Gender Detection using OpenCV

Preeti Jani
Computer Engineering Sardar Patel Institute of Technology, Mumbai, India
preeti.vinakray@spit.ac.in

Pradnya Topale
Computer Engineering Sardar Patel Institute of Technology, Mumbai, India
pradnya.topale@spit.ac.in

Soham Vaidya
Computer Engineering Sardar Patel Institute of Technology, Mumbai, India
soham.vaidya@spit.ac.in

Abstract- In today's technologically advanced world, many researches have been conducted to reduce the gap between the human brain and machine. Automation in various processes of workplaces is becoming common and constantly growing. In recent years, extracting maximum information from human faces is gaining attention worldwide. Detecting a person's gender and age from facial features like skin structure, ears, eyes, nose, etc. has found its uses in fields like surveillance, forensics, social media and marketing intelligence. Our proposed work is a Transfer learning based approach VGG16 model that is trained for the task of predicting gender and age of a person based on the input image of the human face. It is built using Keras and OpenCV. We have trained our model on the Adience Benchmark Gender and Age Detection dataset and obtained an accuracy of %.

Index terms - *OpenCV, VGG16, gender estimation, age estimation, face detection*

I. INTRODUCTION

Over the past several years, the quantity of data, mainly images that are getting uploaded to the Internet has grown at a large rate. This recent growth in the data has encouraged computer scientists to solve problems in computer vision. This has many applications such as a sentiment analysis system that analyzes human sentiments based on facial features, movie recommendation system based on the human age, etc. Until now only the analysis of images like, how many faces are in the picture and where the faces are located are carried out. But now the research on the characteristics which the face possesses is ongoing and has a broader scope. The goal of this project does exactly that by attempting to detect the age and gender of the faces in an image. Research on this technology is ongoing and has a very broad scope. Its applications have high potential which can make an impact on society. Detecting age and gender from an image is a challenging problem than many other tasks in the field of computer vision. For predicting anything we need to train our model with sample data. The difficulty lies in the nature of the sample data that is needed to train these types of systems. In this era of the internet for general object classification tasks, we often have access to millions of images. But the data needed for supervised learning should be labeled data that means the images should be labeled with age and gender. Finding this type of data is challenging since they are very small in number when compared to labeled data. For labeling data, the real problem is that we don't have access to some personal information of the people like their date of birth and, they may not be accurate. So, for our work in detecting age and gender, we are using the Adience Benchmark Gender and Age Detection dataset. This dataset contains 26,580 facial images with age and gender labels across 5 files..

Through our work, we propose a system using the VGG16 that will detect the age and gender of the person through an input of the human face. We first analyzed the dataset using various plots such as bar graphs, pie graphs and noted down some observations. Following this, we performed data pre-processing and labeled the genders and age columns. The age column was labeled into 8 age groups ranging from 0 to 60+. We built our model and then trained it on the Adience dataset and achieved an overall accuracy of 93.84%.

In the following sections, we have discussed: Section II presents the literature survey related to our study, Section III briefs about Computer Vision, Section IV presents the Methodology, Section V elaborates our results and Section VI concludes our paper.

II. LITERATURE SURVEY

In this section, various works that have developed speech emotion recognition systems have been studied.

Akash B. N et al. in [1] proposed a Deep Learning model using Convolutional Neural Networks (CNN) and Pytorch framework. They trained their model on the labeled IMDb-wiki dataset that contains more than 5 lakh images. They first identified a human in the image, processed it using OpenCV framework to get the output label for gender and age. Similarly in [2], they have also used CNN and OpenCV for facial recognition on the OIU-Adience dataset. For gender and age detection, they used the CAFFE model, a framework in Deep Learning and Protocol Buffer Files for data serialization. Abhinav Banerjee et al. [3] used the Local Binary example Histogram (LBPH) strategy and Convolution Neural Network (CNN) to extricate the facial highlights of face pictures with low computational unpredictability for face acknowledgment and sex assessment. The testing constant video and acknowledgment rate was sped up by CNN. They achieved an accuracy of 63% using LPBH. The [4] proposed a Convolutional Neural Network (CNN) based architecture for detection of age & gender classification. They labeled the input images as 8 labels of age and 2 labels of gender. They used the Viola-Jones algorithm for preprocessing - feature extraction that serves as the input to the model. They used the IMDB wiki dataset. Along with this, they used the FER-2013 test set for predicting emotions wherein they obtained a state-of-the-art accuracy of 70.47%.

