International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 11 | November - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Real world Gender and Age Prediction using CNN

M Sri Charan Saichandra, Gnana Ganesh, Thenmozhi M
SRM Institute of Science and Technology Chennai, India

Abstract

The real world gender and age prediction is opined to develop to deal multitude of real world problems and scenarios where in there is an exceedingly great usage for the validation of a person and to authenticate the identity of a person. This project would be developed using a renowned deep learning algorithm – CNN (Convolutional Neural Network) to enhance the precision rate. To this end, we design a simple convolutional network architecture that can be used and deployed using adequate number of pooling layers so as to get sterling precision to the extent possible, training the model using publicly available dataset.

Keywords

deep learning, convolutional neural network, computer vision, prediction.

1.Introduction

Most often than not deep learning is deemed to be a subset of machine learning which in turn is again a subset of artificial intelligence. The algorithms built atop it have the capability to discern the disparate things in the vicinity of the machine which resemble the same as the intricate human nature to perceive the environment around them.

Deep learning is being driven atop the artificial neural network methodology, comprising of the umpteen number of layers in the network. These layers can perform multiple number of operations such as the representation and the abstraction that make sense of the images, sound and text. These neural networks are concoction of layers of nodes analogous to the functioning of human brain consisting a large number of neurons which receive uncountable signals from other neurons which are in turn better interpreted.

The convolutional neural network based model works with a multiple layers wherein signals are transmitted from one layer to another layer in feed-forward fashion. In this project we majorly use this methodology to discern human faces where each layer passes its own insights of image from the previous layers and cascade it to the forward layer. Here majorly the human facial characteristics and features are identified and compared to categorize into the appropriate category.

This projects creates a foundation for innumerable real world cases and scenarios wherein there is a need for facial validation and authentication. Its applications encircles many such situations like the the aadhar verification, biometric, and proffered its assistance a lot to many other domains.

2. Related Works

Before starting and going ahead with our proposed project, we shall first discuss all the deep learning algorithms, evaluate the corrective and suitable algorithm to develop our project and examine the gaps in each algorithm.

2.1 Literature Survey

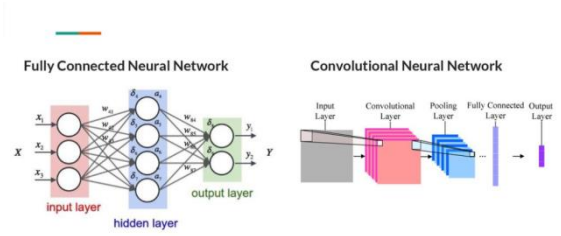
In [1] the research paper by Fatima K. Faek, Sallahaddin University, Zanko Street, Erbil, Kurdistan Region of F.R. Iraq, they mainly dealt with the Speech-based recognition of gender and age using SVM (simple vector machine), and the outcome was that the result depended on the tone of the person's voice. The convolutional neural network model type and feed-forward mechanism were both used in [2] the research paper by Gil Levi and Tal Hassner, Department of Mathematics and Computer Science, The Open University of Israel. The study's conclusion was to feed the network with a face image that had been cropped to 227 227 around the centre. KNN (Kth closest neighbour classifier) was used to build the model in [3] SHUBHAM KUMAR TIWARI's Galgotias University, Uttar Pradesh, with some assistance from CNN. As a result, rather than categorising a person's specific age, classification now refers to their age range.

3 Proposed Methodology

The model development usually requires a well sophisticated and colossal dataset consisting of large number of pictures which are basically human faces which engulf the entire age range and sex categories. The dataset comprises of the images of people of the age groups, typically all the ages and people in the each range belonging to both the genders. The dataset also covers the people belonging to different ethnic groups, varying in complexion form very light-complected to dark-complected. The images are labelled as 0 and 1 for the male and female genders followed by the age of the person in the picture. This helps training the model. The dataset is downloaded from the kaggle website.

The dataset is well categorized as shown in the above pictures according to respective ages followed by genders.

3.1 Proposed Architecture



In our model we preferred to have 5 convolutional layers in the architecture. A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.

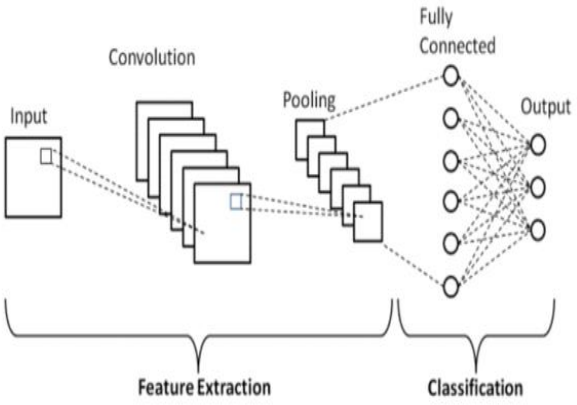
Three different layers in CNN:

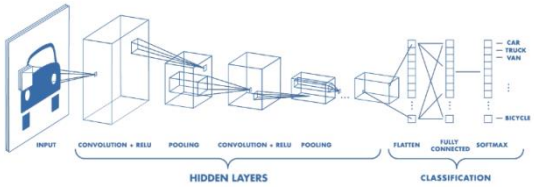
1. Convolutional Layer - The first layer utilised to extract the different features from the input photos is this one. Convolution is a mathematical process that is carried out at this layer between the input image and a filter of a specific size, MxM. The dot product is taken between the filter and the input image's components with regard to the filter's size by sliding the filter over the input image (MxM). The result is known as the Feature map, and it provides details about the image, including its corners and edges. This feature map is later supplied to further layers to teach them additional features from the input image. CNN's convolution layer transmits the outcome to the following layer. The spatial link between the pixels is preserved thanks to convolutional layers of CNN.

