

# Age & Gender Detection using Convolutional Neural Network

Rohit Kumar Gupta, Shivaprasad M B & Dr. S. Srividhya  
Information Science & Engineering BNM Institute of Technology,  
Bangalore Bangalore, India

**Abstract-** The human race has progressed to the point that the twenty-first century represents the commencement of inconceivable achievements. The aforementioned technologies can be used to determine a person's age and gender just by looking at them through a camera, image, or video. The study will examine the various approaches available, as well as which one is the most accurate and how they all fit together. It will also emphasise its significance and how it can be put into practise to improve our daily lives. Furthermore, by overcoming the difficulty of accuracy and time, to obtain the most effective predictions and results. Furthermore, the map for how this technology might be used to societal benefit looks at a wide range of applications, including security services, CCTV surveillance and policing, and dating applications.

## 1. INTRODUCTION

Images and videos are the things that keep everyday chores running in today's society, from security surveillance to looking at cute dog pictures. However, these images and videos can assist in bringing about a change in how people function, as well as assisting authorities in working more efficiently and conveniently. Learning how businesses can use the prediction model that will be developed, as well as how it can be integrated into our daily lives, can help make the world a safer place.

The model will be able to determine a person's age and gender/sex merely by glancing at a picture or video, which may be CCTV surveillance footage [1].

The issue arises: why is it required?

Let's take an example to address this question: A fugitive is on the loose after stealing \$5 million from Pacific Standard Bank. Due to several fabricated identities, his identity has yet to be proven. The police, on the other hand, have a good sense of the person's age and gender. This algorithm can be used as a visual surveillance model in surveillance cameras within a one-mile radius of the bank, and everyone who meets the age and gender requirements will be thoroughly screened.

## 2. COMPUTER VISION

Computer vision assists computers by allowing them to look at/see, figure out, and identify digital images and movies in the same way that humans do. The difficulties it experiences are primarily due to biological vision and the advantage humans have over machines as a result of millions of years of development our civilization has through. This is something the computer struggles with since it fails to comprehend biological vision. Computer

vision is primarily concerned with capturing, processing, analyzing, and comprehending digital images in order to extract data from the actual environment and provide symbolic or numerical information for decision-making. Object recognition, video tracking, motion estimates, and image restoration are all part of this process.

## 3. ARTIFICIAL INTELLIGENCE

Artificial intelligence is the intelligence that machines possess in order for them to adapt to changing environmental conditions and make decisions in order to maximise their chances of success [2]. Artificial intelligence applications are growing by the day, including "optical character recognition, handwriting recognition, audio recognition, face identification, robotics automation, and many others." Artificial intelligence deals with intelligence behavior depicted by machines or any character in video game is usually what is called playing against the CPU [3]. Intelligent reaction of user requests in developing and providing services to the user of normal machines is being looked upon these days a lot [4].

## 4. METHODOLOGY

### A. Convolutional Neural Network

A convolutional neural network (CNN) is a type of deep neural network (DNN) used for image recognition and processing, as well as natural language processing. A CNN, also known as a ConvNet, is a neural network that comprises input and output layers as well as several hidden layers, many of which are convolutional [5]. Convolutional neural network have a great significance in image classification. The neural network cannot be fully connected for image recognition for a computer system or any model [6]. CNN are multilayer perceptrons that have been regularised. In the field of image classification, convolutional neural networks are quite useful. For picture recognition by a computer system or any model, the neural network cannot be fully coupled. The reason for this is that a picture often has a huge dimension and a large number of pixels, making fully connected neural networks difficult to employ because the number of weights required to train a model would be extremely large. For example, a  $40 \times 40$  pixel image would require  $40 \times 40 \times 3$  weights, or 4800 weights, which is still doable, but the image must be at least  $200 \times 200$  pixels in size or larger. However, utilising fully connected neural networks to recognise a  $200 \times 200$  image, the required

weights would be 120000, which is not possible. As a result, the most feasible strategy is to employ convolutional neural networks, where each node is connected to three nodes in the next layer. As a result, the number of weights and neurons necessary to train the model in CNN is lower than in fully connected neural networks.

