IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

SILENT SOUND TECHNOLOGY

[1]CH. Meghams,[2]M. Mounika,[3]M. Tejesh,[4]Yelithoti Sravana Kumar,

[1,2,3]U. G Scholars, Department of ECE, N S Raju Institute of Technology, Sontyam, Visakhapatnam, AP, India

[4]Assistant professor, Department of ECE, N S Raju Institute of Technology, Sontyam, Visakhapatnam, AP, India.

ABSTRACT:* *Without using real audio, silent sound technology uses visual interpretation of lip, mouth, and face movement to comprehend words or speech. This job is challenging because different dictions and speech articulations are used by different people. Keeping this in mind, this project demonstrates the effectiveness of machine learning by developing an automatic lip-reading system using deep learning and neural networks. Two different CNN models were trained on a portion of the dataset. Based on how well the taught silent sound Technology models predicted words using DL, they were assessed. A web application for real-time word prediction used the top performing algorithm. Based on this precision, this can be applied further and in many more uses.

KEYWORDS:* *Convolution Neural Network (CNN), Deep Learning (DL), Image Processing, Silent Sound.

I. INTRODUCTION

Silent sound Technology is a recent topic which has been a problematic concern to even expert lip readers. There is a scope for silent sound Technology to be resolved using various methods of machine learning. Silent sound Technology is a skill with salient benefits. Enhancement in silent sound Technology technology increases the possibility to allow better speech recognition in noisy or loud environments. A prominent benefit would be developments in hearing aid systems for people with hearing disabilities. Similarly, for security purposes, a silent sound Technology system can be applied for speech analysis to determine and predict information from the speaker when the audio is corrupted or absent in the video **[1].**

With the variety of languages spoken around the world, the difference in diction and relative articulation of words and phrases. It becomes substantially challenging to create a computer program that automatically and accurately reads the spoken words solely based on the visual lip movement of the speaker. Even the expert lip readers are only able to estimate about every second word **[3].** Thus, utilizing the capabilities of neural networks and deep learning algorithms two architectures were trained and evaluated. Based on the evaluation, the better performing model was further customized to enhanced accuracy. The model architecture with an overall better accuracy was implemented in a web application to devise a Realtime lip-reading system.

II. OBJECTIVE

The main objective of our paper is to help the people who are unable to speak but wish to speak and also help the people to talk the people in mobile phones who are in the crowd.

III. LITERATURE SURVEY

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS’12). Curran Associates Inc., USA, 10971105 [1].

Ahmad BA Hassanat. 2011. Visual Speech Recognition, Speech and Language Technologies, Prof. Ivo Ipsic (Ed.), ISBN: 978-953-307-322-4, InTech [2].

Abiel Gutierrez, their best model was the Fine-Tuned VGG+LSTM baseline. Data augmentation proved to be helpful only in instance of unseen people. Their baseline outperforms LSTM+CNN architecture. They achieved validation accuracy very close to 75% and test accuracy of 59% [3].

End to End Sentence Level Silent Sound Technology Yannis M Assael, 2016 Their model Lipnet achieved

95.2% accuracy in sentence level lipreading over human lipreaders. [4].

Yiting Li, Yuki Takashima, Tetsuya Takiguchi, and Yasuo Ariki. 2016. Silent sound Technology using a dynamic feature of lip images and convolutional neural networks. In 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). IEEE, 1–6 [5].

IV. SYSTEM MODEL

The proposed technique works based on Convolutional Neural Network which is a part of Deep Learning. Deep learning is a branch of machine learning which is completely based on artificial neural networks, as neural network is going to mimic the human brain so deep learning is also a kind of mimic of human brain. In deep learning, we don't need to explicitly program everything . The concept of deep learning is not new. It has been around for a couple of years now. It's on hype nowadays because earlier we did not have that much processing power and a lot of data. As in the last 20 years, the processing power increases exponentially, deep learning and machine learning came in the picture.