2. Pooling Layer - This layer's main goal is to lower the convolved feature map's size in order to save on computational expenses. This is done independently on each feature map and by reducing the links between layers. Essentially, it is a summary of the features produced by a convolution layer. The greatest component in Max Pooling is obtained from the feature map. The average of the components in a predefined sized Image portion is determined via average pooling. Sum Pooling computes the total sum of the components in the predefined section. Typically, the Pooling Layer acts as a link between the FC Layer and the Convolutional Layer. This CNN approach allows the networks recognise the features on their own by generalising the characteristics extracted by the convolution layer. This assists in decreasing computations within a network.

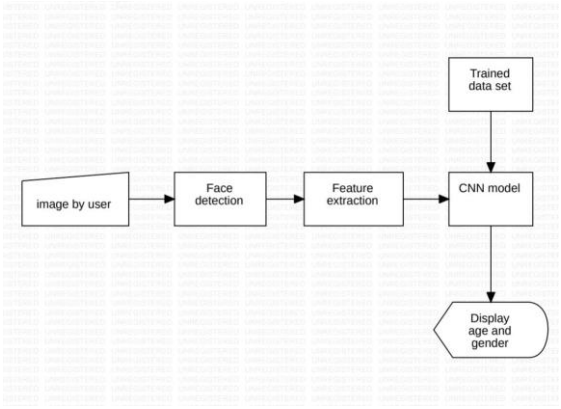
3. Fully Connected Layer - To connect the neurons between two layers, the Fully Connected (FC) layer, which also includes weights and biases, is utilised. These layers make up the final few layers of a CNN architecture and are often positioned before the output layer. The input image from the layers below is flattened and supplied to the FC layer in this. The flattened vector is then put through a few additional FC layers, where the standard operations on mathematical functions happen. The classification procedure starts to take place at this point. Because two fully connected layers will function better than one connected one, two layers are connected. These CNN layers lessen the amount of human oversight.

Feature extraction and classification are the 2 major functions of CNN:





The flow of model is as follows where image is uploaded by the user for prediction where face is detected and features are extracted. This is then sent to trained model for categorising it.



Libaries Needed:

Tensorflow

Numpy - numeric python

openCV - computer vision operations

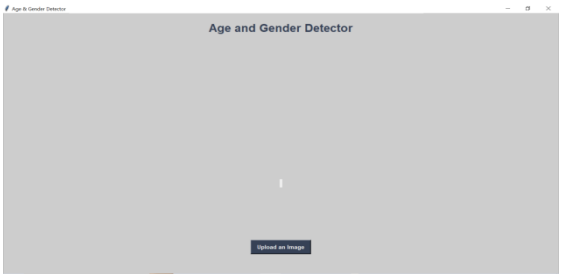
matplotlib – plotting the results

sklearn-for confusion matrix and classification report

-2 fully connected layers of 64 neurons each followed by relu and dropout layer.

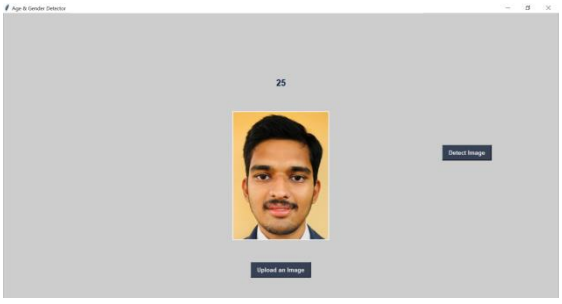
4. Implementation

GUI –



The GUI part developed using the Tkinter library in python where the image can be uploaded by the user to view the results.

5. Result



Persons in fig 1 and fig 2 are same but shot at different times when the ages are 16 and 21 respectively. Based on this genuine information one can clearly understand the predicted age of a person being very close to actual one. The size of the image above is (48, 48, 3).

Conclusion

In this report we demonstrated how age and gender of a person can be predicted using the convolutional neural network based model following a deep learning strategy. The anomalous situation of our project is when people upload a picture containing certain part of body from chest other than face alone. The accuracy in this case fell by noticeable extent. So it is desiderated to upload a picture consisting mostly the person's face.

References

[1] Fatima K.

Faek Sallahaddin University Zanko Street, Erbil, Kurdistan Region of F.R. Iraq

[2] by Gil Levi and Tal Hassner,

Department of Mathematics and Computer Science, The Open University of Israel

[3] SHUBHAM KUMAR TIWARI

Galgotias University , UP