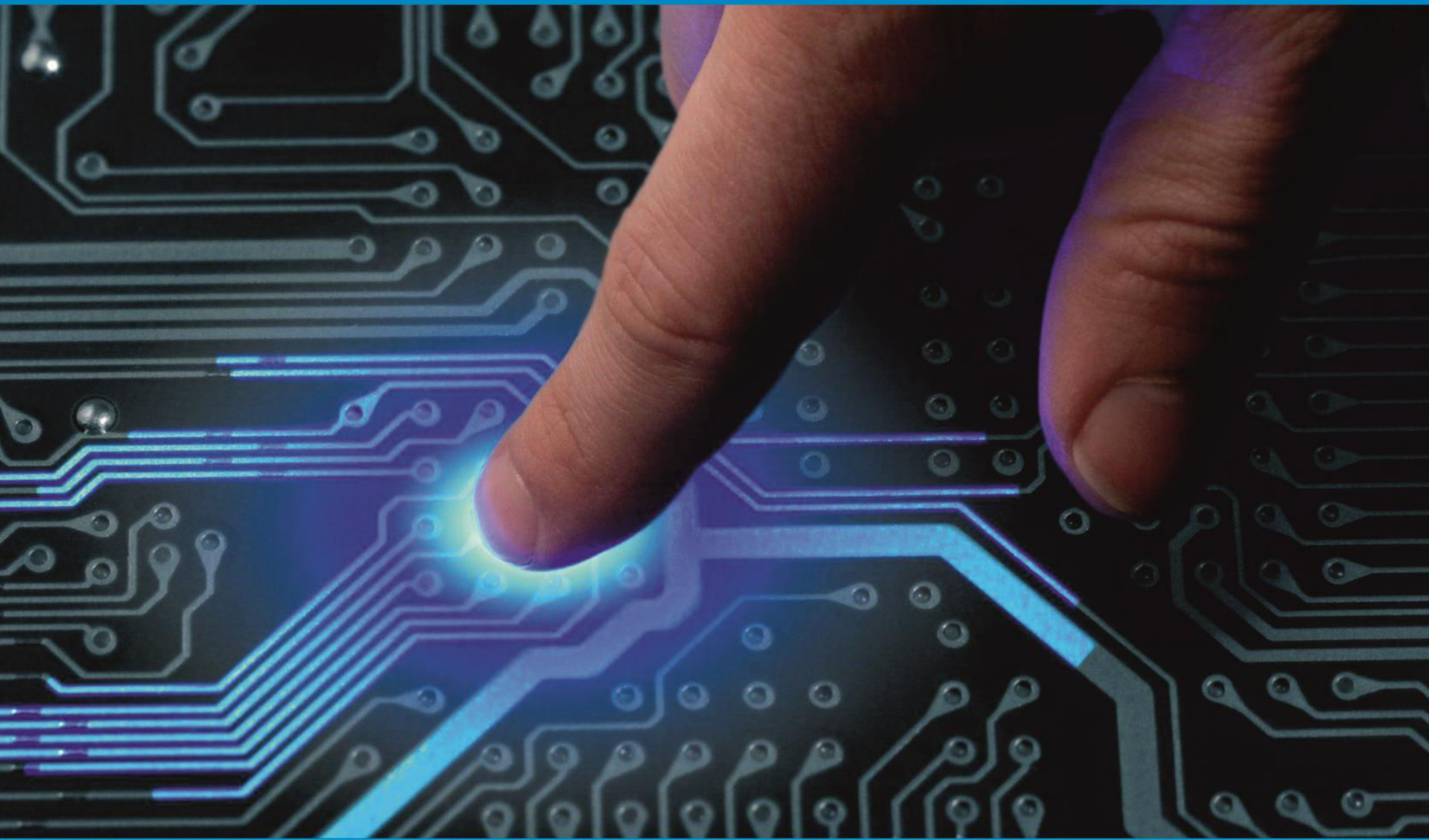
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 4, April 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904095 |

Facial Recognition for Border Security using MTCNN

Krishna Mistry, Purvi Parmar , Miloni Shah , Kinjal Bagal, Sumit Parmar

Diploma Student, Dept. of I.T., Zagdu Singh Charitable Trust's Thakur Polytechnic, Mumbai, India

Lecturer, Dept. of I.T., Zagdu Singh Charitable Trust's Thakur Polytechnic, Mumbai, India

