Grenze International Journal of Engineering and Technology, June Issue

Social Distancing and Face Mask Detection using GoogleNet

Shraddha Vijay Mune[1] and Dr. Sandeep Pande[2]
[1]Student, School of Computer Engineering, MIT Academy of Engineering Alandi,Pune,Dist:Pune,Maharashtra, India,412105
Email: shraddha.mune@mitaoe.ac.in
[2]Sr. Assistant Professor, School of Computer Engineering, MIT Academy of Engineering, Alandi, Pune, Dist:Pune, Maharashtra India,412105
Email: sandeep.pande@mitaoe.ac.in

***Abstract*— According to statistics bought by using the World Health Organization, the world pandemic of COVID-19 has severely impacted the World and has now contaminated greater than 8,000,000 human beings worldwide. Wear face masks and to follow the protocols for safe social distancing maintenance are the security rules that want to be observed in public locations to prevent human beings from the spread of the virus. To develop safeguarded surroundings with COVID-19 free cases that contribute to the safety of the people, we propose a computer vision-based an environment-friendly method that aims the monitoring human beings in real-time in public locations to detect face masks on people's faces and social distancing in a safe manner by implementing the model which will monitor real- time activities of humans and detect breakage of rules through the camera. After detection of the breach, the system generates a red alert or alarm to prevent contact between humans.**

***Index Terms*— Deep Learning, Google Net, Face mask detection, social distancing.**

I. Introduction And Background

The spread of the Pandemic Disease named COVID-19 has created a most serious health crisis in the world that has had a deep impact on human creatures and the way we identify our daily lives and our world. Coronavirus has resulted in transmission from person to person however, the transmission of coronavirus causes coronavirus disease 2019 (COVID-19) can be through an asymptomatic carrier with no covid symptoms. To date, no document is clinically approved for the remedy or medicines that are in the fight against COVID-19.

It has spread instantly throughout the world and has resulted in person-to-person transmission which cause the death of 4,905,094 people and Affected 241,132,266 people worldwide in India the death rate is 452,454 and the Affected rate is 34,094,373 people until now, bringing large health, economic, environmental, and social challenges to the entire human population. Now, WHO recommends that people must put on face masks to keep away from the danger of virus transmission and additionally suggested that a distance between two people should be at least 2m [2] and should be maintained between individuals to prevent person to person spread of disease.

Furthermore, many hospital environments and public providers require human beings to use the service provided by them only if they put on masks on faces and follow secure rules of social distancing. Therefore, detection of face masks and monitoring secure social distance has become an essential computer vision [3] task to help society.

Grenze ID: 01.GIJET.10.2.118

The strategy is to stop the spread of the coronavirus through monitoring in real-time if the person has put on face masks and has followed the rule of secure social distancing in public places.

II. Literature Survey

Covid cases have no specific medical treatment invented yet. If the cases are increases, to help the society to maintain the protocols of maintaining social distancing and detecting face masks various research has been done. Prevention of COVID-19 is better than suffering through it. To predict the face mask of human beings some experiments were performed, and models were implemented. For the detection of face masks, an approach has been developed by Ashutosh Balakrishnan, et.al [6] in which the model detects the face of the human by using Semantic Segmentation in an image. The model classifies every pixel of an image as a face image and a non-face image and then detects that segmented part. The extraction of features and prediction of an image is executed using VGG 16 architecture. The system contains the Viola-Jones Detector which improves the performance of real- time images taken for face detection of humans, but it was observed that it didn't succeed to identify the real-world problems and it was affected by various parameters like the brightness of the face and dot orientation. It was also unsuccessful to work well on images that are taken in darkness and images which are not frontal.

Sabbir Ejaz, et al. have used the algorithm MTCNN for the detection of face masks, the system was created with the method that the recognition of face should maintain the accuracy with different varieties of wearing masks, and when the result was checked for train and test images it was found that the combination of masked faces and non-masked faces gave an accuracy average of 82.48 test accuracy percentage. The Approach by N Ozkaya, et al [8] was to create a model that would generate Face masks and also eyes, mouth, and nose from fingerprints. As fingerprints and faces are connected.

