

NVIDIA Wins NeurIPS Awards for Research on Generative AI, Generalist AI Agents

Groundbreaking work among 60+ NVIDIA Research projects at high-profile conference on machine learning, computer vision and more.

Author: Isha Salian

Two NVIDIA Research papers — one exploring diffusion-based generative AI models and another on training generalist AI agents — have been honored with NeurIPS 2022 Awards for their contributions to the field of AI and machine learning.

These are among more than 60+ talks, posters and workshops with NVIDIA authors being presented at the NeurIPs conference, taking place this week in New Orleans and next week online.

Synthetic data generation — for images, text or video — is a key theme across several of the NVIDIA-authored papers. Other topics include reinforcement learning, data collection and augmentation, weather models and federated learning.

“AI is an incredibly important technology, and NVIDIA is making fast progress across the gamut — from generative AI to autonomous AI agents,” said Jan Kautz, vice president of learning and perception research at NVIDIA. “In generative AI, we are not only advancing our theoretical understanding of the underlying models, but are also making practical contributions that will reduce the effort of creating realistic virtual worlds and simulations.”

Diffusion-based models have emerged as a groundbreaking technique for generative AI. NVIDIA researchers won an Outstanding Main Track Paper award for work that analyzes the design of diffusion models, proposing improvements that can dramatically improve the efficiency and quality of these models.

The paper breaks down the components of a diffusion model into a modular design, helping developers identify processes that can be adjusted to improve the performance of the entire model. The researchers show that their modifications enable record scores on a metric that assesses the quality of AI-generated images.

While researchers have long trained autonomous AI agents in video-game environments such as Starcraft, Dota and Go, these agents are usually specialists in only a few tasks. So NVIDIA researchers turned to Minecraft, the world's most popular game, to develop a scalable training framework for a generalist agent — one that can successfully execute a wide variety of open-ended tasks.

Dubbed MineDojo, the framework enables an AI agent to learn Minecraft's flexible gameplay using a massive online database of more than 7,000 wiki pages, millions of Reddit threads and 300,000 hours of recorded gameplay (shown in image at top). The project won an Outstanding Datasets and Benchmarks Paper Award from the NeurIPS committee.

As a proof of concept, the researchers behind MineDojo created a large-scale foundation model, called MineCLIP, that learned to associate YouTube footage of Minecraft gameplay with the video's transcript, in which the player typically narrates the onscreen action. Using MineCLIP, the team was able to train a reinforcement learning agent capable of performing several tasks in Minecraft without human intervention.

Also at NeurIPS is GET3D, a generative AI model that instantly synthesizes 3D shapes based on the category of 2D images it's trained on, such as buildings, cars or animals. The AI-generated objects have high-fidelity textures and complex geometric details — and are created in a triangle mesh format

used in popular graphics software applications. This makes it easy for users to import the shapes into 3D renderers and game engines for further editing.

Named for its ability to Generate Explicit Textured 3D meshes, GET3D was trained on NVIDIA A100 Tensor Core GPUs using around 1 million 2D images of 3D shapes captured from different camera angles. The model can generate around 20 objects a second when running inference on a single NVIDIA GPU.

The AI-generated objects could be used to populate 3D representations of buildings, outdoor spaces or entire cities — digital spaces designed for industries such as gaming, robotics, architecture and social media.

At the most recent CVPR conference, held in New Orleans in June, NVIDIA Research introduced 3D MoMa, an inverse rendering method that enables developers to create 3D objects composed of three distinct parts: a 3D mesh model, materials overlaid on the model, and lighting.

The team has since achieved significant advancements in untangling materials and lighting from the 3D objects — which in turn improves creators' abilities to edit the AI-generated shapes by swapping materials or adjusting lighting as the object moves around a scene.

The work, which relies on a more realistic shading model that leverages NVIDIA RTX GPU-accelerated ray tracing, is being presented as a poster at NeurIPS.

Another accepted paper at NeurIPS examines a key challenge with pretrained language models: the factual accuracy of AI-generated text.

Language models trained for open-ended text generation often come up with text that includes nonfactual information, since the AI is simply making correlations between words to predict what comes next in a sentence. In the paper, NVIDIA researchers propose techniques to address this limitation, which is necessary before such models can be deployed for real-world applications.

The researchers built the first automatic benchmark to measure the factual accuracy of language models for open-ended text generation, and found that bigger language models with billions of parameters were more factual than smaller ones. The team proposed a new technique, factuality-enhanced training, along with a novel sampling algorithm that together help train language models to generate accurate text — and demonstrated a reduction in the rate of factual errors from 33% to around 15%.

There are more than 300 NVIDIA researchers around the globe, with teams focused on topics including AI, computer graphics, computer vision, self-driving cars and robotics. Learn more about NVIDIA Research and view NVIDIA's full list of accepted papers at NeurIPS.

Original URL: <https://blogs.nvidia.com/blog/2022/11/28/nvidia-neurips-research/>