International Journal of Innovations in Engineering and Technology (IJIET)
http://dx.doi.org/10.21172/ijiet.154.02

Gender Identification from Facial Features

D.Gowtami Annapurna, Dantu Lakshmi Pujita, Dindi Mounika
Department of Computer Science and Engineering
Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, Andhra Pradesh,India

Abstract-Humans are capable of determining an individual's gender relatively easily using facial attributes. Although it is challenging for machines to perform the same task, in the past decade incredible strides have been made in automatically making predictions from face images. Gender is one of the key facial attributes that plays a very prominent role in social interactions, making gender estimation from a single face image which is an important task in intelligent applications, such as access control, human-computer interaction, law enforcement, marketing intelligence and visual surveillance, etc. The project identifies or detects the gender from the given face images. The face is detected from the image using a convolution neural network like the Multi-task Cascade Convolution Neural Network (MTCNN) where as in the video the faces are detected using the HaarCascade. The project has been motivated by problems like lack of security, frauds, child molestation, robbery, criminal identification. However, it is still challenging in real applications due to motion blur, object occlusion and extreme illumination in videos. We Propose the CNN Framework which first detects and extracts the Human face images in videos and then feeds them to the network which classify the gender of the detected face.

Keywords- MTCNN, Haarcascade, SmallerVGGNet.

I. INTRODUCTION

Gender detection is a non-trivial computer vision problem for identifying the gender of the faces in images. The most fundamental application that is used in face recognition technology is gender detection. Many companies like Facebook, Amazon, Google and other technical companies have their respective different implementations of this gender detection. The software must be able to detect it first, before they can recognize. Gender of a person in each cluster is estimated using aggregation of predictions for individual photos.This consists of two steps, first one is to identify the faces in the image/video. After that the features will be extracted. Second is to classify the type of gender.Detection of faces from the images can be done by using the MTCNN whereas the Haarcascade is used for the face detection in videos. This Face detection is a computer vision problem that involves finding faces in photos. After the detection, the features will be extracted from the detected faces. Based on the extracted features, by using the smaller VGG algorithm the gender will be classified. For the classification of gender in videos and images we used database like the wiki_crop, imdb.

II. PROPOSED ALGORITHM

2.1.Multi-Task Cascade Convolutional Neural Network (MTCNN)

There are N number of deep learning methods have been developed for the face detection purpose. In those one of the more popular approaches is "*Multi-Task Convolutional Neural Network*", simply called MTCNN. It achieved the state-of-the-art results on a range of benchmark datasets, and because it is capable of also recognizing other facial features such as eyes and mouth, called landmark detection. The MTCNN consists of three stages. In the first stage, candidate windows are produced quickly through the shallow CNN. Then it refines the windows in order to reject the large number of non-faces windows that are present in the image through a more complex CNN. Finally, it will use a more powerful CNN in order to refine the results and output facial landmarks positions.This MTCNN network uses a cascade structure which consists of three networks; first the image/video will be rescaled to the range of different sizes which is called as an image pyramid, then the first model which is Proposal Network (P-Net) proposes the candidate facial regions, the second model which is Refine Network (R-Net) filters the bounding boxes in the image, finally the third model that is Output Network (O-Net) proposes facial landmarks which are detected in the image. This MTCNN model is a multi-task network because each of the three models that are present in the cascade which are P-Net, R-Net and O-Net will be trained on the three tasks, for example making the three types of the predictions. They are classification of face, bounding the box regression and finally the localization of facial landmarks. These three network models are not connected together. The outputs of the previous stages will be fed as

http://dx.doi.org/10.21172/ijiet.154.02

the input for the next stages. This can be used to perform the additional processing between the stages. For example, non-maximum suppression (NMS) will be used as the filter for the candidate bounding boxes that are proposed by the first stage P-Net which will be provided as the input for the second stage that is the R-Net model.

