

International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164



Chief Editor

Dr. J.B. Helonde

Executive Editor

Mr. Somil Mayur Shah

ABSTRACT

Correct utterances to read Quran for beginners its important matter, especially for listeners with a hearing impairment. There are rules of utterances to learn Quran and they need a software system to tell them if they utter correctly. For that, we built lip-reading model, the model localizes the lips efficiently. Visual speech information plays an important role in lip-reading under noisy conditions.

We present in this study a classification model for some rules of Al-Tajweed as we depended on Machine Learning -Cascade Object Detector (Viola-Jones Algorithm), HOG features, a multiclass SVM classifier and Aggregate Channel Features (ACF) object detector for features extraction. We uses Matlab to train classifiers using a pre-trained convolutional neural network (CNN) for classifying images from the video stream of some Rules of Holy Quran. CNN acquires multiple convolutional filters, used to extract visual features essential for recognizing mouth. CNNs produce highly accurate recognition results.

KEYWORDS: Mouth detection, RCNN, CNN, HOG + SVM classifier, (ACF), Viola-Jones.

1. INTRODUCTION

Since verbal communication is the principal method of conveying information between humans, the possibility of communicating with computers through simple interaction presents an opportunity to profoundly change the way humans interact with machines. Humans' speech precipitation is enhanced by seeing the speakers' face and lips. Several researchers have demonstrated that the primary visible articulators (teeth, tongue, and lips) provide useful information with regard to the place of articulation. Researchers have recently developed audio-visual speech recognizers, which have been proven robust to acoustic noise. If audio-visual recognition systems are to be effective, they must be capable of tracking the lips (inner or outer contour, or both). Deaf and hearing-impaired individuals use lip-reading as their primary source of information for speech communication. Many of them are newcomers to Islam and are unable to understand the correct utterance of the Quran with Al-Tajweed on their own. Learning Al-Tajweed must be done through listening to a good sheikh who has knowledge of Al-Tajweed as to correcting the recitation and repeating it. Deaf and hearing-impaired individuals use lip-reading as their primary source of information for speech communication. A lot of work focusing on audio-visual speech recognition (AVSR) trying to find effective ways of combining visual information with existing audio-only speech recognition systems (ASR). Therefore, we present in this study a classification model for some Al-Tajweed rules as we depended on Machine Learning & Deep Learning Visual information plays an important role especially in noisy environments or for the listeners with hearing impairment. [34].

An automatic speech-reading, or lip-reading, system as part of an audio-video speech processing (AVSP) system is expected to improve ASR in noisy conditions and can thus lead one step closer to more natural human-computer interactions. [Safaa *et al.*, 10(9): September, 2021].

Overview work

First record video, make framing for it to extract images then processing the image extraction for the word then compared with database entries and classification (CNN) to recognizing the corresponding word.

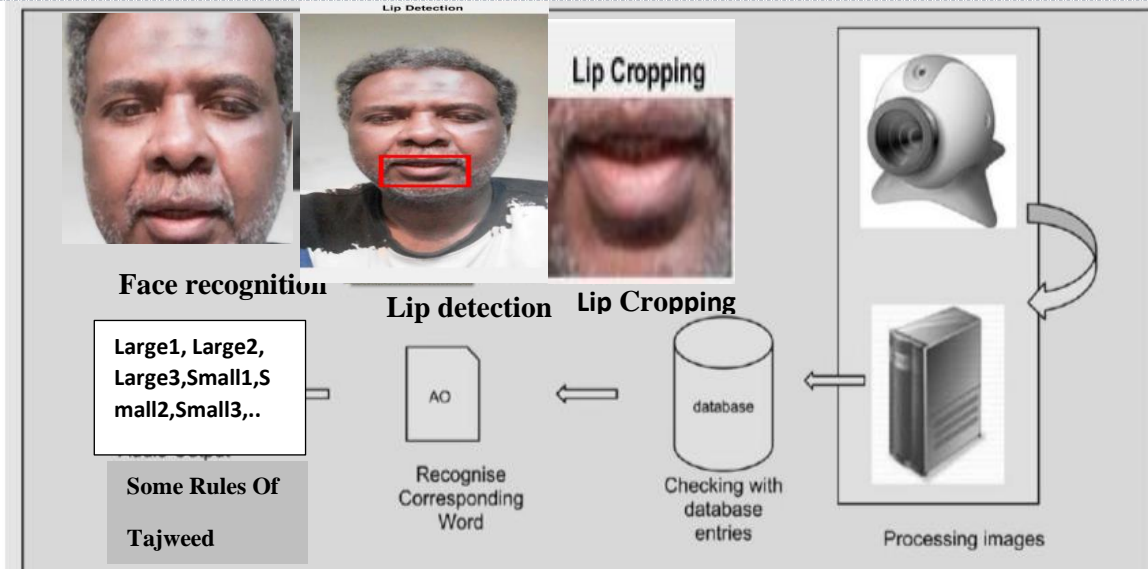


Figure (1.1) Block Diagram of identification word

2. CLASSIFICATION WORKS

2.1 Case Study

Case Study In this section, we illustrate the implementation of a specific case study on the model proposed in this paper. The implementation targets the Arabic visual figures recognition, with the aid of Matlab image processing toolbox.