ABSTRACTThis paper provides an overview of face recognition for border security. The security department in any field has security verification by checking for the documents but there are scenarios where people create fake id. In government restricted zones, that is near the border where trespassing is strictly not allowed. In all these situations to improve the security we propose a method in which the security check or continuous border tracking will be done by Face Recognition. Face Recognition method is achieved by using deep learning techniques. In which a model will be trained on the facial images of people and continuous detection will be done and if unknown person is identified then it will be notified. In border area people are given access to use aminations of certain level, if anyone using other level than the allocated one it will be detected. The system uses a MTCNN-Multi-task Cascaded Convolutional Networks**.**

KEYWORDS: Face Recognition;convolutional network; MTCNN; Border Security; image processing

I. INTRODUCTION

Face recognition is a process of recognizing a face of a respective person using a vision system. Face Recognition method is achieved by using deep learning techniques It is a task of identifying an already detected object as a known or unknown face. Humans interact with each other mainly through speech, but also through body gestures, to emphasize certain parts of their speech and to display their emotions. One of the important ways humans can display their emotions is through facial expressions which is key to non-verbal communication between humans. Facial expression recognition can be implemented in all computer interfaces.

In the current time, it is easy for the trespasser to enter a restricted area. This makes security an important factor. For this reason of security, we have proposed a system of face recognition in the restricted governmental areas. Biometric authenticity is catching everyone's attention due to its uniqueness for every individual. Biometric authentication are fingerprint, face, iris etc. In this paper we will be using face recognition as it is most widely usable as well as widely acceptable in current times

In this paper we are using MTCNN for facial recognition as it gives accurate results under facial recognition. It is compatible with all the available operating system including all the best browsers. It demonstrates impressive result on face detection and alignment. The user must have a camera in order to capture the real image and process it. The system, which we will propose, will try to capture the trespasser trying to pass through the governmental restricted area such as border. Since the man power available is limited, the process of verifying known and unknown officers is quite difficult. However, biometric authentication had made this task simpler.

II. RELATED WORK

II.I MTCNN

MTCNN is use to break the data or task into three levels and thereby builds a pipeline

Level-1: P-Net: This stage produces a candidate window by a shallow convolutional network.

Level-2:R-Net: The main goal of this stage is to dismiss as many non-face windows as possible.

Level-3:O-Net: This stage basically uses a complex network which will further use to refine the output of R-net.

MTCNN can run in real time and even on small devices. This makes it widely acceptable and useful

II.II Techniques for biometrics

Biometric authenticity is gaining a lot of attention due to its uniqueness for every individual. Some of the various biometric authentications are fingerprint, hand geometry, iris, face and palm. In this paper, we are using face

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904095 |

recognition as it's the most popular, easily usable and widely acceptable [3].Biometric Technology is used for security.The various biometric techniques are using face,fingerprints,iris and voice. Fingerprint recognition, can utilize a number of approaches to classification, based on minutiae which are a reproduction of epidermal friction skin ridges found on the palm side of the fingers and the palms ,soles of the feet and thumbs,. We can use them for authentication purpose, the facial recognition system record face images through a digital video camera and then analyze facial characteristics like the distance between the eyes, nose, mouth, and jaw edges and Voice recognition is commonly used to operate a device, or write without having to use a mouse,keyboard, or press any buttons,or perform commands.

II.III**Techniques for face recognition**

The different techniques used in facial recognition include PCA[1], SVM[1], LDA[2], CNN and MTCNN.Principal component analysis(PCA) is a statistical approach used for reducing the number of variables in face recognition. In PCA, every image in the training set is represented as a linear combination of weighted eigenvectors called eigenfaces. Support Vector Machine (SVM) is a algorithm that generates a decision surface separating the two classes. This allows us to construct face-recognition algorithms. For face recognition, we re-interpret the decision surface to produce a similarity metric between two facial images. Convolutional Neural Network (CNN) is a multi-layer network trained to perform a specific task using classification. It basically includes 4- 8 layers with image processing tasks incorporated into the design.Transfer learning of a trained CNN model that is Alex Net is done for face recognition.For facial recognition and authentication of the user, Multi Task Cascaded Convolutional Neural Network is been used. MTCNN is a deep learning algorithm and is found to be efficient in analysing images because they use relatively little pre-processing compared to other image classification algorithms [4]. Face detection is one of the mostly studied problems in vision. Viola and Jones [5] proposed a cascaded face detector, which is the first time to apply Hair-Like features in AdaBoost for training cascaded classifiers. It has a practical performance on real-time with higher accuracy than before, but is not able to effectively handle non-frontal faces and faces in the wild.

