iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 9 **Issue:** VIII **Month of publication:** August 2021

**DOI:** https://doi.org/10.22214/ijraset.2021.37414

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com*

# Text Recognition Using Silent Speech

Dr. Girish Kumar[1], Akhila K[2], Harshitha Reddy RV[3], Kavya[4], Nanditha A[5]
[1, 2, 3, 4, 5]*Research Scholar, VTU, Karnataka*

***Abstract: Our objective is to identify the characters from the quite speech of the English language. We tend to focus on the lip region to recognize the characters spoken clearly in the video. Our contribution is: foremost, this model is developed by using a pipeline method form absolutely automatic information assortment from the video. Though this, it generates a data set that is spoken by the individuals. Secondly, it is developed by using the machine learning algorithm Convolution Neural Network (CNN) that learns the lip motion. Thirdly, Convolution network turn out the efficient result by examining the video and also the data set.***

## I. INTRODUCTION

Lip reading is used to understand or explore speech without understanding it. Text recognition in videos is a research area which develops a computer system with the ability to automatically read from the images and videos the text content visually embedded in the background. The lip movement is a continuous process with a time interval, visual content can be represented by consecutive frames. Therefore, we proposed neural networks Convolution of Neural Network (CNN) and Long Term Short Memory (LSTM). We extract the lip region in the video or image is done by convolution neural network model. Finally, the text recognition will be done.

## II. LITERATURE SURVAY

As many People who are Dumb and Deaf who are suffering from there disability in order to come up with solution in
In[1] the author discussed about the developing the pipeline for recognitionation of words and query expansion technique that spotting 35% higher average precisiom over the recognition.
[2] In this paper he disused about the extraction of the features from image based on automatic lip reading.
In [3] the author discussed about identifying the character from the movement of mouth and skin area that which it is difficult and challenging task using convolutional Neural Network (CNN). The goal of this is to work is to improve performance of lip reading.
In [4] the author explained about how to frame the sentences being spoken in the video with or without audio.
[5]In this paper the author discussed about the input for the application is will be the raw video it needs to detect the face and mouth portion and understand the alphabet feature patterns of the lip movement by differentiate width and height in pixels of lip.

### *A. Implementation*

It involves 3 steps:

*1)* Convolution Neural Network (CNN)
*2)* Long Term Short Memory (LTSM)
*3)* Multimodal Model

Convolutional Neural Network (CNN): Is a type of Neural network model that uses to extract representational for the image content. CNN will take the image's raw pixel data. raw pixel are that recognize the lip movement of human in the video. This identifies the words from that of the lip movement. It involves of two section

*a)* First section is that it recognize the face of human in the video. During the features detection the network will perform a series of convolution and pooling operations. The pooling function is to reduce the spatial size of the representation in the network. This pooling reduces the dimension by capturing the lip region and increases the scaling.
*b)* Second section that involves identifying set of data based the image probability.