### CNN ARCHITECTURE

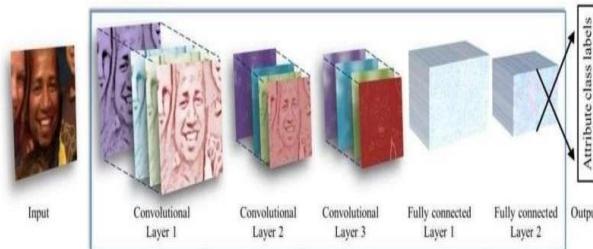


Figure 4.1 CNN Architecture

Three convolutional layers are followed by a corrected linear operation and a pooling layer in the network. The first two layers also use local response normalisation for normalization [7]. The first Convolutional Layer has 96 filters with a resolution of 7\*7 pixels, the second Convolutional Layer has 256 filters with a resolution of 5 pixels, and the third and final Convolutional Layer has 384 filters with a resolution of 3 pixels. Finally, two completely connected layers with 512 neurons each are added. The network processes all three colour channels directly. Images are initially rescaled to 256\*256 pixels and then sent to the network with a crop of 227\*227 pixels. Following that, the three convolutional layers are defined as follows.

- The first convolutional layer applies 96 filters of size 3\*7 \*7 pixels to the input, followed by a rectified linear operator (ReLU), a max pooling layer that takes the maximum value of 3 3 regions with two-pixel steps, and a local response normalisation layer[8].
- The second convolutional layer, which contains 256 filters of size 96\*5\*5 pixels, processes the preceding layer's 96\*28\*28 output. The same hyper parameters as before are used for ReLU, a max pooling layer, and a local response normalisation layer.
- Finally, the third and final convolutional layer applies a set of 384 filters of size 256\*3\*3 pixels to the 256\*14\*14 blob, followed by ReLU and a max pooling layer.
- The first fully connected layer, which has 512 neurons and receives the output of the third convolutional layer, is followed by a ReLU and a dropout layer.
- A second fully connected layer receives the first fully connected layer's 512-dimensional output and comprises 512 neurons, followed by a ReLU and a dropout layer.
- A third, completely connected layer that corresponds to the ultimate age or gender groups.

Finally, the last fully connected layer's output is fed into a soft-max layer, which assigns a probability to each class.

### B. Libraries

#### • OpenCV

OpenCV stands for Open Source Computer Vision and is a free computer vision and machine learning library [5]. This package is mostly used for real-time picture and video preprocessing, as well as advanced analytics. It also works with a variety of Deep Learning frameworks, including TensorFlow, Caffe, and PyTorch. OpenCV is organised in such a way that it comprises a large number of shared or static libraries.

#### • TensorFlow

Google designed and released TensorFlow, a Python library for rapid numerical computing. It is a foundation library that is mostly used to directly develop deep learning models, while wrapper libraries are occasionally used to facilitate the process that is built on top of TensorFlow.

#### • Keras

Keras is a high-level Tensorflow Application Programming Interface (API) that provides the essential abstraction and major building blocks for the creation and distribution of machine learning solutions, as well as a very fast iteration rate. Keras assists engineers and researchers by encouraging them to fully utilise Tensorflow's scalability as well as its cross-platform capacity, which would not have been achievable without Keras. TPU may also be run on a huge cluster of GPUs, and you can export the Keras models to run in Google Chrome or any other browser, as well as on any smartphone or mobile device.

