Deep-learning based Gender Classification using built-in webcams and the effect of lighting conditions

Muneeb Ahmed
University of Kashmir,
North Campus.
muneebpandith@gmail.com

Mohsin Ahmad Sheikh
University of Kashmir,
North Campus.
sheikhmohsin204@uok.edu.in

Reyaz Ahmad
University of Kashmir,
North Campus.

Zaid Zahoor
University of Kashmir,
North Campus.

***Abstract*—Gender classification is a subset of the classification problems in the domain of machine-learning. This work aims to solve the gender recognition problem in humans by transfer-learning approach to re-train the existing Inceptionv3 CNN model, on face-images. We collect 5000 face-images manually and from Internet to build a new dataset and then show experimental data to analyze the performance of the CNN against variations in the distance from the camera illumination (lighting conditions). The system achieves a 96.7% accuracy in correct labelling of gender (male or female) of ages 13-65.**

Keywords—CNN - Convolutional Neural Network.

I. INTRODUCTION

One of the essential characteristics of an ‘intelligent system’ is the ability to recognize and identify objects, and to improve the performance on future tasks by the ‘learning’ process. Machine-learning in itself is a multi-dimensional process. Most recently, architectures based on the idea of layered artificial neural networks have improved the state of the art across a wide range of diverse tasks including gender classification. Image Classification, in general, is one of the primary aspects of computer vision. More or less, the correct recognition actually depends on the image features itself than manual rules and algorithms. Hence the data (called learning-data) plays a key role in solving this problem. In fact, the importance of data is one of the fundamental aspects of ‘learning’ and improving the ‘performance’. Besides, there is a wide range of applications that come up with the general ‘classification’ problem in machine-learning, one of which is recognizing gender (male/female) as in humans, given some differentiating features. The input features of the learning-data vary in type and size. Hence, there are systems that try to solve this problem on variable types of learning-data. Image classification on the other hand, uses specific image-features as a measure to train the classifier and do the classifying process on future data. Now, gender classification can be taken as a subset of the classification problem. Since, we have finite number gender-classes in humans, male or female. Hence, various persons having similar types of features can be grouped together.

Over the years, researchers have studied the gender classification problem in different ways [1]. From clustering algorithms to neural networks, there have been many studies with different outputs of accuracy. Researchers, over the years have developed powerful neural networks that work on image-classification problem in general with significant amount of accuracy. One such model is the Inceptionv3 model, from Google. It is trained on the ImageNet database of images over 1 million in number and has the tendency to classify images across a 1000+ classes.

However, models like these are not trained for some specific subset of image-classification problem, like gender classification. So, one has to start from scratch to develop a model for a particular classification problem. Another alternative to this, which we employed, is applying ‘transfer-learning’ on the existing model to solve the gender classification problem, on face-images, which has yielded us significant results. The underlying rule is to expand the capabilities of a trained model to solve a specific subset of the image-classification problem. Our study, can boost the performance of a wide range of applications including-advertisements, Data Mining, smart human-computer interface, face image database management, Video-surveillance, to name some.

II. WORKFLOW OF THE SYSTEM

We intend the Gender Classification in humans from face-images to be a real-time process. This broadly is a two-step process; comprising Face-Detection and; the gender classification itself. Face-Detection, can be treated as a separate machine-learning problem. In addition, segmentation of face in an image becomes difficult because of the variations in pose, illumination, occultation, expression, and so forth. In this study, a webcam serves as the input device to the system as it captures a stream of images from the environment. The video-stream is then refined by framing it into discrete image sequences. We implemented a trained classifier which uses HAAR-features as principal extracted features for image-segmentation [2]. When it is trained on face-images, it is capable to output the bounding box of the face in real-time prediction-phase. The extracted bounding box, containing the face, is then passed to the re-trained CNN model for gender classification.

Webcams usually have lesser resolution and cannot extract all colours and details, generally. Besides, our retrained model for gender classification is hardware independent. The later processing of it allows us to optimize the model which decreases its size and then it can be ported to smart mobile devices as well, with limited processing power. Yet, in this setting, achieving a significant performance on gender classification gives a slight leverage to the system over the software modules that solve the same problem on mobile devices.

