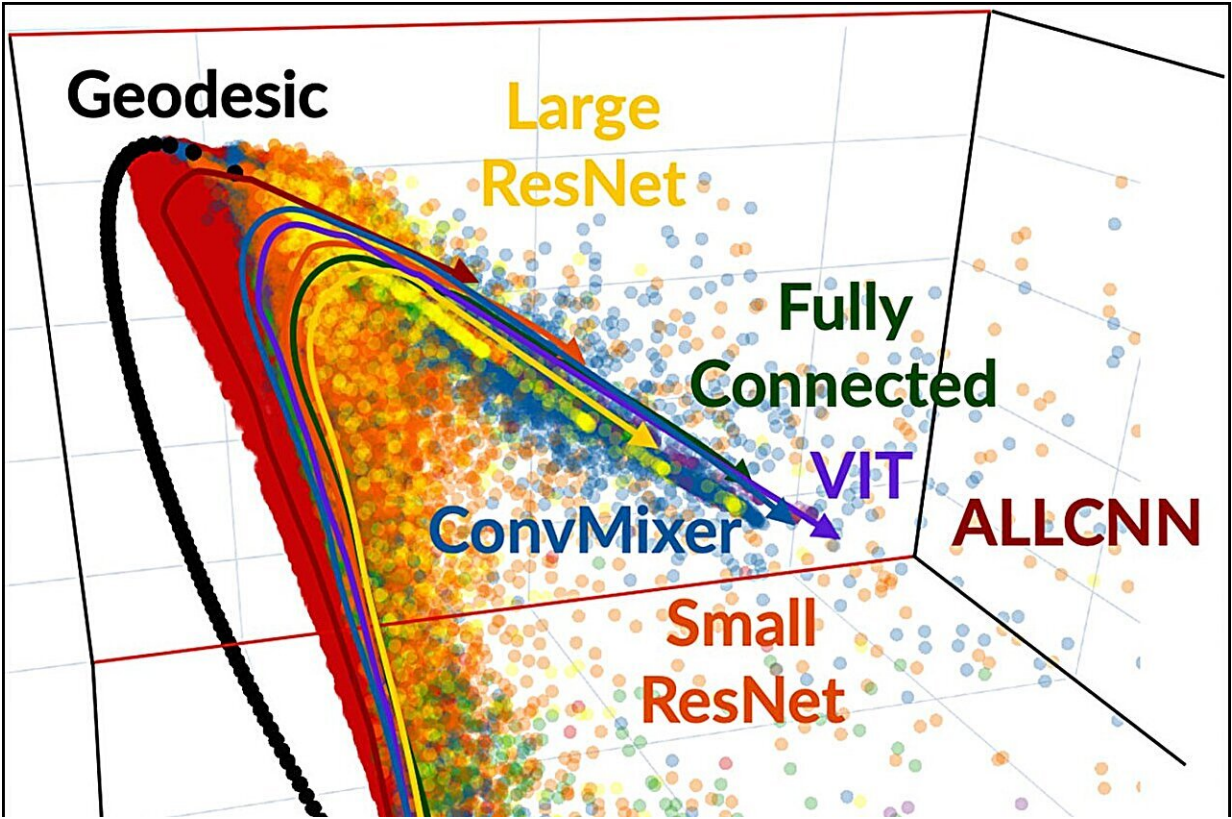


Replica theory shows deep neural networks think alike

March 12 2024, by David Nutt



This data visualization shows the geodesic training trajectory of different deep neural networks as they advance from total ignorance to full certainty. Credit: Cornell University

How do you know you are looking at a dog? What are the odds you are right? If you're a machine-learning algorithm, you sift through thousands

of images—and millions of probabilities—to arrive at the "true" answer, but different algorithms take different routes to get there.

A collaboration between researchers from Cornell and the University of Pennsylvania has found a way to cut through that mindboggling amount of data and show that most successful [deep neural networks](#) follow a similar trajectory in the same "low-dimensional" space.

"Some neural networks take different paths. They go at different speeds. But the striking thing is they're all going the same way," said James Sethna, professor of physics in the College of Arts and Sciences, who led the Cornell team.

The team's technique could potentially become a tool to determine which networks are the most effective.

The group's paper, "The Training Process of Many Deep Networks Explores the Same Low-Dimensional Manifold," is [published](#) in the *Proceedings of the National Academy of Sciences*. The lead author is Jialin Mao of the University of Pennsylvania.

The project has its roots in an algorithm—developed by Katherine Quinn—that can be used to image a large dataset of probabilities and find the most essential patterns, also known as taking the limit of zero data.

Sethna and Quinn [previously used](#) this "replica theory" to comb through cosmic microwave background data, i.e., radiation left over from the universe's earliest days, and map the qualities of our universe against possible characteristics of different universes.

Quinn's "sneaky method," as Sethna called it, produced a three-dimensional visualization "to see the true underlying low-dimensional

patterns in this extremely high-dimensional space."

After those findings were published, Sethna was approached by Pratik Chaudhari of the University of Pennsylvania, who proposed a collaboration.

"Pratik had realized that the method we had developed could be used to analyze the way deep neural networks learn," Sethna said.

Over the course of several years, the researchers collaborated closely. Chaudhari's group, with their vast knowledge and resources in exploring deep neural networks, took the lead and found fast methods for calculating the visualization, and together with Sethna's group they worked to visualize, analyze and interpret this new window into machine learning.

The researchers focused on six types of neural network architectures, including transformer, the basis of ChatGPT. All told, the team trained 2,296 configurations of deep neural networks with varying architectures, sizes, optimization methods, hyper-parameters, regularization mechanisms, data augmentation and random initializations of weights.

"This is really capturing the breadth of what there is today in machine learning standards," said co-author and postdoctoral researcher Itay Griniasty.

For the training itself, the neural networks examined 50,000 images, and for each image determined the probability it fit into one of 10 categories: airplane, automobile, bird, cat, deer, dog, frog, horse, ship or truck. Each probability number is considered a parameter, or dimension. Therefore, the combination of 50,000 images and 10 categories resulted in a half-million dimensions.

Despite this "high-dimensional" space, the visualization from Quinn's algorithm showed that most of the neural networks followed a similar geodesic trajectory of prediction—leading from total ignorance of an image to full certainty of its category—in the same comparatively low dimension. In effect, the networks' ability to learn followed the same arc, even with different approaches.

"Now we can't prove that this has to happen. This is something that's surprising. But that's because we've only been working on it for two decades," Sethna said. "It's inspiring us to do more theoretical work on [neural networks](#). Maybe our method will be a tool for people who understand the different algorithms to guess what will work better."

Co-authors include doctoral student Han Kheng Teoh and researchers from the University of Pennsylvania and Brigham Young University.

More information: Jialin Mao et al, The training process of many deep networks explores the same low-dimensional manifold, *Proceedings of the National Academy of Sciences* (2024). [DOI: 10.1073/pnas.2310002121](#)

Provided by Cornell University

Citation: Replica theory shows deep neural networks think alike (2024, March 12) retrieved 27 April 2024 from <https://techxplore.com/news/2024-03-replica-theory-deep-neural-networks.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.