Tracking Deforestation with Neural Networks

Yiqiao Yin

Scientia

TRACKING DEFORESTATION WITH NEURAL NETWORKS

Forests are key to Earth's biodiversity and the global ecosystem, hosting 80% of the planet's plant biomass and two-thirds of all mammal species. However, they are also in decline; each year, around five million hectares of woodland are lost to logging and wildfires. Tracking this loss is essential to controlling and mitigating deforestation, but this requires careful interpretation of satellite imagery. **Yiqiao Yin**, **Kieran Pichai**, **Benjamin Park** and **Aaron Bao** have developed a model to automatically identify forests from these images. Their new approach is already showing huge promise.

Seeing the Forest for the Trees

A few thousand years ago, forests covered well over half of the world's habitable land. Since then, human activity has removed a third of that forest – around 2 billion hectares. Most of this deforestation has taken place since 1900, as the demand for lumber and land for agriculture have soared. Now another, more indirect threat to forests is emerging.

As the climate warms due to rising atmospheric carbon, extreme weather events are becoming more common. This includes cold and wet periods, but also intense droughts and blistering heat waves, during which a single spark can ignite and destroy vast areas of woodland.

There are efforts to counter this loss by reforesting wild environments – planting trees and protecting them and their ecosystems. Tracking the progress of these efforts is essential to their long-term success, and accurately measuring the balance of growth and loss of woodland across the globe is necessary to understand and mitigate the effects of deforestation on the economy, local ecosystems, and the global biosphere.

The only practical way to do this tracking is through aerial photography, primarily with satellite images. However, this requires careful interpretation of such images to distinguish a forest from grassland or cropland. Furthermore, this needs to be accurate both in the long-term, to track the rate of global deforestation, and in the short-term, to track a wildfire as it burns. Yiqiao Yin, Kieran Pichai, Benjamin Park and Aaron Bao have developed a new technique to automate this process, offering a fast and accurate way to monitor forests around the world.

Understanding a Branching Model

The team's innovative approach splits the problem into two stages: first, their algorithm *segments* the satellite image, identifying which regions are forest and which are not. This produces a black-and-white overlay called a mask, which highlights the forested areas by blacking out the non-forested areas. This output is intuitive and easily understood, but the underlying mechanism is complex.

Secondly, a *classification* algorithm determines whether the land that was imaged is more than 50% forest. These two approaches work together to quickly and accurately provide a

visual overview of forest coverage, and a numerical analysis of the extent of that coverage.

Yin, Pichai and their team use a convolutional neural network to achieve this dual interpretation of forest images. To understand this, we'll also take a dual approach. First, what is a neural network?

Inspired by the structure of the brain, a neural net consists of 'neurons' that can be activated, and connections between those neurons. The neurons are organised into layers, with an input layer for data at one end and an output at the other. Each connection has a strength,

called a *weight*, which determines how large an influence the neuron has on the connected neuron in the next layer.

We *train* a neural network by providing it with data that's already been analysed. That way, we can give it feedback if it makes a mistake, so that it can automatically be adjusted to improve its prediction. After many, many repetitions, the network can accurately analyse the training data, and hopefully, similar novel data.

For example, a neural network could be trained to identify handwritten numbers. Each pixel of the scanned image is sent to a neuron in the input layer, and then is processed by intermediate layers. Finally, all being well, a single one of the ten output neurons would activate, signalling that the scanned digit was a 6.

Second, what makes a neural network *convolutional*? This can be hard to intuitively understand, but the underlying ideas aren't as difficult as they seem.

Suppose you have an old, grainy, grey-scale image. You might want to blur it slightly, to get rid of the grain. The simplest way to do this is to create a new image where each pixel is defined as the average of the old pixel and its surrounding pixels. For example, a black pixel could be encoded as a 0 (representing 0% brightness), while a white pixel is 100 (100% brightness). If a black pixel is surrounded by 8 white pixels, then in the new image, we'd replace it with a pixel of 88.8% brightness, 800 divided by 9. This would blur out any black specks on the photo.