III. PROPOSED METHOD

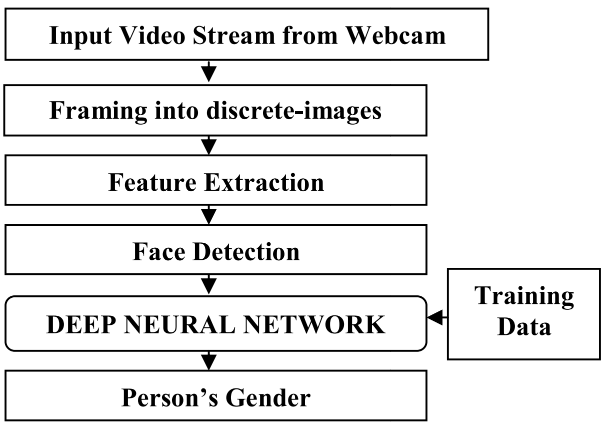

FIGURE I. WORKING PROTOTYPE OF THE SYSTEM

1. CNN Model Architecture

The model used for the system is the Inception V3 [7], which is trained on 1 million+ images of the ImageNet dataset. The layers of the model are broadly grouped into two categories: feature extraction part with convolution layers and classification part with fully-connected and soft-max layers. The schematic architecture of the model, used to solve the gender classification in humans, is shown as;

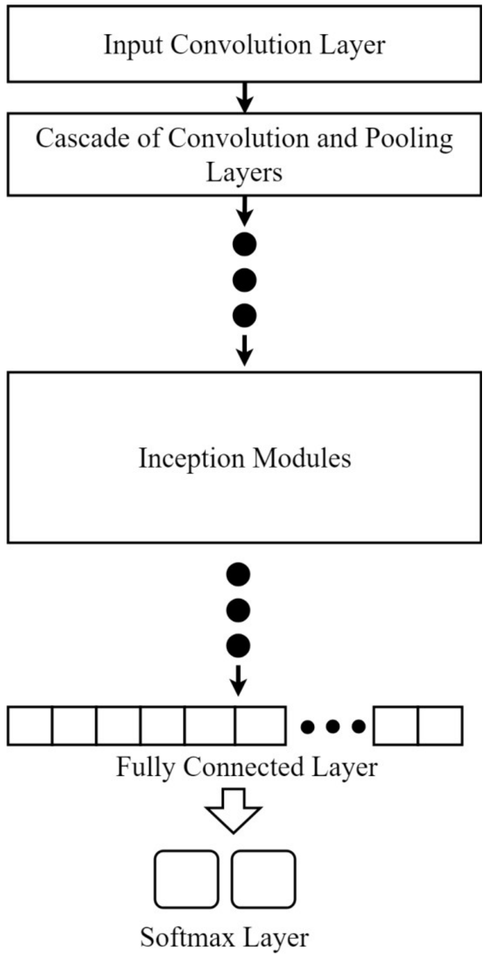

FIGURE II. ARCHITECTURE OF INCEPTION V3 MODEL

2. Pre-processing

Building a standard dataset from scratch is a cumbersome process. The foundation training set used is the IMDB dataset [6]. We increase the frequency of targeted subjects in this dataset by combining it with our own dataset containing 5000 face-images from the Internet along with a custom set of variable face-images collected manually. The total number of face images in our training set is about 15000 after processing .The samples have been kept in the ratio (7500 male : 7500 female) with varied poses, illumination and environment.

Total number of images for retraining	10000 images from IMDB dataset [6] 3000 images from Internet 2000 images of manual dataset of Race: Kashmiri
Number of male faces	7500
Number of female faces	7500
Age group of the dataset	13-65 years

TABLE I. PROPERTIES OF THE DATASET

3. *Re-training the model*

Now, customizing the architecture is needed for solving the subjected problem. Gender recognition is a binary-classification problem. So, we have modified the penultimate fully-connected layer of the model which will contain 1024 perceptrons followed by a soft-max layer with two perceptrons, each representing a gender, either male or female. Transfer-learning is a beneficial process as it involves 're-training' of the existing models on custom classes of data-sets on the pen-ultimate layer of the neural network, while preserving most of the properties of deep-layers of the actual model, thus incorporating custom as well as original information.