Deep Learning is a subset of Machine Learning that is based on artificial neural networks (ANNs) with multiple layers, also known as deep neural networks (DNNs). These neural networks are inspired by the structure and function of the human brain, and they are designed to learn from large amounts of data in an unsupervised or semi-supervised manner **[4].**

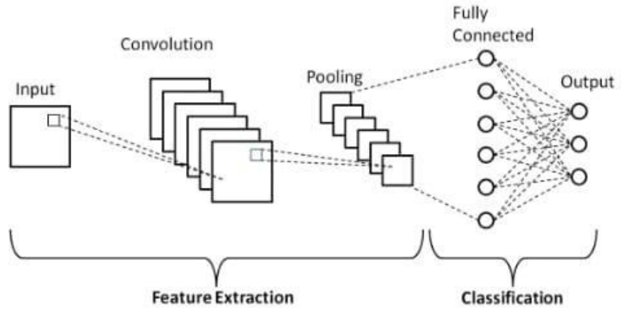


Fig. 1 Working of the CNN model

Convolutional Neural Network (CNN) is a class of Neural Network system in a standard multi-layered network. The layers comprise of a single or more layer connected in a multiple connection series. CNN is capable of utilising the local-connectivity in high dimensional data such as datasets composed of images and videos. This feature permits the applicability of CNN in the field of computer vision and speech recognition. A basic form of CNN has four significant parts: convolutional layer, activation function, pooling layer, and fully connected layer. Convolutional layer uses a set of learning filters to learn **[5].**

V. IMPLEMENTATION

The parameters from the input data. The activation function is a non-linear transformation function that defines the output of one node which is then the input for the next layer of neurons. In the pooling layer, the amount of parameters and computation in the network is reduced to control overfitting by decreasing the spatial size of the network. Fully connected layer takes the input volume from the convolutional layer or pooling layer to transform the result from the feature learning part to output **[2].**

This work provides an analysis of the employment of the temporal sequences using models like Hidden Markov Models and Recurrent Neural Networks (RNN), which is less capable of adjusting with the motion of the image. This lessens the predicted accuracy of the trained models. However, the debate for using CNN is its applicability of use in the moving subject with higher precision.

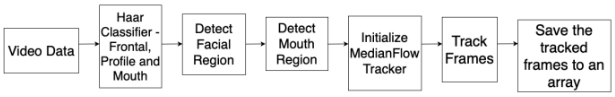

Video Data
Haar Classifier - Frontal, Profile and Mouth
Detect Facial Region
Detect Mouth Region
Initialize MedianFlow Tracker
Track Frames
Save the tracked frames to an array

Fig. 2 Lip Detection Process

In this proposed system Our aim is to design an autonomous Silent sound Technology system to translate lip movements in real-time to coherent sentences. We will use deep learning to classify lip movements in the form of video frames to phonemes. Afterward, we stitch the phonemes into words and combine these words into sentences

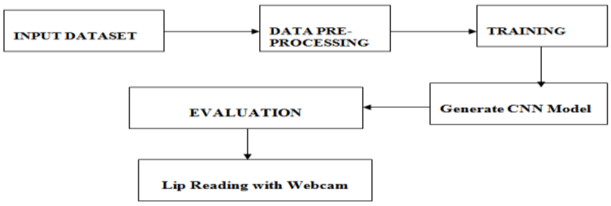


Fig. 3 Data Flow Diagram of System

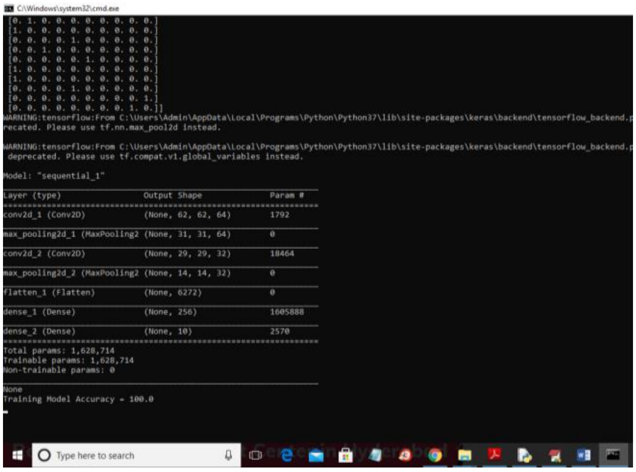
This overall lip detection process is used as datasets and to train the Deep Learning Model. Later the system works as shown in Data Flow Diagram (DFD).

VI. TESTING AND IMPLEMENTATION RESULTS

The purpose of this testing is to exercise the different parts of the module code to detect coding errors. After this the modules are gradually integrated into subsystems, which are then integrated themselves too eventually forming the entire system. During integration of module integration testing is performed. The goal of this is to detect designing errors, while focusing the interconnection between modules. After the system was put together, system is performed. Here the system is tested against the system requirements to see if all requirements were met and the system performs as specified by the requirements. Finally accepting testing is performed to demonstrate to the client for the operation of the system. Some Testing Strategies are:

- Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration.
- Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.
- System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.
- Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Following are some Images Taken from results and while implementing.