Initially, real video dataset has been generated for frontal visual face that consists of 29 samples for one male. Each sample is expanded in the different cases for four tajweed rules, and each rule utterance is repeated 4 times for the same speaker. This yields a total dataset of 116 records for the 4 rule of tajweed. Moreover, the dataset was generated with no noise for the speaker to simulate real life situation of word recognition. A laptop camera was used for the recording mission, and the generated video format was "avi" with resolution of (360*640) at 62 frames/second and average video length of 30 second.

3. DESCRIPTION OF MODULES

3.1 Framing

Load the vedios ,reading the vedios frame by frame ,resize each frame by (360*640) , avedio in Matlab its rotate so we must to rotate it to be state , rotate each frame by 90 degree , save each frame in specific folders

3.2 Segmentation of face

The face is segmented from the input image that is initially whatever the video that is recorded by the camera will be fragmented into the frames and this frames will be given as inputs for segmenting the face.Loading the detector (Vision.cascade object Detectoe()), Loading the directory containing all the frame of the face images, the face region will be cropped , it will be resize all the images to specific size (258*258) to be the same for the classifier SVM to be trained ,the face region are extracted and save it in specific folder .



Figure (3.26) Example of Face region

4 FEATURE EXTRACTION

4.1 Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients, also known as HOG, is a feature descriptor we used it for feature extraction, load the segmentation face regions images you want to calculate HOG features of. Split the data randomly 70% for training, 30% for testing, two new feature vectors was saved (feature_train, feature_test)

4.2 SVM Train

The HOG feature are extracted from each single image in training and validation set and saved in the features_test and features_train respectively, selecting the labels name (name of the folders) SVM will be train using (error correction output code) fitcecoc to solve multi selection problem. evaluate the model to see performance of the model, plot the confusion matrix, calculate the accuracy then calculate the precision.

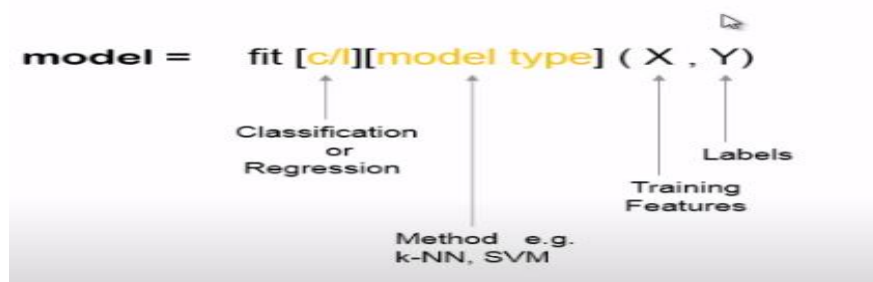


Figure (3.27) Support vector machine model

4.3 Local Binary Pattern (LBP)

Local Binary Pattern, is a feature descriptor we used it for feature extraction, load the segmentation face regions images you want to calculate LBP features of. Split the data randomly 70% for training, 30% for testing, two new feature vectors was saved (feature_train, feature_test)

4.3.1 LBP Train

The LBP feature are extracted from dataset in training and validation set and saved in the features_test and features_train respectively, then applied the features_test and features_train. The DCT block computes the unitary discrete cosine transform (DCT) of each channel in the M-by-N input matrix, u, selecting the labels name (name of the folders) SVM will be train using (error correction output code) fitcecoc to solve multi selection problem. evaluate the model to see performance of the model, plot the confusion matrix, calculate the accuracy then calculate the precision.

The model for two data set are well trained, best accuracy, very good data.