III. EXPERIMENTAL RESULTS

Experiment is carried out on six new subjects (three males and three females) by varying the parameters viz., distance from camera and lighting conditions. The angle of the laptop lid is kept at 110 degrees having an in-built webcam of 1.3megapixel resolution. The description of parameters is given as:

Parameter 1: Distances from camera is varied as below:

- 0.5 m
- 1 m
- 2 m

Subject	Distance from camera	Prediction Probability	Average % Error
MALE	0.5 m	0.9182	8.28% average
	1 m	0.8355	12.89% average
	2m	0.8247	17.21% average

TABLE III. VARYING DISTANCE FROM CAMERA ON MALE SUBJECTS

Subject	Distance from camera	Prediction Probability	Average % Error
FEMALE	0.5 m	0.92	8.0% average
	1 m	0.883	11.7% average
	2m	0.844	15.6% average

TABLE IV. VARYING DISTANCE FROM CAMERA ON FEMALE SUBJECTS

Parameter 2: Lighting condition is varied as below:

- Daylight Indoor (Sunlight through the windows)
- Lowlight Indoor (Room illuminated by a CFL lamp)
- Daylight Outdoor (External Summer Daylight)

Subject	Lighting Condition	Prediction Probability	Average % Error
MALE	Daylight Outdoor	0.79159	21.03% average
	Lowlight Indoor	0.8741	8.91% average
	Daylight Indoor	0.91285	8.45% average

TABLE V. VARYING LIGHTING CONDITION ON MALE SUBJECTS

Subject	Lighting Condition	Prediction Probability	Average % Error
FEMALE	Daylight Outdoor	0.865	13.5% average
	Lowlight Indoor	0.8828	11.72% average
	Daylight Indoor	0.928	7.2% average

TABLE VI. VARYING LIGHTING CONDITION ON FEMALE SUBJECTS

IV. CONCLUSION

Gender detection systems limit the search of an imposter to half of the space in many recognition and security systems where the ultimate goal is the identification of a person [5]. In this paper we presented a way to solve the gender recognition problem by using transfer-learning approach of deep-learning. The experimental results show that the distance from the camera is an important factor in recognition. Prediction probabilities are better when distance of subject from the camera is 0.5m (than 1 m and 2 m). The reason behind this is that, we get more exact pixel information of subjects when a subject is closer to the camera. Also there are better results when subject is in indoor environment. In an indoor environment, there are not much detailed backgrounds as compared to an outdoor environment where there are plenty of colourful details in the background. We also found that (under same lighting conditions, distance from camera, the gender recognition in female subjects yields higher values of probability than male subjects. Female faces are almost alike in facial features, whereas male faces have variations (e.g., variations in beard, hairstyle, etc.) which as such do not exist in female subjects.

III. REFERENCES

[1] Feng Lin, Yan Zhuang, "Human Gender Classification: A Review".

[2] Paul Viola and Michael Jones, "Robust Real-Time Face Detection". International Journal of Computer Vision 57(2), 137–154, 2004.

[3] Sarthak Gupta, "Gender Detection using Machine-learning techniques and Delaunay Triangulation", International Journal of Computer Applications (0975 – 8887) Volume 124 – No.6, August 2015.

[4] Mücahit Büyükyılmaz, Ali Osman Çıbıkdiken, "Voice Gender Recognition Using Deep Learning".Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA 2016)

[5] Musaed Alhussein, Zulfiqar Ali, Mohammad Imran and Wadood Abdul, "Automatic Gender Detection Based on Characteristics for Vocal chords". Hindawi Publishing Corporation Mobile Information Systems Volume 2016, Article ID 7805217, 12 pages

[6] R. Rothe, R. Timofte, and L. Van Gool, "Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks," *Int. J. Comput. Vis.*, vol. 126, no. 2–4, pp. 144–157, 2018.

[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2015.