Project Report

Title of the project   : Silent Speech Recognition
Project id   : SPS_PRO_3451
Challenge Name   : IBM Hack Challenge 2023
Date of submission   : 4th september,2023

Description:

The proposed solution aims to enhance lip reading accuracy and robustness by combining 3D Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRUs).

Plan of action to implement the proposed lip reading model:

1)install and import dependencies
2)Building data loading Functions
3)Creating Data Pipeline
4)Designing the Deep Neural network
5)Setting up Training options and train
6)Making prediction

Technologies I used are:

1)Deep Learning
2)Machine Learning
3)Python
4)Python libraries
5)Flask

6)opencv
7)Tensorflow
8)Advance Ai Models
9)IBM Watson Studio

Aim:

The aim of this project is to develop a accurate silent speech recognition system that can convert visual lip movements into text or speech without the need for audio input.This technology can benefit people with hearing impairments and enable silent communication.

Meet the requirements:

Improved Lip Reading Accuracy: By combining 3D CNNs and GRUs, the model can effectively capture both spatial and temporal information from lip movements. This comprehensive understanding allows the system to make more accurate predictions of the spoken words, benefiting individuals with hearing impairments who rely on lip reading as well as in scenarios where audio information is compromised.

Generalization: Deep learning techniques, such as 3D CNNs and GRUs, have shown strong generalization capabilities when trained on large and diverse datasets. This ensures that the proposed lip reading model can be applied to a wide range of scenarios.

1)Dependencies I included in this project

I used open cv to preprocess my data,to read my data.
Mathplotlib to see the results of preprocessed videos.
Image io to create gifs of frames I created.

gdown to download dataset where I have given google drive link. tensorflow used to build a deep learning network.

2)Building data loading functions

I extracted dataset from google drive using gdown and downloaded. I used face detector to extract lips frames and can dlib to isolate mouth.I used static slicing function to isolate mouth.After completing load video function creating,I defined vocabulary that used in this dataset.And to functions to convert from char to num and num to char.A function to load alignments from dataset and another function to attach alignments and video simultaneous.Setting a frame to identify mouth to read the data and ploted it using matlab plot.

3)Creating data Pipeline

Now given path of dataset.In each batch giving two sets of data with 75 frames in each data.After I created gif using imageio of data.With imageio we can see each frame as gif.Used image.minsave function to convert numpy to array tnto a gif.

4)Designing deep neural network

I used LSTM Neural Network.I used models,layers,optimizers,callbacks from tensorflow to design.I imported densor layer,dropout layer,bi directional layer,Maxpool 3D,activation and many more layers from tensorflow keras.i passed data through convolution with an relu activation with maxpool layer.

5)Setting up Training options and making prediction

I defined CTC loss function.Trained the model on the preprocessed dataset

using appropriate loss functions and optimization algorithms.

```
Downloading...
From: https://drive.google.com/uc?id=1vWscXs4Vt0a_1IH1-ct2TCgXAZT-N3_Y
To: /content/checkpoints.zip
100%|████████████| 94.5M/94.5M [00:00<00:00, 320MB/s]
1/1 [==============================] - 0s 282ms/step
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ REAL TEXT
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ PREDICTIONS
1/1 [==============================] - 1s 1s/step
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ REAL TEXT
[<tf.Tensor: shape=(), dtype=string, numpy=b'bin red at s nine again'>]
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ PREDICTIONS
[<tf.Tensor: shape=(), dtype=string, numpy=b'bin red at s nine again'>]
```