Akash Manish [9] has explained Algorithms like linear CNN (Convolutional neural network), VGG-16 CNN, and MobileNet CNN architecture and have done comparative studies between them. The result found that MobileNet takes less time compared to both linear CNN and VGG-16 CNN and was proved efficient. ResNet, MobileNet, and AlexNet were used as a pre-trained model for face mask detection of humans. It was observed that the technique which was proposed can achieve high accuracy (98.2%) when it is implemented with ResNet50. The author Moein Razavi[10] has implemented the model which uses R-CNN for Face mask detection and used Euclidian distance to calculate the distance between humans. The model correctly detected workers who were too close and didn't accurately detect the physical distancing.

Narinder Singh, et.al [12] have created the model which uses the architecture of YOLO v3 and techniques of the deep sort to detect the human objects and identify the distance between two objects or humans. Whereas the author Adina Rahim [11] has proposed a system that used YOLO (You Only Look Once) v4 architecture to detect the objects and find the distance between the centroid of the objects. It was found that the performance of YOLO v4 architecture [11] was high and the YOLO v3 model [12] approach was found to be sensitive to the locations which are spatial of the camera. The same method can be fine-tuned with the corresponding field of view for better adjustment.

A. S. Tolba, et.al [13] in the model Multi-View Recognition of face have used the SURF algorithm. The system takes the Multiview images which are given to the SURF algorithm then the output is given to the classifier to generate the Result. The system was also tested for face images with different angles and poses, so it could handle the occlusion issue.

Sandeep Gupta, et.al [14] have proposed a model for face mask detection which uses CNN and Alex Net Algorithm. C2D-CNN is used for face recognition and AlexNet is used for image Classification. The model architecture was generated in such a way that the test dataset is given to the CNN and C2DPCA for Extraction of features, it is observed that the output is given for the calculation of mahala Nobis distance and Euclidean distance. After which feature fusion takes place this output generates a result of Face identification and then classification is done.

Mohamad loey [15] has proposed a model which detects medical face mask detection and uses ResNet-50 architecture with YOLO v2. The ResNet-50 model with YOLO v2 has performed better in face detection of the medical mask and proved the effectiveness of the proposed model in the detection of the face of medical masked and found that the model is effective for medical masked faces. The author A. S. Tolba, et.al [16] have proposed a model which has a detector for a face mask that results in high performance, and based on deep learning definition it has been proposed that it isn't computationally more complex. The extraction of features of the model has been improved by residual and modules regression.

Recently various algorithms of deep learning have been used for Face detection and to calculate the distance between two human objects. In [9], VGG-16 and CNN algorithms were less efficient than MobileNet and found that the model architecture of YOLO v4 is more efficient than model architecture YOLOv3 in [11] and [12]. The designed model mentioned is used only for Mask detection or checking the social distancing between two human beings. The author D. G. Dondo [17] has Explained the GoogleNet Architecture, and Methodology of GoogleNet and mentioned that for detection and classification, it ranked First in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). They Concluded that the GoogLeNet method can be used for computer vision for improving neural networks. The main pros of this method are a quality gain due to an increase in computational requirements compared to shallower and narrower architectures. It doesn't use bounding box regression used by another architecture. To improve the accuracy rate, using fewer resources and cost, increase computational gain and detecting the face mask, and calculate the distance between two human objects with real-time images, the proposed system uses GoogleNet for the detection of face masks and person or object detection and calculate the distance between the centroid of the object for checking the distance between two human objects.

III. Proposed System

To improve the rate of accuracy, perform the calculation of distance between two human objects and detection of face mask, we are proposing a model based on AI for detection of social distancing that can detect those human beings, who are keeping a safe distance from each other by taking live video images through the camera. We are also detecting whether a person is wearing a mask on the face or not. The system will also implement an alarm alert if the rules aren't followed. GoogLeNet is used for the detection of face masks and human object detection. When the human objects are detected, the proposed system will calculate the distance between them and an alarm alert will be generated.

