2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.9377041

Gender and Age detection using Deep Learning

Aryan Saxena
Department of Computer Science
Amity School of Engineering and Technology
Amity University, Noida
Uttar Pradesh, India
aryan.saxena094@gmail.com

Prabhangad Singh
Department of Computer Science
Amity School of Engineering and Technology
Amity University, Noida
Uttar Pradesh, India
prabhangadchopra99@gmail.com

Shailendra Narayan Singh
Department of Computer Science
Amity School of Engineering and Technology
Amity University, Noida
Uttar Pradesh, India
snsingh@amity.edu

Abstract - Mankind has advanced technology to the extent that the 21st century is the crack of the dawn for unimaginable achievements. The aforementioned technology can be used for our benefit in identifying one's age and even their gender just based on their glimpse from a camera, image and even a video. This research paper will methodically chalk out the whole procedure, multiple methodologies and algorithms that can be used, which one is the most accurate and how it all comes together. It will also stress on its importance and how it can be implemented to benefit our day to day life. The paramount objective of the paper is to build a gender and age detector that can approximately guess the gender and age of the face of an individual in a picture using Deep Learning on the audience dataset. Moreover, to get the most effective predictions and result by overcoming the problem of accuracy and time. Moreover, the map for the ways this technology can be used to our benefit and look at the huge spectrum where it can be implemented: ranging from security services, CCTV surveillance and policing to dating applications, matrimonial sites.

Keywords - Convolutional Neural Network, Deep learning, Computer vision, Artificial Intelligence, Machine Learning.

1. INTRODUCTION

In today's world, images and videos are the things that keep everyday tasks going, from security surveillance to looking at images of cute dogs. But these images and videos can help bring a change in how people function and can help the authorities to work in a more efficient and convenient way. Learning how firms can use the prediction model that will be developed and how it can be implemented into our daily lives and can help make the world a safer place.

The prediction algorithm that will be implemented will work in a way so that our model will be able predict the age and gender/sex of a person just by looking at an image or a video which can be CCTV surveillance footage [2].

But the question comes to mind: *Why is it needed?*

Let me answer this question with a simple example: A fugitive stole $5million from Pacific Standard Bank, and is on the loose. His identity has still not been confirmed because of multiple forged identities. But the police does have an idea about the age and the gender of the person. This algorithm can be employed as a visual surveillance model in the surveillance cameras in the 1mile radius from the bank, and everyone who matches the age and the gender specification will be checked thoroughly.

In multiple cases around the world, the security forces do not have an accurate description about the fugitive they're looking for, and in these cases our algorithm can work and help the forces.

Apart from this, there is a huge range of applications for this algorithm:

Access Control in matrimonial and dating services so that this model will help the firms confirm the age and gender of the users and make sure the users are authentic and aren't forging their age or faking their identity, intending to harm others. It would be way easier for the firms to track down and narrow down their search for the users using a false identity.

2. LITERATURE REVIEW

Author	Year	Methodology
Zachi I. Attia	2019	trained CNNs using standard 12-lead ECG signals [1]
Zagoruyko	2016	Wide Residual Networks [17].
Sungjoon Choi	2013	Model Compression of DBN-ANN and DBN-R, based on the deep learning framework [8].
Kaiming He	2016	Deep Residual Learning For Image Recognition [16].
James H Cole	2016	Gaussian processes regression and CNN to predict chronological age [4].
Hinton G.E	2012	Image net Classification With Deep CNN [18].

A Toshev	2014	Human Pose Estimation using Deep Neural Networks [10].

3. 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 civilzation 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.

4. ARTIFICIAL INTELLIGENCE

Artificial intelligence is the intelligence possessed by machines so that they can adapt according to the environmental conditions and are able to take decisions in order to maximise the possibility of success [7]. Application of artificial intelligence is increasing day by day including "optical character recognition, handwriting recognition, speech recognition, face recognition, robotics, automation and many more."

