

Decades of computer vision research, one 'Swiss Army knife'

April 1 2016, by Allison Linn



Mary Bellard, left, and Anne Taylor, right, were part of the team that developed Seeing AI, backed by decades of computer vision research.

When Anne Taylor walks into a room, she wants to know the same things that any person would.

Where is there an empty seat? Who is walking up to me, and is that person smiling or frowning? What does that sign say?

For Taylor, who is blind, there aren't always easy ways to get this information. Perhaps another person can direct her to her seat, describe her surroundings or make an introduction.

There are apps and tools available to help visually impaired people, she said, but they often only serve one limited function and they aren't always easy to use. It's also possible to ask other people for help, but most people prefer to navigate the world as independently as possible.

That's why, when Taylor arrived at Microsoft about a year ago, she immediately got interested in working with a group of researchers and engineers on a project that she affectionately calls a potential "Swiss Army knife" of tools for visually impaired people.

"I said, 'Let's do something that really matters to the blind community,'" said Taylor, a senior project manager who works on ways to make Microsoft products more accessible. "Let's find a solution for a scenario that really matters."

That project is Seeing AI, a research project that uses computer vision and natural language processing to describe a person's surroundings, read text, answer questions and even identify emotions on people's faces. Seeing AI, which can be used as a cell phone app or via smart glasses from [Pivthead](#), made its public debut at the company's Build conference this week. It does not currently have a release date.

Taylor said Seeing AI provides another layer of information for people who also are using mobility aids such as white canes and guide dogs.

"This app will help level the playing field," Taylor said.

At the same conference, Microsoft also unveiled CaptionBot, a demonstration site that can take any image and provide a detailed description of it.

Very deep neural networks, natural language processing and more

Seeing AI and CaptionBot represent the latest advances in this type of technology, but they are built on decades of cutting-edge research in fields including computer vision, image recognition, natural language processing and [machine learning](#).

In recent years, a spate of breakthroughs has allowed computer vision researchers to do things they might not have thought possible even a few years before.

"Some people would describe it as a miracle," said Xiaodong He, a senior Microsoft researcher who is leading the image captioning effort that is part of [Microsoft Cognitive Services](#). "The intelligence we can say we have developed today is so much better than six years ago."

The field is moving so fast that it's [substantially better](#) than even six months ago, he said. For example, Kenneth Tran, a senior research engineer on his team who is leading the development effort, recently figured out a way to make the image captioning system more than 20 times faster, allowing people who use tools like Seeing AI to get the information they need much more quickly.

A major a-ha moment came a few years ago, when researchers hit on the idea of using [deep neural networks](#), which roughly mimic the biological processes of the human brain, for machine learning.

Machine learning is the general term for a process in which systems get better at doing something as they are given more training data about that task. For example, if a computer scientist wants to build an app that helps bicyclists recognize when cars are coming up behind them, it would feed the computer tons of pictures of cars, so the app learned to recognize the difference between a car and, say, a sign or a tree.

Computer scientists had used neural networks before, but not in this way, and the new approach resulted in big leaps in computer vision accuracy.

Several months ago, Microsoft researchers Jian Sun and Kaiming He made another big leap when they unveiled a new system that uses very deep [neural networks](#) – called [residual neural networks](#) – to correctly identify photos. The [new approach](#) to recognizing images resulted in huge improvements in accuracy. The researchers shocked the academic community and won two major contests, the ImageNet and [Microsoft Common Objects in Context](#) challenges.

Tools to recognize and accurately describe images

That approach is now being used by Microsoft researchers who are working on ways to not just recognize images but also write captions about them. This research, which combines image recognition with natural language processing, can help people who are visually impaired get an accurate description of an image. It also has applications for people who need information about an image but can't look at it, such as when they are driving.

The image captioning work also has received [accolades for its accuracy](#) as compared to other research projects, and it is the basis for the capabilities in Seeing AI and Caption Bot. Now, the researchers are working on expanding the training set so it can give users a deeper sense

of the world around them.

Margaret Mitchell, a Microsoft researcher who specializes in [natural language processing](#) and has been one of the industry's leading researchers on image captioning, said she and her colleagues also are looking at ways a computer can describe an image in a more human way.

For example, while a computer might accurately describe a scene as "a group of people that are sitting next to each other," a person may say that it's "a group of people having a good time." The challenge is to help the technology understand what a person would think was [most important, and worth saying](#), about the picture.

"There's a separation between what's in an image and what we say about the image," said Mitchell, who also is one of the leads on the Seeing AI project.

Other Microsoft researchers are developing ways that the latest [image recognition](#) tools can provide more thorough explanations of pictures. For example, instead of just describing an image as "a man and a woman sitting next to each other," it would be more helpful for the technology to say, "Barack Obama and Hillary Clinton are posing for a picture."

That's where Lei Zhang comes in.

When you search the Internet for an image today, chances are high that the search engine is relying on text associated with that image to return a picture of Kim Kardashian or Taylor Swift.

Zhang, a senior researcher at Microsoft, is working with researchers including Yandong Guo on a system that uses machine learning to identify celebrities, politicians and public figures based on the elements of the image rather than the text associated with it.

Zhang's research will be included in the latest vision tools that are part of Microsoft Cognitive Services. That's a set of tools that is based on Microsoft's cutting-edge machine learning research, and which developers can use to build apps and services that do things like recognize faces, identify emotions and distinguish various voices. Those tools also have provided the technical basis for Microsoft showcase apps and demonstration websites such as how-old.net, which guesses a person's age, and Fetch, which can identify a dog's breed.

Microsoft Cognitive Services is an example of what is becoming a more common phenomenon – the lightning-fast transfer of the latest research advances into products that people can actually use. The engineers who work on Microsoft Cognitive Services say their job is a bit like solving a puzzle, and the pieces are the latest research.

"All these pieces come together and we need to figure out, how do we present those to an end user?" said Chris Buehler, a software engineering manager who works on Microsoft Cognitive Services.

From research project to helpful product

Seeing AI, the research project that could eventually help visually impaired people, is another example of how fast research can become a really helpful tool. It was conceived at last year's //oneweek Hackathon, an event in which Microsoft employees from across the company work together to try to make a crazy idea become a reality.

The group that built Seeing AI included researchers and engineers from all over the world who were attracted to the project because of the technological challenges and, in many cases, also because they had a personal reason for wanting to help visually impaired people operate more independently.

"We basically had this super team of different people from different backgrounds, working to come up with what was needed," said Anirudh Koul, who has been a lead on the Seeing AI project since its inception and became interested in it because his grandfather is losing his ability to see.

For Taylor, who joined Microsoft to represent the needs of blind people, it was a great experience that also resulted in a potential product that could make a real difference in people's lives.

"We were able to come up with this one Swiss Army knife that is so valuable," she said.

Provided by Microsoft

Citation: Decades of computer vision research, one 'Swiss Army knife' (2016, April 1) retrieved 26 April 2024 from <https://phys.org/news/2016-04-decades-vision-swiss-army-knife.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.