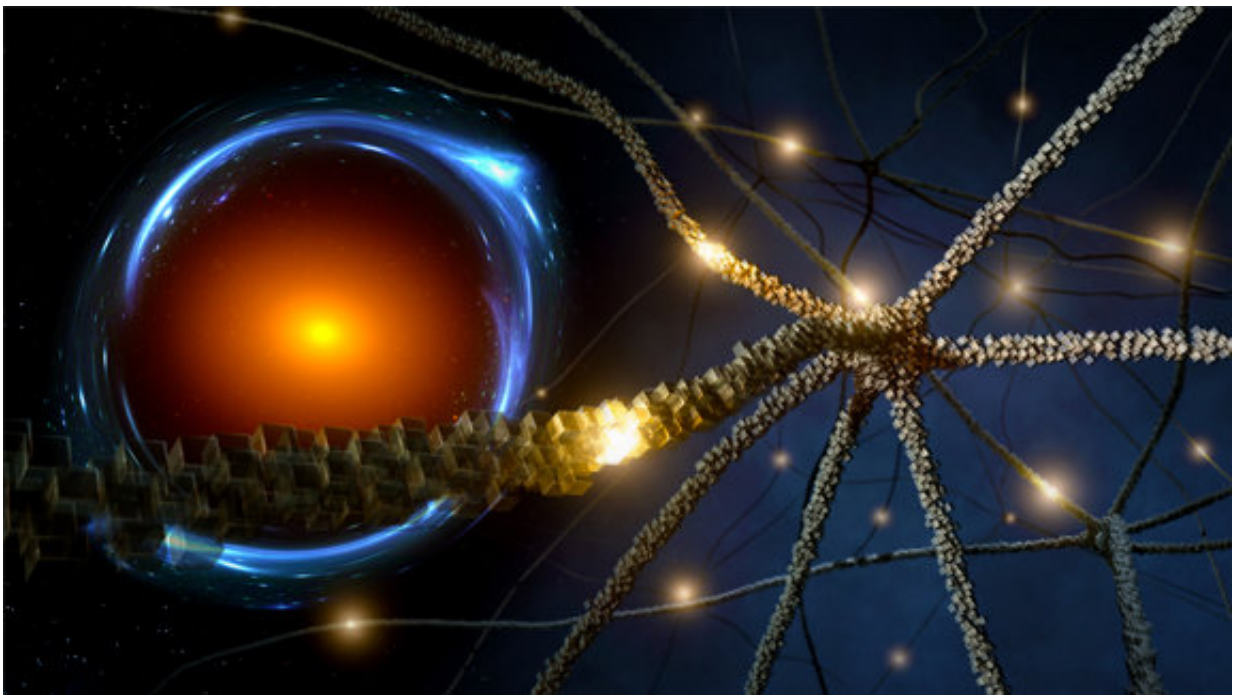


Machine learning and deep learning programs provide a helping hand to scientists analyzing images

November 9 2017



Scientists have for the first time used deep learning to analyze complex distortions in spacetime, called gravitational lenses. This method was 10 million times faster than traditional analyses. Credit: Greg Stewart/SLAC National Accelerator Laboratory

Physicists on the MINERvA neutrino experiments at the Department of

Energy's Fermilab faced a conundrum. Their particle detector was swamping them with images. The detector lights up every time a neutrino, a tiny elementary particle, breaks into other particles. The machine then takes a digital photo of all of the new particles' movements. As the relevant interactions occur very rarely, having a huge amount of data should have been a good thing. But there were simply too many pictures for the scientists to be able to analyze them as thoroughly as they would have liked to.

Enter a new student eager to help. In some ways, it was an ideal student: always attentive, perfect recall, curious to learn. But unlike the graduate students who usually end up analyzing physics photos, this one was a bit more – electronic. In fact, it wasn't a person at all. It was a computer program using [machine learning](#). Computer scientists at DOE's Oak Ridge National Laboratory (ORNL) brought this new student to the table as part of a cross-laboratory collaboration. Now, ORNL researchers and Fermilab physicists are using machine learning together to better identify how neutrinos interact with normal matter.