Artificial Intelligence deals with:

- ❏ Intelligent behavior depicted by machines or any character in a video game is usually what is called playing against the CPU [13].
- ❏ Search engines such as Google should be capable enough to come up with quick responses to the user requests.
- ❏ Movable robots or machines in dangerous situations can respond smartly and act according to the circumstances and protect the individual at risk.
- ❏ Developing control systems which can respond smartly to a wide variety of problems.
- ❏ Most of the major AI research companies have already given up on the thought of coming up with real machines which have the ability to perceive.
- ❏ Intelligent reaction of user requests in developing and providing services to the user of normal machines is being looked upon these days a lot [10].
- ❏ To come up with procedures that can be on the same level with a human intellectual (to have the skill to envision, perceive, contemplate and analyse, being self aware is the major goal of "Artificial Intelligence")

5. METHODOLOGY

A. CONVOLUTIONAL NEURAL NEWORK

A Convolutional Neural Network is a deep neural network (DNN) widely used for the purposes of image recognition and processing and NLP. Also known as a ConvNet, a CNN has input and output layers, and multiple hidden layers, many of which are convolutional [5]. In a way, CNNs are regularized multilayer perceptrons. Convolutional neural networks have a great significance in image classification. The neural network cannot be fully connected for image recognition for a computer system or any model [14]. Reason being that an image generally has a large size with umpteen number of pixels which makes it difficult for us to use fully connected neural networks as the number of weights required to train a model would be quite high. For example, an image of size 40*40 pixels will requires 40*40*3 weights (the number 3 here signifies the RBG i.e. Red Blue Green protocol) that means 4800 weights which is still possible but here, the image is of size minimum of 200*200 or greater than that. However, to recognise an image of size 200*200 using fully connected neural networks, the required weights would be 120000 (200*200*3), which is not feasible. Therefore, the practical approach is to use convolutional neural networks as in CNN a particular node is connected to three nodes of the next layer (red , blue and green) unlike the fully connected where a node is connected to all the nodes of the next layer. Consequently, required number of weights are lesser and neurons to train the model in CNN that in fully connected neural networks.

CNN configuration generally consist of four layers:

- ❏ Convolution layer
- ❏ ReLU layer
- ❏ Pooling layer
- ❏ Fully connected layer [15].

B. LIBRARIES USED

- OpenCV

OpenCV is short for Open Source Computer Vision and in it is an open-source Computer Vision and Machine Learning library [6]. This library is mainly used in processing the image and video real-time and is also used for boasting analytics. Moreover, it also supports the multiple Deep Learning frameworks like TensorFlow, Caffe, and PyTorch.
OpenCV has a structure which means that the package includes many shared or static libraries.

- 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.

- 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 or 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.

- Argparse

This module helps the user to write command-line interfaces which are user-friendly and easy to grasp. Program specifies the arguments that are needed to be run, and this module, Argparse will make out a way to parse the commands of sys.argv. The module will automatically create and generate the help needed and usage messages and send issue errors if there is a clause in the argument which is invalid or is causing a problem.

- 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.
However, this module does not provide support for complex numbers and functions cannot be run on complex numbers. In order to do that, there is another module labeled 'cmath' which helps provide the support for all the complex number functions that might be needed. The reason for this distinction is that most of the users don't require complex numbers and dont understand it. So, the basic math module has excluded it from the module. Most of the users perform basic and advanced mathematical functions on integers itself. Moreover, it's easier to receive an exception in a case of 'if-else' or any conditional statement rather than getting a complex result and the programmer would have to determine the result in order to run the conditional statement.

C. DATASET

In order to get the most effective results with high accuracy an appropriate dataset was found from kaggle.com. Our models would be trained on this dataset to get effective results. It is 1 GB file which consists of 26,580 photos of 2,284 individuals in the eight age groups of (0 – 2), (4 – 6), (8 – 12), (15 – 20), (21 – 24), (25 – 32), (38 – 43), (48 – 53), (60 – 100). These images are procured from Flickr albums and distributed under official license of CC i.e. Creative Commons. There are a plethora of pictures for practical conditions including appearance, expressions, poses, noises and lighting (either of bright and dark).
This particular dataset was chosen because of the spectrum of photos it contained. With the help of this dataset, more diverse data can be fed which will in turn help the algorithm to work better and more efficiently.

