Age and Gender Recognition using Convolution Neural Network

Akash. B. N¹, Akshay. K Kulkarni², Deekshith. A³, Gowtham Gowda⁴
Department of Computer Science and Engineering
The National Institute of Engineering, Mysuru, India

Abstract:
This paper proposes an automatic age and gender recognition system based on a human face by using a convolution neural network. In recent years extracting information from the human face is one of the active research topics. The Human face consists of facial features like eyes, ears, nose, chin, etc. Which can be analyzed as per our requirements. Age and gender recognition start primarily with face detection. The detection is a technique in which various factors are recognized based on the input and according to the requirement. Our system makes use of the Py Torch framework and consists of several deep convolutional neural networks that are trained to recognize the face from an input image and predict its age and gender. The deep convolutional neural networks are not only inexpensive but also provide good results.

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. You may have seen recent applications of this. A sentiment analysis system that analyzes human sentiments based on the facial feature and there are several other applications like movie recommendation system based on the human age. 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. Researches on this technology are ongoing and have 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 the nature of the sample data that is needed to train these types of systems. In this era of 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 IMDb-wiki dataset. The IMDb dataset contains 460,723 facial images with age and gender labels and the Wikipedia dataset includes 62,328 images. So, All-in-all we have got access to over 5 lakh images with labels.

II. RELATED WORK

I. Age and Gender classification

Gil Levi and Tal Hassner
Conference Location: Boston, MA, USA
Date of Conference: 7-12 June 2015
In this work, they have shown that by using deep-trained neural networks, performance of the system can be significantly improved. They had designed a simple neural network architecture that can be applied to the prediction system even though the amount of training data is limited. In this approach they tried to close the gap that existed between automatic face recognition capabilities and those of age and gender detection methods. They had tested their system on the Audience benchmark. This dataset is basically images that are uploaded to Flickr from mobile phones before any filtering. This system showed more accurate results than most of the state-of-art systems.

II. Age and gender classification of human faces for analyzing human behavior

Xiaofeng Wang, Azliza Mohd Ali, Plamen Angelov
Conference Location: Exeter, UK 2014
Date of Conference: 21-23 June 2017
Human behavior cannot be predicted, it can be affected by emotions or the surrounding environment. So, to understand human behavior and to perform autonomous detection of anomalous human behavior they have introduced a method to classify age and gender from the human face. This work may be implemented for several applications like investigating any suspicious or any abnormal activities which can be used to prevent any crime from happening. Here they have used the concept of supervised learning. Support vector machine algorithm is used here for classification purposes. Also, the main approach here is the recently introduced concept called transfer learning which is based on the deep learning concept. A pre-trained deep network is used for feature extraction from the face and then it is fed into the support vector machine classifier. Since the availability of the labeled data is very low, they built a dataset named Gay face and applied the proposed method on that dataset. The proposed method was very robust, and they achieved great results. For age classification they achieved an accuracy of about 80.17% and for gender classification an accuracy of about 90.33% which will make a great impact on analyzing human behavior.

III. EXISTING WORK

Over the past several years, the studies are undergoing on the areas of facial feature extraction and predicting age and gender based on that. To address this concern scientists had come up with various approaches, and each of these...
approaches solved some critical problems that were raised in this field. For predicting age and gender accurately, even some of the minor differences in the images should be extracted carefully. Extracting facial features to that extent is challenging and only a few approaches focused on solving this problem. Those minor differences include size of the eyes, ears, mouth and the distances between them. Most of early methods have focused on images which were maintained in lab conditions (like maintaining ideal lighting, angle of the image, etc.). Very few methods have addressed the difficulties that arise when applying these methods to real world images.

IV. PROPOSED WORK

In this proposed system, we will develop a deep learning model using Convolutional neural networks and Pytorch framework. The model receives the image, passes it through different layers by reducing the size in each layer. We then train our model using the IMDb-wiki dataset which contains over 5 lakh images of human beings which include different ethnicity, color, and many more factors. The dataset also provides a label for each image. Once the model is trained, we can use the model for testing. We first detect the presence of a human in each image. Then, we process the image using the Open CV framework to obtain all the human faces present in the image. Each face is then processed by a developed deep learning model to get the output label which essentially gives us the age and gender of each person in the image.