"Most of the scientific work that's being done today produces a tremendous amount of [data](#) where basically, you can't get human eyes on all of it," said Catherine Schuman, an ORNL computer scientist.

"Machine learning will help us discover things in the data that we're collecting that we would not otherwise be able to discover."

Fermilab scientists aren't the only ones using this technique to power scientific research. A number of scientists in a variety of fields supported by DOE's Office of Science are applying machine learning techniques to improve their analysis of images and other types of scientific data.

Teaching a Computer to Think

In traditional software, a computer only does what it's told. But in machine learning, tools built into the software enable it to learn through practice. Like a student reading books in a library, the more studying it does, the better it gets at finding patterns that can help it solve a big-picture problem.

"Machine learning gives us the ability to solve complex problems that humans can't solve ourselves, or complex problems that humans solve well but don't really know why," said Drew Levin, a researcher who works with DOE's Sandia National Laboratories.

Recognizing images, like those from experiments like MINERvA, is one such major problem. While humans are great at identifying and grouping photos, it's difficult to translate that knowledge into equations for computer programs.

Speeding up Analysis

In the past, creating image-recognition programs was incredibly complex. First, programmers identified every single type of feature in the image they wanted to analyze. Using this list of features, they then made rules for the program to follow. For the neutrino experiments, those rules included describing all of the possible angles a proton could travel. Because scientific images can involve thousands of variables, the process was so slow that many astrophysics experiments had scientists analyze the images by hand instead. Unfortunately, that too was a slow and laborious process.

But machine learning eliminates the vast majority of that work. Programmers create a set of examples that tell the program how to broadly do the analysis, such as processing an image. The program then works to "understand" the data and come up with the rules. It's the difference between telling a student how to add together objects one by

one each time and teaching the principles behind arithmetic.

After the programmer finishes creating the program, he or she then supplies it with large amounts of sample data. The program creates its rules, processes the data, and spits out an answer. In the beginning, these predictions may seem random. As the program takes more data into account, it revises its equations. Those equations then come up with more accurate answers.

To Supervise or Not to Supervise?

Training a machine-learning program can be either "supervised" or "unsupervised."

In supervised learning, the program receives input data as well as output data that gives the "right" answer. Like a student self-scoring a test with an answer key, the program checks to see how its result differs from the correct one. It then tweaks its calculations to get a little closer the next time. Programs that classify images, like identifying whether a photograph is of a star or a galaxy, need to use supervised learning. Scientists can also use supervised learning for creating programs that analyze relationships between variables, such as how the position of a star affects its brightness.