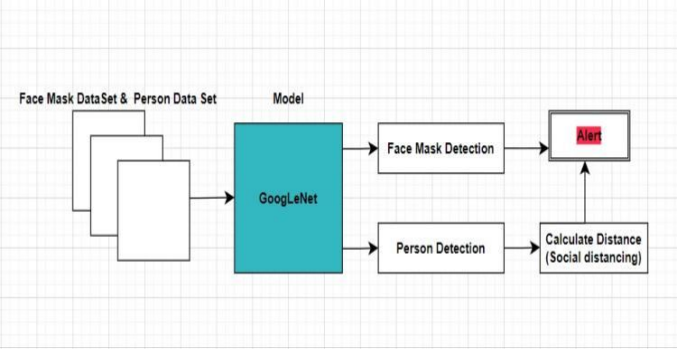

Fig.1 Proposed Architecture of Social distancing and Face mask detection using GoogLeNet.

i. Dataset and Data Analysis.

To train the GoogleNet model the Face Mask dataset and Person Dataset are used for the approach of facemask detection and social distancing. The database used for the approach is the Masked Faces (MAFA) dataset and human detection data set and online images of people in crowded areas.

In [6] the dataset used Dataset named Multi Human Parsing containing about 5000 images, each with at least two persons. Out of these, 2500 images were used for training and validation while the remaining were used for testing the model. Using the semantic model, gave an accuracy of 93.88%. In [7] the dataset used is the AR face database, MFD database, and IIIT Disguise Face Database. The dataset consists of a combination of masked and non-masked face images. To train the model they used various combinations like masked and non-masked faces, only masked faces, and only non-masked faces. They found the Average train accuracy of 99.78 % and test accuracy of 82.48% using the CNN approach. Whereas in [10] the dataset was taken from the MakeML website that contains 853 images that contain one or multiple normal faces which have different illumination and poses. The images used were annotated with faces with a mask, without the mask, and the wrong position of mask-wearing. 1000 other images were added to increase the training data. Using these datasets, the model gave an average of 90% accuracy using the Faster R-CNN approach for face mask detection.

In [15] the Medical Masks Dataset (MMD) and Face Mask Dataset (FMD) are used for the detection of face masks. The MMD dataset contains an image count of 682 which has over 3k medical masked images and FMD contains

853 images. ResNet-50 with YOLO v2 for medical mask detection gave an average precision value of 81%. The dataset of the medical facemask is given in fig 2.

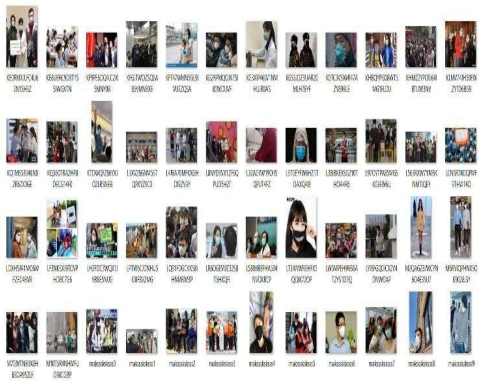

Fig.2 The medical facemask dataset.

In [16] AIZOO face mask dataset and the Moxa3K face mask dataset are used for the face mask detector. The AIZOO dataset contains 8000 images of wider faces and faces with masks and the Moxa3K dataset contains 3000 images which gave a mean average precision of 67.04 percent. The face mask dataset by larxel has 853 images which have three classes they are without mask faces, mask worn incorrectly, and with a mask. Fig 3. Shows the facemask dataset.

Fig 3. Face mask dataset by larxel.

