

Visual Speech recognition for Sinhala language using CNN

Jayarathne W. M. U.^{*1}, Perera W. A. S. C.¹, Ketheesan T.²

¹*Vavuniya Campus of the University of Jaffna*

²*Faculty of Technology University of Jaffna*

Visual Speech Recognition (VSR) is an essential tool that is facilitating to understand the speech from the video by the visually impaired people. Moreover, VSR play an important role in analyzing the CCTV footage for a crime investigation where the audio is not available. On the other hand, VSR system for Sinhala language still under research not explored largely. Hence in this research, a preliminary research work is carried out to understand the suitability of convolutional neural network (CNN) to recognize the Sinhala character from the image which contain the mouth region. The proposed methodology train the CNN with the help of lip pose features and corresponding character label. The architecture of the CNN employees' three convolution layers, two fully connected layers and one max pool layer. There is no data set available publicly for Sinhala language visual speech recognition and for the evaluation of the system, own data set was created for five Sinhala characters that has phonetics sound a, e, i, l, m. The data set was augmented to increase the feature domain and the outliers are removed to overcome the ambiguity. The system was trained with fifteen images and tested with ten images, those are containing the lip pose when pronounce five sounds. For the evaluation purpose the confusion matrix is analyzed and the accuracy was determined by the F_1 score. The F_1 score is calculated using the precision and recall and found 0.83, it means that the proposed methodology performs well.

Key words: *CNN, Sinhala, character, visual*

*Corresponding author: upalakshi@gmail.com