

PCMag editors select and review products independently. If you buy through affiliate links, we may earn commissions, which help support our testing.

Home > News > AI

What Is Computer Vision?

Computer vision is one of the hottest areas of computer science and artificial intelligence research, but it can't yet compete with the power of the human eye. Here's why.

By [Ben Dickson](#)

February 9, 2020





Table of Contents



When you look at the following image, you see people, objects, and buildings. It brings up memories of past experiences, similar situations you've encountered. The crowd is facing the same direction and holding up phones, which tells you that this is some kind of event. The person standing near the camera is wearing a T-shirt that hints at what the event might be. As you look at other small details, you can infer much more information from the picture.





Photo by Joshua J. Cotten on Unsplash

But to a computer, this image—like all images—is an array of pixels, numerical values that represent shades of red, green, and blue. One of the challenges computer scientists have grappled with since the 1950s has been to create machines that can make sense of photos and videos like humans do. The field of *computer vision* has become one of the hottest areas of research in computer science and artificial intelligence.

Decades later, we have made huge progress toward creating software that can understand and describe the content of visual data. But we've also discovered how far we must go before we can understand and replicate one of the fundamental functions of the human brain.

A Brief History of Computer Vision

In 1966, Seymour Papert and Marvin Minsky, two pioneers of artificial intelligence, launched the Summer Vision Project, a two-month, 10-man effort to create a computer system that could identify objects in images.

To accomplish the task, a computer program had to be able to determine which pixels belonged to which object. This is a problem that the human vision system, powered by our vast knowledge of the world and billions of years of evolution, solves easily. But for computers, whose world consists only of numbers, it is a challenging task.

At the time of this project, the dominant branch of artificial intelligence was symbolic AI, also known as rule-based AI: Programmers manually specified the rules for detecting objects in

images. But the problem was that objects in images could appear from different angles and in various lighting. The object might appear against a range of different backgrounds or be partially occluded by other objects. Each of these scenarios generates different pixel values, and it's practically impossible to create manual rules for every one of them.

Naturally, the Summer Vision Project didn't get far and yielded limited results. A few years later, in 1979, Japanese scientist Kunihiko Fukushima proposed the neocognitron, a computer vision system based on neuroscience research done on the human visual cortex. Although Fukushima's neocognitron failed to perform any complex visual tasks, it laid the groundwork for one of the most important developments in the history of computer vision.

The Deep-Learning Revolution

In the 1980s, French computer scientist Yan LeCun introduced the convolutional neural network (CNN), an AI system inspired by Fukushima's neocognitron. A CNN comprises multiple layers of artificial neurons, mathematical components that roughly imitate the workings of their biological counterparts.

When a convolutional neural network processes an image, each of its layers extracts specific features from the pixels. The first layer detects very basic things, such as vertical and horizontal edges. As you move deeper into the neural network, the layers detect more-complex features, including corners and shapes. The final layers of the CNN detect specific things such as faces, doors, and cars. The output layer of the CNN provides a table of numerical values representing the probability that a specific object was discovered in the image.

Top layers of neural networks detect general features; deeper layers detect actual objects (Source: arxiv.org)

LeCun's convolutional neural networks were brilliant and showed a lot of promise, but they were held back by a serious problem: Tuning and using them required huge amounts of data and computation resources that weren't available at the time. CNNs eventually found commercial uses in a few limited domains such as banking and the postal services, where they were used to process handwritten digits and letters on envelopes and cheques. But in the domain of object detection, they fell by the wayside and gave way to other machine-learning techniques, such as support vector machines and random forests.

In 2012, AI researchers from Toronto developed AlexNet, a convolutional neural network that dominated in the popular ImageNet image-recognition competition. AlexNet's victory showed that given the increasing availability of data and compute resources, maybe it was time to revisit CNNs. The event revived interest in CNNs and triggered a revolution in deep learning, the branch of machine learning that involves the use of multi-layered artificial neural networks.