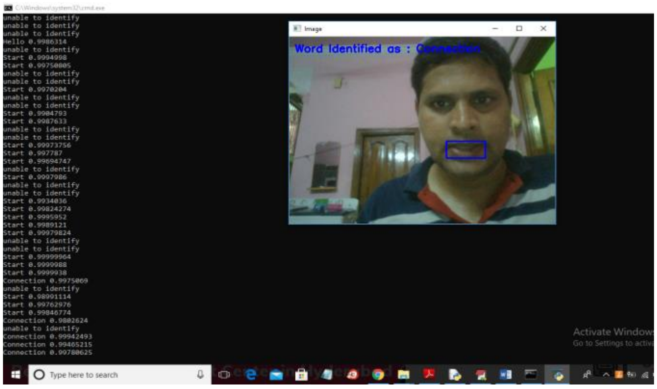


```
[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
[0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]
[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
[0. 0. 0. 0. 0. 0. 0. 0. 1. 0.]]
WARNING:tensorflow:From C:\Users\Admin\AppData\Local\Programs\Python\Python37\lib\site-packages\keras\backend\tensorflow_backend.p
recated. Please use tf.nn.max_pool2d instead.

WARNING:tensorflow:From C:\Users\Admin\AppData\Local\Programs\Python\Python37\lib\site-packages\keras\backend\tensorflow_backend.p
 deprecated. Please use tf.compat.v1.global_variables instead.

Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 62, 62, 64)        1792
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 31, 31, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 29, 29, 32)        18464
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 14, 14, 32)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 6272)              0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               1605888
_________________________________________________________________
dense_2 (Dense)              (None, 10)                2570
=================================================================
Total params: 1,628,714
Trainable params: 1,628,714
Non-trainable params: 0
_________________________________________________________________
None
Training Model Accuracy = 100.0
```

Fig. 4 CNN model training is done on dataset and we got 100% accuracy



```
unable to identify
unable to identify
unable to identify
Hello 0.9986314
unable to identify
Start 0.9994998
Start 0.99750805
unable to identify
unable to identify
Start 0.9970204
unable to identify
unable to identify
Start 0.9984793
Start 0.9987633
unable to identify
unable to identify
Start 0.99973756
Start 0.997787
Start 0.99694747
unable to identify
Start 0.9997986
unable to identify
unable to identify
Start 0.9934036
Start 0.99824274
Start 0.9995952
Start 0.9989121
Start 0.99979824
unable to identify
unable to identify
Start 0.99999964
Start 0.9999988
Start 0.9999938
Connection 0.9975069
unable to identify
Start 0.98991114
Start 0.99762976
Start 0.99846774
Connection 0.9882624
unable to identify
Connection 0.99942493
Connection 0.99465215
Connection 0.99788625
```

Fig. 5 Image which shows our Trained model is identifying the words

VII. CONCLUSION

To predict silent sound Technology application will take images from webcam and then apply HAAR CASCADE files to detect face and mouth and then detected mouth will be input to CNN to identify word based on lips movement.

	Y. Sravana Kumar, M. Tech, (Ph. D) working as Assistant Professor in ECE department of NS Raju Institute of Technology having 12 years of experience with knowledge of VLSI and Embedded Systems.
	M. Mounika, Studying B. Tech in Electronics and Communication Engineering at N S. Raju Institute of Technology, Visakhapatnam.
	CH. Meghams, Studying B. Tech in Electronics and Communication Engineering at N S. Raju Institute of Technology, Visakhapatnam.
	M. Tejesh, Studying B. Tech in Electronics and Communication Engineering at N. S. Raju Institute of Technology, Visakhapatnam.

VIII. REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). 265–283.

[2] Gary Bradski and Adrian Kaehler. 2008. Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.". [3] L. Chen. 2016. keras,js. https://github.com/transcranial/keras-js [4] François Chollet. 2015. Keras documentation. keras. io (2015).

[3] Joon Son Chung and Andrew Zisserman. 2016. Silent sound Technology in the wild. In Asian Conference on Computer Vision. Springer, 87–103.

[4] Sanghoon Hong, Byungseok Roh, Kye-Hyeon Kim, Yeongjae Cheon, and Minje Park. 2016. Pvanet: Lightweight deep neural networks for real-time object detection. arXiv preprint arXiv:1611.08588 (2016).

[5] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015).