AI-Based English to American Sign Language (ASL) Translation Enabled by the Largest Continuous ASL Dataset

Jiang, Yehong (School: The Nueva School)

The goal of this project is to develop an automatic American Sign Language (ASL) production system that generates high-quality ASL captions for multimedia content, making it accessible to the deaf community. To provide an end-to-end solution, the project overcomes challenges in three areas: dataset, AI modeling, and ASL presentation. Despite significant progress in natural language translation through deep learning, ASL translation has been hindered by the lack of large-scale datasets. I developed a system that generates English-to-ASL sentence-pair datasets. The current dataset has 131-hour video, 5x and 1.6x bigger than Microsoft's MS-ASL and Facebook's How2Sign. To ensure data accuracy, I created a technology to standardize the ASL skeleton poses, such as head and shoulder positions, resulting in a 3x reduction in training error. The dataset continues to grow. I plan to publish it to promote deep-learning research in ASL. As ASL has different grammars, word-to-word translation cannot be applied. I developed a Transformer neural network that enables English-to-ASL translation without intermediate gloss representation. The network was enhanced with Keypoint Scaling and Normalization techniques, which improved its ability to distinguish fine joint structures and restore small movements. And multi-modality was employed to enhance the results obtained from the Transformer neural network. Using a medium size dataset and corresponding vocabularies, the training accuracy reached 80%. To make ASL translations more human-like, the neural-network-generated pose sequences are mapped to an armature model in Blender to create photorealistic 3D-rendered ASL signers. This system successfully generates ASL captions for YouTube video, online classes and movie clips.

Awards Won: