International Journal of Research in Engineering, Science and Management
Volume 6, Issue 3, March 2023
https://www.ijresm.com | ISSN (Online): 2581-5792

Gender and Age Detection using Deep Learning

Divyanshu Singh[1*], Akansh Bhatnagar[2], Sunil Kumar[3]
[1,2]Student, Department of Information Technology, Meerut Institute of Engineering and Technology, Meerut, India
[3]Department of Information Technology, Meerut Institute of Engineering and Technology, Meerut, India

***Abstract*: Most applications now require automatic gender and age prediction, especially with the advent of social media and social platforms. Furthermore, existing technologies' performance on actual imageries is pointedly absent, exclusively when associated to the mammoth rises in performance freshly testified for the related endeavor of facial recognition. In this paper, we demonstrate how deep-CNN can be used to learn representations (CNN). The five phases of the predicted technique are facial recognition, environment removal, face alignment, numerous CNN, and voting systems. This model is tested on the recent Audience-Face benchmark face dataset for gender detection and age approximation, and it is implemented using Python software.**

***Keywords*: Convolutional neural networks (CNN), Deep learning, Human face recognition, Computer vision, Region of interest (ROI).**

1. Introduction

Gender and age are two most important attributes of face and plays a very significant role in community and social gathering and interactions. Making gender detection and age approximation is a very important task for most of the intelligent applications, such as human-computer interaction, access control, marketing intelligence, visual surveillance and law enforcement etc and that too using a single face frame. Past approaches for the estimation or classification of these attributes from different images of face depends on the differences in features of facial dimensions. In the last few years, the face recognition techniques have shown that a significant amount of growth could be made by using Convolutional neural network (CNN). This research paper suggests an idea for the use of multiple CNN for gender and age classification. The system includes three CNN layers separated from one other, and they are linked together through a system of voting which depends on the majority for the prediction of the final class. The 3 Convolutional layers used are:

- CNN layer 1- consists of 96 nodes and a kernel size 7
- CNN layer 1- consists of 256 nodes and a kernel size 5
- CNN layer 1- consists of 384 nodes and a kernel size 3

Every CNN in the system contains a particular layer of convolutional with varying depth and has a different architecture.

In this paper, we have attempted to reduce the difference between the age and gender estimation methods and those of automatic face recognition capabilities. In this methodology, the input is taken in a form of an image frame and an algorithm will run to estimate the age and detect the gender of person in that image frame. For age approximation, we have divided the age into some groups in form of ranges such as [(0-2), (4-6), (8-12), (15-20), (25-32), (38-43), (48-53), (60-100)] and the frame will fall in one of the given range groups. And the gender will be either female or male. Our model is tested on the audience benchmark dataset for gender and age classification of face images which were unfiltered. This model has been tested twice, one for the classification of age and other for the classification of gender.

Automatically estimating the age of a person and detecting his/her gender is significant in many fields of work. A variety of approaches exists in this area of computer vision, mostly based on machine learning, to accomplish this task. It would be beneficial for the businesses as the shopkeeper or the owner of the mart would be able to analyse the age and gender of people coming for buying products in his shop. So, he would be able to fulfil the demands of different age groups and gender effectively and efficiently. It would result in better productivity of the business as well as the society at large.

2. Literature Review

In order to better understand how we will develop a new CNN structure in our study, we will examine some prior research that emphasizes the importance of age and gender recognition. CNN's organizational structure has an impact on the performance of recognition or prediction.

Karen Simon and Andrew Zisserman [1] devised the VGG-16 structure for CNN implementation, which included 13 convolutional layers and three fully connected layers. All of the filters they used were 3*3 in size, and they discovered that the best image recognition occurred when the depth of the structure was between 16 and 19 layers.

He et al. [2] presented a 152-layer CNN structure that overcomes the complexity issue that develops as network depth grows by learning the network using residual learning. After empirically obtaining the best results for facial image identification accuracy, they built this structure with 1000 layers without any problems. ResNet-152 was the name given to the newly created CNN.

Deep residual networks have been shown in the past to grow in depth and include hundreds of layers while still functioning properly [5].

The issue with this type of network, however, was the latency

*Corresponding author: divyanshusingh270801@gmail.com

and time loss caused by the layers' frequent repetition. They provided a revolutionary design known as the ResNet block, which is based on decreasing the depth of the net while increasing the breadth of the residual net. They carried out experiments to prove that this structure was more effective than the previous one; this network was dubbed the Wide Residual Net.

Deep Alex Krizhevsky et al. [7] trained CNN to classify over one million photos into thousands of classes, and they used the Dropout approach to reduce overfitting inside the fully linked layers. The CNN was built with twelve layers, including three fully linked and eight convolutional layers.

