SPEECH RECOGNITION USING DEEP LEARNING

Dr. Narendrababu Reddy G

Speech Recognition using Deep Learning

Dr. Narendrababu Reddy G,
Assistant. Professor, Dept of CSE,
G Narayanamma Institute of Technology and Science Shaikpet, Hyderabad.

PREFACE

Lip reading is a method of determining speech by looking at the movement of lips. Audio visual speech recognition (AVSR) is an approach that uses image processing abilities in lip reading to assist speech recognition systems. This system focuses on multimodal sensor architectures with deep learning for audio-visual speech recognition on the wild scenarios, particularly when the audio is corrupted by noise. It is a combination of both audio part and visual part, which implies integration of both lip reading and speech recognition processes working separately. In this research work, we propose new improvements for AVSR models by incorporating data augmentation techniques to generate more data samples for building the classification models. For the data augmentation techniques, we utilise a combination of conventional approaches (e.g., flips and rotations), as well as new approaches, such as generative adversarial networks (GANs). To validate the approaches, we used augmented data from well known datasets (LRS2-Lip Reading Sentences 2 and LRS3) in the training process and testing was performed using the original data. This framework enhances the performance of the AVSR framework in the wild for any datasets.

CONTENT