If this blur was too strong, we could instead use a *weighted average* by making the surrounding pixels matter less than the central pixel. This is achieved with a *kernel*, which is a small grid of numbers telling the computer how to weight the average. In the first example, the kernel would be a 3x3 grid, with each cell just containing the number 1, since no pixels are weighted. If we wanted a less harsh blur, we could replace the central cell in the kernel with a 3, meaning the central pixel matters three times more.

By carefully adjusting that kernel, you can process images in many ways. The most important use for Yin, Pichai and their colleagues is *edge detection*. By using asymmetrical kernels, the output pixel will only have a high value if there's a large difference on opposite sides of the central pixel. Processing the whole image like this gives an outline of all the edges, borders and boundaries in the images, without any detail inside those borders.

The team's convolutional neural network combines these two approaches by allowing the neural network to employ convolutional methods. When the neural network is trained using images of forests that have already been analysed, the model can learn to use edge detection to find the boundaries of forested areas and focus on this data. This drastically simplifies the amount of data that the neural net must process, without discarding the salient patterns.

Future Growth

The team plans to continue improving their results by adjusting the model, and training it on a larger and more detailed dataset. They also hope to include further classifications, beyond 'dense' and 'sparsely forested'. With such improvements, the model developed by Yin, Pichai and their colleagues could become a critical tool in managing wildfires by tracking the spread of fire in real-time, which could be used to organise the evacuation of residents or to co-ordinate fire crews and helicopters to tackle the fire in critical regions.

As it stands, the team's final model is an excellent benchmark. It's excelled in testing, correctly identifying an image as densely or sparsely forested in more than 80% of tests, and correctly identifying which parts of the image are forested about 80% of the time. This is a drastic increase in model accuracy, allowing us to reliably monitor the health of Earth's forests in the face of climate change, and track the success of important reforestation programs.

Meet the researchers

Yiqiao Yin
LabCorp
Princeton, NJ

Yiqiao Yin received an MA from Columbia University in 2019 for his machine learning research, focused on predicting relapse in breast cancer patients. His model achieved an accuracy of 92%, improving on the industry standard by 20–30%. In 2022, Yin began work at LabCorp in New Jersey, where he leads research projects tackling the AI challenges of developing drugs, employing techniques such as convolutional neural networks and long-short term memory.

E: Eagle0504@Gmail.com

W: https://www.yiqiao-yin.com/

Kieran Pichai
Menlo School
Menlo Park, CA

Kieran Pichai is a student at Menlo School, California. He is passionate about solving environmental issues using computational approaches. In addition to his studies, he also works as a team to build machine-learning algorithms and other software. Pichai is also passionate about teaching, and has taught many classes in computer science, mathematics and robotics. In one of his educational projects, he created a computer science curriculum for high-achieving students in under-resourced schools.

E: kieran.pichai@gmail.com

Aaron Bao
The Harker School
San Jose, CA

Aaron Bao is a student at The Harker School, California, where he studies computer science, mathematics and several other subjects. His dedication to his studies has been recognised with the 2022 Harker Love of Learning Award. In addition to computer science and maths, Bao also studies congressional debate, and has attended many national debate tournaments. He volunteers at the Tahirih Justice Center, which helps immigrant survivors of gender violence.

E: aaron.bao64@gmail.com

Benjamin Park
Menlo School
Menlo Park, CA

Benjamin Park is currently a student at Menlo School, California, who specialises in computer science, physics and mathematics. He is particularly interested in the use of AI as a tool for tackling climate change. In addition to his studies, Park is also a competitive pianist. In both 2017 and 2022, he won first prize in the CAPMT Northern California Piano Competition. He is also an athlete, and regularly competes in golf tournaments.

E: benjamin.park@menloschool.org

FURTHER READING

Kieran Pichai, Benjamin Park, Aaron Bao, and Yiqiao Yin, Automated Segmentation and Classification of Aerial Forest Imagery, Analytics, 2022, 1(2), 135-143.