II.IV**Deterministic face embedding**

We must convert face images into data that the algorithm can understand i.e., numerical data. In order to calculate measurements system requires facial features and landmarks

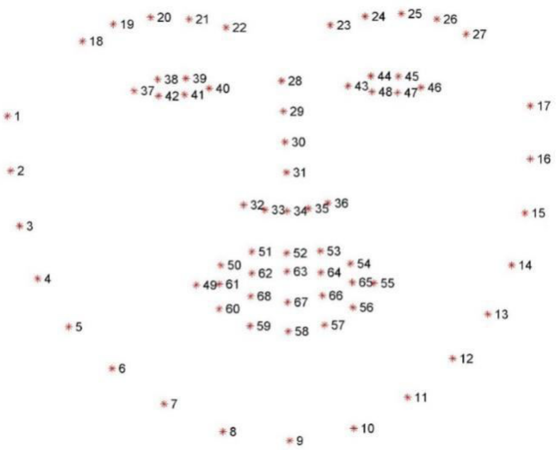


Fig. 1 is a visualization of 68 facial landmarks, also known as facial key points

Face embedding is created where you are converting a face image into a numerical data.The data is then represented into latent space as a vector.The closer the embedding the more likely they are of same person. But the accuracy often depends on the clarity of the input data. Furthermore, these models are often tested. Quality of the image may be low or portion covered by the face in the image may not be proper.Often in such cases, system that uses deterministic face embeddings may suffer. If we analysepixels directly, really dark images and really light images of the same person will have totally different pixel values.So, replacing pixels with gradient will work in such case. If we consider the direction of the input image, both really dark images and really bright images will end up with the same exact representation.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904095 |

II.V Data augmentation

Data augmentation is the method of expanding the size of data used for training a prototype. For reliable predictions, the deep learning prototypes often require a much of teaching data, which is not always available. Therefore, the existing data is augmented in order to make a better generalized prototype. Although data augmentation can be applied in numerous domains, it's frequently used in computer vision. Various techniques are flipping (twisting the image upright or parallel), rotation (turns the image by a specified degree), cropping (object appear in unlike positions in unlike proportions in the image) shearing (move one fragment of the image like a parallelogram). dash in, dash out. Convolutional Neural Networks (CNN) need a vast number of images for the prototype to be instruct effectively. This helps to expand the performance of the prototype by generalizing greater and thereby decreasing overfitting. Most relevant data sets for categorization and object detection data sets have a some thousand to millions of images.

III. PROPOSED METHODOLOGY

For face detection, this paper usesMulti-task Cascaded Convolutional Network (MTCNN). This model consists of three convolutional networks namely P-Net, R-Net and O-Net. Whileretaining real-time performance It is able to perform many face-detection benchmarks.

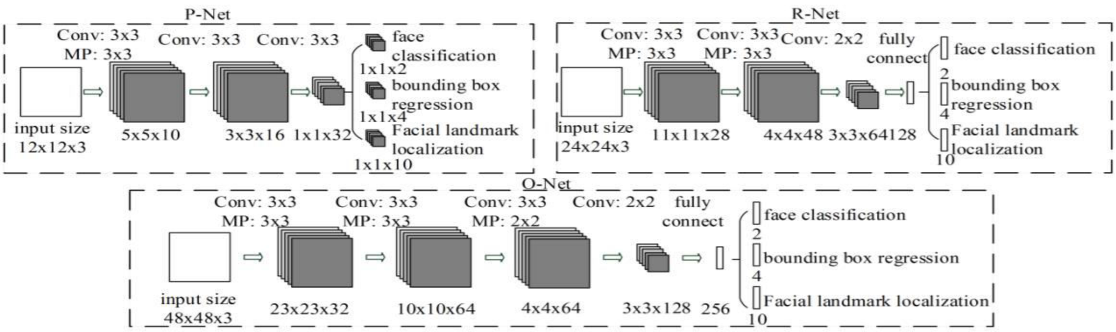


Fig.2.MTCNN Structure

Stage 1: The first and the foremost to do is to pass in an image to the program.In this model, we will create different copies of the same image in different sizes to search for different sized faces within the image.

