Review of face mask detection systems

Aditya Nayak[1], Kunal Yeola[2], Sham Nagmote[3], Vasundhara Gaikwad[4], Prof. Mohini G. Devikar[5], Prof. Shubhangi R. Khade[6]

[1,2,3,4]Students, Dept. Of Computer Engineering, MES College of Engineering Pune, Maharashtra, India
[5,6]Professor, Dept. Of Computer Engineering, MES College of Engineering Pune, Maharashtra, India

---

Abstract- *Face Mask Detection is an emerging area of research during pandemic times in such situations we need to ensure proper mask-wearing to avoid the spread of infections and reduce the caseload on the medical facilities. In this paper, we discuss various methods and algorithms to detect the face mask by reviewing various papers. The methods used images and videos as input for detection. The methods discussed in this paper uses Machine learning, MTCNN, YOLOv3, FaceNet and many more such algorithms for helmet detection, face detection and mask detection. Using all these methods it will be easy to differentiate among people wearing a mask and not wearing a mask.*

Key Words: MTCNN, YOLOv3, FaceNet, Object Detection, Viola Jones Algorithm, Video Analytics, Haar like feature, Helmet Detection, Circle Hough Transform

1. INTRODUCTION

During the COVID-19 epidemic, it is necessary for everyone to wear a mask in order to effectively prevent the spread of virus. Still, some people refuse to wear a mask. Hence we are trying to make a system that will check if a person is wearing a mask or not. In order to make this idea successful, we have studied different research papers on face detection, helmet detection and mask detection which will be a great help for our future research. We have reviewed various papers related to our project domain. In this paper, we have written our views about the papers we have reviewed.

2. LITERATURE REVIEW

[1] This paper talks about the use of MTCNN for the detection of masked faces. Face recognition is a promising area in the field of computer vision. Some devices use Face recognition as an alternative to a fingerprint scanner. CNN has the ability to learn valuable features by itself. The author used IIIT-Delhi masked face images dataset and applied data augmentation to enlarge the dataset so that reliability and efficiency can be improved. They used a pre-trained Multi-task Cascaded Convolutional Neural Network (MTCNN) for the detection of faces from the dataset. MTCNN outperforms many other face-detection tools. It works in 3 stages. First, it creates multiple copies of the images of different scales. This is called an image pyramid. The first stage is called the P-Net or Proposal Network. It introduces candidate facial regions. The second stage is the R-Net or refinement network. It refines the bounding boxes. The third and final stage is O-Net or Output Network. It Determines the final landmarks on the image. In image post processing, the images are cropped and resized according to the FaceNet Specification i.e. 160x160. A pre-trained FaceNet Model was used as a baseline for deep networks. It used 22 deep convolutional layers. A large number of images of masked and unmasked faces were used to train the model. The classification was done with the help of the Support Vector Machine (SVM). The results of this methodology were promising. It gave accuracy up to 98.50% in some datasets and cases.

[2] The paper proposes using two components: i) a deep neural network for identifying single or more than one riders on a motorbike by using the YOLOv3 model and ii) another neural network for detecting whether the rider has worn a helmet or not. In this system, the traffic surveillance system provides input to the model and the video frames are given as input to the CNN for detecting helmets on the riders.

Initially, the YOLOv3 is used for detecting the motorbike and the riders. The YOLOv3 model is an improved version of YOLO which was developed by J. Redmon. The model can detect huge sets of classes, among them only two classes i.e. person and motorbike are detected. The boxes are drawn around the target to localize the objects. The network predicts 4 coordinates; bx, by which are the center coordinates and bw, bh are width, height respectively of the focused target. The overlapping area between motorbike and person is taken from the bounding boxes to determine whether the person is a motorbike rider or not. Determine the Euclidean distance from the center coordinates of the two bounding boxes of a person to the motorbike. If the distance is within the bounding box of the motorbike, then it is understood that the targeted person is the rider of that motorbike. The CNN model is then used to identify and classify whether the rider is wearing a helmet or not. For this, the top one-fourth part of the identified motorbike rider is sent as input from the output received from the YOLOv3 model. The CNN model consists of five layers of which the input layer takes the input from the input image and passes the image through consecutive convolutional layers where each layer transforms the image using specific features and sends it to the next layer. Each layer filters the input image given and extracts the required features with plenty of differentiating attributes to distinguish the target object from other objects. After these five layers, additional two layers are added which are connected. Depending on the image extracted, the softmax classifier classifies the object to predict classes with probability distribution as wearing a Helmet or not wearing a helmet. The CNN predicts bounding boxes along with class probabilities for accuracy of prediction. In the detection process, the input image is divided into an N×N grid. This grid is responsible for object detection of any kind of object that falls into that grid's cell. Each bounding box consists of 4 measures: px, py, w, h where (px, py) coordinates represent the center of the box relative to the bounds of the grid cell. The height (h) and width (w) are predicted relative to the whole image.

The proposed paper explains in detecting single, multiple riders or basically all riders of a motorbike who are not wearing helmets from traffic surveillance videos. The First YOLOv3 model has been used for motorcyclist detection. Then, the proposed lightweight convolutional neural network detects the wearing of a helmet or no helmet for all motorbike riders. This project performs better than other CNN based helmet detection methods and can be extended in the future to detect more complicated cases of several riders including child riders. As YOLOv3 and CNN models detect a person's face accurately from a given image and can tell whether a person is wearing a helmet or not, so one can also use these models to determine if a person is wearing a face mask or not.

