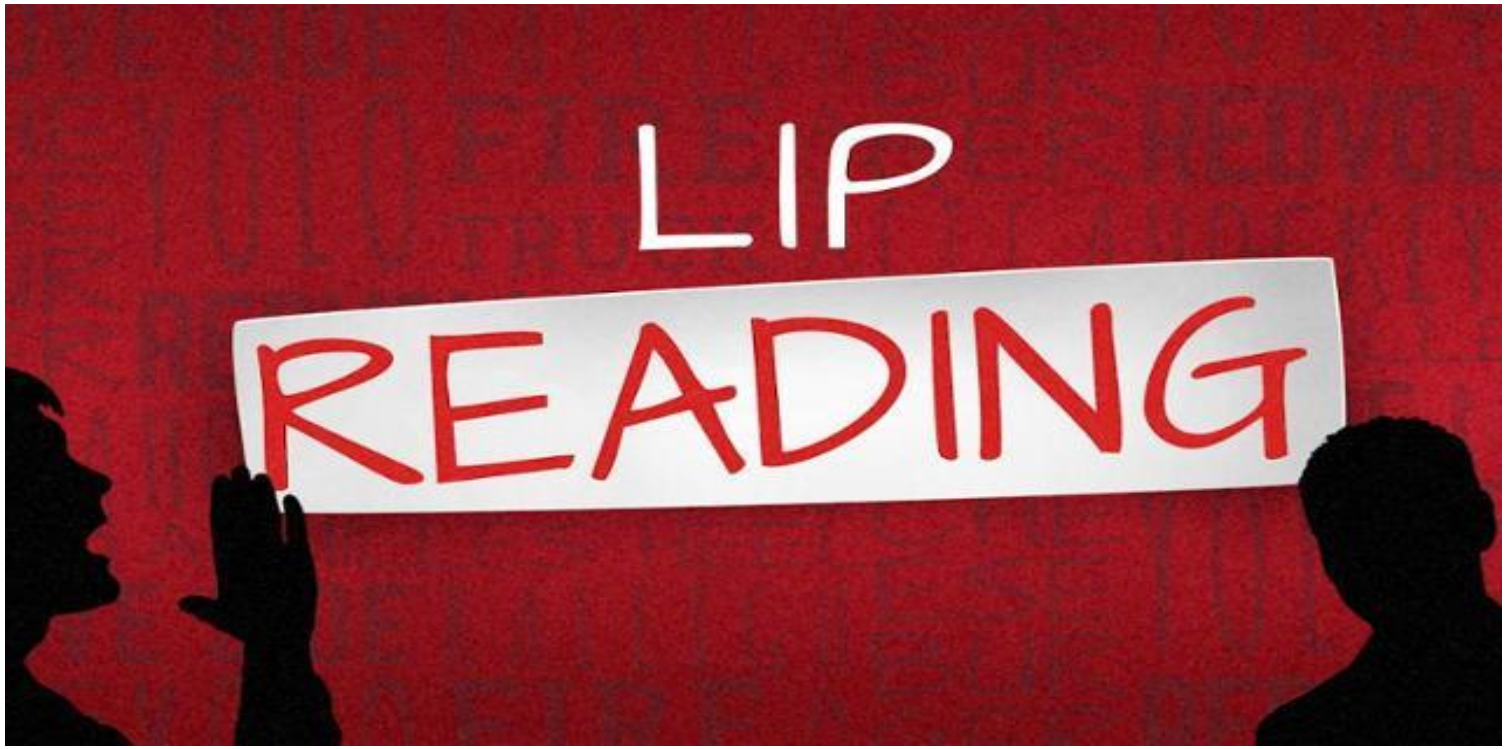


Big Data Content Analytics,

Spring Semester

Professor: C.Papageorgiou

Student: Karala Aikaterini, I.N: p2821810

In our assignment we are going to train a model, which is able to recognize a word or a phrase which told from specific speaker at specific time.

Below, you can see the word “PLEASE”

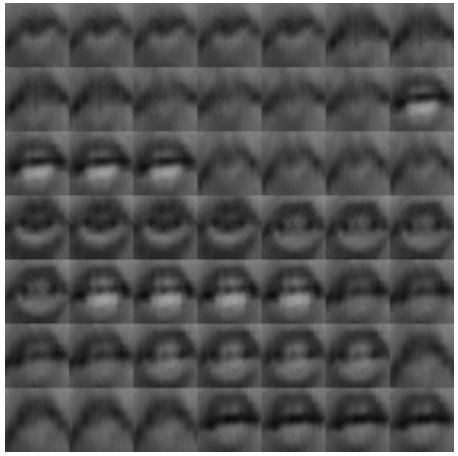
Our database is MIRACL-VC1 (downloaded from kaggle) which provides 3000 different videos, and 15 different speakers (5 men and 10 women) are talking in front of the camera . These specific words and phrases are represented below:

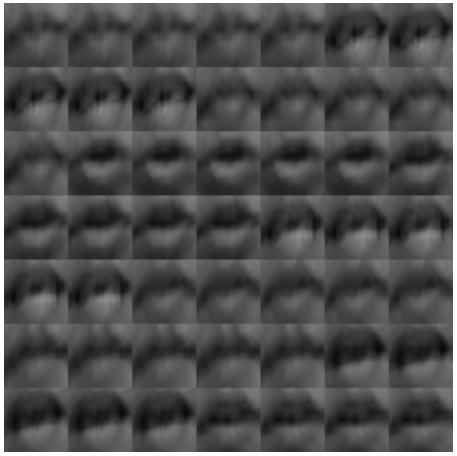
ID	Words	ID	Phrases
1	Begin	1	Stop navigation.
2	Choose	2	Excuse me.
3	Connection	3	I am sorry.
4	Navigation	4	Thank you.
5	Next	5	Good bye.
6	Previous	6	I love this game.
7	Start	7	Nice to meet you.
8	Stop	8	You are welcome.
9	Hello	9	How are you?
10	Web	10	Have a good time.

An example as far as our database is depicted below:

Preprocessing

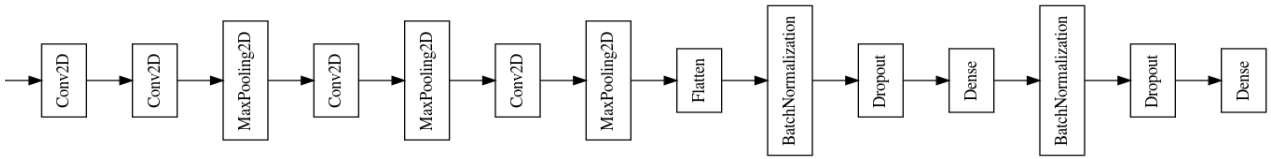
- In the first step we cropped the image and finally focuses on speakers' lips only using the classifiers and dlib libraries.
- Secondly, we order all these cropped images to only one and we have different frames from which image
- We also use Keras library in order to train our model
- A sequence of different frames is depicted below :





Method

- Our architecture is represented below:



- Our architecture is consisted of four levels , which includes max pooling layer which reduces dimension layers.
- The batch normalization layer converts its inputs in order to have the mean equals to zero and the median equals to one.

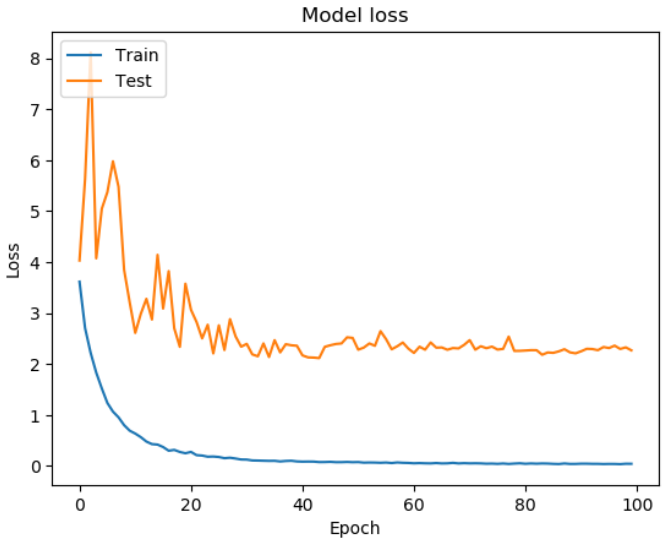
Train the model

- ✓ We divided our database into three parts : 1) train 2) test and 3)development. We used the Stochastic Gradient Descent .

- ✓ We wrote our code in Python and we visualize our outputs with use of matplotlib library.

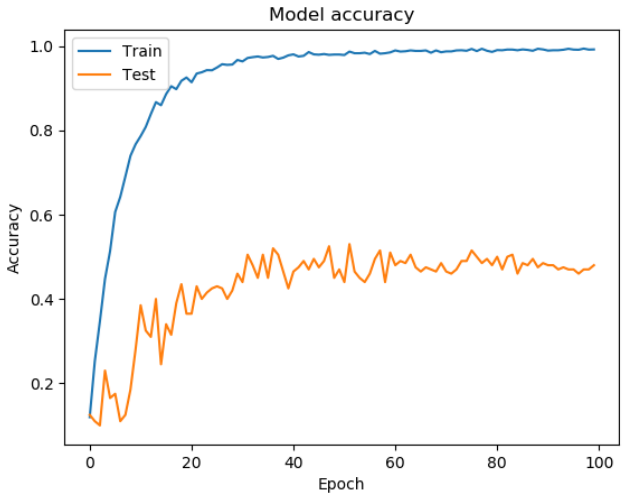
Outputs

- Out diagram with model loss is represented:



Outputs

- Out diagram with model accuracy is represented:




BabUlJannah
THANK YOU
FOR YOUR
TIME