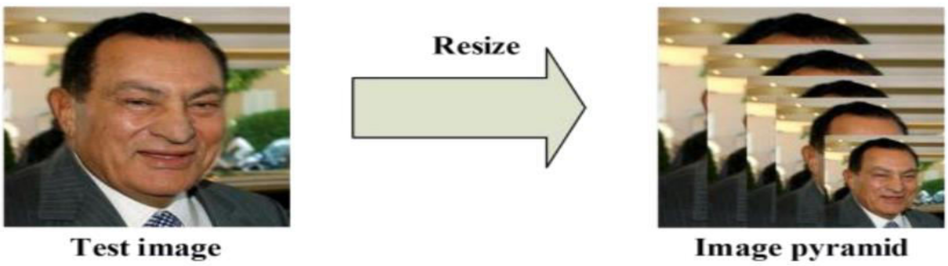


Fig.3.Image Pyramid

We have a 12 x 12 stage 1kernel for each scaled copy that will gothrough every part of the image scanning for faces. It begins in the top left corner, a portion of the image from (0,0) to (12,12). This section of the image is then passed to P-Net. If it notices a face, it returns the co-ordinates of a bounding box. It will then execute one more time the process with part (0+2a,0+2b) to (12+2a, 12+2b), moving the 12 x 12 kernel 2 pixels right side or down side at a time. The stride is known as the shift of 2 pixels or the number of pixels the kernel moves by every time.
Having a stride of 2 is helpful in order to reduce computation complexity without significantly sacrificing accuracy. Since faces in most images are significantly greater than two pixels, it’s highly improbable that the kernel will fail to notice a face merely because it shifted 2 pixels.The only drawback is that we have to recalculate all indexes that are related to the stride. For instance, if the kernel detects a face after moving one step to the right, the output index will tell us the top left corner of that kernel is at (1,0). However, because the stride is 2, we have to multiply the index by 2 to get the accurate coordinate: (2,0).Each kernel will be smaller relative to a greater image, so it will be able to find

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904095 |

smaller faces in the greater-scaled image. Similarly, the kernel will be greater relative to a smaller sized image, so it will be able to find larger faces in the smaller-scaled image.

After passing in the image, we need to generate multiple scaled copies of the image and pass it into the first neural net — P-Net — and assemble its output.The weights and biases of P-Net have been instructed so that it outputs a relatively accurate bounding box for every 12 x 12 kernel. However, the network is more confident about some boxes in comparison to others. Thus, we need to analyze the P-Net output to get a list of confidence levels for each bounding box, and remove the boxes with lower confidence.

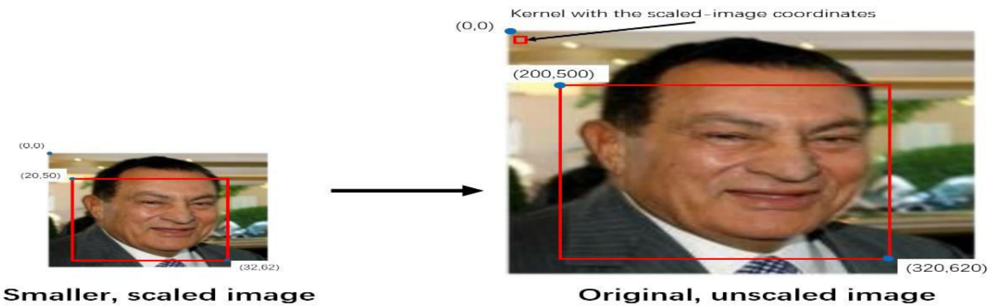


Fig.4.Standardizing kernel coordinates by multiplying it by the scale

After we've chosen out the boxes with higher confidence, we will have to standardize the coordinate system, converting all the coordinate systems to that of the actual, "un-scaled" image. Their coordinates will be based on the smaller image since most of the kernels are in a scaled-down image.

Still, there are many bounding boxes left, and most of them overlap each other. There is a method which is helpful in reducing the bounding boxes called as Non-Maximum Suppression or NMS.

In this particular program, NMS is done by initially sorting the bounding boxes (and their respective 12 x 12 kernels) .This is done by using their confidence or score. In case of taking the most confident box in the network, in some of the models the NMS takes the largest bounding box.Afterwards, we have to calculate the area of three things which are area of each of the kernels, the overlapping area between each kernel and the kernel who has the highest score. The NMS returns a list of "surviving" bounding boxes after the kernels that overlap a lot with the high scoring kernel gets deleted.

