setting for understanding the learning process and facilitating the improvement of AI robots. Robotic systems, which rely on trial and error, can be taught in an environment where they cannot cause physical harm. In addition, "many of us are excited about the doors that these sorts of virtual worlds open for doing experiments on humans to understand human perception and cognition. There's the possibility of creating these very rich sensory scenarios, where you still have total control and complete knowledge of what is happening in the environment."

McDermott, Gan, and their colleagues are presenting this research at the conference on Neural Information Processing Systems (NeurIPS) in December.

Behind the framework

The work began as a collaboration between a group of MIT professors along with Stanford and IBM researchers, tethered by individual research interests into hearing, vision, cognition, and perceptual intelligence. TDW brought these together in one platform. "We were all interested in the idea of building a virtual world for the purpose of training AI systems that we could actually use as models of the brain," says McDermott, who studies human and machine hearing. "So, we thought that this sort of environment, where you can have objects that will interact with each other and then render realistic sensory data from

them, would be a valuable way to start to study that."

To achieve this, the researchers built TDW on a video game platform called Unity3D Engine and committed to incorporating both visual and auditory data rendering without any animation. The simulation consists of two components: The build, which renders images, synthesizes audio, and runs physics simulations; and the controller, which is a Python-based interface where the user sends commands to the build. Researchers construct and populate a scene by pulling from an extensive 3D model library of objects, like furniture pieces, animals, and vehicles. These models respond accurately to lighting changes, and their material composition and orientation in the scene dictate their physical behaviors in the space. Dynamic lighting models accurately simulate scene illumination, causing shadows and dimming that correspond to the appropriate time of day and sun angle. The team has also created furnished virtual floor plans that researchers can fill with agents and avatars. To synthesize true-to-life audio, TDW uses generative models of impact sounds that are triggered by collisions or other object interactions within the simulation. TDW also simulates noise attenuation and reverberation in accordance with the geometry of the space and the objects in it.

Two physics engines in TDW power deformations and reactions between interacting objects—one for rigid bodies, and another for soft objects and fluids. TDW performs instantaneous calculations regarding mass, volume, and density, as well as any friction or other forces acting upon the materials. This allows machine learning models to learn about how objects with different physical properties would behave together.

Users, agents, and avatars can bring the scenes to life in several ways. A researcher could directly apply a force to an object through controller commands, which could literally set a virtual ball in motion. Avatars can be empowered to act or behave in a certain way within the space—e.g.,

with articulated limbs capable of performing task experiments. Lastly, VR head and handsets can allow users to interact with the virtual [environment](#), potentially to generate human behavioral data that machine learning models could learn from.

Richer AI experiences

To trial and demonstrate TDW's unique features, capabilities, and applications, the team ran a battery of tests comparing datasets generated by TDW and other virtual simulations. The team found that neural networks trained on scene image snapshots with randomly placed camera angles from TDW outperformed other simulations' snapshots in image classification tests and neared that of systems trained on real-world images. The researchers also generated and trained a material classification model on audio clips of small objects dropping onto surfaces in TDW and asked it to identify the types of materials that were interacting. They found that TDW produced significant gains over its competitor. Additional object-drop testing with [neural networks](#) trained on TDW revealed that the combination of audio and vision together is the best way to identify the physical properties of objects, motivating further study of audio-visual integration.

TDW is proving particularly useful for designing and testing systems that understand how the physical events in a scene will evolve over time. This includes facilitating benchmarks of how well a model or algorithm makes physical predictions of, for instance, the stability of stacks of objects, or the motion of objects following a collision—humans learn many of these concepts as children, but many machines need to demonstrate this capacity to be useful in the real world. TDW has also enabled comparisons of human curiosity and prediction against those of machine agents designed to evaluate social interactions within different scenarios.

Gan points out that these applications are only the tip of the iceberg. By expanding the physical simulation capabilities of TDW to depict the real world more accurately, "we are trying to create new benchmarks to advance AI technologies, and to use these benchmarks to open up many new problems that until now have been difficult to study."

The research team on the paper also includes MIT engineers Jeremy Schwartz and Seth Alter, who are instrumental to the operation of TDW; BCS professors James DiCarlo and Joshua Tenenbaum; graduate students Aidan Curtis and Martin Schrimpf; and former postdocs James Traer (now an assistant professor at the University of Iowa) and Jonas Kubilius Ph.D. Their colleagues are IBM director of the MIT-IBM Watson AI Lab David Cox; research software engineer Abhishek Bhandwalder; and research staff member Dan Gutfreund of IBM. Additional researchers co-authoring are Harvard University assistant professor Julian De Freitas; and from Stanford University, assistant professors Daniel L.K. Yamins (a TDW founder) and Nick Haber, postdoc Daniel M. Bear, and graduate students Megumi Sano, Kuno Kim, Elias Wang, Damian Mrowca, Kevin Feigelis, and Michael Lingelbach.

More information: ThreeDWorld: A Platform for Interactive Multi-Modal Physical Simulation is available as a PDF at openreview.net/pdf?id=db1InWAwW2T

Provided by Massachusetts Institute of Technology

Citation: Generating a realistic 3D world (2021, December 6) retrieved 27 April 2024 from <https://techxplore.com/news/2021-12-realistic-3d-world.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.