ISSN 2278-3091
Volume 10, No.3, May - June 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse171032021.pdf
https://doi.org/10.30534/ijatcse/2021/171032021

Multi Model Approach to Extract Human Features in Real Time

Raj Baldania[1], Barkha Bhavsar[2]
[1]Researcher, LDRP Instituteof Technology & Research,gandhinagar-382015, Gujarat,India
[2]Assistant Professor, LDRP Instituteof Technology & Research, gandhinagar-382015,Gujarat, India

ABSTRACT

Face detection and analysis systems has been growing in last few years for various applications. Since the hardware performance increase in last few years, use of Deep Learning, Convolution Neural Network, Face detection, Face analysis techniques is increasing and day by day developed models are breaking accuracies of previous models and research in various tasks. Facial analysis system with age, gender and emotion recognition have been proposed with good accuracies for real-time and non-real time both. The present research paper focuses to provide a robust system architecture for age, gender and emotion recognition in real time which can be use in commercial, healthcare, and many more industries. To achieve this a literature survey is done on the same topic with previous researches to compare their results. The final model architecture proposed in this research paper is efficient and fast and provides accurate results as compare to previous researches

Key words: Facial Expression Recognition; Feature Extraction; VGGFace, Deep Learning, Convolution Neural Network, Age Recognition, Gender Recognition

1. INTRODUCTION

In previous researches, studies have used various techniques to know about human feeling by measuring stress, voice and retinas. Same way for more accurate data of what a human being is feeling can be measure by recognizing the emotions by detecting their face with the help of computer vision and machine learning. It is difficult to detect voice of particular human in a crowded environment and also to scan their body for stress which will take more time than face detection. For this only solution to measure the feelings and thinking of human is to recognize face and extract features. Face detection and recognition both have only one difference that is in face detection the computer detects the faces but in recognition the computer detects the face and tries to recognize that person by scanning the face in existing database. In real world, there are so many number of applications and scope regarding this research. One can use this to measure that how many people likes their advertisement in a street screen having a camera above it. We can measure then number of happy and sad audience in a live show in real time and many more applications.

In this research, we consider using face, gender, age and emotion data for human recognition in a visible light camera environment. When the images are captured by camera installed in outdoor or indoor environment, the system will detect face using a pre-trained CNN model and classification of age, gender and emotion will be done using proposed pre-trained CNN model.

2. SUMMARY OF RELATED WORK

Table 1: Summary of related work (Comparative analysis)

Papers	Datasets	Methods	Accuracy
DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks [1]	A large dataset of 4 million images and 600k images labelled with real age	Deep CNN, Detailed proposed method not mentioned	Top -1 accuracy for age prediction is 61.3 %, accuracy for emotion recognition is 76.1 % and for gender it is 91%
Audience Analysis System on the Basis of Face Detection, Tracking and Classification Techniques [2]	Image databases MORPH [3] and FG-NET [4] and their own image database gathered from different sources	Haar cascade for face detection, AF-SVM algorithm [5] for Gender Recognition, Hierarchical approach using Binary Classifier for	Not mentioned by author

	which consisted of 10,500 face images	Age Estimation	
Age, gender and emotion detection using CNN [6]	FER2013 dataset [7]	Haar Cascade classifier [8] for face detection, CNN for emotion detection, ResNet [9] for Age and Gender Classification	95 % accuracy for age and gender prediction. Accuracy rate for emotion recognition is not mentioned
A Convolutional Neural Network for Real-time Face Detection and Emotion & Gender Classification [10]	IMDB-WIKI age and gender dataset [11], FER emotion recognition dataset	Haar cascade for face detection, basic CNN architecture, Back propagation	Accuracy rate for gender prediction is 95% and for emotion recognition it Is 66%
Real Time System for Facial Analysis [12]	IMDB-WIKI dataset, CVPR2016 LAP challenge dataset [13] , AffectNet emotion dataset [14]	SSD [15] for face detection and MobileNet [16] for age, gender and emotion recognition	Accuracy rate for gender is 88.3 % and for emotions it is 55.9%

Table 1 shows a literature survey of all the previous papers which uses different approaches or methods to evaluate and extract the facial attributes for age, gender and emotion classification. Some papers mentioned real time approach and some of them are non-hybrid as not all them had extracted age, gender and emotion together.

3. PROPOSED METHOD

The proposed model is present in this chapter after a complete review of the main concepts, methods and techniques which were considered in this research paper. First the overview of datasets used are explained and then the proposed system architecture will be introduced.