The facemask detection video dataset contains 4357 video frames and 4357 annotations. The dataset contains multiple images with or without masks walking within the college environment. Each annotated frame contains multiple persons who have a unique identification, bounding boxes, and information of its class. Fig 4. shows the facemask detection video dataset.

Fig.4 The facemask detection video dataset.

ii. GoogleNet

In 2014, various models were created like VGG but in the contest of ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14) for classification of image and object detection, GoogleNet ranked First. They Concluded that to improve neural networks for computer vision GoogLeNet method can be used. The main pros of this method are a quality gain due to an increase in computational requirements compared to shallower and narrower architectures. To improve the accuracy rate, fewer resources usage and cost, increase computational gain and face mask detection and calculation of social distance between two human objects with real-time images. Hence, the approach used for detection of face masks and person, or object detection is GoogleNet which calculates the distance between the centroid of objects for checking the distance between two persons.It is a type of inception architecture that is based on a convolutional neural network (CNN). The Inception modules utilize, the network which is chosen for multiple convolutional uses its filter sizes in each block. It is an Inception network that stacks the inception modules of GoogLeNet above each other, layers of max pooling stride 2 to halve the resolution of the grid. Fig 4 shows the layers GoogleNet Architecture.GoogLeNet Model contains 22 Layers of Architecture. It has nine inception layers and two auxiliary classifiers. The model has used Max pooling and Average pooling. The inception module and auxiliary module have different layers explained below.

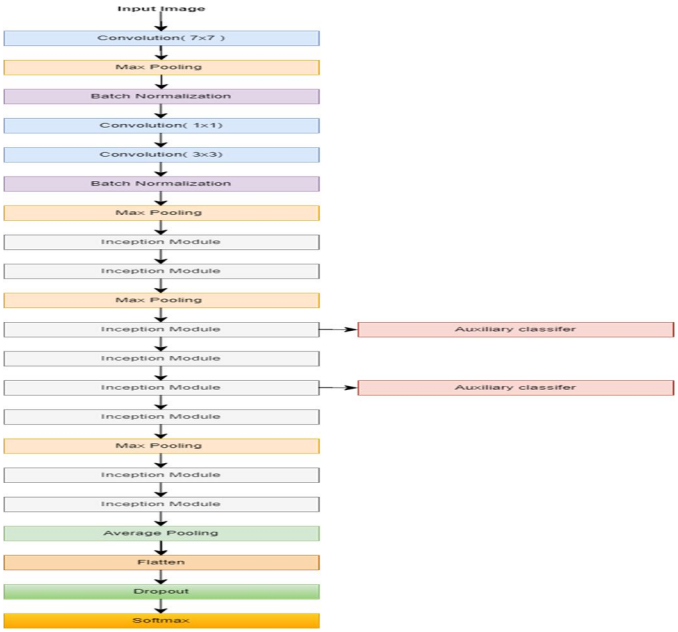

Fig 4. Layers of GoogleNet Architecture

- *Inception Module*

In Googlenet, the Inception module is used which has 1×1, 3×3, 5×5 convolution, and 3×3 max pooling and it is performed in a parallel way at the input and the output of these are stacked together to generated the final output. These different convolution filters of different sizes are used to handle objects at multiple scales better. Fig 5. is attached below which shows layers of the inception module.

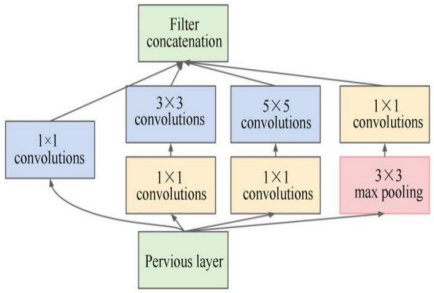

Fig 5. Inception Module

- *Auxiliary classifier*

In Googlenet, an Auxiliary classifier is used which has an Average pooling layer, 1x1 convolution, and dropout regularization with a dropout ratio of 0.4.Fig 6 below shows the diagram of Auxiliary classifier.

