



No, computer, it's not a whirlpool

Do you see what I see?

Teaching computers to understand the world as we see it will help them understand us. **Aviva Rutkin** takes a closer look

“WHAT am I doing now? How about now? And now?” University students have been popping GoPro video cameras on their heads and filming a first-person view of their daily lives, then asking a computer to interpret it.

Vain though it may sound, the exercise has a point. Researchers want artificial intelligences to understand us better – and teaching them to see the world through our eyes is a good place to start.

“It allows us to indirectly tap into human minds,” says Gedas Bertasius at the University of Pennsylvania in Philadelphia. “It allows us to reason much more accurately about human behaviour – the connection between what we see and how we’re going to act.”

Bertasius and his colleagues are building EgoNet, a neural network

system that tries to predict what objects someone might be interested in. Volunteers had to annotate videos of their day-to-day lives frame-by-frame, to show where their attention was focused in each scene. They then fed the footage into a computer and asked EgoNet over and over again to tell them what they were doing. That data helped train it to make predictions, picking out things that a person was about to touch or look at more closely.

EgoNet examines the world through two lenses, backed by separate neural networks. One looks for objects likely to stand out to someone – because of being brightly coloured, say, or being centrally placed in the scene. The other estimates how each object might relate to that person. Is it within reach? Is it oriented in a way that allows the person to

interact with it easily? You are more likely to pick up a coffee mug if its handle is pointed towards you, for example. Similarly, someone wanting to use a computer will approach it keyboard-first.

The team tested EgoNet on footage that included people cooking, kids playing, and a dog running in a park. It still has some way to go before it can rival a

“One drawback of training an AI on a diet of fiction is that it thinks everyone lives inside a soap opera”

human, but Bertasius hopes a version of the system might be useful in healthcare, perhaps helping doctors diagnose unusual behaviour patterns in children.

In a separate project called Augur, researchers at Stanford

University in California are also trying to get computers to understand what’s happening in first-person video. But instead of learning from annotated footage, Augur was brought up on a very different data set: 1.8 billion words of fiction taken from Wattpad, an online writing community.

Events to chew over

Fiction is a great resource for making predictions about human behaviour, says Ethan Fast, who leads the Augur project. “Fiction describes the breadth of human life.” Stories also tend to have a narrative structure that provides a logical sequence of events for a computer to chew over. “This thing happens, then this thing happens, then this,” says Fast.

When Augur identifies an object in a scene, it mines what it has read to guess what a person might do with it. If it spots a plate, for example, it infers that someone is probably planning to eat, cook or wash up. If you wake up and look at your alarm clock, Augur should guess that you are about to get out of bed.

One drawback of relying on fiction is that it has given Augur a dramatic bent. If a phone rings, it thinks you are about to start swearing and will throw the phone against a wall. Tweaking the system using more mundane scenarios will help teach Augur that not everyone lives inside a soap opera.

The Stanford team will present Augur at the CHI conference in San Jose, California, next month. They will also show off its first application, a Google Glass app that plays a soundtrack for your life, choosing songs it deems appropriate for whatever it perceives the wearer to be doing.

The team thinks future versions could help in other ways, such as screening calls when it recognises that you’re busy, or reminding the wearer of their shopping budget when it catches them eyeing the checkout counter. ■