Posted on MAY 11, 2017 by SAMANTHA ZEE

Whose Sign Is It Anyway? AI Translates Sign Language Into Text

Deaf people can't hear. Most hearing people don't understand sign language.

That's a communication gap AI can help bridge, Syed Tousif Ahmed, a research assistant at the Rochester Institute of Technology's Future Everyday Technology Lab, explained at the GPU Technology Conference this week.

Ahmed and his colleagues are using computer vision, machine learning and embedded systems to transform American Sign Language into words that can be read on a video screen.

"Bridging this gap means a hearing person can interview a deaf person, or someone who is hard of hearing, via Skype or Google Hangout," Ahmed said. "They can hold a meeting or do job interview, and just communicate in a natural way."

Syed Ahmed, research assistant at Rochester Institute of Technology, speaking at GTC 2017.

Real-Time Video Captioning

Ahmed detailed how to build a complete video captioning system focused on American Sign Language using deep neural networks. The goal: a messaging app that would let a hearing person reply through automatic speech recognition, and a deaf person reply through a video captioning system.

"Another application could be an ASL learning app, where those using American Sign Language would be able to evaluate their proficiency through the video captioning," Ahmed

said. "Wouldn't it be great to get a score so you know that your sign language is acceptable?"

Using TensorFlow, Ahmed developed a neural network for his sequence to sequence network, which learned the representation of a sequence of frames to decode the information into a sentence that describes an event in the video. The images are encoded, processed into a feature vector and then decoded.

The system's additional features include caption-generation, a data input pipeline and use of the open-source Seq2Seq encoder-decoder framework to create the models. The system is then deployed on embedded platforms, like the NVIDIA Jetson TX2, for real-time captioning of live videos.

Each aspect of the system, from interpreting lip reading to physical motions, is layered upon another to help ensure future communication is effortless for everyone.

Raw video and captions used in training.

CATEGORIES: Deep Learning

TAGS: Education | GTC 2017 | Jetson