# Video-based Human Age and Gender Recognition using Deep Convolutional Neural Networks

[1]Prof.Abhilasha Kulkarni, [2]Radhika Khiste, [3]Suteja Patil,[4]Shreeya Tahasildar,[5]Shreeya Tambolkar

[1]Assistant Professor, [2,3,4,5]BE Students
[1,2,3,4,5]Department of Computer Engineering,
[1,2,3,4,5]Marathwada Mitra Mandal's College of Engineering, Pune, India.

***Abstract :*** In this paper, we propose an automatic age and gender recognition system from a live-video stream. Haar cascade classifier is used for face detection while for age and gender recognition,VGG16 is used. VGG-16 is a famous deep-Convolutional Neural Networks(CNN) architecture.VGG-16 network is trained on ImageNet dataset which has over 14 million images and 1000 classes, and achieves 92.7% top-5 accuracy. It surpasses AlexNet network by replacing large filters of size 11 and 5 in the first and second convolution layers with small size 3x3 filters. Transfer learning is implemented for iterative transfer of knowledge and better recognition with greater accuracy. IMDB-WIKI dataset is used for training and testing of the model which is the largest dataset of human faces with gender, name and age information. This real-time prediction model is characterized by more accuracy when compared to the publicly available methods. As face analysis is a challenging task because of variations in images having different postures, lighting ,angles and expressions, this model successfully overcomes all the barricades and proves to be the most accurate while using less computational resources.

***IndexTerms*** **- real-time, VGG-16, AlexNet, Transfer learning.**

## I. INTRODUCTION

Age and gender recognition have been one of the most demanding application of deep learning. It can be implemented using various deep learning algorithms. One of the most common algorithm used is Convolutional Neural Networks(CNN). This recognition model has wide set of applications like biometrics, surveillance, human-robot interaction and many other computer vision implementations. Computer vision is the study of how a computer understands digital images and videos. This project can be installed at bars, movie theatres, hostels, public restrooms, ATM's, hospitals and other recommendation systems.

Haar Cascade is a machine learning object detection algorithm. It is used to identify objects in an image or video. So in this project we are using it for detecting the faces. A cascade function is used for the same which is trained from a lot of positive and negative images. Based on this training it is then used to detect the objects in other images. The algorithm is implemented in OpenCV and is good at detecting edges and lines.

Convolutional Neural Networks is known for image classification problems which are sometimes unsolvable by even humans. One of such famous Convolutional Neural Network architecture is VGG-16. Its name comes from the fact that it has 16 layers. VGG-16 network consists of Convolutional layers, Max Pooling layers, Activation layers, Fully connected layers. There are 13 convolutional layers, 5 Max Pooling layers and 3 Dense layers which sums up to 21 layers but only 16 weight layers. Conv 1 has number of filters as 64 while Conv 2 has 128 filters, Conv 3 has 256 filters while Conv 4 and Conv 5 has 512 filters. It was trained on 4 GPU's for 2-3 weeks. It is primarily used as a baseline feature extractor.