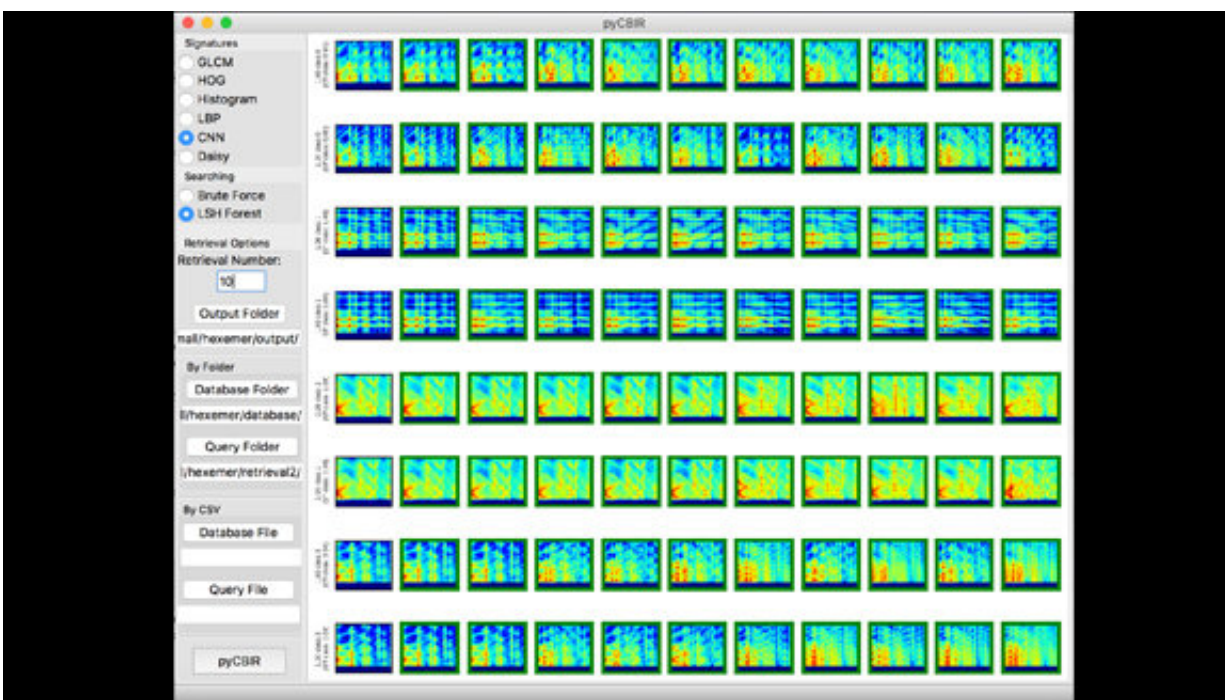
But supervised learning requires data with the answer clearly labeled. For many experiments, labeling the data the program would need for training purposes could take so long that the scientists might as well just analyze it themselves.

"Beginning with unlabeled data is a challenge," said Daniela Ushizima, a researcher at DOE's Lawrence Berkeley National Laboratory (Berkeley Lab), who develops machine learning tools. Unlabeled data particularly pose an issue when researchers are interested in a rare event.

That's where unsupervised learning comes into play. Unsupervised learning requires the program to find patterns itself without the "correct" answers. It's the computer version of independent study. Fortunately, these programs can still group types of data, such as similar images from particle detectors.

Creating a Brain: Deep Learning

While machine learning itself is useful, deep learning takes the concept to the next level. Deep learning is a form of machine learning that uses a neural network – software inspired by human brains.



The pyCBIR deep-learning tool can help researchers match their images to similar ones already in the database. This is an analysis of images from X-ray scattering from the Advanced Light Source. Credit: Daniela Ushizima, Lawrence Berkeley National Laboratory

Each deep learning program is made of a series of very simple units networked together. By grouping the units into hierarchical layers and stacking those layers, programmers create powerful programs. Each layer of units is like a separate team in a factory assembling an intricate puzzle. The earliest teams process basic features. In images, these would be edges and lines or even points. They then pass that analysis along to later teams or deeper layers. The deeper layers put the simple features together to create more complex ones. For an image, this could be a texture or a shape. The final layer spits out an answer. For MINERvA, this final answer may include a variety of information, including where the neutrino collided and what particles resulted from the collision.

As the program learns, it doesn't necessarily change the equations as it would in a simpler machine-learning program. Instead, it subtly changes the relationships between the units and layers, shifting connections from one to another.

What Machine Learning Can Do For You

Grouping and identifying images is one of the most promising uses for machine learning. Back in 2012, a deep-learning program could identify photos in a specific database of images with a 20 percent error rate. Over the course of only three years, scientists improved deep-learning programs so much that a similar program in 2015 beat the average human error rate of 5 percent.

"There's a lot of image-based science that can benefit from deep learning," said Tom Potok, leader of ORNL's Computational Data Analytics group.

For image recognition that requires special expertise, machine learning can provide even bigger benefits. "These techniques are extremely efficient at finding subtle signals" like small shifts in particle tracks, said

Gabe Perdue, a Fermilab physicist on the MINERvA experiment.

