Social Distancing and Face Mask Detection Using Open CV

Majji Ramachandro[1](✉), Ala Rajitha[1], Dasari Madhavi[2], Jagini Naga Padmaja[3], and Ganesh B. Regulwar[1]

[1] Information Technology, Vardhaman College of Engineering (Autonomous), Hyderabad, India
{ramachandro.m,ganeshregulwar}@vardhaman.org
[2] Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India
[3] Department of Computer Science and Engineering, Vardhaman College of Engineering (Autonomous), Hyderabad, India
nagapadmaja_14@vardhaman.org

Abstract. In 2019, people are getting sick from the coronavirus. We can only stay safe from the epidemic if we wear masks and stay away from each other. Airports, hotels, hospitals, and train stations, among other places, require users to wear the Mask and stay away from other people. Manually checking people to see if they follow the mask-and-distance rule is hard because it costs a lot. The COVID-19 Face Mask and Social Distancing Detector System uses machine learning to find face masks and social distances at the same time. It does this by combining high-level contextual features with feature maps and an artificial neural network. IP address and CCTV cameras with computer vision would be used in the technology to identify people without masks or social isolation. This solution keeps things safe even when no one is watching. The technology could help hospitals, offices, schools, building sites, airports, and more. People may be safer if they use our face mask and social distance detecting device.

Keywords: Coronavirus · Covid-19 · Face Mask · Machine Learning

1 Introduction

Coronavirus disease has had a significant global effect early 2019. One of the best methods to protect yourself is to always wear a mask in public and keep your distance from other people. Isolation and mask use are the only ways to prevent the spread of covid-19. This project aims to stop the spread of the coronavirus by developing a tracking system that utilizes existing IP address and CCTV cameras in combination with Computer Vision to identify individuals who aren't covering their faces or staying in designated areas.

With the use of computer vision and artificial intelligence, the COVID-19 Face Mask and Social Distancing Detector System (COVID FSD) can identify signs of social

S. Nandan Mohanty et al. (Eds.): ICISML 2022, LNICST 470, pp. 488–498, 2023.
https://doi.org/10.1007/978-3-031-35078-8_40

distance. To determine whether or not a person is wearing a mask, the COVID FSD System uses an Artificial Network. The system may be integrated with any IP or CCTV camera, old or new, to enable the recognition of people even while they are wearing masks. The program may now run automatically and the mask must be worn at all times. This system provides safety and security features without the requirement for user monitoring.Its artificial intelligence software identifies faults such as not wearing a face mask and social distancing. This system may be installed in hospitals, office buildings, government offices, schools and educational institutions, construction sites, manufacturing facilities, and airports, among other locations. The COVID FSD System is easy to operate and requires no technical support.The method guarantees complete consumer confidentiality. In truth, none of the three applications capture photographs; they only produce an alert message when a certain condition is satisfied. During the most vital phase of the pandemic's fight, COVID FSD may be an especially simple-to-use instrument.

Researchers have made important contributions to the fight against COVID-19, and the number of AI models related to COVID-19 is growing quickly. Well-trained AI models can make sure accurate and quick diagnoses or help doctors speed up the process of diagnosing and do less manual work. AI models could find high-risk patients early on, describe the spread of COVID-19, and model how the disease spreads based on training data. AI-based methods could help find new drugs and vaccines. For example, existing drugs could be used in new ways, targets could be tested as vaccines based on a possible mutation model for SARS-CoV-2, and compounds could be tested as possible vaccine adjuvants.

A chatbot driven by artificial intelligence (AI) has already proven useful in the medical field, where it has the potential to help far more patients than a human contact centre could (18). A.I. might control the epidemic by imposing social isolation and lockdowns and by employing heat sensors to search public locations for suspected patients (3, 17). COVID-19 patients have benefited greatly from the use of AI in a variety of settings, including but not limited to: diagnosis; public health; clinical decision making; social control; therapeutics; vaccine development; surveillance; combination with big data; operation of many other core clinical services; and control (3, 18, 19).

The most important things to do to stop the spread of the COVID19 pandemic are quick diagnosis, accurate prediction, better monitoring, and effective treatments. This is because the pandemic is putting a lot of pressure on the few medical resources that are available. A lot of review articles on similar topics have been written. But the results of these studies aren't always the same, and there isn't much research that looks at the use of AI for COVID-19 in a way that is consistent with PRISMA. Most of these studies only talk about things like diagnosis or treatment. So, we did this review to find out how well AI worked for COVID-19 and to describe the main ways AI is used, its possible benefits and drawbacks, and where it might go in the future.

2 Literature Survey

This research combines RestNet50 [1] and YOLOV3 (You Only Look Once) [2] with transfer learning to achieve a balance between resource limits and recognition accuracy in real- time video surveillance in order to identify persons wearing face masks while keeping a safe social distance. Using Open CV and RestNet50, we use neural network models to analyze, Real-Time Streaming Protocol (RTSP) video streams. We use new deep learning algorithms with traditional projective geometry approaches to help meet real-time needs with high prediction accuracy.

If a person violates covid-19 safety rules, the state police control center will be alerted. It facilitates automation while needing a mask and social distance standards. This model's accuracy on a local PC was 85–95%. Deep model object recognition approaches [3] have made great strides in computer vision in recent years and may be better at solving complex situations than shallow models.