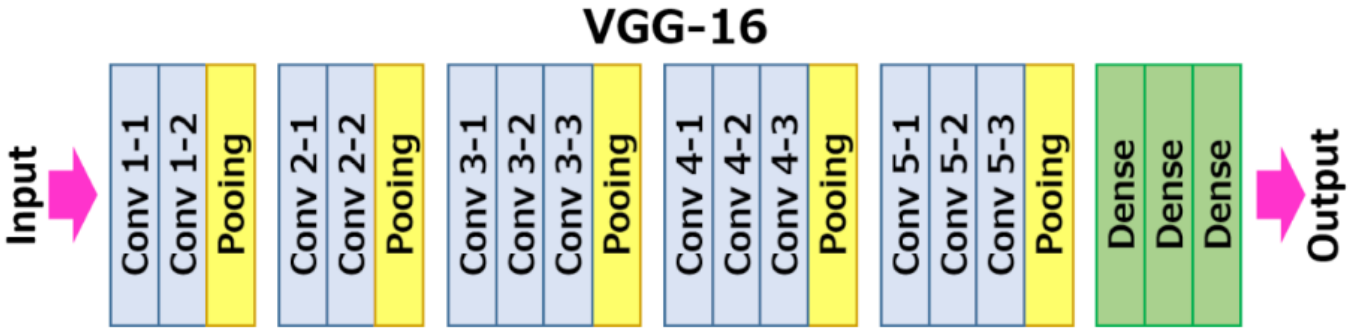


**Figure 1: VGG-16 Layer Definition**

We are using IMDB-WIKI dataset for training and testing of the model. Other publicly available datasets are small in size and often without any age information. But this is the largest publicly available dataset of face images with gender and age labels. It contains 500 thousand+ images with all the meta information. The IMDb dataset contains 460,723 facial images (with gender and age labels) of film stars, predominantly Hollywood actors and actresses, and the Wikipedia dataset includes 62,328 of celebrities from various fields, such as sports, politics, social events, and the film industry. The images in the dataset have a face score, a second face score, age and gender labels on each image. Images with only one frontal face have high face scores and the other have comparatively low face scores. The second face score shows how clearly the second face is shown in the image. All the images are in .jpg format. The total dataset is of size 296GB while the face-cropped version images is of size 7GB.

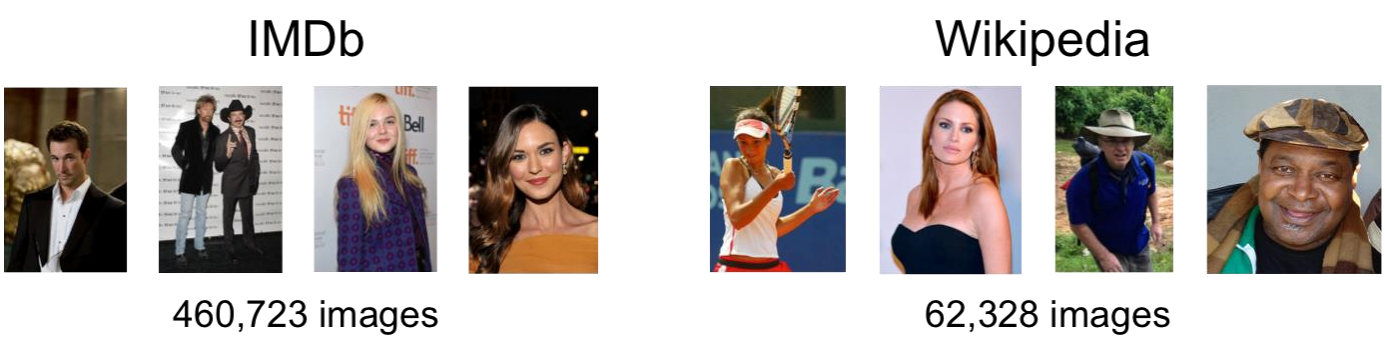


**Figure 2: Faces from IMDB-WIKI Dataset**

## II. PROPOSED METHODOLOGY

First we consider the dataset. The index file of the dataset is stored in Matlab datenum format and contains the date of birth of every person in the image. The date of birth is subtracted from the date when the photo was clicked and it is used to detect the age of the person. This information is useful in training of the prediction model. Next, we clean the data. We discard the images which do not include people. Images with less face score value and also without age information are also discarded. Unnecessary columns are dropped so that the dataset occupies less memory. Now the data having age ranging from 0 to 100 is only considered for preparing the recognition model .

The final data set consists of 22578 instances. It is split into 15905 train instances and 6673 test instances .The pixel values of the images from the training dataset is calculated and transfer learning is used to train them. The input images are normalized and converted in the required vector form. The layer weights of early layers from the pre-trained weights are locked because they could already detect some patterns. All the layers are blocked except the last 3 convolution layers. Then the model is trained where it is checked whether the over-fitting and validation loss is minimum. The process of detecting is very complex and starts by tracking human eyes as they are the most easiest feature to detect on a human face. After the eyes, the algorithm tries to detect eyebrows, the mouth, nose, nostrils and the iris. Once the algorithm has surmised that it has detected a facial region, it validates the same and sends it for further classification. We have used the haar cascade classifier algorithm for detecting the face. First the captured image frames are loaded using OpenCV which by default loads the image into BGR color space. The image is then converted to gray scale as OpenCV face detector expects gray images. Then we load the haar classifiers(downloaded XML files) which considers the image as the input.OpenCV's CascadedClassifier detects the face using detectMultiScale function.