3.1 Datasets

A. FER 2013 dataset

The dataset, used for training the model is from a Kaggle Facial Expression Recognition Challenge a few years back (FER2013). The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples. The public test set used for the leader board consists of 3,589 examples and the class distribution is presented in table 2.

Table 2: Emotion dataset class distribution

Emotions	**Number of faces (%)**
Angry	4953 (14)
Disgust	547 (2)
Fear	5121 (14)
Happy	8989 (25)
Sad	6077 (17)
Surprise	4002 (11)
Neutral	6198 (17)

B.IMDB-WIKI Face Only dataset

The complete data sets from the IMDB-WIKI project are very large — a whopping 272 GB for the data and images from both IMDb and Wikipedia. The IMDB-WIKI project also offers much smaller subsets of data and images for face-only data — 7 GB for IMDb, Face Only images and 1 GB for Wikipedia Face Only images. The Face Only data sets were perfect for our needs. The age and gender class distribution is shown in figure 1.

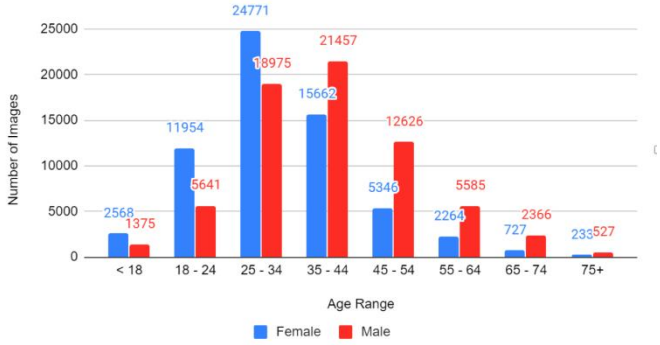

Figure 1: Age and gender class distribution

3.2Proposed method architecture

In the proposed system present in this research paper consists of 2 phases. The pre-trained networks used are VGGFace[17] and Sequential CNN model and the pre-trained weights for age and gender are provided by computer vision researchers of ETH Zurich University for transfer learning purpose. In phase 1, the pre-trained model with weights are loaded for age and gender model. Same way weights and CNN architecture for emotion classification is loaded which was trained on FER 2013 dataset.

Figure 2 is the proposed system architecture which illustrates both phases and working pipeline of the architecture to extract facial features of human in real-time.