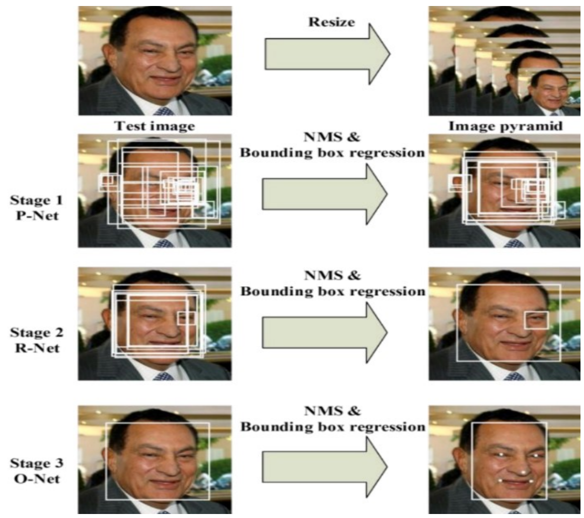


Fig.1MTCNN Face detection in the image

2.2. Haar Cascade Algorithm

The **Haar Cascade** is a classifier which is used for the detection of the object for which it is trained from the sources. By this, better results will be obtained by using high quality images and thereby increasing the number of stages for which the classifier has been trained. HaarCascade classifier is based on the Haar Wavelet technique which is used to analyse the pixels that are present in the image into squares by the function. This technique uses a concept "internal image" to compute the "features" that are detected. Ada boost learning algorithm is used inthe Haar Cascade which is used to select a small number of important features from a large set in order to obtain an efficient result of classifiers and then use the cascade techniques for the detection of face in the images. Haarcascade will be trained by giving some input faces and non faces. In order to detect/recognize the objects with cascade. Haar Cascade classifier is used for the detection of faces in the video.

Fig-2. Haar Cascade face detection in video

III. EXPERIMENT SETUP

3.1 Dataset Generation:

The dataset which is having names of different nationalities for public use was not found, the images in different regions were collected from various online sources such as Wikipedia. We collected different images of celebrities from IMDb and Wikipedia which can be handily available for public use. This is the largest public dataset readily available for gender classification to date. In total we obtained 460,723 face images from 20,284 celebrities from

IMDb and 62,328 from Wikipedia, thus 523,051 in total. While some of the images (from IMDb) contain more than one person in which the photos with the value of strongest face detection below the threshold are used in order to extract the face with a margin. For the pretrained model, 40% margin of the width and height on all four sides have been used.

3.2 Model Architecture:

Our Experiment is having input images that are 96 x 96 with a depth of 3.

Below is our first CONV => RELU =>POOL block.

The convolution layer has 32 filters with a 3 x 3 kernel. We use RELU, the activation function followed by batch normalization. Our POOL layer uses a 3 x 3 POOL size to reduce spatial dimensions quickly from 96 x 96 to 32 x 32.

96 x 96 x 3 input is taken to train our network, dropout used in our network architecture works randomly by disconnecting nodes from the current layer to the next layer. During the training batches, this process of random disconnection helps naturally in order to introduce the redundancy in the model.

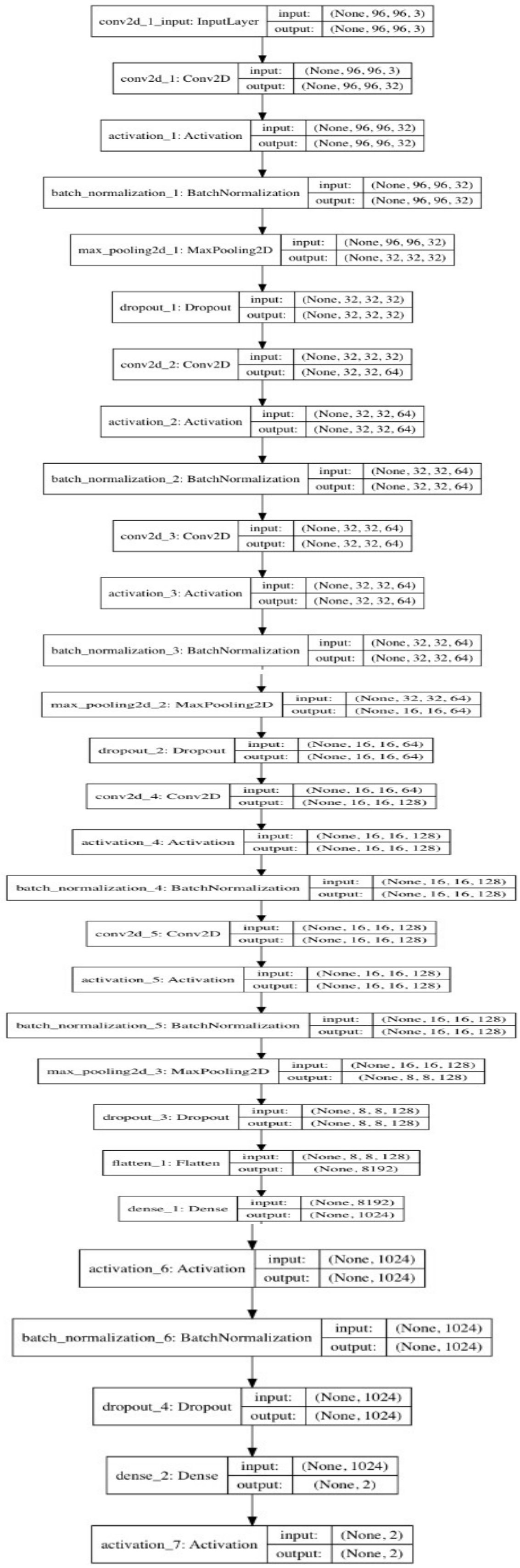
From there we'll add (CONV => RELU) * 2 layers before applying another POOL layer:

Using the multiple CONV and RELU layers together helps reduce the spatial dimensions of the volume thereby allowing us to learn a larger number of features. Then we increase our filter size from 32 to 64. The deeper we go in the network, the smaller the spatial dimensions of our volume, and the more filters we learn. After we decrease max pooling size from 3 x 3 to 2 x 2 to get more approximation.

Dropout is again performed at this stage. Thereafter, add another set of (CONV => RELU) * 2 => POOL,

Again, we increased our filter size to 128 here. And finally, we have a set of FC =>RELU layers and a sigmoid classifier.

The fully connected layer is specified by using Dense (1024) with a rectified linear unit activation and batch normalization. Lastly dropout is performed at a final time. This time we're dropping out 50% of the nodes during training fully-connected and 10-25% in previous layers. After rounding the model value with a sigmoid classifier, it will return the predicted probabilities for each class (Male or Female).


conv2d_1_input: InputLayer
input: (None, 96, 96, 3)
output: (None, 96, 96, 3)
conv2d_1: Conv2D
input: (None, 96, 96, 3)
output: (None, 96, 96, 32)
activation_1: Activation
input: (None, 96, 96, 32)
output: (None, 96, 96, 32)
batch_normalization_1: BatchNormalization
input: (None, 96, 96, 32)
output: (None, 96, 96, 32)
max_pooling2d_1: MaxPooling2D
input: (None, 96, 96, 32)
output: (None, 32, 32, 32)
dropout_1: Dropout
input: (None, 32, 32, 32)
output: (None, 32, 32, 32)
conv2d_2: Conv2D
input: (None, 32, 32, 32)
output: (None, 32, 32, 64)
activation_2: Activation
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
batch_normalization_2: BatchNormalization
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
conv2d_3: Conv2D
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
activation_3: Activation
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
batch_normalization_3: BatchNormalization
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
max_pooling2d_2: MaxPooling2D
input: (None, 32, 32, 64)
output: (None, 16, 16, 64)
dropout_2: Dropout
input: (None, 16, 16, 64)
output: (None, 16, 16, 64)
conv2d_4: Conv2D
input: (None, 16, 16, 64)
output: (None, 16, 16, 128)
activation_4: Activation
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
batch_normalization_4: BatchNormalization
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
conv2d_5: Conv2D
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
activation_5: Activation
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
batch_normalization_5: BatchNormalization
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
max_pooling2d_3: MaxPooling2D
input: (None, 16, 16, 128)
output: (None, 8, 8, 128)
dropout_3: Dropout
input: (None, 8, 8, 128)
output: (None, 8, 8, 128)
flatten_1: Flatten
input: (None, 8, 8, 128)
output: (None, 8192)
dense_1: Dense
input: (None, 8192)
output: (None, 1024)
activation_6: Activation
input: (None, 1024)
output: (None, 1024)
batch_normalization_6: BatchNormalization
input: (None, 1024)
output: (None, 1024)
dropout_4: Dropout
input: (None, 1024)
output: (None, 1024)
dense_2: Dense
input: (None, 1024)
output: (None, 2)
activation_7: Activation
input: (None, 2)
output: (None, 2)

http://dx.doi.org/10.21172/ijiet.154.02

IV. EXPERIMENTAL CONDITION & RESULTS

By evaluating the performance of the proposed algorithm, it is being observed that the quality of image is varied because of the changes made in the illuminating conditions that are dependent on the region or surrounding from where the image has been taken. By the hitting ratios of the gender classification, the output has achieved an accuracy of 93%. From the considered data set, algorithms have shown good robustness and reasonable accuracy for the images taken. The proposed system provides a low complexity and is suitable for real time implementations, such as real time facial recognition. The algorithms have reduced the processing time and their implementation optimization, which has to be performed and detected yet.