D. FRAMEWORK

Here a function facehighlight() is developed, which will get the image from the data set or can take the real time image and detect the peculiare face dimensions in the image using the predefined model named as opencv_face_detector.pbtxt. Firstly, the width and height of the face is recognised using the model. Secondly, One of the most fundamental functions used in deep learning algorithms is introduced while image recognition known as blobfromimage(). This function is used to preprocess the image. The functionality of this method is that it tries to bring your image onto the same level of property as the original image. The researchers who deployed or made these models went through different types of images which they themselves collected and they had their own database. Those images are completely different on the basis of property level like brightness, profile, saturation. So the function blobfomimage() converts the image into this bare minimum format by removing the variables and extra factors and using that image for detection. The main steps involved are mean subtraction, scaling, altering of mean values of model i.e. Red, Green and Blue and then finally channel swapping.

6. EXPERIMENTAL ANALYSIS

```
import cv2
import math
import argparse
```

These libraries have been imported which have been explained earlier in the section of libraries used.

```
def facehighlight(net, frame, conf_threshold=0.7):
    frameOpencvDnn=frame.copy()
    frameHeight=frameOpencvDnn.shape[0]
    frameWidth=frameOpencvDnn.shape[1]
    blob=cv2.dnn.blobFromImage(frameOpencvDnn, 1.0, (300, 300), [104, 117, 123], True, False)

    net.setInput(blob)
    detections=net.forward()
    faceBoxes=[]
    for i in range(detections.shape[2]):
        confidence=detections[0,0,i,2]
        if confidence>conf_threshold:
            x1=int(detections[0,0,i,3]*frameWidth)
            y1=int(detections[0,0,i,4]*frameHeight)
            x2=int(detections[0,0,i,5]*frameWidth)
            y2=int(detections[0,0,i,6]*frameHeight)
            faceBoxes.append([x1,y1,x2,y2])
            cv2.rectangle(frameOpencvDnn, (x1,y1), (x2,y2), (0,255,0), int(round(frameHeight/150)), 8)
    return frameOpencvDnn,faceBoxes
```

This is the facehighlight() function developed by us. This function takes the network and the frame (image) as the parameters in order to effectively execute the function. In this function:

- In order to fetch the facial dimensions, a copy of the frame is generated.
- The copy is converted into a blob using blobfromimage().
- Forward pass to the network is given after setting the input.
- The confidence is defined in the range 0-1 for each value in 0 to 127 for the empty list faceBoxes. Coordinates x1, y1, x2, y2 are obtained when the confidence beats the threshold of 0.7(as defined in the parameter).
- The coordinates are appended to the list faceBoxes.
- Each list of coordinates in the faceBoxes accounts for the rectangles on the image.
- The function returns the copy of frame as well as list of faceBoxes.

```
parser=argparse.ArgumentParser()
parser.add_argument('--image')

args=parser.parse_args()
```

The argparse library is used to create an argument which can be parsed to get the image from the command prompt. The parser parse the argument which holds the image file (image path).

```
faceProto="opencv_face_detector.pbtxt"
faceModel="opencv_face_detector_uint8.pb"

faceNet=cv2.dnn.readNet(faceModel,faceProto)

```

Storage of the neural network file for face detection is done in faceProto which is the proto buffer and faceModel which has the model. Using the predefined readnet() method, the network is loaded using faceModel and faceProto as arguments. The former parameter consists of trained weights and the later holds the configuration of the network.

```
video=cv2.VideoCapture(args.image if args.image else 0)
padding=20
```

VideoCapture() function from openCV library is used to read the image. It either gets the image from argument parser or else it will get a live image using the webcam i.e. VideoCapture(0) by setting padding value as 20.

```
while cv2.waitKey(1)<0:
    hasFrame,frame=video.read()
    if not hasFrame:
        cv2.waitKey()
        break

    resultImg,faceBoxes=facehighlight(faceNet,frame)
    if not faceBoxes:
        print("No face detected")



    cv2.imshow("Detecting age and gender", resultImg)
```

- Now a while loop is used to continuously read the image until any key is pressed. Live Stream from webcam is read and then stored in hasframe and frame. A waitkey is again called followed with break. This is when it is not a live video and an image is read from argparse.
- facehighlight() function is called with face detection network configuration (facenet) and frame as parameters and the output is stored in variables resultimg and faceboxes.
- If the network does not recognise any face then faceboxes list would be empty and hence a message is displayed saying “no face detected”.
- Imshow is used to display the result.