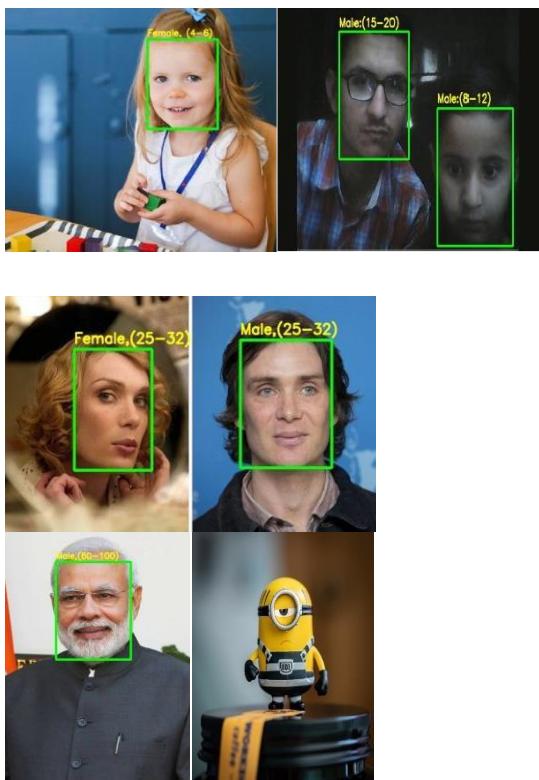
#### • Argparse

This module assists the user in creating user-friendly and understandable command-line interfaces. This module, Argparse, will figure out a way to parse the commands of sys.argv based on the arguments that are required to run the programme. If a cause in the argument is invalid or producing a problem, the module will automatically build and generate the relevant help and usage messages, as well as deliver issue errors.

### C. Dataset

A suitable dataset was selected on kaggle.com in order to obtain the most effective and accurate results. To achieve successful outcomes, our models would be trained on this dataset. It's a 1 GB file with 26,580 photos of 2,284 people in the age groupings of (0-2), (4-6), (8-12), (15-20), (21-24), (25-32), (38-43), (48-53) and (60-100). These photos were taken from Flickr albums and are licensed under the CC (Creative Commons) license. There are numerous photographs for various situations, such as appearance, attitudes, stances, noises, and lighting (either of bright and dark).

## 5. RESULTS



## 6. CONCLUSION & FUTURE SCOPE

Deep Learning and convolutional neural networks will be used to reliably determine a person's gender and age range from a single photograph of a face in the future using this Prototype. The image may be from a certain dataset, could be parsed in a specific way by the prototype, or it could even be real-time if no arguments are processed. The projected gender is binary, and will be classified as Male or Female, and the predicted age will fall into 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)

In light of factors such as the use of cosmetics, lighting, deterrents, and outer appearances, determining an actual age from a picture is extremely difficult. As a result, rather than considering it as a regression problem, it is considered as a classification challenge. In addition, the programme will analyse more photos in order to forecast age and gender. More tests will be carried out in order to obtain accurate gender and age in a real-time image.

The following is a rough outline of how the model would work:

- Detection of facial features
- Gender classification (Male vs. Female)
- Age classification into one of the nine age groups listed above
- We'll use the facebox to show the outcome to the wizard.

The use of this prototype has a wide range of

possibilities in terms of how and where it might be used. As previously discussed in the project, this algorithm will be extremely beneficial to industries with large employee populations, surveillance firms that need to keep track of who is entering, and the government sector, where security forces may use the algorithm to track a suspect or a potential threat to the public.

## REFERENCES

- [1] Levi, Gil, and Tal Hassner. "Age and gender classification using convolutional neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 34-42. 2015.
- [2] Cichy, Radoslaw M., Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. "Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition." *arXiv preprint arXiv:1601.02970* (2016).
- [3] Felbo, Bjørke, Pål Sundsøy, Sune Lehmann, and Yves-Alexandre de Montjoye. "Using deep learning to predict demographics from mobile phone metadata." (2016).
- [4] Toshev, A., and C'Deppose Szegedy. "Human pose estimation via deep neural networks." *CVPR.(Columbus, Ohio, 2014): 1653-1660*.
- [5] Antipov, Grigory, Moez Baccouche, Sid-Ahmed Berrani, and Jean-Luc Dugelay. "Effective training of convolutional neural networks for face-based gender and age prediction." *Pattern Recognition* 72 (2017): 15-26.
- [6] Ito, Koichi, Hiroya Kawai, Takehisa Okano, and Takafumi Aoki. "Age and gender prediction from face images using convolutional neural network." In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 7-11. IEEE, 2018.
- [7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).
- [8] Grassmann, Felix, Judith Mengelkamp, Caroline Brandl, Sebastian Harsch, Martina E. Zimmermann, Birgit Linkohr, Annette Peters, Iris M. Heid, Christoph Palm, and Bernhard HF Weber. "A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography." *Ophthalmology* 125, no. 9 (2018): 1410-1420.