©IJRASET: All Rights are Reserved

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com*

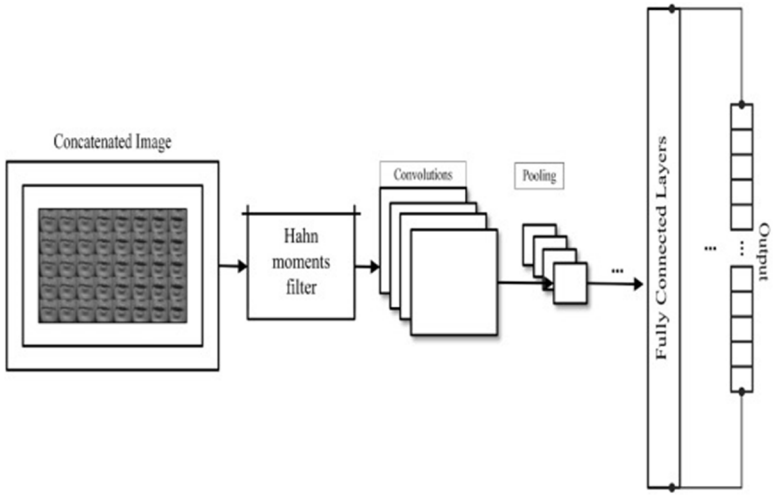


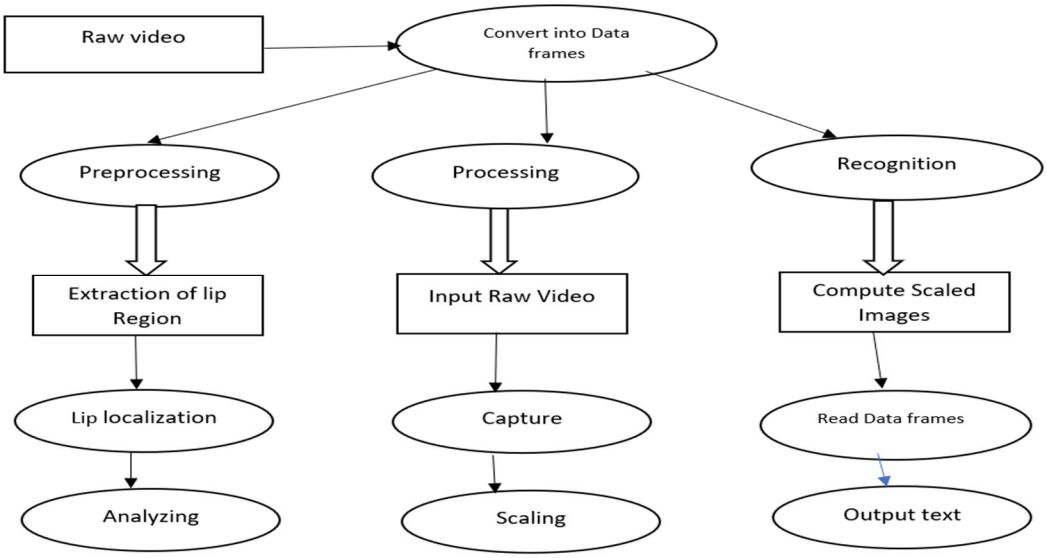

Fig 1: convolution and pooling method

Long Term Short Memory (LTSM): This is the second process it is an artificial recurrent neural network used in the fields of the deep learning. This process that is used for the lip classification tasks and for this type of the application is also keeps how well it may works. This modes that involves of the three gates.

- First gate that is generally an input gate that takes the video images as the input or video by the pixel
- Second gate is the forget gate this gate removes all the unnecessary data, unwanted data and background noise of the video and only the processed image and give the only lip reading.
- Third gate is output gate. The output is in the form that prints the output in text of format.

LSTM networks are suitably for classifying, processing and making the predictions.
Multimodal Model: This is the third process the multimodal model process that consists of visemes. Visemes is any of several speech that looks like same like the lip reading. Recognizing the lip movement that presents many of the challenges. Many of the features used in classification that calculates the shapes of the lips. This recognize each shape will be recorded
This is the method that captures the images that is received it scans only the lips of the human as the motion of the lips will only required. the lips region is scanned based on width and height ratio. the lips shapes and motions are converted into the characters and digits and that is displayed as the output in English language on the screen.

*B. Design (Flow Diagram)*




©IJRASET: All Rights are Reserved

## III. RESULTS

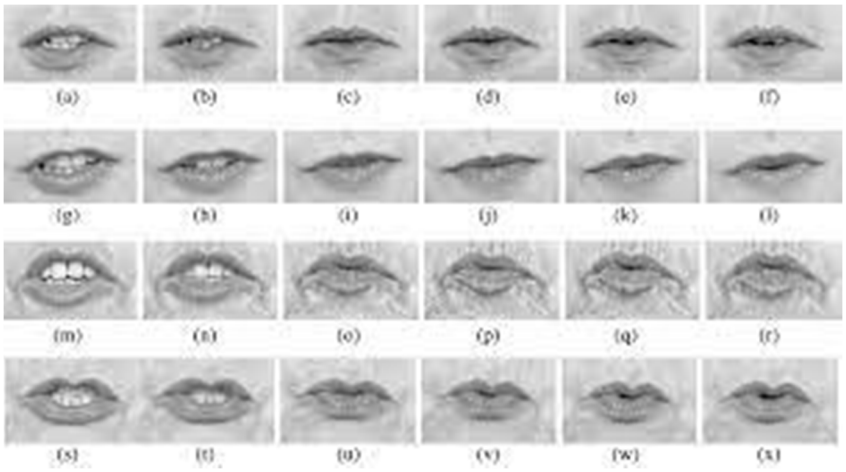



Fig 2: Shapes of Alphabet recorgizing







Fig4: Step by Step removing Unwanted data.