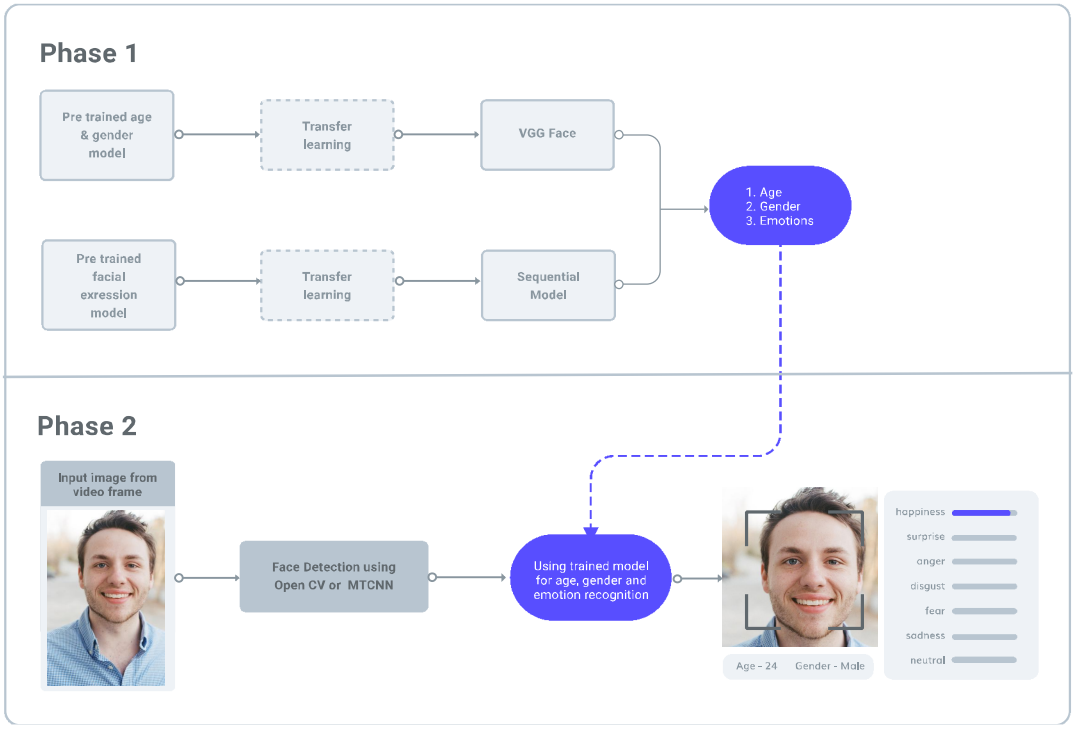

Figure 2: Proposed system architecture

A. Age estimation and gender classification

Age prediction is a regression problem, but researchers define it as a classification problem. There are total 101 classes in the output layer for ages 0 to 100. They applied transfer learning for this using VGG. Similarly, VGGFace is used in this paper because, this model is tuned for face recognition task as there are chances to have outcomes for patterns in the human face.

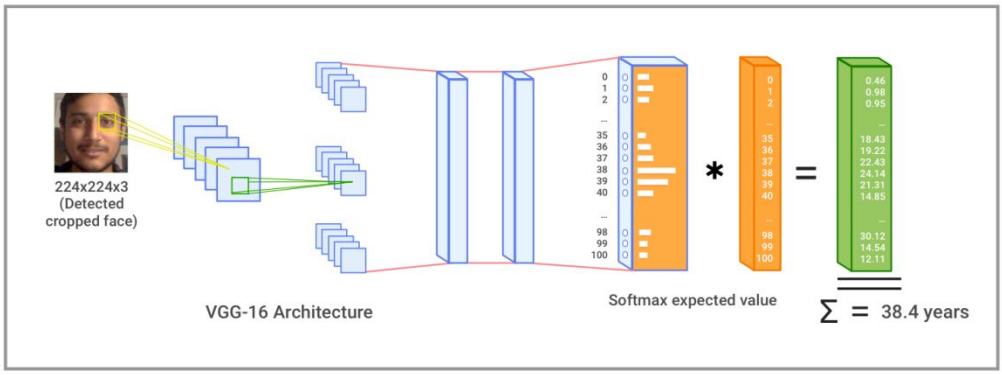

Figure 3: VGGFace architecture for age and gender model.

The layer weights are locked for previous layers as they could already detect some patterns. All layers are freeze except the last 3 convolution layers. The last convolution layer is terminated as well because it has 2622 units and we just need101 (ages from 0 to 100) units for age prediction task.

Afterwards, we added a custom convolution layer which contains 101 units.

Researchers developed an age prediction approach and convert classification task to regression. For this they proposed that multiply each softmax out with its label and summing this multiplication which will be the apparent age prediction. The softmax expected out value is defined as

$$E(O) = \sum_{i=0}^{100} y_i o_i \quad (1)$$

where y_i is the discrete year corresponding to each class i, o_i represents softmax output probabilities and O defined as the 101 dimensional output layer: {0, 1, ……. , 100}.

Age prediction was a difficult task. However, gender prediction is much more predictable as we have to apply just binary encoding to target gender class

```
classes = 2
base_model_output = Sequential()
base_model_output = Convolution2D(classes, (1, 1), name='predictions')(model.layers[-4].output)
base_model_output = Flatten()(base_model_output)
base_model_output = Activation('softmax')(base_model_output)
gender_model = Model(inputs=model.input, outputs=base_model_output)
```

In this way, the pre-trained VGGFace architecture is used to classify gender and predict the age by extracting facial features.

