A Novel Lip-Reading Method Based on Transfer Learning from MobileNet and LSTM Architecture

Ji, Xingyu(Carl) (School: Portsmouth Abbey School)

Children with cleft palate often suffer from speech disorders. Children need guidance from professional speech therapists to teach them to pronounce correctly. Speech therapists and similar organizations are hard to find in many regions of the world. Even with the opportunity to receive surgeries, kids may still not be able to change their way of speaking. I realize the difficulties in completely altering the process of speech therapy, but I can use technology in the field of deep learning to increase the efficiency of it. The aim of my project is to construct a neural network that takes inputs from the patients, determine whether the input matches the correct mouth poses or not, and finally gives the patient feedbacks on each word. One of the most important aspects of speech therapy is to train kids to pose their mouths correctly while speaking. Even though the main process remains personal training with therapists, the duration of training can be shortened if children can practice on their own and gain immediate feedback on their performances. The main approach of this project is using convolutional neural networks and recurrent neural networks to process the data inputs, mostly video clips of patients speaking, and output to the patient if the input corresponds with correct poses that the model is trained on.

Awards Won: