Customising Video Messages using GANS

by Darshan Umapathi

Supervisor Prof. Gerard Lacey

Abstract:

With an exponential rise in video messages on the internet, creating videos and customizing them quickly has become an essential need. Short videos are growing, and they tend to be engaging, for example, reels on Instagram, Facebook, TikTok, etc. However, customizing pre-recorded video is challenging, and modifications require either re-shooting the video with new phrases or using complex editing systems such as Adobe Photoshop or ProCreate to achieve desired results. These methods are expensive and time-consuming. This study proposes a novel ensemble model which helps to automatically translate text input to video to generate realistic lip synchronization output.

The proposed model takes an image or a video as input that is processed using a facial recognition system, and simultaneously the real-time text-to-speech model converts the transcripts to the actor's voice. Both the outputs are fed to Wav2Lip plus GANs model to incorporate lip-sync synthesizes. The output generates a new synthesized video from the text input and customizes messages by altering the text using the same driving video. The proposed approach is effective at incorporating low-resolution images or videos of talking head up to 45 seconds, and it is evaluated using the start-of-the-art lip synchronization model-SyncNet to produce a lip synchronized output video with Lip-Sync Error-Distance of 8.14 and Lip-Sync Error-Confidence of 5.65. The obtained results indicate an accurate and effective model compared to previously proposed techniques.