Gil Levi and Tal Hassaner [3] proposed a CNN structure with five layers—two fully connected and three convolutional. The FERET dataset was used for both the training and test samples. In this structure, two techniques were used for prediction.

The first step was to centre crop the source image, which reduced its dimensions to 227 by 227. The second method was over-sampling, which divided the image into five 256*256 sections. When their experimental findings were compared to those of previous researchers, they demonstrated good gender.

3. Methodology

1) For detection of face, we have a .pb file; this has the trained weights for the system model. We use this for the successful execution of the trained system model. And for gender and age, .prototxt files are used to represent the configuration of the network and .caffemodel file is used for defining the parameters and the internal states of the various layers.
2) For creating the argument parser, we have used the argparse library with the help of which makes the arguments of the image to be inserted into the model by the command prompt.
3) Now, the protocol is initialized; buffer and the model for the age, face and gender.
4) Initialize mean values of the list of ranges for the age classification and list of genders to classify.
5) Use readNet() method for loading the network. One parameter from them holds weights that were already trained and the other holds configuration of the network.
6) Now we have to capture a video stream for classifying the image or stream using a webcam. The padding is set to 20.
7) As long as no is pressed, the stream is read and content into the frames. If it is nor a video, waitKey() is called from cv2.
8) Now, highlightFace() function is called with FaceNet and Frame parameters, and the result is observed i.e. what it returns that will be stored by the names given as faceBoxes and resultImg. Moreover, if the number of faceBoxes is 0, it means that no face is detected.

   The faceNet is a DNN Face Detector model that holds approximately 2.7 MB of disk space.

   - A copy of frame is created and it's weight and height is measured.
   - Now, with the help of shallow copy, create a blob.
   - After setting up the point a forward pass is made to the network.
   - Now, the faceBoxes is a type of an empty list. For every value define the confidence. If the confidence is larger than the threshold confidence, we get the coordinates of the faceBoxes and we append their list on the faceBoxes.
   - Now, a rectangle is placed on the image for every such list of coordinates.
9) For every rectangle, a 4-dimensional blob is created to define the face. And for creating this, we resize scale it, and forward it to the mean value.
10) Now, we pass an input to the model to get the confidence of both the classes after giving a forward pass to the network. The gender of the person in the image is specified by the confidence that is greater.
11) Now, same thing is repeated for the prediction of age.
12) At last, we will add both the age and gender texts to the resultant image and display it with the help of imshow().

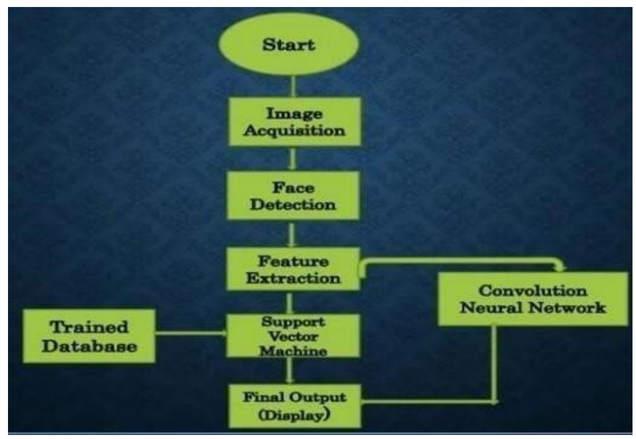

Fig. 1.

4. Proposed System

Correct identification of gender is kind of difficult particularly when the image When the image is taken from a far distance by using haar-like features. To solve this issue, we have used an effective idea of applying the cascaded method. This model uses ROI (Region of interest) as our face. We used frontal face images for training our model which includes external features like glasses, earings, makeup, hairstyle and other accessories.

A. Convolutional Neural Network

CNN or convolutional neural network is an algorithm of Deep Learning which takes the image file as an input and that image file is used by the CNN to adapt and learn about the different features of that image. This practice help the Convolutional Neural Network to know and learn about the various objects and features that are present in that image file. For instance, we can say that the CNN will the distinctive features of a male face that differentiate from that of a female face so when the input image of a male or a female is provided to CNN, then it can easily differentiate between them and gives the output as "male" or "female". Moreover, some of the unique features of CNN that makes it different from other similar kind of algorithms are its capability of preprocessing the data by itself. So, the user does not have to take the stress about accommodating a lot of resources just for the data preprocessing. However, during the start in the initial phase it requires some amount of manual work or efforts but as it proceed towards the progress in the training they are able to adapt and implement the features that they have learned develop their own filters. So, we can say that this algorithm of Deep Learning is a continuously growing and evolving algorithm with data.