7. RESULT

IMAGE TEST 1:

A Line of code is run, where the model is implemented on the given photo, and then the model successfully reads the face dimensions and highlights it using the facebox.

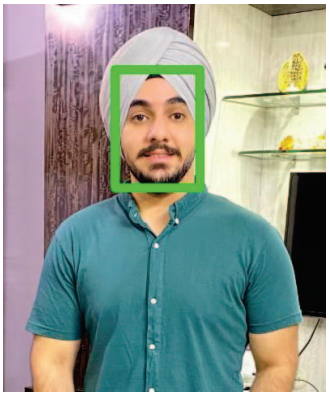


IMAGE TEST 2

When the image of a minion (a cartoon character) is introduced, the model was unable to detect a human face and displays a message saying 'no face detected'. This shows that the model is quite accurate in the process of face detection and does not register redundant faces which do not resemble the human face features.

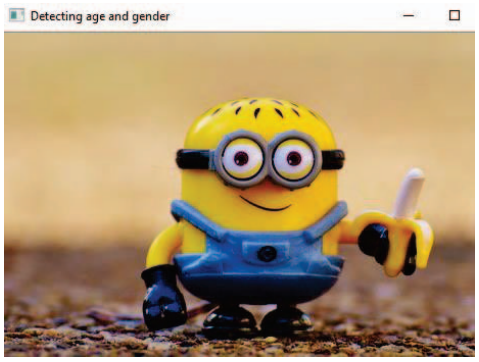
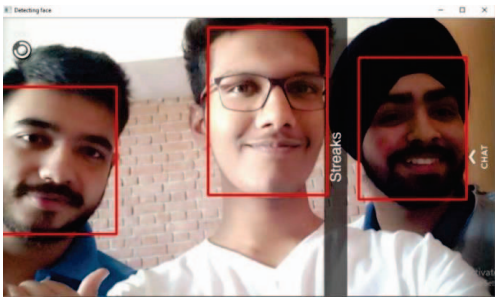


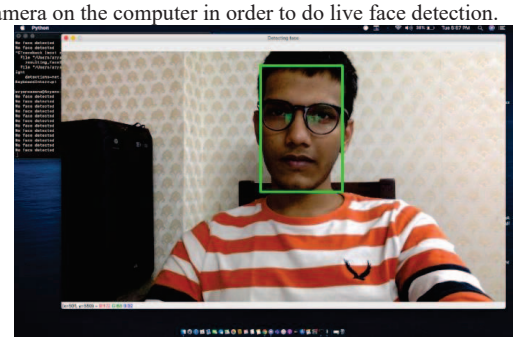
IMAGE TEST 3:

When an image with multiple people is parsed through the model, it accurately detects the face of every single person and depicts it with the help of facehighlight() function.



REAL TIME IMAGE:

In a case, where no argument has been passed for an image, then the model automatically prompts to get access of the web camera on the computer in order to do live face detection.



In the image above, the argument has been parsed and the computer gets access of the webcam and detects the face in real-time.

8. CONCLUSION AND FUTURE SCOPE

Further on with this Prototype, Deep Learning and convolutional neural networks will be implemented in order to accurately identify the gender and range of the age to which a person belongs to, from a single image of a face. The image may belong to a particular dataset, could be singularly parsed through the prototype or may even be realtime in case no argument is parsed. The predicted gender is binary, and will be placed in the categories of 'Male' and 'Female', and the predicted age may be one of the following ranges- (0 – 2), (4 – 6), (8 – 12), (15 – 20), (21-24), (25 – 32), (38 – 43), (48 – 53), (60 – 100) (8 nodes in the final softmax layer).
It is extremely strenuous to precisely figure an accurate age from a picture in the view of elements like use of cosmetics, lighting, deterrents, and outward appearances. And therefore, instead of making it a regression, the consideration is done as a classification problem. Moreover, more images will be parsed through the model in order to predict the age and gender. Further experiments will be done in order to get accurate gender and age in a real time image as well.