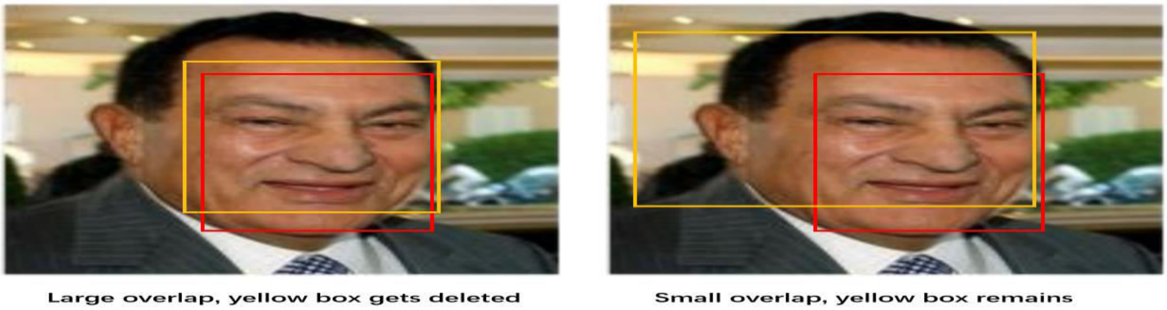


Fig.5.Non-Maximum Suppression

We use NMS two times, once for every scaled image and the second time with all the surviving kernels from each side. We use NMS twice so that we can remove the redundant bounding boxes and also our search will narrow down to one correct and accurate box per box.

If there is more than one face in other images, we will have to discard all the bounding boxes for other faces but since there is only one face in the above image, we can skip this step.Later on, we have to transform the co-ordinates of the bounding box into the co-ordinates of the real image. At this very moment, the coordinates of each and every bounding box is a merit between 0 and 1, with (0,0) as the topmost left section of the kernel which is 12x12 and (1,1) as the rear end rightsection. To transform the bounding box co-ordinates into the standard and real sized image co-ordinates, we have to accumulate by the real image's width and the height.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904095 |

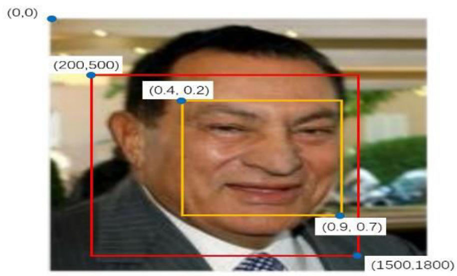


Fig.6.Here, the red box is the 12 x 12 kernel, while the yellow box is the bounding box inside it

The image that we have has a red box which appears for the 24x24 kernel, which is resized back to the original image. We are able to measure the width and height of the kernel: 1500–200 = 300, 1800–500 = 300 (Record how the width and height aren't necessarily 12. That is happening as we're using the coordinates of the kernel in the original image. The width and height we get here of the red box is similar to that of the width and height of the kernel when s ascended back to its actual and real size). Thereafter, we calculate the bounding box coordinates by multiplying 300: 0.4x300 = 120, 0.2x300 = 60, 0.9x300 = 270, 0.7x300 = 210.At the end, we append the topmost left correspondent of the kernel to get the correspondent of the bounding box which will be as: (200+120, 500+60) and (200+270, 500+210) or (320,560) and (470,710).
In the view of the fact that, bounding boxes may possibly not appear to be a square, we then reconfigure the bounding boxes to a square by extending the smaller sides (if the width is shorter than the height, we expand it sideways; if the height is shorter than the width, we expand it vertically or in the upper direction).Finally, we save the coordinates of the and forward it on to stage 2.

Stage 2: Many a times, an image may contain a part of a face peeking in from the side of frame. In such situation, the network shall return a bounding box that is half out of the frame, like John's picture shown below:

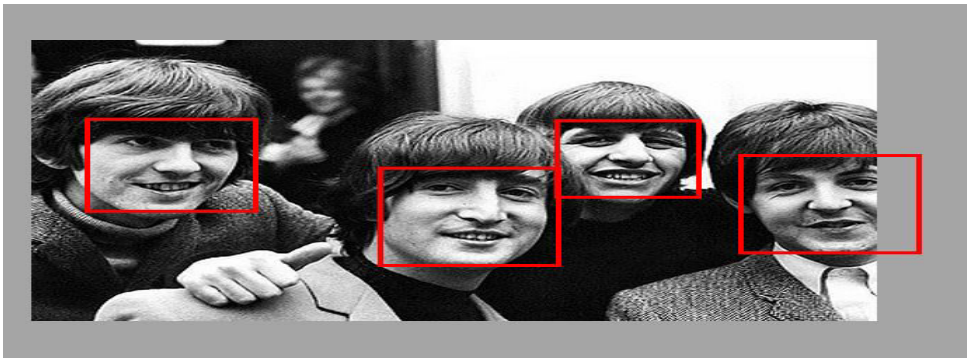


Fig.7.The Beatles and their bounding boxes. John's box is out of bounds and requires padding.

