Real-Time Face-Age-Gender Detection System

Lakshitaa Sehgal[1], Anshal Aggarwal[2]

[1]*University Institute of Engineering and Technology Chandigarh, Panjab University, Chandigarh, India*
[2]*University Institute of Engineering and Technology Chandigarh, Panjab University, Chandigarh, India*

Abstract - *It is crucial to detect a person's gender and age group since keeping track of who enters the building is important from a security and other perspective nowadays. A strict check and knowing the maximum information about people entering a bank, ATM booth, state/country border, data centers, hotels, restaurant, hospital, police station, railway station, airport, etc. is vital. So, automated real-time facial detection with gender and age group recognition becomes the foremost concern. This paper proposes a real-time "face-gender-age detection system" using Raspberry Pi to capture real-time facial images of the people which are then passed into the proposed model for gender and age group prediction using the Convolutional Neural Network (CNN). Using the proposed model, this facial image and basic information is gathered and saved into the database, so they can be easily retrieved and referred to in the future. UTKFace and AGE, GENDER AND ETHNICITY (FACE DATA) CSV are the datasets on which the model has been tested. Additionally, noise is added to the images in the above dataset while testing the model to see how well it performs on noisy images.*

***Key Words*: Human Age and Gender Detection, Convolutional Neural Networks, Deep Learning, Facial Detection, Computer Vision**

1.INTRODUCTION

In recent times where crime is increasing day by day, it becomes important to keep a check and store basic information about the person entering the building, specially public buildings like banks, cyber security sensitive areas, country and state border crossing points, police stations, highly security sensitive government and private buildings hospitals, restaurants, multiplexes, shops etc. Basic information which could be recognized through the appearance of a person is age and gender recognition. So here, to keep a track for security reasons, we have proposed a system which would capture an image of a person and store the image along with detected age and gender in the database of the computer along with the model using Convolutional Neural Network (CNN) for age and gender prediction.

The Raspberry Pi 4[1] and a CCTV camera with NVR/DVR or webcam with the attached system memory are used to capture the real-time videos and storing them, which is then passed through a face detection model, whose output is used as the input for the proposed age and gender detection model. The output of both the face detection model and age and gender detection model is stored into the database system of the computer.

To perform training on the datasets, a deep learning model has been proposed. Different experiments have also been performed on various datasets. Along with training and testing on the original dataset, images have also been trained and tested after adding salt and pepper noise to the images. The proposed model has been trained and tested on the following two datasets:

1) UTKFace Dataset[2]
2) AGE, GENDER AND ETHNICITY (FACE DATA) CSV Dataset[3]

Our Main contributions are:

1) Proposed fully developed system to store basic information (face, age and gender) of a person into the database.
2) Trained and tested images (both with and without adding noise) on a CNN model to detect the age and gender of the person.
3) Performed fine tuning to enhance the robustness of the model.

2. PROPOSED MODEL

This section presents the proposed system and the deep learning model using convolutional neural networks for detecting age and gender of a person.

