

New Scientist Live Limited VIP tickets remaining – Book now

DAILY NEWS 10 August 2017

DeepMind AI teaches itself about the world by watching videos



Lions roar: video clips beat labels for AIs seeking knowledge
OE KLAMAR/AFP/Gettyvide

By **Matt Reynolds**

To an untrained AI, the world is a blur of confusing data streams. Most humans have no problem making sense of the sights and sounds around them, but algorithms tend only to acquire this skill if those sights and sounds are explicitly labelled for them.

Now DeepMind has developed an AI that teaches itself to recognise a range of visual and audio concepts just by watching tiny snippets of video. This AI can grasp the concept of lawn mowing or tickling, for example, but it hasn't been taught the words to describe what it's hearing or seeing.

"We want to build machines that continuously learn about their environment in

an autonomous manner,” says Pulkit Agrawal at the University of California, Berkeley. Agrawal, who wasn’t involved with the work, says this project takes us closer to the goal of creating AI that can teach itself by watching and listening to the world around it.

Learn more at New Scientist Live: See our expert talk on the future of AI

Most computer vision algorithms need to be fed lots of labelled images so it can tell different objects apart. Show an algorithm thousands of cat photos labelled “cat” and soon enough it’ll learn to recognise cats even in images it hasn’t seen before.

But this way of teaching algorithms – called supervised learning – isn’t scalable, says Relja Arandjelović who led the project at DeepMind. Instead of relying on human-labelled datasets, his algorithm learns to recognise images and sounds by matching up what it sees with what it hears.

Learn like a human

Humans are particularly good at this kind of learning , says Paolo Favaro at the University of Bern in Switzerland. “We don’t have somebody following us around and telling us what everything is,” he says.

Arandjelović created his algorithm by starting with two networks – one that specialised in recognising images and another that did a similar job with audio. He showed the image recognition network stills taken from short videos while the audio recognition network was trained on 1-second audio clips taken from the same point in each video.

A third network compared still images with audio clips to learn which sounds corresponded with which sights in the videos. In all, the system was trained on 60 million still-audio pairs taken from 400,000 videos.

The algorithm learned to recognise audio and visual concepts, including crowds, tap dancing and water, without ever seeing a specific label for a single concept. When shown a photo of someone clapping, for example, most of the time it knew which sound was associated with that image.

Sight and sound

This kind of co-learning approach could be extended to include senses other than sight and hearing, says Agarwal. “Learning visual and touch features simultaneously can, for example, enable the agent to search for objects in the dark and learn about material properties such as friction,” he says.

DeepMind will present the study at the International Conference on Computer Vision which takes place in Venice, Italy, in late October.

While the AI in the DeepMind project doesn't interact with the real world, Agarwal says that perfecting self-supervised learning will eventually let us create AI that can operate in the real world and learn from what it sees and hears.

But until we reach that point, self-supervised learning might be a good way of training image and audio recognition algorithms without input from vast amounts of human-labelled data. The DeepMind algorithm can correctly categorise an audio clip nearly 80 per cent of the time, making it better at audio-recognition than many algorithms trained on labelled data.

Such promising results suggest that similar algorithms might be able to learn something by crunching through huge unlabelled datasets like YouTube's millions of online videos. "Most of the data in the world is unlabelled and therefore it makes sense to develop systems that can learn from unlabelled data," Agrawal says.

Journal reference: arxiv.org

Read more: Curious AI learns by exploring game worlds and making mistakes

NewScientist | Jobs

More jobs ►