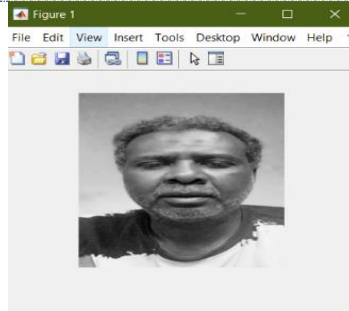


Figure (3.28): LBP operator Feature Extraction

5 SEGMENTATION OF MOUTH

5.1 Machine Learning - Cascade Object Detector (Viola-Jones Algorithm).

Mouth region was extracted from framing images of the vedios for each rules .Loading the detector (Vision.cascadeobjectDetectoe()),Loading the directory containing all the frame of the face images, the mouth regio will be cropped , some images the detector fail to extract the mouth region the tables below contains the detailes of it.

The mouth cropped it will be resize to specific size (227*227) to be the same for the identification CNN to be trained .

Show that the method localizes the lips efficiently, with high level of accuracy (91.15%).

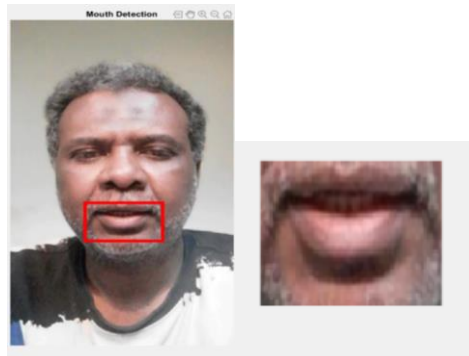


Figure (3.29): Mouth region



Figure (3.31) not cropping Mouth



Figure (3.30): Cropping Mouth region

Table 3.7: Cropping and not cropping Mouth for (Allah Elevating (Moufakham))






Word	Cropping Mouth	Not Cropping Mouth	Sample of images not cropping
الله	230 Images	37 images	
والله	184 images	0	Nothing
أعبدو الله	201 images	3 images	

Table 3.8: Cropping and not cropping Mouth for (Allah Lowering (Moureqeq))

Word	Cropping Mouth	Not Cropping Mouth	Sample of images not cropping
الله	167 images	0	Nothing
قوماً الله	254 images	3 images	
بسم الله	250 images	7images	
قل الله	193 images	39 images	

5.2 Machine Learning Aggregate Channel Features (ACF)

This MATLAB function detects objects within image we using the input aggregate channel features (ACF) object detector,its two parts Training part and testing Detect and label people using aggregate channel features (ACF). This algorithm is based on the peopleDetectorACF function. To use this algorithm, you must define at least one rectangle ROI label. You do not need to draw any ROI labels. [44]

5.2.1 Advantage of using Anchor Boxes

When using anchor boxes, you can evaluate all object predictions at once. Anchor boxes eliminate the need to scan an image with a sliding window that computes a separate prediction at every potential position. Examples of detectors that use a sliding window are those that are based on aggregate channel features (ACF) or histogram of

gradients (HOG) features. An object detector that uses anchor boxes can process an entire image at once, making real-time object detection systems possible.

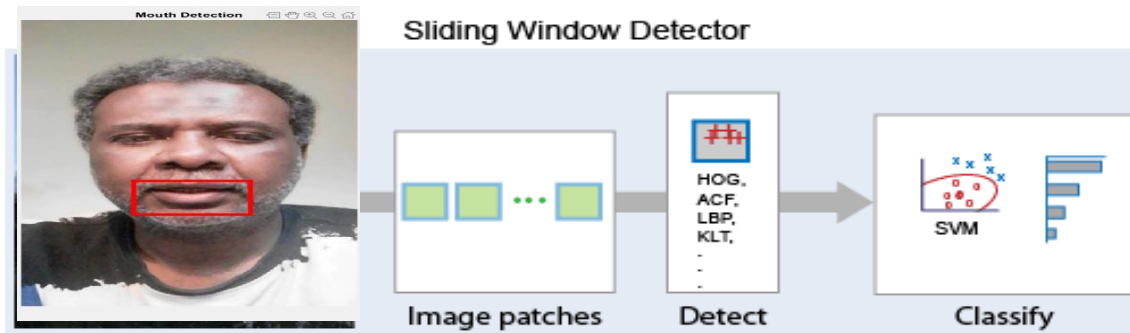


Figure (3.12): Sliding Window Detector

Because a convolutional neural network (CNN) can process an input image in a convolutional manner, a spatial location in the input can be related to a spatial location in the output. This convolutional correspondence means that a CNN can extract image features for an entire image at once. The extracted features can then be associated back to their location in that image. The use of anchor boxes replaces and drastically reduces the cost of the sliding window approach for extracting features from an image. Using anchor boxes, you can design efficient deep learning object detectors to encompass all three stages (detect, feature encode, and classify) of a sliding-window based object detector.[45]

5.2.2 Train Detectors for Mouth

Step 1: Label the ground truths using image labeler app in MATLAB apps

Step 2: select the object images path and ground truth

Step 3: save the Object images path and ground truth into a single table

Train Object Detector

Step 4: this detector uses edge histogram descriptor features and multi decision tree (ensemble i.e. boosting)

Step 5: save the trained object detector model

5.2.3 ACF Object Detector Training

The training will take four stages.