2.1 SYSTEM SETUP

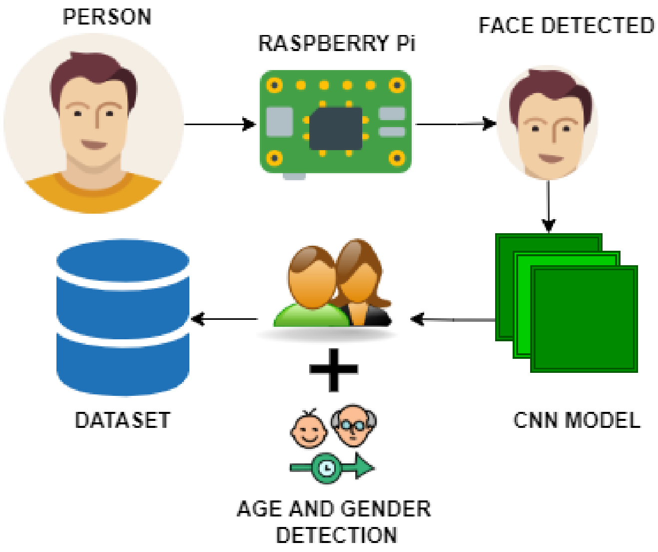

Fig -1: System Setup

The Raspberry Pi 4[1] and a CCTV camera with NVR for IP cameras and DVR for analog cameras or webcam with the attached system memory are used to capture the real-time videos. The hardware setup captures videos, which are then converted to images. A facial recognition model is then run on these images, and its output is then passed into our proposed model. The outcomes of both models are then stored in the database.

2.2 PREPROCESSING

During preprocessing of videos, the OpenCV library extracts images at 72 frames per second. The images are then cropped off using MTCNN[4] for face detection, as it is more accurate than the Haar Cascade classifiers[5] in OpenCV. The final shape of the image after facial detection is 48x48x3 before passing into the proposed model.

2.3 NETWORK ARCHITECTURE

Computer vision tasks, such as image classification and detection, utilize deep, feed-forward artificial neural networks such as convolutional neural networks (CNNs). A CNN is similar to a traditional neural network, but with deeper layers. The neurons of the CNN are arranged in a volumetric manner, including height, width, and depth.

There are three dimensions to the input image(RGB channels): width, height, and depth which is [48x48x 3] in this model. An output of neurons connected to local regions in the input will be computed by the convolutional layer. A layer's parameters consist of a set of learnable filters (or kernels) that are convolved across the width and height of the input volume extending through its depth, computing the dot product of the input and the filter entries. Thus, the network learns filters that trigger when it detects some particular type of feature at some particular spatial position in the input by generating a 2-dimensional activation map of the filter. A function called the Rectified Linear Unit (ReLU)[6] layer will execute elementwise activation, represented below.

$$(x) = \max(0, x)$$

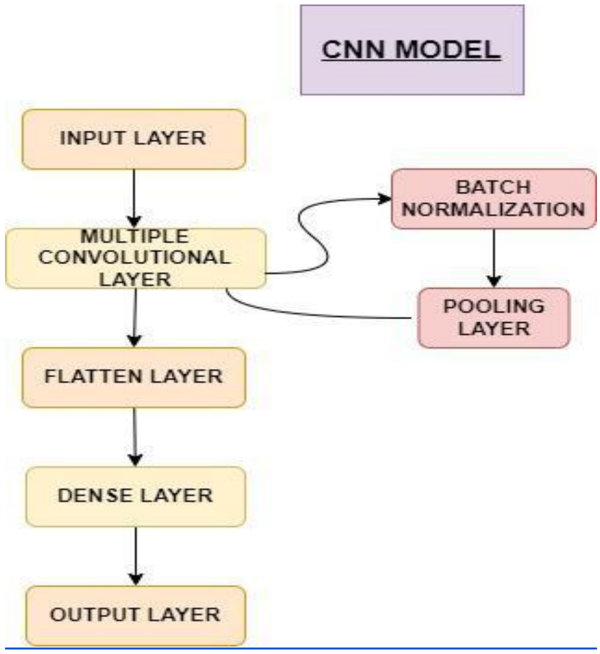

Fig -2: Network Architecture

The maximal activation in a region is output by the pooling layer. The spatial dimensions, such as width and height, are down sampled in this way. The completely connected layer, which is identical to the final layer of a neural network, is the output layer.

This model has been trained using the sigmoid activation function[7] for age determination and ReLU activation function[6] for gender determination. Here, the binary cross-entropy loss function along with the Adam optimizer has been used.

3. EXPERIMENTS PERFORMED

This section elucidates the experimentation performed on the datasets, Model Training and Hyper-Parameter tuning and obtained results.