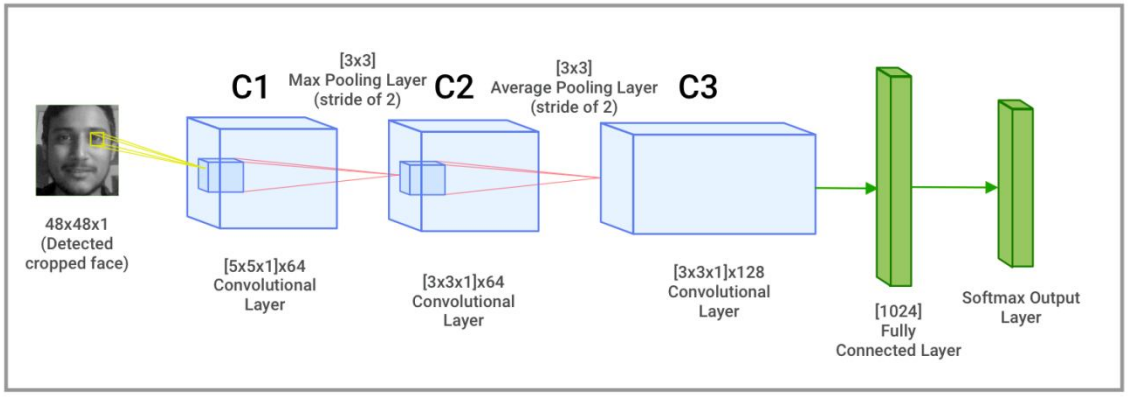

Figure 4: CNN architecture for emotion model.

B. Emotion classification

Each image from FER-12 dataset was stored as 48×48 pixel. The dataset consists of image pixels (48×48=2304 values), emotion of each image. For training the dataset a CNN structure is developed with 3 convolution blocks which is visualized in Figure 4. That is the reason why train and fit generator used. The loss function is cross entropy as the task is multi class classification. The categorical cross entropy is well suited to classification tasks, since one example can be considered to belong to a specific category with probability 1, and to other categories with probability 0. The categorical cross entropy loss function is defined as

$$Loss = -\sum_{i=1}^{output\ size} y_i \cdot \log \hat{y}_i \quad (2)$$

where y_i represents corresponding target value, $\hat{y}_i$re presents the *i-th* scalar value in the model output and output size is the number of scalar values in the model output.

We got the following results as accuracy

```
Test loss: 2.27945706329
Test accuracy: 57.4254667071

Train loss: 0.223031098232
Train accuracy: 92.0512731201
```

Images are already cropped and just face area were focused on in the training set to increase the accuracy. Emotions stored as numerical as labeled from 0 to 6. Keras would produce an output array including these 7 different emotion scores. The prediction results are shown in evaluation section in this paper.

In Phase 2 the models are implemented in real time stream. Here for real time streaming video, a webcam with 480p resolution is used. Although a webcam with 720p is also a better option than this but it is more useful only for commercial purposes. The webcam takes the video stream

input to the main program where all the models and weights are combined.

Steps involved in working of phase 2 are as follow: -

1. The face is detected using OpenCV Haarcascade classifier.
2. The detected face is cropped and freeze for 5 seconds which is the threshold time
3. This cropped face is pre-processed for age and gender recognition to 224x224 with grayscale.
4. Same way the same cropped face is pre- processed for facial expression recognition to 48x48 with grayscale.
5. ThegrayscaleimagesarepassedtowardstheirrespectivemodelsandtheoutputprovidedthroughmodelspassedontoOpenCVtodisplayonthescreenwithcroppedimage.

4. IMPLEMENTATION

The implementation of extracting facial features of human in real-time consists of maintained by a computer system type x64-based as mentioned in chapter 4. The main target of the proposed system is to extract facial attributes emotions, age and gender of a human in real time using transfer learning approach with pre-trained deep learning models. The whole implementation is done in python programming language
Pre-requests for whole system architecture is as follows: -

- tensorflow 2.4.1
- keras 2.4.3
- opencv 4.1.2

In spite of the low-quality images, in addition to they are extracted once by the web cam, the proposed model succeeded to verify them accurately. The main target of the proposed application is to present to which extent it can be utilized the "transfer learning" concept.

1. Face detection using OpenCV Haarcascade classifier

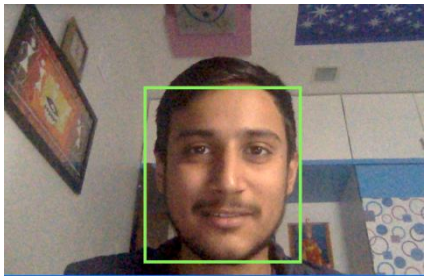

Figure 5: Face detection using webcam in OpenCV

As explained in proposed system that system will detect face from webcam in real-time.

2. Cropped face and freeze (threshold time)

Figure 6: Cropped face from detected area

The detected face is cropped out for pre-processing purpose and at the time of pre-processing and analyzing the facial attributes through pre-trained models and weights the webcam freezes for 5 second which is a threshold time.