1-The model size is 82x142. Train classifier with 130 positive examples and 260 negative examples...Completed. ACF object detector training is completed. Elapsed time is 44.0951 seconds.

2- The model size is 78x137 Train classifier with 127 positive examples and 254 negative examples Elapsed time is 29.7861 seconds.

3- The model size is 87x140 Train classifier with 149 positive examples and 298 negative examples. Completed. Elapsed time is 19.5888 seconds.

6 CONVOLUTIONAL NEURAL NETWORK

Convolution neural network with convolutional and pooling layer works

Suppose we have an input image 60 x 60 x 3

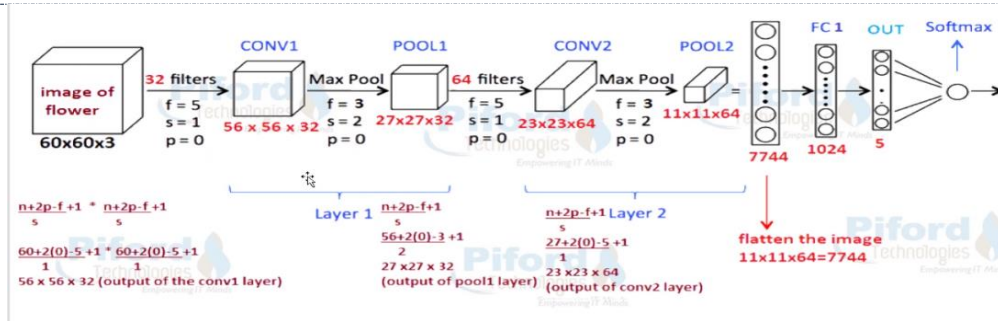


Figure (3.32): CNN Working Progress

7 RCNN: REGION

R-CNN is an object detection framework, which uses a convolutional neural network (CNN) to classify image regions within an image [1]. Instead of classifying every region using a sliding window, the R-CNN detector only processes those regions that are likely to contain an object. This greatly reduces the computational cost incurred when running a CNN. It achieved 90% of accuracy

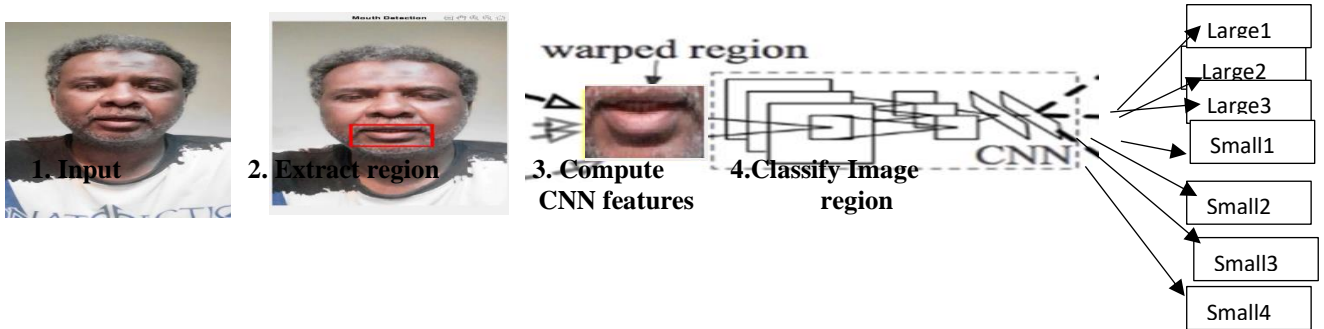


Figure (3.34) RCNN: Region with CNN features

8 DATA BASE

We recorded video recitation for one reader, A laptop camera was used for the recording mission, four rules in Al-tajweed :

- Allah Elevating (mufakhum) there is three cases in Quran each case recorded 4 times,
- Allah Lowering (mouregeq) there is 4 cases in Quran each case recorded 4 times,
- sunny لا there is 12 states in Quran each cases recorded 4 times
- Moony لا there is nine cases in the Quran each cases recorded 4 times.

The total database for reader is 116 videos samples.

Frames rate of each videos is 29.87 – 30.12 Frames /second.