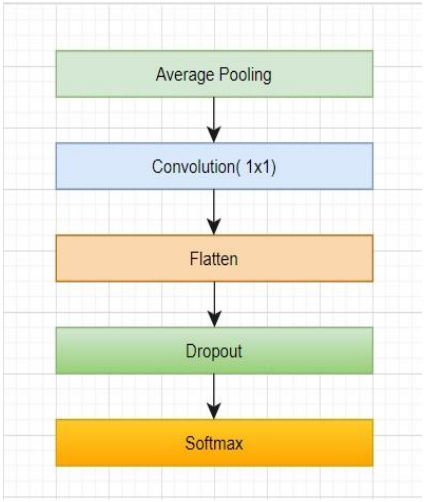

Fig 6. Auxiliary Classifier

- *Max Pooling*

In Google Net , there are four Maximum pooling layer which takes 3x3 pool size and strides of 2.

- *Convolution*

In Google Net, there are 3 Convolutional layers, the first layer of Google Net is Convolutional layer and it takes 7x7 kernal size and strides of 2.The fourth and fifth layer takes 1x1 kernal size and strides of 1 and 3x3 kernal size and strides of 1 respectively.

- *Batch Normalisation*

Batch normalization applies a trans-formation that maintains the mean output close to 0 and the output standard deviation close to 1.

- Average Pooling

In Google Net , there is 1 Average pooling layer which takes 7x7 pool size and strides of 1.

IV. System Architecture

Below is the overall system Architecture for face mask and social distancing using GoogLeNet.

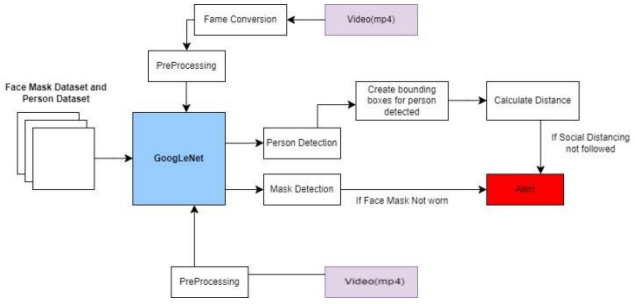

Fig 6. System Architecture.

i. Person Detection and Distance calculation.

For Social distancing, The video is converted into frames. In pre processing, the mapping of the actual distance and camera distance is done.Then the image is given to GoogleNet from the training it will detect the person, For

Every Person detected the boxes are drawn .From the boxes drawn we will calculate the points of centroid. After that from the centroid of person, the distance will be calculated of two persons using Euclidean distance.

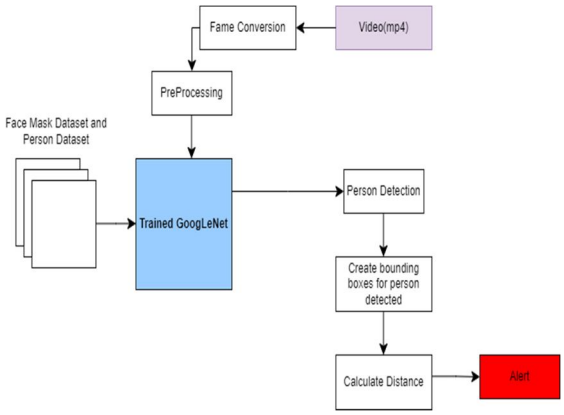

Fig 7. Person Detection and Calculate Social distancing

Suppose x1 and y1 are the centroid points of a person1 and x2 and y2 are the centroid points of the person2 then the distance will be calculated by the formula,