3. Pre-processing of cropped image

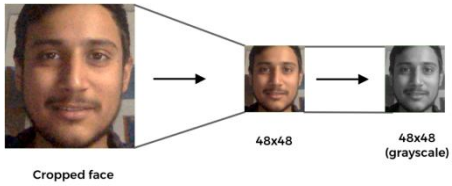

Figure 7: Pre-processing of image for emotion recognition

The cropped-out image is pre-processed for emotion recognition and transformed to 48x48 grayscale image as per the model input size

4. Age and gender pre-process face

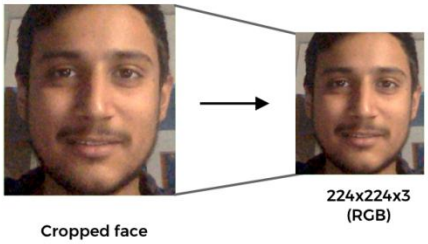

Figure 8: Pre-processing of image for age and gender recognition

Similarly, the cropped-out image is pre-processed for age and gender recognition as eel and transformed to 242x242x3 RGB image as per the model input size.

5. Building model and loading the weights for age and gender recognition

Age and gender dataset is trained on VGGFace model which is a state of art model use to recognize face. To use VGGFace model, the whole model is built as shown in figure with an input shape of 224x224 RGB. After building the model all the pre-trained weights are initialized using: -

- age_model.load_weights('weights/age_model_weights.h5')
- gender_model.load_weights('weights/gender_model_weights.h5')

6. Model construction for emotion

FER 2013 dataset is trained on a sequential CNN model. The whole model is built as shown in figure with an input shape of 48x48 grayscale. After building the model all the pre-trained weights are initialized using: -

- model.load_weights('weights/facial_expression_model_weights.h5'

5. EVALUATION AND COMPARISON

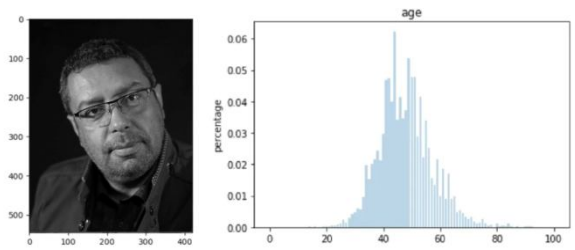

Figure 9: Age estimation on a single image in testing phase.

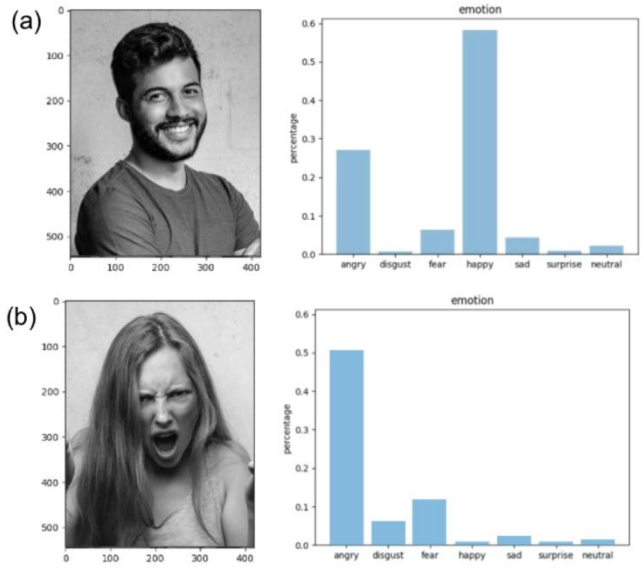

Figure 10: Emotion classification on 2 different images for testing (a) first image; (b) second image.

Figure 9 and 10shows the age estimation and emotion classification result evaluated during testing phase of age and emotion models.

After complete implementation of the proposed system the real-time results are shown below:

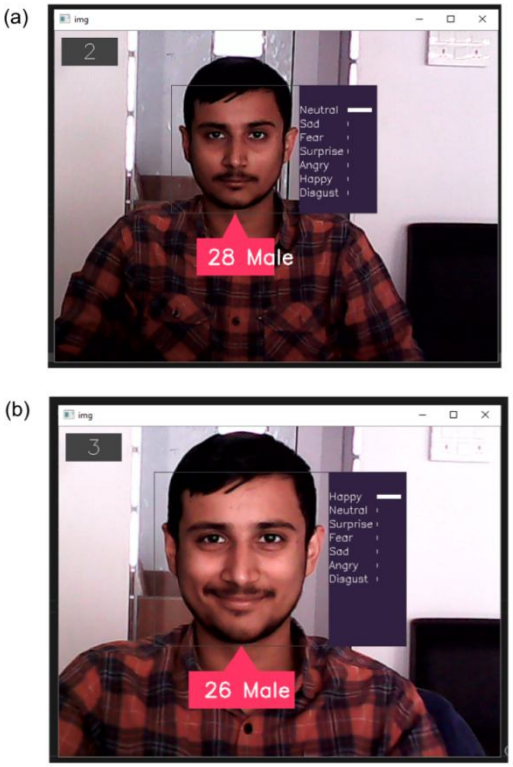

Figure 11: Real-time output using webcam - (a) first image output (b) second image output

Figure 11 shows the output as face analysis providing insights for age, gender and emotions.In both images, it has been observed that gender recognition is very accurate as it's a binary classification problem. On other side emotionsare also accurately recognized as per the facial expressions. The only problem is that the age varies in both of the images as age prediction is a regression problem.Another output results showing multiple faces in single frame is shown in Figure 12.

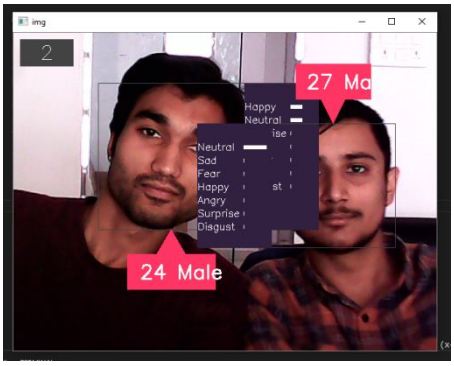

Figure 12: Output of 2 face detected with analysis

As mentioned in literature survey, different methods are there with several approaches to extract facial attributes and all are having results as accuracy tested on datasets. But only one paper "DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks" created a whole system which takes the image frame from live video and provides facial analysis by extracting facial attributes. So here the proposed system is compared to the existing DAGER's

Sighthound system[18]. Following figure show the comparison between an existing system and proposed system.

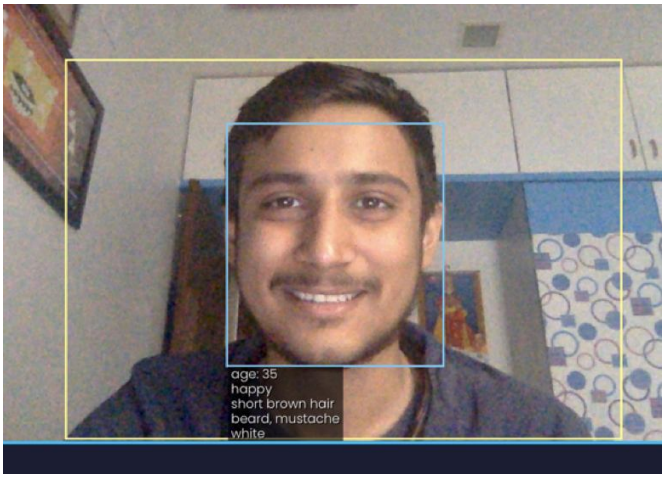

Figure 13: Output from Sighthound's system (existing)

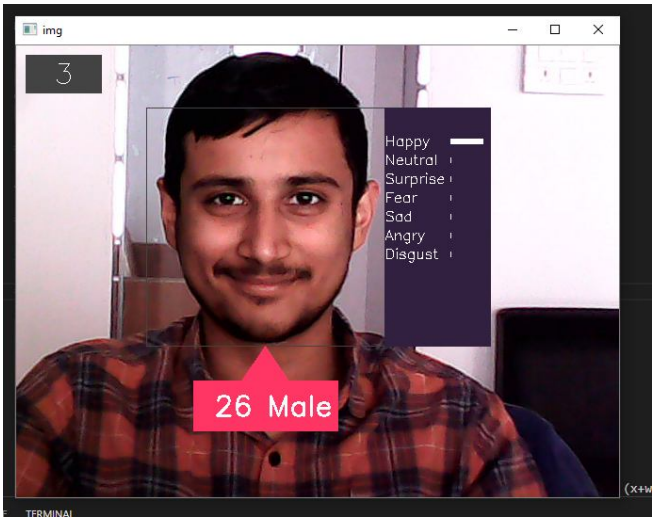

Figure 14: Output from proposed system

Figure 13 shows the output from Sighthound's cloud API for facial attributes and Figure 14 shows the output from proposed system. It is seemed that the Sighthound's proposed system didn't provided the gender recognition output, but instead of gender recognition it provides other attributes such as facial color, hair color, beard and moustache makes quite different from the proposed system.