Age prediction is referred as a regression problem. In regression, outcome is predicted based on the relationship between variables obtained from the dataset. In the same way, age prediction model is constructed which uses the features in the trained dataset to estimate an age. There are 101 classes in it ranging from 0 to 100 years of age. Transfer learning is used to implement the same in keras using inception V3.Keras Applications are deep learning models that are made available alongside pre-trained weights. These models can be used for prediction, feature extraction, and fine-tuning. Inception V3 is a type of Convolutional Neural Networks. It consists of many convolution and max pooling layers. Finally, it includes fully connected neural networks. Keras handles its network structure. Inception is constructed and the weights are loaded automatically after the construction.The weight parameter is specified as ImageNet which helps in providing pre-trained weights for ImageNet challenge.The pre-constructed network structure and the pre-trained weight models are the essentials for running inception V3.VGG-16 is one of the CNN pre-trained weight model on ImageNet dataset. It uses only 3×3 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier. So now VGG-16 works with weights pre-trained on ImageNet. In this way we transferred the learning outcomes for ImageNet winner model InceptionV3 to VGG-16 to predict the age of humans. Transfer learning also involves training along with copying network structure and loading pre-trained weights. For the age prediction model, we have applied partial training. So we can say that we copied the ImageNet structure and the its pre-trained weights to implement VGG-16 in age recognition model. So successfully implemented transfer learning through this process.

Gender prediction is quite easy in comparison to age prediction. This is because unlike age prediction which has 101 classes to categorize, gender prediction has only two- "Male" and "Female". Apparently age prediction was a challenging problem but gender prediction is much more predictable. Binary encoding is applied to target the gender class. Only 2 classes are needed in the output layer to identify male and female. Binary encodings are a special case of categoric features such as gender. The pre-trained weights are loaded and then used for gender classification. Finally the model is prepared is applied on the testing data which helps in validating the system. The final age and gender recognition system is ready to be implemented .

## III. RESULTS

Thus age and gender recognition model has been successfully implemented using haar cascade classifier and VGG-Face Model. The accuracy of the gender model is 95% while that of age is around 90%. The snapshots of the working model can be seen below. VGG-Face Model outperformed all the other conventional methods even while working with such a large sized dataset. It was working in all lighting conditions and can successfully predict the age and gender of humans with varying postures, expressions, features .