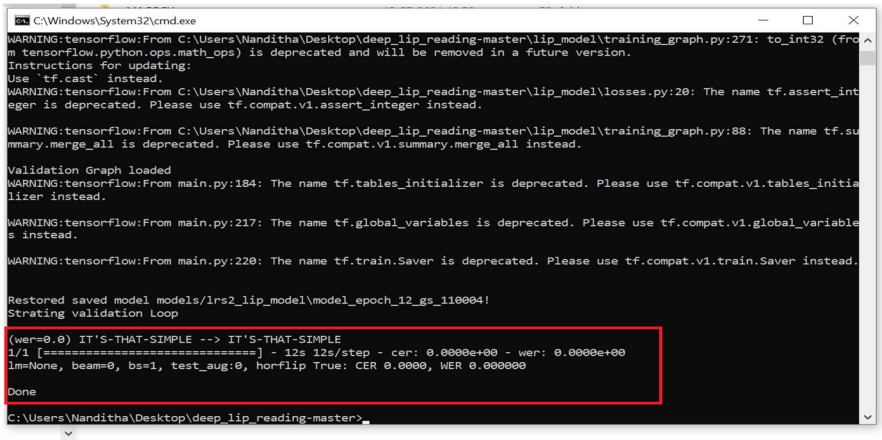



Fig 5: Output of Text

©IJRASET: All Rights are Reserved

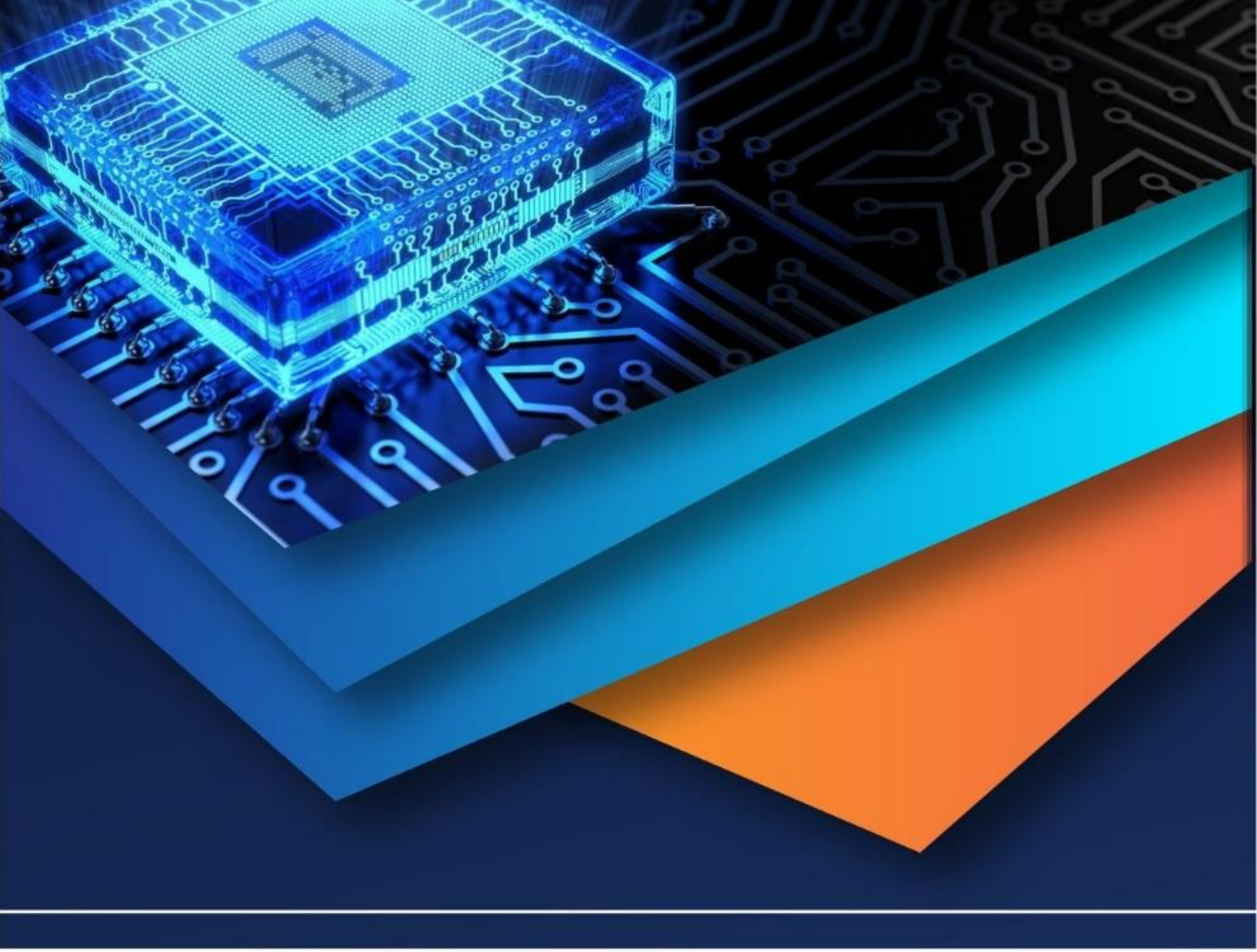
## IV. CONCLUSION

In this Paper, Text Recognizing using silent speech which we have classified different lip movement motion and recognizing text, using CNN that distinguish between the lip portion and the skin area. In which the tool used in neural network. Therefore which it helps for dumb people as they are unable to speak because of there Vocal Disability and deaf people can not able to hear voice .hence this paper provides them application that recognize the text from silent video that we provide.

## REFERENCES

[1] AbhishekJha,Vinay P. Namboodiri,C. V. Jawahar,"Spotting words insilent speech videos: a retrieval-based approach,"08 February 2019.
[2] Yuanyao Lu *, Hongbo Li,“ Automatic Lip-Reading System Based on Deep Convolutional Neural Network and Attention-Based Long Short-Term Memory,“ Article,2019.
[3] Youda Wei, Xiaodong Hu*, “ Text Recognition from Silent Lip Movement Video,“ IEEE 3rd International Conference on Signal and Image Processing, 2018.
[4] TriantafyllosAfouras, Joon Son Chung, Andrew Senior, Oriol Vinyals, Andrew Zisserman,“ Deep Audio-Visual Speech Recognition,“arXiv:1809.02108v2[cs.CV], 22 Dec2018.
[5] Thein Thein,KalyarMyo San,“ Lip Movements Recognition Towards An Automatic Lip Reading System for Myanmar Consonants,“ 2018.
[6] Joon Son Chung∗, Andrew Zisserman,“ Learning to lip read words by watching videos,“ Article 1077-3142,2018.