3.1 DATASET DESCRIPTION

Two datasets have been used for training and testing our proposed model. Following are the two datasets:

1) UTKFace[2]
2) AGE, GENDER AND ETHNICITY (FACE DATA) CSV[3]

UTKFace:

The UTKFace dataset[2] is a huge face dataset consisting of people varying from age 0 to 116 years old. This dataset contains over 20,000 face images with labeling of age, gender, and ethnicity. The images cover large variations in pose, facial expression, illumination, occlusion, resolution, etc.

AGE, GENDER AND ETHNICITY (FACE DATA) CSV:

This dataset comprises data in the form of CSV format. This includes a CSV of facial images that are labeled on the basis of age, gender and ethnicity. This includes 27305 rows and 5 columns.

3.2 NOISE ADDITION

Salt and pepper noise is added to the images to test the robustness and performance of the proposed model in the presence of noise.

4. RESULTS AND OBSERVATIONS

Below are the results obtained after training and testing the proposed model on both the datasets.

UTKFace Dataset

The proposed model, when trained on UTKFace[2] without addition of noise, gave accuracy of 64% on age estimation and 79% on gender estimation on 10 epochs. Then after hyperparameter tuning the best outcome came out to be 79% on age estimation and 93 % on gender estimation on 60 Epochs, 64 Batch Size and 0.001 Learning Rate. After addition of salt and pepper noise the best outcome came out to be 72% on age estimation and 86% on gender estimation on 70 Epochs, 64 Batch Size and 0.0001 Learning Rate.

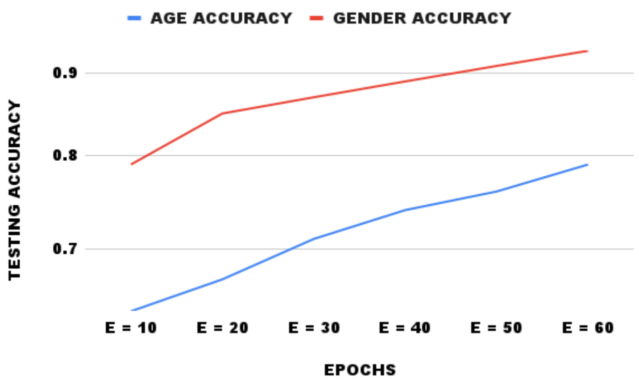

Chart -1: Testing Accuracy (Age and Gender Estimation) vs Epochs for UTKFace (Without Addition of Noises in Images)

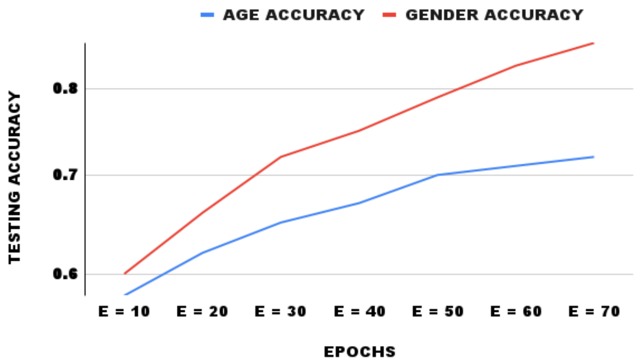

Chart -2: Testing Accuracy (Age and Gender Estimation) vs Epochs for UTKFace (With Addition of Noises in Images)

AGE, GENDER AND ETHNICITY (FACE DATA) CSV

The proposed model, when trained on AGE, GENDER AND ETHNICITY (FACE DATA) CSV[3] without addition of noise, gave testing accuracy of 67% on age estimation and 79.2 % gender estimation on 10 epochs. Then after hyperparameter tuning the best testing outcome came out to be 80.4% for age estimation and 94.2% on gender estimation on 90 Epochs, 64 Batch Size and 0.0001 Learning Rate. After addition of salt and pepper noise the results achieved the best testing accuracy of 70.3% on age estimation and 85.6% on gender estimation on 110 Epochs, 64 Batch Size and 0.0001 Learning Rate.