Thanks to advances in convolutional neural networks and deep learning since then, computer vision has grown by leaps and bounds.

Applications of Computer Vision

Many of the applications you use every day employ computer-vision technology. Google uses it to help you search for objects and scenes—say, "dog" or "sunset"—in your Images library.

Other companies use computer vision to help enhance images. One example is Adobe Lightroom CC, which uses machine-learning algorithms to enhance the details of zoomed

images. Traditional zooming uses interpolation techniques to color the zoomed-in areas, but Lightroom uses computer vision to detect objects in images and sharpen their features when zooming in.

One field that has seen remarkable progress thanks to advances in computer vision is facial recognition. Apple uses facial-recognition algorithms to unlock iPhones. Facebook uses facial recognition to detect users in pictures you post online (though not everyone is a fan). In China, many retailers now provide facial-recognition payment technology, relieving their customers of the need to reach into their pockets.

Advances in facial recognition have also caused worry among privacy and rights advocates, though, especially as government agencies in different countries are using it for surveillance.

Content moderation is another important application for computer vision. Companies such as Facebook must review billions of posts every day and remove images and videos that contain violence, extremism, or pornography. Most social-media networks use deep-learning algorithms to analyze posts and flag those that contain banned content.

RECOMMENDED BY OUR EDITORS

[What Is Artificial Intelligence \(AI\)?](#)

[Artificial Intelligence Has a Bias Problem, and It's Our Fault](#)

[What Is Machine Learning?](#)

Moving on to more specialized fields, computer vision is fast becoming an indispensable tool in

medicine. Deep-learning algorithms are showing impressive accuracy at analyzing medical images. Hospitals and universities are using computer vision to predict various types of cancer by examining x-rays and MRI scans.

Self-driving cars also rely heavily on computer vision to make sense of their surroundings. Deep-learning algorithms analyze video feeds from cameras installed on the vehicle and detect people, cars, roads, and other objects to help the car navigate its environment.

The Limits of Computer Vision

Current computer-vision systems do a decent job at classifying images and localizing objects in photos, when they're trained on enough examples. But at their core, the deep-learning algorithms that power computer-vision applications are matching pixel patterns. They have no understanding of what's going on in the images.

Understanding the relations between people and objects in visual data requires common sense and background knowledge. That's why the computer-vision algorithms used by social-media networks can detect nude content but often struggle to tell the difference between safe nudity (breastfeeding or Renaissance art) and banned content such as pornography. Likewise, it's hard for these algorithms to tell the difference between extremist propaganda and a documentary about extremist groups.

Humans can tap into their vast knowledge of the world to fill the holes when they face a situation they haven't seen before. Unlike humans, computer-vision algorithms need to be thoroughly instructed on the types of objects they must detect. As soon as their environment contains things that deviate from their training examples, they start to act in irrational ways, such as failing to detect emergency vehicles parked in odd locations.

For the moment, the only solution to solving these problems is to train AI algorithms on more and more examples, hoping additional data will cover every situation the AI will face. But as experience shows, without situational awareness, there will always be corner cases—rare situations that confound the AI algorithm.

SITUATIONS THAT CONFOUND THE AI ALGORITHM.....

Many experts believe that we will only achieve true computer vision when we create artificial general intelligence, AI that can solve problems in the same way as humans. As computer scientist and AI researcher Melanie Mitchell says in her book *Artificial Intelligence: A Guide for Thinking Humans*: "It seems that visual intelligence isn't easily separable from the rest of intelligence, especially general knowledge, abstraction, and language... Additionally, it could be that the knowledge needed for humanlike visual intelligence...can't be learned from millions of pictures downloaded from the web, but has to be experienced in some way in the real world."



This AI-powered robot can make 300 pizzas in an hour

Get Our Best Stories!

Sign up for **What's New Now** to get our top stories delivered to your inbox every morning.

Enter your email

Sign Up