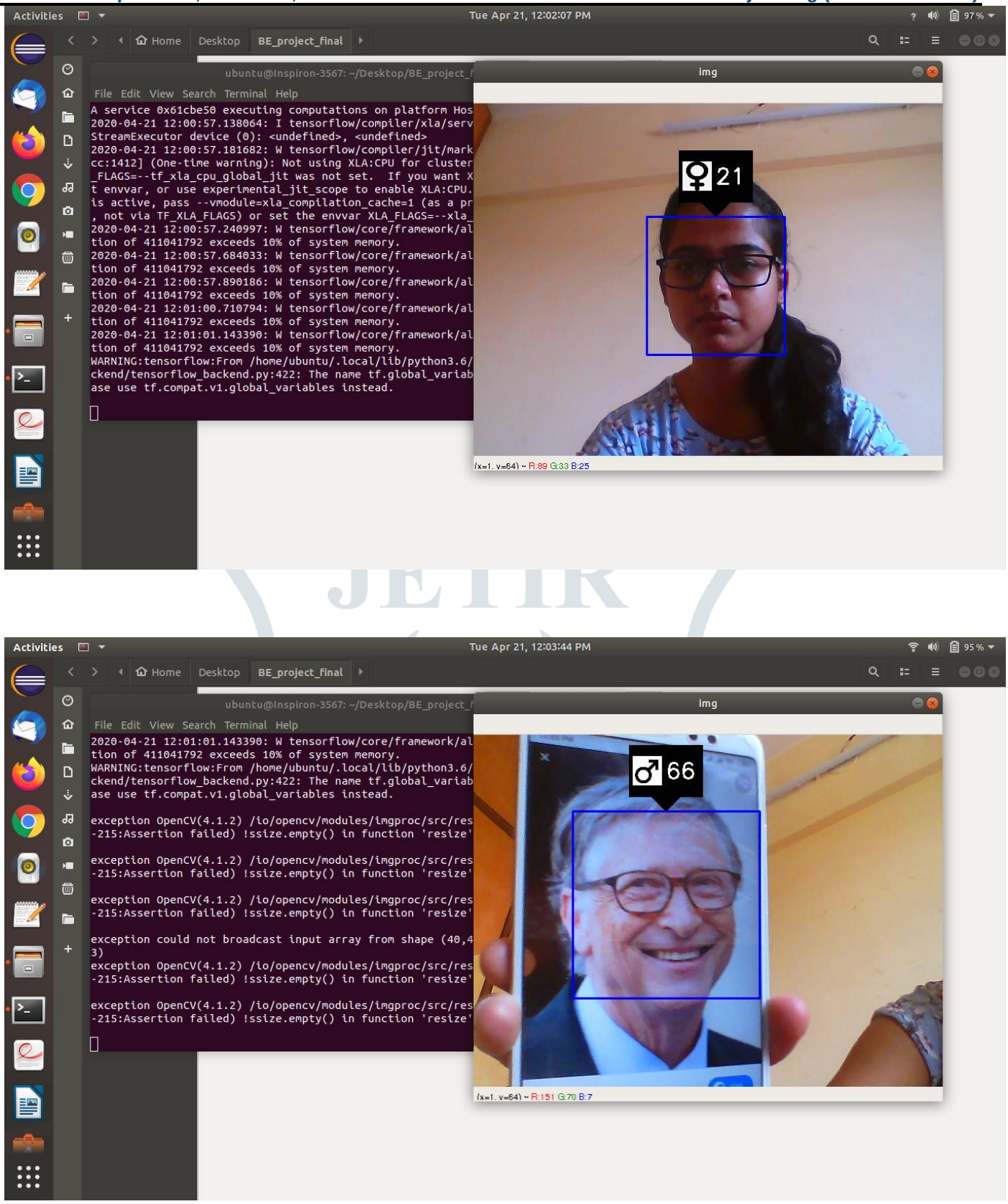


**Figure 3: Age and gender recognition system**

## IV. CONCLUSION

In this paper, we proposed to build an age and gender recognition model using VGG-Face model that was trained on ImageNet. VGG-16 architecture is the most accurate CNN model to build the network. The algorithms used are one of the most accurate methods for age and gender classification. This system works in all working conditions and has been successfully implemented. The database used is the largest known dataset of human faces with meta information. We explored different CNN architectures and various regression and classification methods. In future work, we plan to use more powerful data cleaning methods and improve the accuracy of the model.

## REFERENCES

**[1]** Koichi Ito, Hiroya Kawai, Takehisa Okano,Takafumi Aoki,"Age and Gender Prediction from Face Images Using Convolutional Neural Network",Proceedings, APSIPA Annual Summit and Conference 2018.

**[2]** G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, “Effective training of convolutional neural networks for face-

based gender and age prediction," Pattern Recognition, vol. 72, pp. 15–26, 2017.
**[3]** G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 34-42.
**[4]** J. Beom Ko, W. Lee, S. E. Choi, and J. Kim. A gender classification method using age information. In 2014 International Conference on Electronics, Information and Communications (ICEIC), pages 1–2, Jan 2014.
**[5]** G. Guo and G. Mu, "A framework for joint estimation of age, gender and ethnicity on a large database," Image Vision Comput., vol. 32, no.10, pp. 761–770, Oct. 2014.
**[6]** C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," CoRR, vol. abs/1409.4842, 2014. [Online]. Available: http://arxiv.org/abs/1409.4842.
**[7]** Choi SE, Lee YJ, Lee SJ, Park KR, Kim J. 2011. Age estimation using a hierarchical classifier based on global and local facial features. Pattern Recognition 44(6):1262–1281.
**[8]** J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu, "Age and gender estimation from facial image processing," in SICE 2002. Proceedings of the 41st SICE Annual Conference, 2002, pp. 13-18.