In [5], the preprocessing is applied to the image. Features are extracted from the neural network through the convolution network. Based on the trained models the image is then classified to one of the age classes. Features are extracted from the images for further processing. The features are processed further and sent to the training systems. The databases provide a study of the features and help in completing the face detection for proving the age detection of the person in the image. Paper [6] network is pretrained on an IMDb-WIKI with noisy labels and then fine-tuned on MORPH-II and finally on the training set of the OIU-Adience (original) dataset. The experimental results, when analyzed for classification accuracy on the same OIU-Adience benchmark, show that their model obtains the state-of-the-art performance in both age group and gender classification. It improves over the best-reported results by 16.6% (exact accuracy) and 3.2% (one-off accuracy) for age group classification and also there is an improvement of 3.0% (exact accuracy) for gender classification. In [7], Levi and Hassner proposed a simple convolutional neural network architecture with 5 layers, to perform age and gender prediction. Despite the simplicity of this model, they achieved promising results on Adience benchmark for age and gender estimation.

III. COMPUTER VISION

Computer Vision helps the computers and enables them to look at/see, figure out and identify digital images and videos as a human would. The challenges it faces are mostly because of the biological vision and the advantage humans have over a machine with the millions of years of evolution our civilization went through, this is something the computer faces challenges in, it fails to understand the biological vision. Computer vision majorly involves acquiring, processing, analyzing and understanding digital images to extract data from the real world in order to generate symbolic or numerical information that it uses to make decisions on. This process includes different practices like object recognition, tracking a video, motion estimation and image restoration.

IV. METHODOLOGY

A. Dataset

In this study, we used the dataset named Adience Benchmark Gender and Age Classification having facial images.. The detailed information about it is as follows:

The Adience dataset, published in 2014, contains 26,580 photos across 2,284 subjects with a binary gender label and one label from eight different age groups, partitioned into five splits. The key principle of the data set is to capture the images as close to real world conditions as possible, including all variations in appearance, pose, lighting condition and image quality, to name a few.

It has 5 files each with the following columns -

1. user_id
2. original image
3. face_id

4. age
5. gender
6. x
7. y
8. dx
9. dy
10. tilt_angle
11. fiducial_yaw_angle
12. fiducial_score

B. VGG16

The input to the network is an image of dimensions *(224, 224, 3)*. The first two layers have *64* channels of *3*3* filter size and same padding. Then after a max pool layer of stride *(2, 2)*, two layers which have convolution layers of 256 filter size and filter size *(3, 3)*. This is followed by a max pooling layer of stride *(2, 2)* which is the same as the previous layer. Then there are *2* convolution layers of filter size *(3, 3)* and *256* filters. After that there are *2* sets of *3* convolution layers and a max pool layer. Each has *512* filters of *(3, 3)* size with the same padding.This image is then passed to the stack of two convolution layers. In these convolution and max pooling layers, the filters we use is of the size *3*3* instead of *11*11* in AlexNet and *7*7* in ZF-Net. In some of the layers, it also uses *1*1* pixel which is used to manipulate the number of input channels. There is a padding of *1-pixel* (same padding) done after each convolution layer to prevent the spatial feature of the image.

After the stack of convolution and max-pooling layer, we got a *(7, 7, 512)* feature map. We flatten this output to make it a *(1, 25088)* feature vector.After this there are *3 fully* connected layer, the first layer takes input from the last feature vector and outputs a *(1, 4096)* vector, second layer also outputs a vector of size *(1, 4096)* but the third layer output a *1000* channels for *1000* classes of ILSVRC challenge, then after the output of 3rd fully connected layer is passed to softmax layer in order to normalize the classification vector. After the output of classification vector top-5 categories for evaluation. All the hidden layers use ReLU as its activation function. ReLU is more computationally efficient because it results in faster learning and it also decreases the likelihood of vanishing gradient problem.

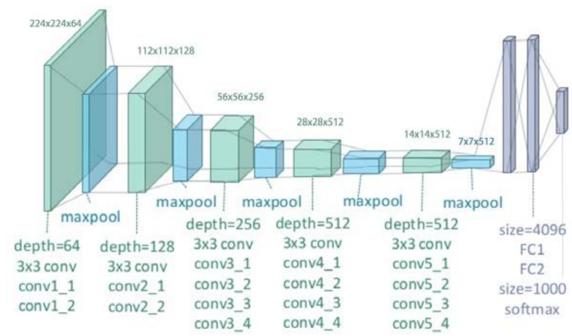


Fig.1: VGG16 Architecture

C. Libraries Used

1. OpenCV

OpenCV is short for Open Source Computer Vision and it is an open-source Computer Vision and Machine Learning library. This library is mainly used in processing the image and video real-time and is also used for boasting analytics. Moreover, it also supports multiple Deep Learning frameworks like TensorFlow. Caffe, and PyTorch.

OpenCV has a structure which means that the package includes many shared or static libraries.

2. Tensorflow

TensorFlow is a library in Python which is used for quick numerical computation and is created and released by google. It is a foundation library and it is mainly used to create deep learning models directly, and sometimes by using wrapper libraries that are used to simplify the process that is built on top of tensorFlow.

