How deep learning is bringing automatic cloud detection to new heights

December 12 2018, by Susan Bauer

Clouds come in all shapes and sizes, and now a deep learning model can help detect the finer details in cloud data. Credit: Adriel Kloppenburg on Unsplash

Kids lying on their backs in a grassy field might scan the clouds for images—perhaps a fluffy bunny here and a fiery dragon over there.

Often, atmospheric scientists do the opposite—they search data images for the clouds as part of their research to understand Earth systems.

Manually labeling data images pixel by pixel is time-consuming, so researchers rely on automatic processing techniques, such as cloud detection algorithms. But the algorithms' output is not as exact as the scientists want it to be.

Recently, researchers at the Department of Energy's Pacific Northwest National Laboratory teamed up to find out if deep learning—a distinct subset of machine learning—can do a better job at identifying clouds in lidar data than the current physics-based algorithms. The answer: A clear "yes." The new model is much closer to the answers scientists arrive at but in just a fraction of the time.

Lidar is a remote sensing instrument that emits a pulsed laser and collects the return signal scattered back by cloud droplets or aerosols. This return signal provides information about the height and vertical structure of atmospheric features, such as clouds or smoke layers. Such data from ground-based lidars are an important part of global forecasting.

Earth scientist Donna Flynn noticed that, in some cases, what the algorithms detected as clouds in the lidar images did not match well with what her expert eye saw. The algorithms tend to overestimate the cloud boundaries.

"The current algorithm identifies the clouds using broad brushstrokes," says Flynn, a co-principal investigator on the project. "We need to more accurately determine the cloud's true top and base and to distinguish multiple cloud layers."

Upgrade initiated

Until recently, computing power limited [artificial neural networks](#), a type of deep learning [model](#), to a small number of computational layers. Now, with increased computing power available through supercomputing clusters, researchers can use more computations—each building off of the last one—in a series of layers. The more layers an artificial neural network has, the more powerful the deep learning network.

Figuring out what those computations are is part of the model training. To start, the researchers need properly labeled lidar data images, or "ground truth" data, for the training and testing of the model. So, Flynn spent many long hours hand-labeling images pixel by pixel: cloud or no cloud. Her eye can distinguish the cloud boundaries and cloud versus an aerosol layer. She took 40 hours—the equivalent of a full work week—to label about 100 days of lidar data collected at the Southern Great Plains atmospheric observatory, part of DOE's Atmospheric Radiation Measurement user facility, in Oklahoma.

Given how time and labor intensive the hand-labeling process is, PNNL computational scientist and co-principal investigator Erol Cromwell used learning methods that required minimal ground truth data.

The model learns through self-feedback. It compares its own performance against hand-labeled results and adjusts its calculations accordingly, explains Cromwell. It cycles through these steps, improving each time through.

Cromwell will be presenting the team's findings at the Institute of Electrical and Electronics Engineers [Winter Conference on Applications of Computer Vision](#) in January.

Goal achieved

With the training, the deep learning model outperforms the current algorithms. The model's precision is almost double and much closer to what a human expert would find—but in a fraction of the time.

The next steps are to evaluate the model's performance on lidar data collected at different locations and in different seasons. Initial tests on data from the ARM observatory at Oliktok Point in Alaska are promising.

"An advantage of the deep learning model is transfer learning," says Cromwell. "We can train the model further with data from Oliktok to make its performance more robust."

"Reducing sources of uncertainty in global model predictions is especially important to the atmospheric science community," says Flynn. "With its improved precision, deep learning increases our confidence."

She adds: "Plus, it gives us more time to be outside looking at real clouds !"

Provided by Pacific Northwest National Laboratory

Citation: How deep learning is bringing automatic cloud detection to new heights (2018, December 12) retrieved 27 April 2024 from https://phys.org/news/2018-12-deep-automatic-cloud-heights.html