Length of videos =00:00 03-00:00:04 second

Frame width = 1280

Frame height = 720

Channels = 2(stereo)

Audio sample rate = 48.00 KHZ

Readers used in the system = 1443 samples, used 70% of data for training, 30 % of data for testing

8.1 Data Base Preparation for Training and Testing

Case Study In this section, we illustrate the implementation of a specific case study on the model proposed in this paper. The implementation targets the Arabic visual figures recognition, with the aid of Matlab image processing toolbox.

Initially, real video dataset has been generated for frontal visual face that consists of 29 samples for one male. Each sample is expanded in the different cases for four tajweed rules, and each rule utterance is repeated 4 times for the same speaker. This yields a total dataset of 116 records for the 4 rule of tajweed. Moreover, the dataset was generated with no noise for the speaker to simulate real life situation of word recognition. A laptop camera was used for the recording mission, and the generated video format was “avi” with resolution of (360*640) at 62 frames/second and average video length of 30 second.

Create two data set total images for data set_1 = 1582 images, the total images for data set_2 = 4824 images , all the images and classes are loaded ,partition the data into training set and validation set(test set) , 70% for training ,30% for testing , detail’s below in the tables ,.the no of features extraction are the same (34596 features).

Table 3.3: Words for (Allah Elevating (Moufakham))

Word	Length	Frame rate	No of images after framing
الله	00:00:03	30.23 frames/second	257 images
والله	00:00:02	29.94 frames/second	184 images
أعبدو الله	00:00:02	30.13 frames/second	207 images

Table 3.4: Words for (Allah Lowering (Moureqeq))

Word	Length	Frame rate	No of images after framing
الله	00:00:02	30.0 frames/second	167 images
قوماً الله	00:00:03	30.12 frames/second	271 images
بسم الله	00:00:03	30.23 frames/second	257 images
قل الله	00:00:02	30.15 frames/second	239 images

Table 3.5: Words for (sunny) لام

Word	Length	Frame rate	No of images after framing
الطامة	00:00:04	30.12 frames/second	385 images
الثواب	00:00:03	30.23 frames/second	234 images
الصابرين	00:00:04	30.18 frames/second	226 images
الرحمن	00:00:03	29.97 frames/second	189 images
الضالين	00:00:02	29.97 frames/second	372 images
النور	00:00:04	30.12 frames/second	230 images
الدين	00:00:03	30.13 frames/second	212 images
السماء	00:00:03	30.01 frames/second	226 images
الظالمين	00:00:02	30.14 frames/second	229 images
الزكاة	00:00:02	30.14 frames/second	178 images
الشمس	00:00:02	30.02 frames/second	114 images
الليل	00:00:02	29.87 frames/second	120 images

Table 3.6: Words for (moony) لام

Word	Length	Frame rate	No of images after framing
الاول	00:00:03	30.23 frames/second	199 images
البر	00:00:03	30.23 frames/second	213 images
الغنى	00:00:02	30.11 frames/second	169 images
الحكيم	00:00:03	30.17 frames/second	231 images
الكبير	00:00:02	30.23 frames/second	150 images
الودود	00:00:03	30.16 frames/second	187 images

الخبير	00:00:02	30.12 frames/second	221 images
الفتاح	00:00:02	30.23 frames/second	203 images
نو الجلال	00:00:03	30.18 frames/second	221 images

9 IMPLEMENTATION AND RESULT

The 7 rules data set _1 are trained and identification achieved

- 1- for Allah Elevating (mufakhum) there is 3 cases in Quran each case recorded 4 times accuracy = 100% , Recall = 99.5000, Precision =100 , elapsed time is 35 min and 18 Sec , Epoch = 20 of 20, Iteration 260 of 260

