DeepLetters: A Convolutional Long Short-Term Memory (CNN-LSTM) Approach to Fingerspelling Translation

White, Riley (School: Stone Bridge High School)
Maheshwari, Saarthak (School: Stone Bridge High School)

Although many language-to-language conversion technologies have been optimized to automate translation at a fast and accurate rate, automating the conversion of sign language to text has been an ongoing challenge because sign language is not a written language. With over 2 million Americans classified as deaf, it is important to diminish the communication obstructions between the hearing and hard-of-hearing communities. Using artificial intelligence, new strides in the translation of sign language to text are being made. Past researchers have attempted to mitigate this problem using artificial intelligence, but their translators were inaccurate outside of training and testing data. This paper presents a unique convolutional neural network long short-term memory (CNN-LSTM) approach to translating American Sign Language (ASL) fingerspelling. The GoogLeNet previously trained on various hand images was retrained on thousands of static fingerspelling images, acting as a character-based model. This retrained CNN was linked into a LSTM network in order to better handle time-sequence data. Individual video frames were given to the CNN, the output being feature vectors given to the LSTM. The LSTM acts as a letter-based model, and was trained using fingerspelling video data. Two accuracy indexes were created: one from the static hand image network and one from the LSTM. The static hand image network was able to achieve 89.76% accuracy. The LSTM network data are pending.