What are Convolutional Neural Networks? video transcript

OK, pop quiz. What am I drawing?

I'm going to make three predictions here.
Firstly, you think at your house, you'd be right? Secondly, that that just came pretty easily to you, it was effortless. And thirdly, you're thinking that I'm not much of an artist, and you'd be right on all counts there.

But how can we look at this set of geometric shapes and think, Oh, how? If you live in a house, I bet it looks nothing like this. Well, that ability to perform object identification that comes so easily to us does not come so easily to a computer, but that is where we can apply something called convolutional neural networks to the problem. Now, a convolutional neural network, or a CNN, is an area of deep learning that specializes in pattern recognition.

My name is Martin Keane, and I work in the IBM Garage at IBM. Now let's take a look at how CNN works at a high level. Well, let's break it down.

CNN, convolutional neural network. Well, let's start with the artificial neural network part. This is a standard network that consists of multiple layers that are interconnected, and each layer receives some input, transforms that input to something else, and passes an output to the next layer. That's how neural networks work, and CNN is a particular part of the neural network, or a section of layers that say it's these three layers here.

And within these layers, we have something called filters. And it's the filters that perform the pattern recognition that CNN is so good at. So, let's apply this to our house example now. If this house were an actual image, it would be a series of pixels, just like any image. And if we zoom in on a particular part of this house, let's say we zoom in around here, then we would get, well, the window. And what we're saying here is that a window consists of some perfectly straight lines. Almost perfectly straight lines. But, you know, a window doesn't need to look like that window could equally look like this, and we would still say it was a window.

The cool thing about CNN is that, using filters, CNN could also say that these two objects represent the same thing. The way they do that, then, is through the application of these filters.

So, let's take a look at how that works. Now, a filter is basically just a three-by-three block. And within that block, we can specify a pattern to look for. So, we could say, let's look for a pattern like this, a right angle in our image. So, what we do is we take this filter, and it's a three-by-three block here. We will analyze the equivalent three-by-three block up here as well. So, we'll look at first of all, these first. Group of three-by-three pixels, and we will see how close are they to the filter shape? And we'll get that numeric score, then we will move across one, come to the right and look at the next three-by-three block of pixels and score how close they are to the filter shape. And we will continue to slide over or vote over all of these pixel layers until we have not every three-by-three block.

Now, that's just for one filter. But what that will give us is an array of numbers that say how closely the image matches filter, but we can add more filters, so we could add another three-by-three filter here. And perhaps this one looks for a shape like this. And we could add a third filter here, and perhaps this looks for a different kind of right-angle shape. If we take the numeric arrays from all of these filters and

combine them together in a process called pooling, then we have a much better understanding of what is contained within this series of pixels.

Now that's just the first layer of the CNN. And as we go deeper into the neural network, the filters become more abstract, and they can do more. So, the second layer of filters perhaps can perform tasks like basic object recognition. So we can have filters here that might be able to recognize the presence of a window or the presence of a door or the presence of a roof.

And as we go deeper into the CNN to the next leg, well, maybe these filters can perform even more abstract tasks, like being able to determine whether we're looking at a house or we're looking at an apartment or whether we're looking at a skyscraper. So you can see the application of these filters increases as we go through the network and can perform more and more tasks.

And that's a very high-level, basic overview of what CNN is. It has a ton of business applications. Think of OCR, for example, for understanding handwritten documents. Think of visual recognition and facial detection and visual search. Think of medical imagery and looking at that and determining what is being shown in an imaging scan. And even think of the fact that we can apply a CNN to perform object identification for badly drawn houses. If you have any questions, please drop us a line below, and if you want to see more videos like this in the future, please like and subscribe.

Thanks for watching.