[7] ZiadThabet,AmrNabih,KarimAzmi,YoussefSamy,GhadaKhoriba,MaiElshehaly,"Lipreading Using a Comparative Machine Learning Approach,"978-1-5386-5083-7/18/$31.00 IEEE, 2018.
[8] Joon Son Chung,AndrewSenior,OriolVinyals,Andrew Zisserman,“ Lip Reading Sentences in the Wild,“ IEEE.
[9] Leticia Ria Aran, Farrah Wong and Lim Pei Yi,“ A Review on Methods and classifiers in lip, “ IEEE 2nd International Conference on Automatic Control and Intelligent Systems, 21 October 2017
[10] Girish Kumar D, Rajashree V Biradar, V C Patil, “Multiple Authority Based Data Fragmentation Technique for Providing Secure Storage in Cloud”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019.
[11] D. Girish Kumar D, Rajashree V. Biradar, and V. C. Patil, “A Novel Technique for Data Integrity Using Block Verification in Cloud Computing”, Journal of Computational and Theoretical Nanoscience Vol. 15, 1–3, 2018
[12] Rajashree V Biradar , Girish Kumar D, “Blockcloud: Integrating Blockchain Security Features in Cloud Computing”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-4, April 2020
[13] D. Girish Kumar,Dr. Rajashree V. Biradar, Dr.V.C. Patil,” A Study on Various Cryptographic Techniques Used in Cloud Computing” Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 03-Special Issue, 2018.
[14] Girish kumar D, Rajashree V Biradar, dr. V C Patil, “ A Review on Secure Storage Using Bidirectional Verification Techniques in Cloud Computing” International Journal of Engineering & Technology, 7 (4.36) (2018) 511-514
[15]

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)