The convolutional Neural Network will work in 3 convolutional layers for this project as shown below:

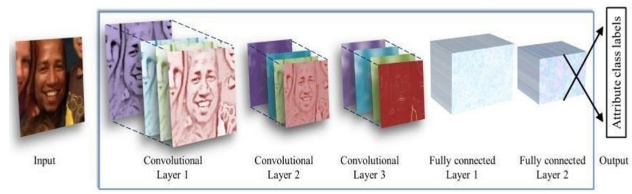

Fig. 2.

B. OpenCV

Open-source computer vision (OpenCV) is a open source library for computer vision and machine learning. This library is used for processing the real time images and videos. Apart from processing the images and videos, it also boasts the analytical capabilities. Like Caffe, PyTorch and TensorFlow this library also supports the deep learning framework.

Features of OpenCV Library:

- It can write and read the frames.
- It can process frames.
- Using, OpenCV, we can save as well as capture videos.
- Using this, we can detect some specific objects in frames and videos such as ears, eyes etc.
- It can also help in detection of feature.
- Using OpenCV, we can remove the background, estimate the motion and perform object tracking in frames and videos i.e., we can analyze the videos.
- The origin of OpenCV is C++. Also, some of the features of Python and Java are also there. OpenCV is capable to run on a range of Operating Systems like Linus, windows, FreeBSD Open BSD, Net BSD OSx etc.

C. The Audience Dataset

In this deep learning project, we have used the audience dataset.

Fig. 3.

5. Results and Discussion

The results for the respective inputs have been shown below. For the classification of age, we have compared the accuracy when the result of the algorithm is exactly same as desired age group or when it differs by one adjacent age group. We have increased the amount of dataset for training our model to improve the accuracy and performance of our model. Also, for some non-living creatures it will give the result as “No face detected”. Some of the outputs are given below:

Fig. 4. Input image-1

Fig. 5. Output-1

Here, the model first detecting the face in the image then it is specifying the gender and the age group of the person in the image.

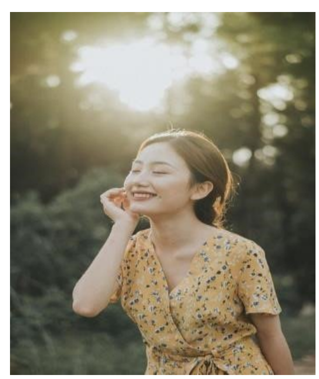

Fig. 6. Input image-2

Fig. 7. Output-2

Here, the model first detecting the face in the image then it is specifying the gender and the age group of the person in the image.

Fig. 8. Input image-3

```
PS C:\Users\LENOVO\Downloads\gad> py gad.py --image kid1.jpg
Gender: Male
Age: 4-6 years
```

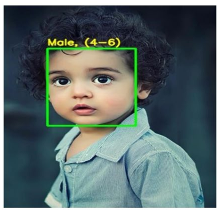

Fig. 9. Output-3

Here, the model first detecting the face in the image then it is specifying the gender and the age group of the person in the image.

Fig. 10. Input image-4

```
PS C:\Users\LENOVO\Downloads\gad> py gad.py --image minion.jpg
No face detected
```

Fig. 11. Output-4

Since the model is unable to detect the human face in the given picture so it is giving the output as “No face detected”.

6. Conclusion and Future Scope

In this paper, we propose a model for gender and age classification that makes use of multiple sub-CNN and other machine learning techniques. Each sub-CNN is evaluated independently, and the final model is evaluated using voting mechanism. The idea behind using different sub-CNN separately and then voting mechanism is that we can get a diverse representation of facial features by using different sub-CNN. Thus, it helped in performing a more accurate classification of age prediction. Also, experimentation have shown that by using multiple CNN model we can get a lower error rate than a single CNN model. The methodology includes grouped approaches and calculations whereby the deep learning is the prime component in usage designs in this model.

References

[1] G. Panis and A. Lanitis, “An overview of research activities in facial age estimation using the FGNET aging database,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015.

[2] G. Antipov, S. A. Berrani, and J. L. Dugelay, “Minimalistic CNN-based ensemble model for gender prediction from face images,” Pattern Recognit. Lett., 2016.

[3] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui, “Longitudinal face modeling via temporal deep restricted Boltzmann machines,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.

[4] X. Li, Y. Xu, Q. Lv, and Y. Dou, “Affine-transformation parameters regression for face alignment,” IEEE Signal Process. Lett., 2016.

[5] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, “AgeDB: The First Manually Collected, In-the-Wild Age Database,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2017.

[6] A. Jourabloo and X. Liu, “Pose-Invariant Face Alignment via CNN-Based Dense 3D Model Fitting,” Int. J. Comput. Vis., 2017.

[7] R. Rothe, R. Timofte, and L. Van Gool, “Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks,” Int. J. Comput. Vis., 2018.