$$\text{Distance}((x1, y1), (x2, y2)) = \sqrt{[}\ (x2 - x1)2 + (y2 - y1)2]$$

ii. Mask Detection

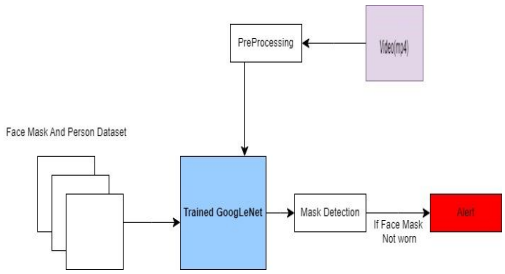

Fig 8. Mask Detection using GoogLeNet

The live video is read and each frame is given for pre-processing, where the image is converted to a gray image. The gray image is given to the haar cascade file to detect the faces. For Each face in the image, the face is given to a trained Google net model to predict the face mask present or absent.
If the face doesn't have a mask then the alarm will be generated and the output image is shown which rectangle boxes with label mask and without a mask.

V. RESULT

When the image is given to a trained google Net model, it will predict the face mask present or absent. And the system will generate an alarm if the mask is absent. Along with that the rectangle box on faces with label mask or not a mask. Below is the face mask detected image.

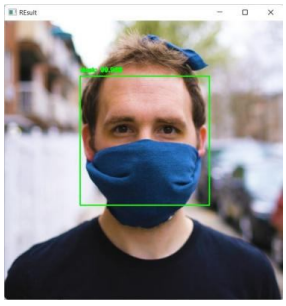

Fig 9. Face Mask detection Output

When the video is converted into frames and each frame is given to google Net then, it should predict the person and the system should calculate the distance between humans if social distancing is violated then the alarm should be generated. And Red, Green, and Yellow bounding boxes will be displayed representing high, no, and low-risk resp. Below is the output of Social distancing and person detection.

Fig 10. Person Dectection and social distancing Output

VI. Result Analysis And Performance Analysis

The GoogleNet testing results are compared with CNN(Convolutional Neural Network) with same Dataset of images that is MAFA(Masked Faces) and Person detection Dataset were used for testing ,Both the GoogleNet and CNN model was trained and It is found that the testing result of GoogleNet gave 95 percent accuracy for Face Maskdetection and CNN gave 85 % for FaceMaskDectection.And the testing results for person detection gave accuracy of 96% and CNN person detection gave accuracy of 89%.The GoogleNet test model has precision value of 0.9333 , Recall value of 0.933and f-1 measure score of 0.9 for FaceMask , Whereas the CNN test model gave precision value of 0.86, Recall value of 0.76 and f-1 measure score of 0.8.The GoogleNet person detection train model gave precision of 0.96 , Recall value of 0.95 and F1 measue of 0.88 whereas CNN Person detection for social distancing gave Precision value of 0.88,Recall value of 0.8 and F-1 measure of 0.8.Below is the confusion matrix graph for GoogLeNet facemask, and CNN FaceMask.GoogLeNet person detection for social distancing and CNN person detection for social distancing.

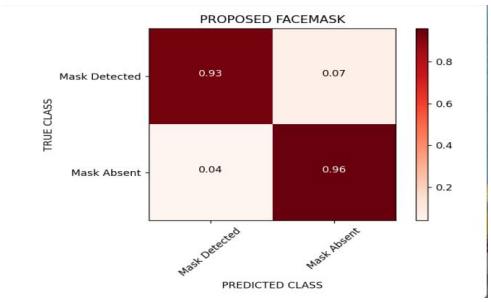

Fig 11. Proposed face mask confusion matrix.

GoogLeNet FaceMask confusion matrix has 0.93 True Positive(TP) which is actually masked and also predicted masked,0.07 (False Positive) which predicted non-mask but has worn mask.0.04(False Negative) which predicted mask but is masked. and 0.96 True Negative(TN) which has predicted Non-mask and is a non mask.

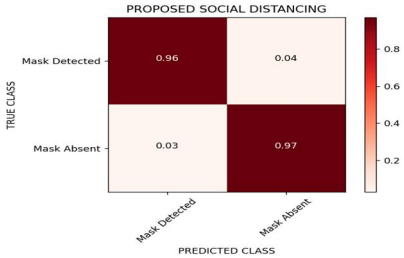

Fig 12. Proposed Person Detection Confusion Matrix.