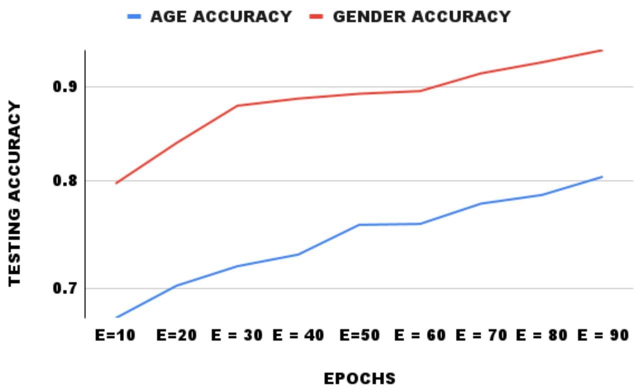

Chart -3: Testing Accuracy (Age and Gender Estimation) vs Epochs (Without Addition of Noises in Images)

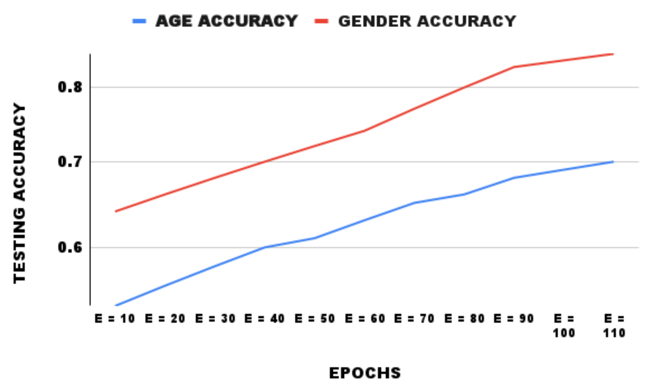

Chart -4: Testing Accuracy (Age and Gender Estimation) vs Epochs (With Addition of Noises in Images)

4. CONCLUSIONS

This paper puts forth an end to end face-gender-age detection system and a deep learning model using Convolutional Neural Network (CNN) which is trained and tested on various different datasets. It captures real time images and then detects faces. The output of facial detection is further passed into the proposed CNN model which has been trained for estimating age and gender. These results are then stored in the database of the computer system for further use.

We hope that this system will be incorporated in real life in the near future. So, the results obtained demonstrate that the model works fine for both images without noises and containing noises and can be put into real time use.

REFERENCES

[1] Maksimovic, Mirjana & Vujovic, Vladimir & Davidovi´c, Nikola & Milosevic, Vladimir & Perisic, Branko. (2014). Raspberry Pi as Internet of Things hardware: Performances and Constraints.

[2] https://www.kaggle.com/jangedoo/utkface-new

[3] https://www.kaggle.com/nipunarora8/age-gender-and-ethnicity-face-data-csv

[4] N. Zhang, J. Luo and W. Gao, ”Research on Face Detection Technology Based on MTCNN,” 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), 2020, pp. 154-158, doi: 10.1109/ICCNEA50255.2020.00040.

[5] L. Cuimei, Q. Zhiliang, J. Nan and W. Jianhua, ”Human face detection algorithm via Haar cascade classifier combined with three additional classifiers,” 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), 2017, pp. 483-487, doi: 10.1109/ICEMI.2017.8265863.

[6] arXiv:1803.08375 [cs.NE]

[7] Pratiwi, Heny & Windarto, Agus & Susliansyah, S. & Aria, Ririn & Susilowati, Susi & Rahayu, Luci & Fitriani, Yuni & Merdekawati, Agustiena & Rahadjeng, Indra. (2020). Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks. Journal of Physics: Conference Series. 1471. 012010. 10.1088/1742-6596/1471/1/012010.

[8] Alqadi, Ziad. (2018). Salt and Pepper Noise: Effects and Removal. International Journal on Electrical Engineering and Informatics. 2.