Abstract

Communication is still difficult for people who have difficulty hearing, especially since there aren't trained interpreters available. Visual conversation is possible with sign language, but more automated technologies are needed to make it easier for everyone to use. Many thanks to convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, two new ideas in the field of deep learning, they have both shown promise in tasks that require recognizing gestures.CNNs are better at getting spatial information from images than LSTMs are at finding temporal connections between video frames. The LSA64 dataset has 3,200 videos of 10 people making 64 different signs. This study uses this dataset to suggest a CNN-LSTM model for recognizing Argentinian Sign Language. Aligning time and making sure frames are all the same as preparation steps. The hybrid architecture includes CNN for extracting spatial features and LSTM for modeling temporal sequences. This makes it possible to have a strong understanding of both space and time. We compared how well the CNN and LSTM models worked on their own to how well the combination model worked. The hybrid CNN-LSTM model showed promise in automatic sign language recognition by doing better than both separate models in the experiment, with an F1-score of 92.9% and an accuracy rate of 92.9%.

Keywords—sign language, motion detection, deep learning, CNN, LSTM