Google Net Person Detection for social distancing confusion matrix has 0.96 True Positive(TP) which is actually a person and also predicted person 0.04 (False Positive) which predicted a non-pedestrian but is a person.0.03(False Negative) which predicted a person but is non-pedestrian. and 0.97 True Negative(TN) which predicted the person is absent and is absent.

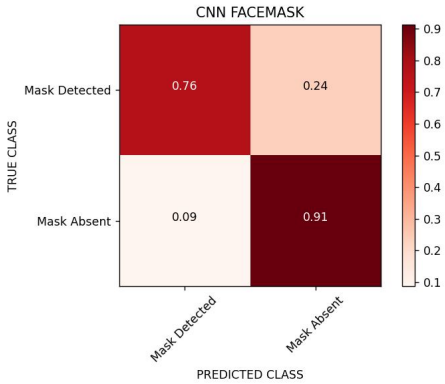

Fig 13. CNN Face Mask detection confusion matrix.

CNN FaceMask confusion matrix has 0.76 True Positive(TP) which is actually masked and also predicted masked,0.24 (False Positive) which predicted non-mask but have worn mask.0.09(False Negative) which predicted mask but is masked. and 0.91 True Negative(TN) which has predicted Non-mask and is a non mask.

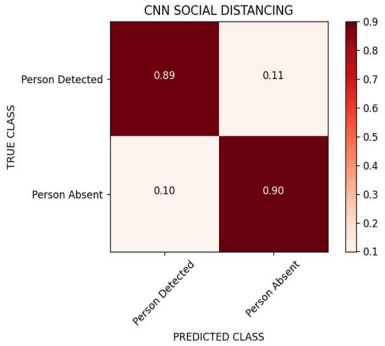

Fig 14. CNN Person detection confusion matrix.

CNN Person Detection for social distancing confusion matrix has 0.89 True Positive(TP) which is actually a person and also predicted person,0.11 (False Positive) which predicted nonpedestrian but is a person.0.09(False Negative) which predicted a person but is nonpedestrian. and 0.91 True Negative(TN) which predicted person is absent and is absent.

VII. CONCLUSION

The proposed idea of Maintaining the protocol of Social distancing and detection of face masks using GoogLeNet will improve security by saving time and decreasing the spread of Covid.In this paper, we have proposed GoogleNet model which will detect FaceMask and Persons. For social distancing the video is converted into frames and given to the GoogleNet which will predict the person and boxes are drawn around the person and from the centroid of two boxes.The distance will be calculated between each person in the frame.And Red,Green,Yellow bounding boxes will be displayed representing high, no, low risk resp.And if the person fails to wear mask and fails to follow social distancing rule alarm will be generated.To gain the accuracy the GoogLeNet model is descovered and the GoogLeNet model is compared with CNN model and after comparision It is found that the testing result of GoogleNet gave 95% percent accuracy for Face Maskdetection and CNN gave 85 % for FaceMaskDectection.And the testing results for person detection gave accuracy of 96% and CNN person detection gave accuracy of 89%.This proposed system can be used in real world if the cases increase again.

REFERENCES

[1] Chen Wang,Peter W Horby,Frederick G Hayden, George F Gao "A novel coronavirus outbreak of global health concern" www.thelancet.com Vol 395 February 15, 2020.

[2] Laura Matrajt, Tiffany Leung "Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease" pmc article doi: 10.3201/eid2608.201093. Epub 2020 Apr 28.

[3] Shixing Chen; Caojin Zhang; Ming Dong; Jialiang Le; Mike Rao "Using Ranking-CNN for Age Estimation" 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[4] Shashi Yadav "Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID- 19 Safety Guidelines Adherence" International Journal for Research in Applied Science & Engineering Technology (IJRASET) 2020

[5] Sachin Sudhakar Farfade, Mohammad Saberian, Li-Jia Li "Multi-view Face Detection Using Deep Convolutional Neural Networks" Computer Vision and Pattern Recognition 2015

[6] Toshanlal Meenpal, Ashutosh Balakrishnan, Amit Verma "Facial Mask Detection using Semantic Segmentation", 4th International Conference on Computing, Communications and Security (ICCCS) 2019.