This paper proposed is about gender classification from different facial features by using these below interfaces

1)Image 2)Webcam 3)Videos

(a)

(b)

(c)

Fig 4. (a) Output of image (b) Output for webcam(c) output for video

Fig 5. Plot of Loss and Accuracy on Training and Validation Datasets

From plot of accuracy we can see that the model could be trained a little more as the accuracy for both datasets is rising in last epochs. We can see the training of datasets on model is not yet overloaded as both the datasets showing equal skill. From the loss, we see that the model is showing equal performance on both train and validation datasets. If the Original image is not matched with predicted image, then the change will be shown in the graph.

V. CONCLUSION

This Paper describes a practical facial feature extraction system to determine the gender of the persons in the image/video, which combines good accuracy of the face feature extraction and gender classification with robustness to images of different quality and picture taking conditions. This method has achieved an accuracy of 93%. In this paper, the process is composed of two phases: feature extraction and gender classification. The proposed system is having a less amount of complexity and is very suitable in real time implementations, such as real time facial animation. So, from this we can conclude that gender recognition of the persons in the videos and images with minimum probability have been considered and the result obtained is shown as output.

REFERENCES

[1] Ng, C., Tay, Y., Goi, B.M., Recognizing human gender in computer vision: A survey, in: PRICAI 2012: Trends in Artificial Intelligence. volume7458 of Lecture Notes in Computer Science, pp. 335–346, 2012.
[2] Shakhnarovich, G., Viola, P.A., Moghaddam, B., A unified learning framework for real time face detection and classification, in: Proceedings of the Fifth IEEEInternational Conference on Automatic Face and Gesture Recognition, pp. 16–, 2002.
[3] Leng, X., Wang, Y., Improving generalization for gender classification., in: ICIP, pp. 1656–1659, 2008.

[4] Mustafa E. Yildirim, J. S. Park, J. Song and B. W. Yoon "General Classification Based on Haar Cascade" presented at International Journal of Computer and Communication Engineering, Vol. 3 N0. 2, March 2014.
[5] Eman Fares Al Mashagba, Computer Science Department, Zarqa University, Zarqa, Jordan "Real-Time Gender Classification by Face" presented at (IJACSA) International Journal of Advanced Computer Science and Application Vol. 7, No. 3,201.
[6] Emon Kumar Dey, Mohsin Khan & Md Haider Ali "Computer Vision-BasedGender Detection from Facial Image" presented at International Journal ofAdvanced Computer Science. Vol. 3, No. 8, Aug 201.
[7] Vandna Singh, Dr. Vinod Shokeen, Mr. Bhupendra Singh "Comparison of Feature Extraction Algorithm For Gender Classification from Facial Images" presented at International Journal of Engineering Research & Technology (IJERT) Vol.2, May 201.
[8] Vladimir Khryashchev, Andrey Priorov, Lev Shmaglit and Maxim Golubev "Gender Recognition via Face Area Analysis" presented at Proceedings of the World Congress on Engineering and Computer Science 2012 Vol I, October 2012, San Francisco, USA.
[9] Anushri Jaswante, Dr. Asif Ullah Khan, Dr. Bhupesh Gour "Gender Classification Technique Based on Facial Features using Neural Network" presented at International Journal of Computer Science and Information Technologies, Vol. 4, 2013.
[10] Zhiguang YANG, Ming Li, Haizhou AI "An Experimental Study on Automatic Face Gender Classification".
[11] Aakash Rastogi, Manish Aneja, Mayur Dhingra, Naman Gupta, Anshu Malhotra, 2018. "Gender Classification using Facial Images" presented at IndiaCOM 2018, BVICAM March 2018.
[12] Lemley, Sami Abdul-Wahid, Dipayan Banik, Razvan Andonie "Comparison of Recent Machine Learning Techniques for Gender Recognition from Facial Images" presented at MAICS 2016.
[13] O. M. Parkhi, A. Vedaldi, A. Zisserman, "Deep Face Recognition," in British Machine Vision Conference, 2015.
[14] Lemley, Sami Abdul-Wahid, Dipayan Banik, Razvan Andonie "Comparison of Recent Machine Learning Techniques for Gender Recognition from Facial Images" presented at MAICS 2016.
[15] Mr. Raghvendra, Prof.Sandeep Sahu "Gender Recognition & Age Prediction" presented at International Journal of Engineering and Computer Science Vol. 3September 2014.