Another main difference it can be seen that the age result is having a difference and proposed system is having more accurate age prediction than the existing one.

5. CONCLUSION

It is concluded from the results that the proposed system is having good accuracy in emotion and gender as compare to age recognition. It might be possible because of low quality image from webcam is responsible for minor inaccuracy in age. However, the results of proposed system are better than the existing one (Sighthound).There are limitations like number of multiple faces which are unable to get detect, inaccuracy in age prediction and so on. But there are solutions which can solve this limitation as well.For multiple face detection, a library called MTCNN [19] can be use which is having good accuracy as compare to OpenCV. For age prediction, the model can be train again on age bucket datasets to provide age group results like child, adult, young adult and so on.

ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their valuable and insightful comments. We believe their comments significantly improved the quality of this manuscript.

The research activities described in this paper were funded by LDRP Institute of Technology and Research, Gandhinagar, Carloman Systems, Ahmedabad, Gujarat, India.

REFERENCES

1. Afshin Dehghan, Enrique G. Ortiz, Guang Shu, Syed Zain Masood, "**DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks**," 2017.
2. Vladimir Khryashchev, Alexander Ganin, Maxim Golubev, Lev Shmaglit, "**Audience Analysis System on the Basis of Face Detection, Tracking and Classification Techniques**," in Proceedings of the International MultiConference of Engineers and Computer Scientists , Hong Kong , 2013.
3. K. Ricanek, T. Tesafaye, "**MORPH: a longitudinal image database of normal adult age-progression**," 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp. 341-345, 2006.
4. "**The FG-NET Aging Database**," 2010. [Online]. Available: http://www.fgnet.rsunit.com/, http://wwwprima.inrialpes.fr/FGnet/.
5. V. Khryashchev, A. Priorov A, L. Shmaglit, M. Golubev,, "**Gender Classification for Real-Time Audience Analysis System**," in Proceedings of the 15th Conference of Open Innovations Association FRUCT, Saint-Petersburg, 2014.
6. Manasa SB, Jeffy.S.Abraham, Anjali Sharma, Himapoornashree KS, **"Age, gender and emotion detection using CNN**," International Journal of Advanced Research in Computer Science , vol. 1, no. 1, pp. 68-70, May 2020.
7. Wofram Research, "**FER-2013**," [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data.
8. Viola and Jones, "**Rapid object detection using a boosted cascade of simple features**," Computer cool Vision and Pattern Recognition, 2001.

9. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "**Deep Residual Learning for Image Recognition," 10 12 2015**.
10. Md. Jashim Uddin, Dr. Paresh Chandra Barman, Khandaker Takdir Ahmed, S.M. Abdur Rahim, Abu Rumman Refat, Md Abdullah-Al-Imran, "**A Convolutional Neural Network for Real-time Face Detection and Emotion & Gender Classification**," OSR Journal of Electronics and Communication Engineering , vol. 15, no. 3, pp. 37-46 , May 2020.
11. "**IMDB-WIKI Dataset**," [Online]. Available: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/.
12. Janne Tommola, Pedram Ghazi, Bishwo Adhikari, Heikki Huttunen, "**Real Time System for Facial Analysis**," in Computer Vision and Pattern Recognition .
13. "**2016 Looking at People CVPR Challenge**," [Online]. Available: https://gesture.chalearn.org/2016-looking-at-people-cvpr-challenge.
14. Ali Mollahosseini, Behzad Hasani, Mohammad H. Mahoor, "**AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild**," IEEE Transactions on Affective Computing, 2017.
15. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, "**SSD: Single shot multibox detector**," in Computer Vision and Pattern Recognition, 2015.
16. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, "**Mobilenets: Efficient convolutional neural networks for mobile vision applications**," 2017.
17. Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, "**VGG Face Descriptor**," [Online]. Available: https://www.robots.ox.ac.uk/~vgg/software/vgg_face/.
18. Sighthound, "**Sighthound cloud api**," [Online]. Available: https://www.sighthound.com/products/cloud.
19. Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao, "**Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks**," IEEE Signal Processing .