NEUROMORPHIC COMPUTING: SMART ENVIRONMENTS THAT RESPOND TO YOUR GESTURES

VIDEO TRANSCRIPT

Hi I'm Karthik Narain and I lead Accenture cloud first we help businesses and governments optimize their performance and bring new ideas to life. Today I've been asked to explain cloud computing at three levels of complexity to three different people. A child, a college student, and an expert.
Cloud computing is a model that enables resources to be available on demand, configurable, and that can be shared. For organizations to be cloud first, they need to re-imagine their businesses as though they were reborn in the cloud. This would mean their products and services, the way they interact with their customers and partners, and how their employees work with each other need to be reimagined by leveraging the power of cloud.

Speaker 1: People use gestures to communicate with each other all the time. With the pandemic at the forefront of everyone's mind, there's been a renewed push for touchless interfaces. So for example, instead of having to touch kiosks at the store, at the mall, at the movie theater, what if you could use natural gestures to browse through pages, scroll, or make selections?

This summer, we hosted Kenneth Stewart, a graduate student in Emre Neftci's Lab from UC Irvine to work with Andreea Danielescu, a researcher at Accenture Labs. Kenneth and Andreea will now highlight work they're doing using neuromorphic computing paired with a neuromorphic camera to create a robust gesture recognition capability.

Kenneth Stewart: Everyone uses gestures and body language when talking to each other. So what if we could interact with technology as easily as you interact with other people? To select the movie you want to see at the theater or in your home, or to find out how to get to your favorite store in the mall. We believe that neuromorphic computing can play an important role in making this kind of interaction common place and are working on developing new algorithms to achieve this.

Automatic recognition of full body gestures, like waves, has been challenging because people can make the same gesture many different ways, and people have no problem recognizing these different gestures as having the same meaning because our brains adapt to the subtle differences in these gestures. To conventional AI, these gestures are completely different. Using neuromorphic computing, we can create AI models that, like our brains, can adapt to the subtle differences in people's gestures using synaptic plasticity. This feature of the Loihi research processor enables continuous on-chip learning and doesn't require the large amount of data that other AI approaches require.

Our approach combines the best of both conventional and neuromorphic hardware. State-of-the-art gesture recognition algorithms use deep convolutional neural networks on GPU's. So we first train a model using a deep spiking convolutional neural network and Loihi simulator on a GPU.

transfer the model to the Intel Loihi and retrain the last layer of the network using local learning rules that can learn from new input data in real time as seen here. We also leverage an event based camera, which is effectively a Silicon retina that mimics how human retinas see what's around us. The Silicon retina can detect changes up to 30,000 times faster than an RGB camera with virtually no motion blur, making it ideal for seeing dynamic movements such as gestures.

Here's an example. First, I teach Loihi how I wave, then I try waving in a different way with a larger, faster motion. After seeing only one example of both types of waves, Loihi can recognize both as waving gestures. We can use the same concept to easily recognize different people's waves and compare the similarity of those waves using variational autoencoders or VAEs. VAEs can be used to learn relationships between data points and complex data, such as gestures. The data can be encoded into a latent space representation that can be used to analyze the relationships between data points. The representation captures relationships such as nearby pixels being organized into objects, such as right hand versus a left hand. By calculating the latent space representation of our gesture set, we can compare similarities between people's gestures.

Here, we use a hybrid spiking neural network and artificial neural network VAE. The spiking network encodes gestures, and the artificial neural network decodes the latent space representation into an image for visualization.

Because VAE's are an unsupervised learning method, we could learn from any new data that is provided after we train our model without needing a label. This allows us to do online, self-supervised learning of new people's gestures by comparing the latent space representation of the new gestures to existing gesture data and applying a label automatically for learning. With this new approach, using neuromorphic computing, people will be able to naturally interact with technology through gestures to watch TV, browse the directory at the local mall and much more.