For each and every bounding box, we will create an array of same size and copy these pixel values to the newly created array. If the box is out of bounds, we will only copy the part of the image to the newly created array and fill remaining part with 0. In the image above, the newly created array for John's face will be having pixel values in the left side of box, and some of the columns of 0s near the right edge. This process of filling the array with 0s is basically called as padding.
After performing padding, we resize them to 24 x 24 pixels, and try to normalize them to values almost between -1 and 1. Now, the pixel values are between 0 to 255 (called as RGB values). By differentiating each pixel value by half of 255(127.5) and dividing it by 127.5, we can keep their values between 1 and -1
Now we have bundles of 24 x 24 images i.e., arrays with as many possible numbers of boxes that survived Stage 1, although each one of them has been resized and normalised. We can now feed it to R-Net and can gather its output.
The output of R-Net's is quite similar to that of P-Net: It mainly has all the coordinates of the new, more accurate boxes, and the confidence level of each and every bounding boxes.
Now, we get rid of the boxes with lower confidence, and will be performing NMS on every box. This is done to eliminate further redundant boxes. We need to convert the coordinates of new boxes to the standard coordinates, as the coordinates are based on the P-Net bounding boxes. Once the standardizing coordinates is done, reshaping the boxes to square is done in order to pass it on the next net i.e., O-Net.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904095 |

Stage 3:Prior to process in the bounding boxes from R-Net, the boxes which are out of bounds,firstly we have to pad them. Later, that we recompute the boxes to 48 x 48 pixels, we can proceed in the bounding boxes into O-Net.
The outputs of O-Net are moderately different from that of P-Net and R-Net. O-Net present 3 outputs: 1)the correspondent of the bounding box (out [0]),2) the correspondent of the 5 facial landmarks (out [1]), and3) the confidence level of each box (out [2]).
And again, we get rid of the boxes with slightly lower confidence levels and then standardize the coordinates of bounding boxes and the coordinates of facial landmark. In the end, we run them through the last NMS. At this moment, there should only be one bounding box for each face in the image.

IV. Simulation Results

In this paper we propose a model that can identify faces.First and important step for face recognition is face detection. Face detection is use to detect faces in images.It is a part of object detection and can be used in many areas such as security,bio-metrics,etc. The proposed model can recognize the faces in different light conditions. Since, the camera input does not have stable light source and thus, variations in light conditions will not affect the prediction of the model. The technique used to detect faces is MTCNN which helps to accurately locate the face from a distance. The accuracy obtained of the model is 80%. This will help to recognize and identify faces in all type of conditions and varied camera quality.

V. Conclusion and Future Work

The computational models, which were implemented in this project, were taken after extensive research and a successful testing result.Because of its passive nature,facial recognition is often chosen over other biometric identification technique. Facial recognition devices used to take trespasser images do not require any physical contact which helps to increase the to acceptability. Facial recognition will scale a greater height for border control security in near future as, it is easier to quickly and accurately identify trespasser. Facial recognition has all the key attributes of an end-to-end identity management system with the huge advantage of convenience, speed, passiveness, accuracy and efficiency. We hope to see the developed and widespread use of this technology, making our lives easier and most importantly secure.

References

1. Muzammil Abdulrahman, Alaa Eleyan, “Facial expression recognition using Support Vector Machines”, 23nd Signal Processing and Communications Applications Conference (SIU), IEEE, 2015.
2. Yuan Wei, “Face Recognition Method Based on Improved LDA”, 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), IEEE, vol. 2, pp. 456 - 459, 2017.
3. GittipatJetsiktat, SasipaPanthuwadeethorn, SuphakantPhimoltares, “Enhancing User Authentication of Online Credit Card Payment using Face Image Comparison with MPEG7-Edge Histogram Descriptor”, International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), IEEE, 2015.
4. Guosheng Hu, Yongxin Yang, Dong Yi,et. al.,When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition, IEEE International Conference on Computer Vision Workshop, 2015.
5. Viola, Paul, and M. Jones. "Robust object detection using a boosted cascade of simple features." Proc Cvpr 1(2001):I-511-I-518 vol.1.