Deep Learning and its Applications Project Presentation on Animated News Anchor Group-15

Prashant Shekhar Gupta, Amit Rajain, Shubham Kumar

Indian Institute of Technology Mandi

April 18, 2019

Problem Statement

There has been a lot of advancements in technology till date. For most of the things we do, there is already an automated tool to do those things. But still there are some fields where we need humans to do any particular job. You must have seen robots to carry out some jobs. You must have seen them talking. But have you seen any robot talking and syncing his lips in the same way as humans? Reading out loud something with the same lip movement as humans is still a topic to be researched about. If one can achieve this, this would be a whole new revolution in this era of technology.

Motivation

As this is a very generalized problem, we can see its application in various fields.

1. **Making the dubbed movie more realistic**
2. **Dubbing the cartoon without hard coding**
3. **Dubbing Movies and Anime**
4. **Animated Character for News Anchoring**

Challenges

Challenges which we can face in this project when we go further are :-

1. **A proper dataset on which we need to train our model**
2. **Detecting lips of the subject in the video**
3. **Detaching lips from the video**
4. **Pasting the trained lips in the video**
5. **The network where the training will happen**
6. **The algorithm from which the training will be done**

Dataset

Our approach for dataset relies on the following pipeline:

During Training

1. **Collecting videos from youtube**
2. **Extracting their lips and labelling them to their corresponding Sound Waveform**
3. **Training the model on the dataset**

During Testing

1. **Collecting videos from youtube**
2. **Give the audio of video as input to the model and generating the lip video. After that, calculating the accuracy by comparing it to the original lip movement in the video**
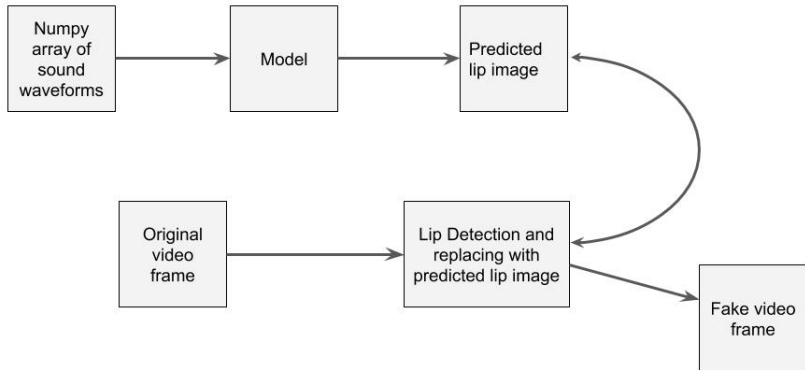
Proposed Methodology

Our approach for methodology relies on the following pipeline:

1. **Collecting video from various sources**
2. **Extracting audio waveforms and video frames in different numpy array**
3. **Audio waveform acts as the input features and video frames are respective ground truth**
4. **Training the model on this dataset.**

Training
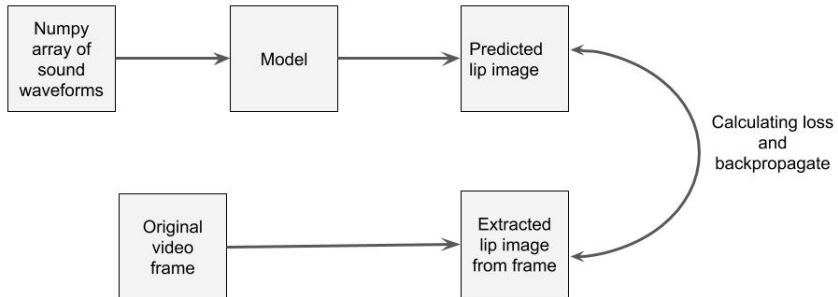
Video Modification

Methodology Explored

1. **We explored that our project is very much analogous to "Audio to Text converter".In both the project input is audio but the outputs are video(our project) and text.The video will be tokensied word by word and will be labelled with its corresponding audio, and then model will be trained with the data set.**

Result and discussion

1. **Till now we have prepared the dataset and we are working on the model which will be used for training.**

Conclusion and future work

1. **After completing the training we will be testing our algorithm on different videos and will try to improve the efficiency of conversion.**