V. SYSTEM FLOW

There are 2 steps in building this model. They are:
1. **Model design:** The actual deep learning model is designed using convolutional neural networks.
2. **Training the model:** The model is trained using a large set of images taken from the dataset.

VI. SYSTEM IMPLEMENTATION

In this implementation first, we need to extract the face from an image provided by the webcam. That is done by using Python’s OpenCV module. Face Detection using Haar feature-based cascade classifiers which is a machine learning-based approach is an effective object detection method proposed by [4] Paul Viola and Michael Jones. There will be a lot of positive and negative images and the classifier is trained based on it. It is then used to detect the face in other images. For age and gender prediction we have used our proposed system architecture through our experiments. System architecture is shown in Figure 1. The schematic diagram for our CNN model is shown in Figure 2. We have two convolutional layers and three linear layers. There is a module for the neural network called PyTorch’s neural network MODULE class. Each of our layers extends this module. There will be two primary items encapsulated inside, which are forward function definition and a weight tensor for each layer. There as on we are specifying our layers as attributes inside our Network class that the training process starts the Network learns the weight values and then they are updated to the weight tensor inside each layer.

![Figure 1. System Architecture](https://example.com/figure1.png)

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.

- Convolutional layers
  - in_channels
  - out_channels
  - kernel_size
- Linear layers
  - in_features
  - out_features

![Figure 2. CNN diagram](https://example.com/figure2.png)

Now we have finished building our model. We will now move into training our model. For training our model, first step is Forward Propagation. The process of transforming an input tensor to an output tensor is called Forward Propagation. For predicting the results from our network, we have used our forward method. The network will return a prediction tensor and that tensor contains a prediction value for our input. The tensor’s shape will be 1 x 10. This means that the length of the first axis is one while the length of the second axis is ten. This means that, in our batch there will be one image and that image will have ten prediction classes. For most of the classes...
probabilities were near to 10%, and this is true since we get 10 prediction classes for an image and network is guessing based on those ten classes. So Basically, in our work, we have performed these 4 steps

- Prepare the data
- Build the model
- Train the model
- Analyze the model’s results

VII. RESULT AND DISCUSSION

To extract features from the face and to test the results we pass our images to a pre-trained network. We use two sets of images. One is for training and another one is for testing purposes. When we consider one dataset which consists of several subsets and each of those subsets describes a class. Next, we have resized our images to make it suitable for further operations that are carried out in the network. All the features have been obtained after running the pre-trained deep-learning network. Each feature in each image has 4096 dimensions. We have obtained an average accuracy of 94% for age classification. The confusion matrix is used to tabulate the classifications per class.

![Confusion matrix for Age.](image1)

The same process will be used for gender classification as well. The same dataset will be utilized for gender classification. We have obtained an average accuracy of 94% for gender classification. The confusion matrix for gender classification is shown below.

![Confusion matrix for gender.](image2)

The accuracy may vary marginally for every execution. This new method of using CNN for predicting age and gender proposed in this paper have shown great results. Therefore, for predicting age and gender from an image, this method can be applied successfully.

VIII. CONCLUSION AND FUTURE WORK

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. Taking this example which is related to the problem of the face recognition system. We have trained our deep CNN 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, CNN 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.

IX. REFERENCES


[5]. Gil Levi and Tal Hassner, “Age and Gender Classification using Convolutional Neural Networks”, 2015 IEEE.

[6]. Ari Ekmekji,” Convolutional Neural Networks for Age and Gender Classification”, Stanford University.

[7]. Sepidehsadat Hosseini, Seok Hee Lee, Hyuk Jin Kwon, Hyung Il Koo, Age and Gender Classification Using Wide Convolutional Neural Network and Gabor Filter, 2018 IEEE.