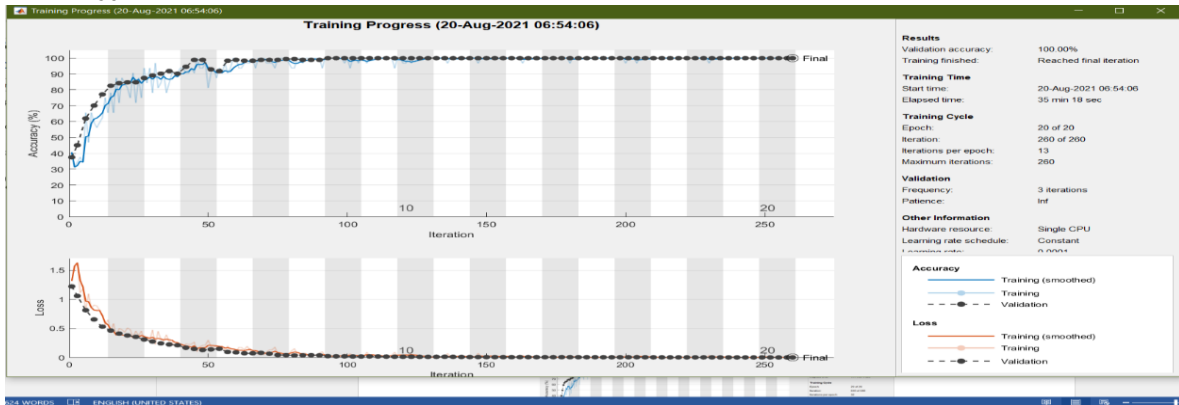


Figure of CNN Training Progress Allah Elevating (mufakhum)

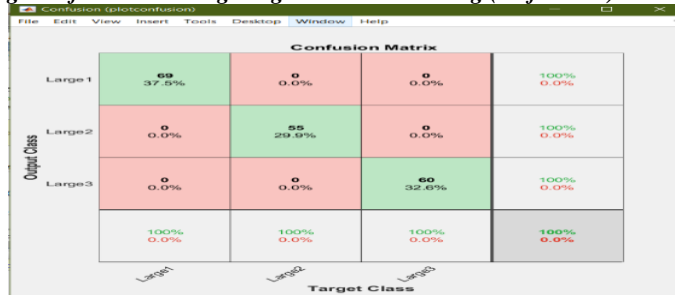


Figure Confusion Matrix of CNN Training Progress Allah Elevating (mufakhum)

- 2- for Allah Lowering (moureqeq) there is 4 states in Quran each state recorded 4 times accuracy = = 98.84%, Recall = 98.5088, Precision = 99.0213, elapsed time is 50 min and 25 Sec , Epoch = 20 of 20, Iteration 360 of 360.

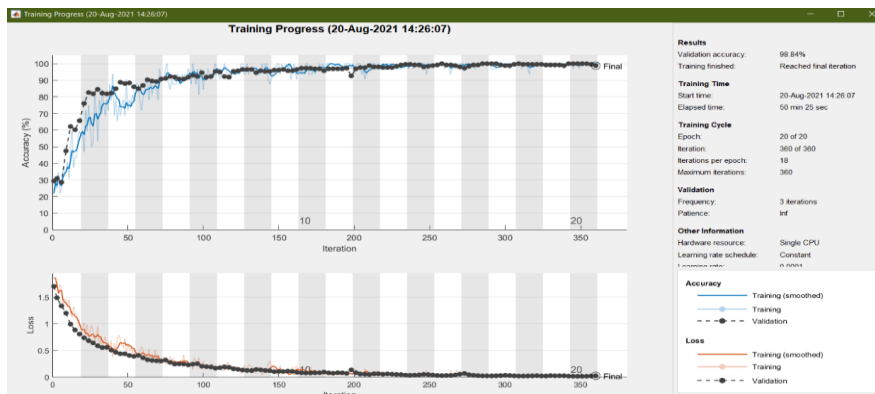


Figure of CNN Training Progress Allah Lowering (moureqeq)

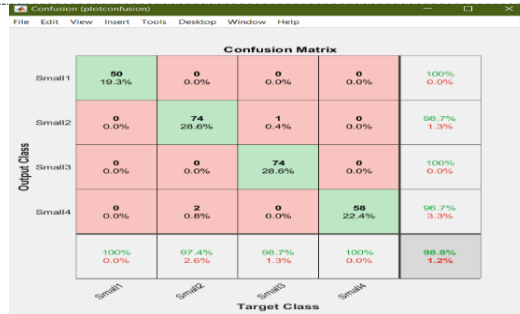


Figure of CNN Training Progress Allah Lowering (moureqq)



Figure of R- CNN Training Progress Allah Lowering (moureqq)

10 RESULTS

Many difficulties have been faced for any VSR system to detect and localize such regions to capture the related visual information, such as the fact that we cannot read lips without seeing them first. Therefore, lip localization is an essential process for any VSR system. The lips and mouth region are the visual parts of the human speech production system; these parts contain the majority of visual speech information; challenges in capturing the related visual information, such as pose and lighting variations, and blurring lips when pronouncing some letters.