While Fermilab physicists are using deep learning to understand neutrinos, other scientists are using it to understand images from sources as diverse as telescopes and light sources.

Spotting when a very large object is warping our view of a galaxy can help astronomers understand unknown phenomena like dark matter and dark energy. But it can take expert astronomers weeks to analyze a single image. This rate is fine for current equipment, which has only captured a few hundred images of this happening. But when the Large Synoptic Survey Telescope goes online in 2022, astronomers predict it will photograph tens of thousands of these galaxies.

To get ready, scientists at SLAC National Accelerator Laboratory have already developed a deep-learning program to tackle it. First, they spent a day feeding about half a million real and simulated images of galaxies into the eager student's electronic brain. The program then analyzed a combination of real images from the Hubble Space Telescope and simulated images in a few seconds. This analysis was 10 million times faster than previous methods and just as accurate. It even provided data that the previous methods didn't, like measurements of how much mass was warping the images.

Other scientists are using machine learning to sort and organize images. Most databases of scientific images are difficult to search or limited to images with in-depth descriptions. But Ushizima thought she had a better way. She imagined something like Google's Image Search, where you can upload an image and have Google find others like it.

"Instead of looking for experiments in terms of keywords or a mathematical model, we would have a more concrete way to retrieve results: We input an image," she said

Using a DOE Early Career Research Program award, she and graduate students Flavio Araujo and Romuere Silva developed a deep-learning tool called pyCBIR. The program can tell researchers how similar their images are to ones already in its database. Currently, Ushizima and her fellow researchers are working with the Advanced Light Source, a Office of Science user facility, to analyze many of its images. With the program analyzing the content from millions of images, scientists can now rank and organize experimental data without needing to rely on filename or other textual information. While sifting through massive amounts of unlabeled data could take days or even longer, the pyCBIR software allows scientists to find relevant images in seconds.

Image analysis is just one application of deep learning for science. Scientists are using machine learning to identify extreme weather events in earth system simulations. They're also using it to predict flaws in new metal alloys and analyze millions of cancer drug results.

Tackling Future Challenges

But machine and deep learning aren't panaceas for scientific research. One of the biggest challenges is ensuring that the programs are providing the correct answers. In deep learning, programmers can only see calculations that are happening in the first and last layer. In the classroom of deep learning, there's no way to ask the student about its thought process. In addition, if there are inaccuracies in the training data, the deep-learning program will amplify it.

"You might worry if there's some bias or some mistake coming from these simulations used to train these [machines](#)," said Perdue.

Usually, [scientists](#) can recognize when results they receive are wrong or at least different from what they expected. In the case of MINERvA, neutrinos only move and interact in certain ways. If the [images](#) show

something different, they need to double-check the machine. Scientists also understand what kinds of problems can arise from the programs they use to conduct the analysis and how to fix them. But because programs used for [deep learning](#) are so different from traditional ones, they throw a wrench in that institutional knowledge.

The ORNL team helping Fermilab analyze the MINERvA data is hoping to solve some of those challenges. They're using three different technologies to design a powerful, accurate deep- learning program. Using a specialized computer that processes quantum information, they hope to design the best structure for the program. They'll then use ORNL's fastest supercomputer, Titan, to create the best arrangement of units and connections to maximize accuracy and speed. Lastly, they'll run the program on a brain-like piece of hardware. In addition to MINERvA, they plan to use this program to analyze data from ORNL's Spallation Neutron Source, a Office of Science user facility.

Whether in [neutrino experiments](#) or cancer research, machine learning offers a new way for both researchers and their electronic students to better understand our world and beyond.

As Prasanna Balaprakash, a computer scientist at DOE's Argonne National Laboratory, said, "Machine learning has applications all the way from subatomic levels up to the universe. Wherever we have data, machine learning is going to play a big role."

Provided by US Department of Energy

Citation: Machine learning and deep learning programs provide a helping hand to scientists analyzing images (2017, November 9) retrieved 19 April 2024 from <https://phys.org/news/2017-11-machine-deep-scientists-images.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.