[3] This paper proposed a new technique of helmet detection which combines two methods in order to make the detection rate better. Those two methods are i)Haar like feature and ii) Circle hough transform. By using these methods the system detects whether a person is wearing a full helmet or half helmet. When the system receives video input it first separates the images from video then uses a Haar like feature for detection of a full helmet. As we know the human face is full of contrast (e.g. eye region is darker than the cheek region), Haar like feature uses these contrasts to encode the human face, nose, mouth, eyebrows, right eye, left eye. This paper has used 14 feature prototypes to encode the features which include Edge features(4), Center-surround features(2), Line features (8). For each image of 24 x 24 sub-window, there are more than 1,17,000 rectangular features so to select only specific rectangles weak learning algorithm have been used. To boost the performance of classification they have used the AdaBoost classifier. And to increase the detection efficiency they have used a cascade classifier which also reduces the computation time radically. For detection of half helmet circle hough transform methods have been used by authors. This method not only detects the circular shapes but also any kind of shapes in the given picture which makes it easy to locate helmets, and hence it makes it possible to detect half helmets. This paper has overcome different issues which was raised before while detecting full and half helmets. They have tested this algorithm in real-time and the results are very positive.

This paper proposed a new technique of masked face detection by taking the help of video analytics which combines four steps in order to make the detection rate better. The four steps are: i) Distance from camera ii) Eye line detection iii) Facial part detection iv) Eye detection. Video analytics deals with the detection of people and events like walking, falling, standing at the camera.

[4] This paper uses a technique called Analog Devices Inc.'s Cross Core Embedded Studio (CCES) in addition to HOGSVM for person detection it also describes the method of Histogram of Oriented Gradients (HOG) which is a feature set based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. As compared to the best Haar wavelet-based detector it gives good results for person detection, reducing false-positive rates relatively. The main idea is to detect whether a person is wearing a mask or not. So if a person is detected but their face is not detected, then it can be considered that a person is wearing a mask. But this will be also true in a situation like a person facing in the direction opposite to the camera, in such a situation it will detect the person but not their face and will give the wrong output. Therefore to deal with such scenarios, it is important to find out if a person is coming towards or going away from the camera.

To determine whether a person is approaching a camera or going away from a camera the author has discussed four steps in this paper. As it is a four steps execution. The first step is the Distance from the camera method. This method is used to see if a person is approaching the camera or going away from a camera. As the decreasing distance between a person and a camera indicates that the person is approaching the camera and face detection can be triggered. The second step is the eye line detection method. This method helps to find out the valley in horizontal histogram projection. If the eye line is detected, face detection can be applied to see if the person is wearing a mask or not. The third step facial part detection. In this method, the author has used Viola Jones's algorithm to detect facial parts like nose, mouth, eyes, eyebrows, etc. This algorithm results in a very high true detection rate and a very low false positive rate which will be shown in the cases where a person is not wearing a mask. If any person is wearing a mask or his/her face is covered with cloth or hand, then in such cases the detection of the face might not take place, or face detection will take place but either nose or mouth will not be detected indicating it as a mask. The final and most important step is to find out the eyes and then trigger the face detection using the eye detection method. If the person is not wearing a mask, eyes will be detected and face detection can thus be applied. When a person is wearing a mask, eye detection returns true but face detection returns false indicating it is a mask.

This paper has stated that in video analytics, the false detection rate is maximum in eye line detection algorithms as well as in eye detection algorithms. The reason is that eye line detection and eye detection will detect very small parts of the image. For images with poor resolution, it will result in false detection. For facial part detection, the execution time is maximum as compared to all other steps as it deals with face detection followed by face parts detection which is a complex algorithm. This paper has a detailed explanation of how to detect a face mask and the authors have tested these above steps in real time and the results are quite practical and satisfactory.

3. CONCLUSIONS

In this paper, we have discussed four research papers about face recognition, mask detection and helmet detection. Each of these papers uses a different kind of algorithms, different techniques, different approaches but their goal is the same to detect a face, facial features like eyes, nose, eyebrows and to find out whether the face of a person is covered with a mask or not. Similar goes for helmet detection. After doing a deep study of all the algorithms we have concluded that each of these techniques have their own pros and cons but as compared to the other algorithms YOLOv3 and CNN algorithms give better results with more accuracy and are more successful in real life.

REFERENCES

[1] M. S. Ejaz and M. R. Islam, "Masked Face Recognition Using Convolutional Neural Network," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/STI47673.2019.9068044.

[2] M. Dasgupta, O. Bandyopadhyay and S. Chatterji, "Automated Helmet Detection for Multiple Motorcycle Riders using CNN," 2019 IEEE Conference on Information and Communication Technology, Allahabad, India, 2019, pp. 1-4, doi: 10.1109/CICT48419.2019.9066191.

[3] P. Doungmala and K. Klubsuwan, "Helmet Wearing Detection in Thailand Using Haar Like Feature and Circle Hough Transform on Image Processing," 2016 IEEE International Conference on Computer and Information Technology (CIT), Nadi, 2016, pp. 611-614, doi: 10.1109/CIT.2016.87.

[4] G. Deore, R. Bodhula, V. Udpikar and V. More, "Study of masked face detection approach in video analytics," 2016 Conference on Advances in Signal Processing (CASP), Pune, 2016, pp. 196-200, doi: 10.1109/CASP.2016.7746164.