In this research has been done to the performance of a face detection to detect mouth to extract visual words system by making use of feature extraction with Histogram of Oriented Gradients (HOG). Features extractions +support vector machine classifier (SVM), Local Binary Patterns. (LBP) and R-CNN an object detection framework, which uses a convolutional neural network (CNN) to classify, mouth in an image regions [Safaa et al., 10(9): September, 2021]

It mainly consists of four parts, namely face detection, mouth detection and cropping, and feature extraction and classification. Face detection represents how the model detect a face and a mouth, determines the successive algorithms of detection and recognition. The most useful and unique features of the face image / mouth image are extracted in the feature extraction phase. In the classification, the mouth image is determine the correct word (word of the rule). The accuracy of the system for face detection is above 100% by Histogram of Oriented Gradients (HOG). Features extractions +support vector machine classifier (SVM) and the Local Binary Patterns algorithm, the accuracy of the system is above 100%.

Moreover, Cascade Object Detector (Viola-Jones Algorithm), Show that the method localizes the lips efficiently, with high level of accuracy (91.15%).

The accuracy of the system is low scores detection the mouth region by Aggregate Channel Features (ACF). The accuracy of the system is (90%) with RCNN good results, localizing the lips efficiently.

Many difficulties has been faced for any VSR system to detect/localize such regions to capture the related visual information, such as we cannot read lips without seeing them first. Therefore, lip localization is an essential process for any VSR system . The lips and mouth region are the visual parts of the human speech production system; these parts hold the most visual speech information difficulties to capture the related visual information, i.e. pose and lighting variations, , and lips occlusions , Blurring lips when pronouncing some letters.

In future to improve the our system performance, Vision Transformers (ViTs) techniques can be combined with Convolutional neural networks (CNNs), in the system for video-based visual speech recognition on real time.

We choose MATLAB version 9.5.0.944444 (R2018b) as our programming environment as it offers many advantages. It contains a variety of image processing.

REFERENCES

- [1] Mary Hepburn parsons, "The reading of Speech from the lips", Gallaudet college Kendall green Washington, 9, 1900.
- [2] D.G. Stork and M.E. Hennecke, "Speech reading by Humans and Machines", Berlin, Germany: Springer, 1996.
- [3] W. H. Sumby and I. Pollack, "Visual contributions to speech intelligibility in noise," Journal of the Acoustical Society of America, 26:212–215, 1954.
- [4] Jie yang,Alex waibel, "Tracking Human Faces in Real -Time ", Pittsburgh,pennsy ivania 15213,1995.
- [5] Robert august Kaucic Jr, "Lip tracking for Audio-visual Speech recognition",Merton college University of oxford ,1997 .
- [6] Sharmila Sengupta , Arpita Bhattacharya , Pranita Desai , Aarti Gupta "Automated Lip Reading Technique for Password Authentication", New York, USA,212 .
- [7] Md. Hasan Tareque1, Ahmed Shoeb Al Hasan, "Human Lips-Contour Recognition and Tracing", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 3, No. 1, 2014.
- [8] Alin Chițu, Léon J.M. Rothkrantz, Zaidi Razak, ZulkifliMohd Yusoff,
- [9] Kuniaki Noda , Yuki Yamaguchi , Kazuhiro Nakadai , "Lipreading using Convolutional Neural Network", Japan,2014 .
- [10]Ikrami A. Eldirawy, "Visual Speech Recognition", Islamic University of Gaza, May 2011.
- [11] T. Chen and R.R. Rao, "Audio-Visual Integration in Multimodal
- [12] T. Chen and R.R. Rao, "Audio-Visual Integration in Multimodal
- [13] T. Chen and R.R. Rao, "Audio-Visual Integration in Multimodal Communication" , Proc. IEEE, vol. 86, no. 5, pp. 837-852, May 1998
- [14]David G.Stork,Marcus E. Hennecke, "Speech reading by Humman and Machines",Ricoh California Research center 2882 Sand Hill Road # 115 MenloPark, 94025-7022,USA .
- [15]Audrey R. Nath and Michael S. Beauchamp , "A Neural Basis for Interindividual Differences in the McGurk Effect, a Multisensory Speech Illusion", PMC3196040, 20 Jul 2012.
- [16]K. Neely. "Effect of visual factors on the intelligibility of speech," Journal of the Acoustical Society of America, 28(6):1275–1277, 1956.
- [17]C. Binnie, A. Montgomery, and P. Jackson, "Auditory and visual contributions to the perception of consonants," Journal of Speech Hearing and Research, 17:619–630, 1974.
- [18]D. Reisberg, J. McLean, and A. Goldfield, "Easy to hear, but hard to understand:A lipreading advantage with intact auditory stimuli, " In B. Dodd and R. Campbell, "Hearing by Eye, " pages 97–113. Lawrence Erlbaum Associates, 1987.
- [19]K. P. Green and P. K. Kuhl, "The role of visual information in the processing of place and manner features in speech perception," 45(1):32–42, 1989.