[7] Md. Sabbir Ejaz*, Md. Rabiul Islam, "Masked Face Recognition Using Convolutional Neural Network", International Conference on Sustainable Technologies for Industry 4.0 (STI), (2019).

[8] N Ozkaya, S Sagiroglu, "Intelligent face mask prediction system", IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), (2008).

[9] Akash Manish "Comparative Analysis of Convolutional Neural Network Architectures for Real Time COVID-19 Facial Mask Detection" (2021)

[10] Moein Razavi "An Automatic System to Monitor the Physical Distance and Face Mask Wearing of Construction Workers in COVID-19 Pandemic" (2021)

[11] Adina Rahim, Ayesha Maqbool, Tauseef Rana, "Monitoring social distancing under various low light conditions with deep learning and a single motionless time of flight camera", (2021).

[12] Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal and Gaurav Rai "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques "2021

[13] M. Y. Shams, A. S. Tolba, S. H. Sarhan "A Vision System for Multi-View Face Recognition" (2016)

[14] Sandeep Gupta , S.V.N. Sreenivasu , Kuldeep Chouhan , Anurag Shrivastava , Bharti Sahu , Ravindra Manohar Potdar "Novel Face Mask Detection Technique using Machine Learning to control COVID'19 pandemic"

[15] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and Nour Eldeen M. Khalifa "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection"(2021)

[16] D. G. Dondo, J. A. Redolfi, R. G. Araguás, and D. Garcia, "Application of deep-learning methods to real time face mask detection," IEEE Latin America Transactions, vol. 19, no. 6, pp. 994–1001, 2021.

[17] Christian Szegedy1, Wei Liu2, Yangqing Jia1, Pierre Sermanet1, Scott Reed3"Going Deeper with Convolutions" IEEE Conference Paper

[18] Y. Chen, M. Hu, C. Hua et al., "Face mask assistant: detection of face mask service stage based on mobile phone," IEEE Sensors Journal, vol. 21, no. 9, Article ID 11084, 2021.

[19] X. Fan, M. Jiang, and H. Yan, "A deep learning based light-weight face mask detector with residual context attention and Gaussian heatmap to fight against COVID-19," Ieee Access, vol. 9, Article ID 96964, 2021.

[20] B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID- 19," Sensors, vol. 20, no. 18, 2020.

[21] Choudhary, Manisha, Monika Lokhande, Rushikesh Borse, and Avinash Bhute. "A machine learning approach to aid paralysis patients using EMG signals." In Advanced Data Mining Tools and Methods for Social Computing, pp. 107-125. Academic Press, 2022.

[22] Bhute, Avinash, Rakesh S. Patil, Shreyas Sanghai, Prajakta Yadav, and Sanketkumar Sathe. "Functional Electrical Stimulation for Hand Movement of Person." In Proceedings of International Conference on Machine Intelligence and Data Science Applications, pp. 465-481. Springer, Singapore, 2021. P. S. Sangle, R. M. Goudar and A. N. Bhute, "Methodologies and Techniques for Heart Disease Classification and Prediction," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225673.

[23] Yadav and R. Badre, "Lung Carcinoma Detection Techniques: A Survey," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 2020, pp. 63-69, doi: 10.1109/CICN49253.2020.9242633.

[24] Kasi Uday Kiran, Sandeep Dwarkanath Pande, P. Poorna Priya, K. Kalki Sai, Nandigama Apoorva, M. Geethika, Sk Hasane Ahammad "Smart Weapon Detection System Using Thermal Imaging Through Machine Learning Algorithms", India 2023, DOI: 10.4018/978-1-6684-9189-8.ch021