3. Keras

Keras is basically a high-level Application Programming Interface (API) of Tensorflow, it provides the necessary abstractions and the major building blocks that are needed for the development and shipping of the machine learning solutions along with an extremely high iteration velocity. Keras helps engineers and researchers and pushes the user to take full advantage of the scalability of Tensorflow, along with the cross-platform capability that it holds, which would not have been possible without Keras. TPU can also be run on a large cluster of GPUs, and you can also export the keras models to run in google chrome or any browser or any smartphone or mobile device.

4. Math

This is the most fundamental library which is used for almost all the programs which involve mathematical functions. This module provides access to them which are defined by C-standard.

D. Proposed Methodology

In our work, we have performed the following steps

- Prepare the data
- Use transfer learning to build th model
- Train and compile the model

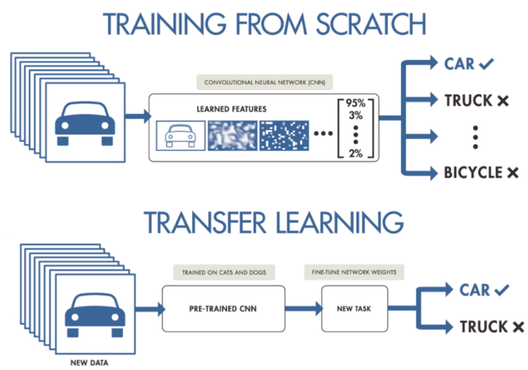
In this implementation first, we have used our proposed system architecture through our experiments. The schematic diagram for our VGG16 model is shown in Figure 1.

We visualized the dataset using pie graphs and bar graphs based on the columns of gender and age. We mapped the age column into 8 age groups such as

1. 0-2
2. 4-6
3. 8-13
4. 15-20
5. 25-32
6. 38-43
7. 48-53
8. 60+

Along with this, we dropped a few data that was not required such as the unknown gender in the gender column and the records that do not have the correct age values.

There are two primary items encapsulated inside, which are forward function definition and a weight tensor for each layer. We are specifying our layers as attributes inside our Network classes that the training process starts. The Network learns the weight values and then they are updated to the weight tensor inside each layer. We pass values for each parameter to the layer's constructor when we are constructing a layer. The convolutional layers contain three parameters and the linear layers contain two parameters.



Now we have finished building our model. We will now move into training our model. We have split the dataset into 80:20 that means the model would be trained on 80% data and the remaining 20% would be used as the testing set.

V. RESULTS

In this section, we present the results of our Gender and Age Detection System. The VGG16 model used in the paper gave an accuracy of **93.84**%.

CONCLUSION

Though many methods have solved the problems that were raised when detecting age and gender, until recently, most of these methods have focused on images which had constraints and limitations, and which were maintained in lab conditions. Such conditions will have an impact on the methods we perform on real-world images on social websites and online platforms. Internet images are abundant. Anybody can have access to the internet which consists of huge collections of real-world images which makes the training process very effective for our machine learning-based systems. Even though for supervised learning we want a large collection of labeled data, this huge availability of data makes our problem simpler. Take this example which is related to the problem of the face recognition system. We have trained our network based on this internet data. From this work we can conclude with two important conclusions. First, even though there is less availability of labeled images for age and gender, VGG16 can be used to provide improved age and gender detection results.

Second, the performance of this system can be improved marginally by using more training data and more elaborate systems.

REFERENCES

[1] Akash. B. N , Akshay. K Kulkarni , Deekshith A. , Gowtham Gowda, “Age and Gender Recognition using Convolution Neural Network ”, International Journal of Engineering Science and Computing, June 2020.

[2] Mahija Kante, Dr. Esther Sunandha Bandaru, Gadili Manasa, Meghana Emandi, Vanarasi Leela Lavanya, “Age and Gender Detection using OpenCV”, International Journal of Advance Research, Ideas and Innovations in Technology, Volume 7, Issue 3 - V7I3-2163, 2021.

[3] Prof. Supriya Mandar Khatavkar , Abhinav Banerjee, Hemanka Sarma, Anurag Sharma, “Gender and age detection using deep learning techniques ”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 09, Sep 2020.

[4] Sidharth Nai, Dipesh Nair, Gautam Nair, Anoop Pillai and Prof Sujith Tilak, “Detection of Gender, Age and Emotion of a Human Image using Facial Features”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 05, May 2020.

[5] Eran Eidinger, Roee Enbar, and Tal Hassner,” Age and Gender Estimation of Unfiltered Faces”, 2014 IEEE

[6] Olatunbosun Agbo-Ajala, Serestina Viriri, "Deeply Learned Classifiers for Age and Gender Predictions of Unfiltered Faces", *The Scientific World Journal*, vol. 2020, Article ID 1289408, 12 pages, 2020. https://doi.org/10.1155/2020/1289408

[7] 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.

[8] K. Ito, H. Kawai, T. Okano and T. Aoki, "Age And Gender Prediction From Face Images Using Convolutional Neural Network", 2018.