- [20] D. W. Massaro, "Integrating multiple sources of information in listening and reading," In *Language perception and production*. Academic Press, New York.
- [21] R. Campbell and B. Dodd, "Hearing by eye," *Quarterly Journal of Experimental Psychology*, 32:85–99, 1980. Yannis M. Assael¹, Brendan Shillingford¹, Shimon Whiteson, "LIPNET: END-TO-END SENTENCE-LEVEL LIPREADING" , 3 Department of Computer Science, University of Oxford, Oxford, UK 1 Google DeepMind, London, UK 2 .
- [22] Alin Chițu and Léon J.M. Rothkrantz, "Automatic Visual Speech Recognition", Delft University of Technology, Netherlands Defence Academy, the Netherlands.
- [23] Achraf Ben-Hamadou, Walid Mahdi, Ahmed Rekić's "An adaptive approach for lip-reading using image and depth data", *Multimedia Tools and Applications* · July 2015.
- [24] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", 3rd Edition. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [25] F. G. Smith, K. R. Jepsen, and P. F. Lichtenwalner, "Comparison of neural network and Markov random field image segmentation techniques" , in *Proceedings of the 18th Annual Review of progress in quantitative nondestructive evaluation*, vol. 11, 1992, pp. 717-724.
- [26] A. Blake and M. Isard, "Active Contours, Springer", 1998.
- [27] J. Shi and J. Malik, "Normalized cuts and image segmentation", in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*. Washington, DC, USA: IEEE Computer Society, 1997, p. 731.
- [28] J. Sethian, "Level set methods and fast marching methods: Evolving interfaces in computational geometry," 1998.
- [29] D. Reisberg, J. McLean, and A. Goldfield, "Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli," In B. Dodd and R. Campbell, "Hearing by Eye" , pages 97–113. Lawrence Erlbaum Associates, 1987.
- [30] G. W. Greenwood, "Training partially recurrent neural networks using evolutionary strategies", *IEEE Trans. Speech and Audio Processing*, 5(2):192–194, 1997.
- [31] E. Owens and B. Blazek, "Visemes observed by hearing impaired and normal hearing adult viewers ", *Journal of Speech Hearing and Research*, 28:381–393, 1985.
- [32] Ikrami A. Eldirawy "Visual Speech Recognition", Islamic University of Gaza, May 2011
- [33] Ahmad Basheer Hassanat, "Visual Words for Automatic Lip-Reading", University of Buckingham United Kingdom, December 2009
- [34] Yuanhang Zhang, Shuang Yang, Jingyun Xiao, Shiguang Shan, Xilin Chen, "Can We Read Speech Beyond the Lips? Rethinking RoI Selection for Deep Visual Speech Recognition", arXiv: 2003.03206v2 [cs.CV] 9 Mar 2020.
- [35] Elena Alionte, Corneliu Lazar, "A Practical Implementation of Face Detection by Using Matlab Cascade Object Detector", *International Conference on System Theory, 2015 19th Control and Computing (ICSTCC)*, Cheile Gradistei , October 14-16, Romania
- [36] Guoying Zhao, Mark Barnard, "Lip-reading with Local Spatiotemporal Descriptors", *IEEE Transactions on Multimedia* · December 2009
- [37] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 886-893, 2005.
- [38] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, "Gradient-based learning applied to document recognition.", *Proceedings of the IEEE*, P. (1998). 86, 2278-2324.
- [39] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A. Y. Ng, "Reading Digits in Natural Images with Unsupervised Feature Learning NIPS Workshop on Deep Learning and Unsupervised Feature Learning", 2011.
- [40] Lindsay I Smith, "A tutorial on Principal Components Analysis", February 26, 2002
- [41] Sushma Ronanki, Sonia Gundu, Rupavathi Baratam, P Mounika, J Rajesh Kumar, "FACE DETECTION AND IDENTIFICATION USING SVM", B.TECH, Electronics and Communication, SSCE, Srikakulam, Andhra Pradesh (India), April 2017 .
- [42] https://www.mathworks.com/videos/introduction-to-deep-learning-what-is-deep-learning--1489502328819.html?s_tid=vid_pers_recs
- [43] <https://www.mathworks.com/discovery/convolutional-neural-network-matlab.html>
- [44] https://www.mathworks.com/discovery/machine-learning.html?s_tid=srchtitle

-
- [45] <https://www.mygreatlearning.com/blog/viola-jones-algorithm/>
[46] <https://www.mathworks.com/help/vision/ref/imagelabeler-app.html>