Tentative process of how the model would function:

- Detection of the facial area
- Classification into the gender (Male/Female)
- Classification of age into one of the 9 age groups mentioned above
- We shall display the resultant on the image using the facebox.

The usage of this particular prototype has a huge spectrum of how and where it can be implemented. Like discussed multiple times in the project prior, this algorithm will be extremely beneficial to industries which have a massive employee size, or to the surveillance firms who need to keep a track of the people entering, or to the government sector where the security forces might use this algorithm to track a suspect or a potential threat to the public.

Some potential applications of the prototype would be:

1. Targeted Advertisements over social media.
2. Visual Surveillance
3. Dating/Matrimonial service.
4. Human-computer interaction

REFERENCES

1. Zachi I. Attia , Paul A. Friedman , Peter A. Noseworthy , Francisco Lopez-jimenez , Dorothy J. Ladewig , Gaurav Satam , Patricia A. Pellikka , Thomas M. Munger , Samuel J. Asirvatham , Christopher G. Scott , Rickey E. Carter , Suraj Kapa Age And Sex Estimation Using Artificial Intelligence From Standard 12-lead Ecgs
2. Levi, G., &Hassner, T. (2015) Age And Gender Classification Using Convolutional Neural Networks. In Proceedings Of The Ieee Conference On Computer Vision And Pattern Recognition Workshops
3. Khalaf, O.I., Abdulsahib, G.M., &Sadik, M. (2018) A Modified Algorithm For Improving Lifetime Wsn.
4. Cole, J.H., Poudel, R.P., Tsagkrasoulis, D., Caan, M.W., Steves, C., Spector, T.D., & Montana, G. (2017) Predicting Brain Age With Deep Learning From Raw Imaging Data Results In A Reliable And Heritable Biomarker. Neuroimage, 163, Pp. 115-124.
5. Antipov, G., Baccouche, M., Berrani, S.A., &Dugelay, J.L. (2017) Effective Training Of Convolutional Neural Networks For Face-based Gender And Age Prediction.
6. Grassmann, F., Mengelkamp, J., Brandl, C., Harsch, S., Zimmermann, M.E., Linkohr, B., & Weber, B.H. (2018) A Deep Learning Algorithm For Prediction Of Age-related Eye Disease Study Severity Scale For Age-related Macular Degeneration From Color Fundus Photography.
7. Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016) Deep Neural Networks Predict Hierarchical Spatio-temporal Cortical Dynamics Of Human Visual Object Recognition.
8. Choi, S., Kim, E., & Oh, S. (2013). Human Behavior Prediction For Smart Homes Using Deep Learning. In Proceedings Of The 2013 Ieee Ro-man
9. Sutton, R.S., &Barto, A.G. (1998) Reinforcement Learning: An Introduction.
10. Toshev, A., &Szegedy, C. (2014). Deeppose: Human Pose Estimation Via Deep Neural Networks. In Proceedings Of The Ieee Conference On Computer Vision And Pattern Recognition
11. Salman, A.D., Khalaf, O.I. And Abdulsahib, G.M. (2019). An Adaptive Intelligent Alarm System For Wireless Sensor Networks.
12. Cole, J.H., & Franke, K. (2017) Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers.
13. Felbo, B., Sundsøy, P., Lehmann, S., & De Montjoye, Y.A. (2016) Using Deep Learning To Predict Demographics From Mobile Phone Metadata. Arxiv-1511.06660
14. Ito, K., Kawai, H., Okano, T., And Aoki, T. (2018) Age And Gender Prediction From Face Images Using Convolutional Neural Network.
15. Simonyan, K., And Zisserman, A. (2014) Very Deep Convolutional Networks For 9 Large-scale Image Recognition
16. He, K., Zhang, X., Ren, S. And Sun, J. (2016) Deep Residual Learning For Image Recognition, Proceedings Of The Ieee
17. Zagoruyko, S. And Komodakis, N. (2016) Wide Residual Networks
18. Krizhevsky, A. Sutskever, I, And Hinton G.E. (2012) Imagenet Classification With Deep Convolutional Neural Network