Deep person identification methods emphasize feature, contextual, and occlusion learning. Deep Learning object identification models [5] fall into two categories: I enjoy R-CNN [6], Fast R-CNN [7], Faster R-CNN (ii) one-stage detectors like YOLOv3 [9]. Two-stage detectors perform detection in phases: the first step calculates solutions and the second classifies them. [10] YOLO considers detection a regression problem and only looks at the picture once.

The Viola – Jones [11] object identification system can identify any thing, however it's most often used to recognize faces since it's more accurate and quicker. The Viola and Jones method is a good example of supervised learning. Zhu [12] also disclosed some further information. It is common practice to employ a neural network-based detector when trying to identify a face. It works best on the front, erect portion of the face. Another framework for face detection was suggested by Li et al. [13, 14] in the form of a Multi-View Face Detector with Surf Capabilities.

When applied to the GTX470, the face identification technique proposed by Oro et al. [15] that makes use of acoustic features increased speed by a factor of 2.5. However, they recently used CUDA, a technology for programming NVIDIA GPUs. When implementing the viola-jones face recognition method on graphics processing units (GPUs), an uneven workload developed. This is in comparison to OpenCL, which is employed in a wide range of calculated components.

Glass et al. (2006) [16] addressed the significance of social differentiation and how, without the use of vaccines or antiviral drugs, the danger of pandemic growth may be gradually lowered by preserving social isolation.

The authors performed extensive research in both rural and urban areas to demonstrate a decline in population increase. Z., Luo [17] examines the identification of individuals with full-face or partial blockage. This approach puts individuals into two groups: those whose hands are covering their faces and those whose faces are obscured by objects. This approach is inapplicable to our situation, which requires recognizing faces covered by masks such as scarves, mufflers, handkerchiefs, etc.

3 Methodology

The suggested method uses Open CV and Keras to train the model. We employ a triangle similarity methodology to measure the distance between persons acquired by cameras in real time in public settings, and we combine bespoke data collection with a transfer compression method to construct a face mask recognition model based on public mask types.

This approach incorporates both face mask recognition and social distance detection. Our approach suggests using an ARMv8 1.5 GHz CPU and 4 GB of RAM equipped machine 4 Model-B as the edge device (Fig. 1).

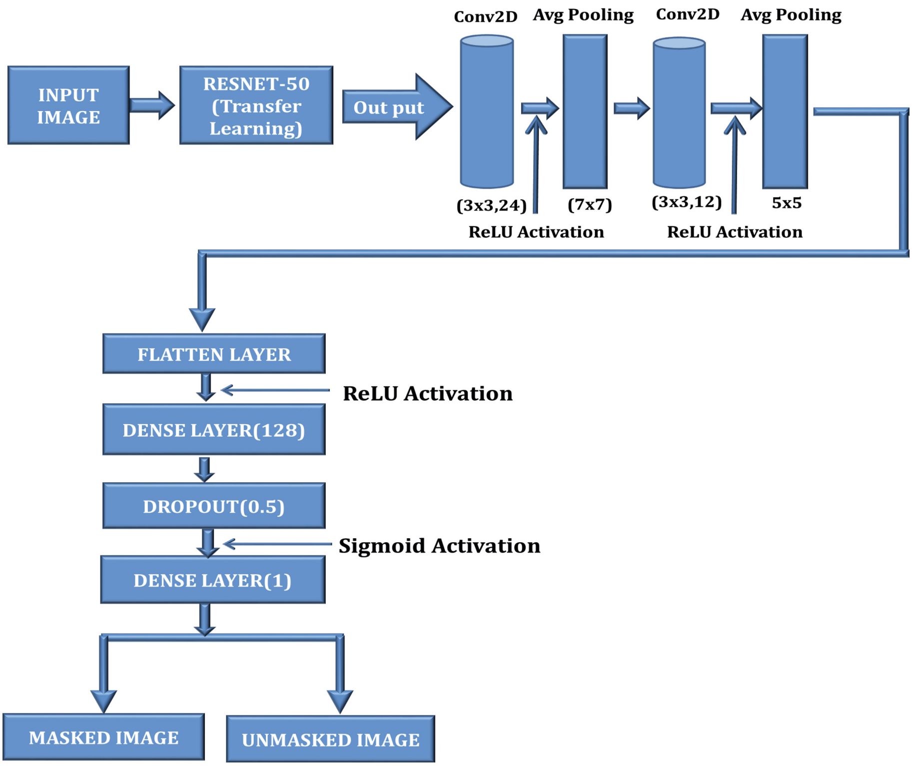

Fig. 1. Stacked ResNet-50 Face Mask Classifier

In this research used various hyper parameter values used. There are various Components like number of layers, nodes epochs, learning rate, activation functions and different optimizers. Tested with these parameters were obtained significant results with different combinations of the hyper meters.

Overview of Hyper Parameters Used

1.	Selecting Number of Layers	: (1,2,3,4)
2.	Fixing Number of Nodes or Neurons	: (2, 4, 8, 16, 32, 64, 128, 256, 512)
3.	Number of Epochs	: 50 (for each generation)
4.	Number of generations	: 50
5.	Learning rate	: 0.01 to 0.1
6.	Chosen the Activation functions	: ReLu [0 to x], Tanh [-1 to 1], Sigmoid [0 to 1]
7.	Chosen optimizers	: Adam, Rmsprop, Adadelta, Adagrad, SGD
8.	Dropout rates	: 0.1 TO 0.8

Number of Layers: Hidden, input, and output were the only traditional neural network layers. These are all the same sort of layer if input data comes from outside and output goes outside.

Number of Nodes or Neurons: Between the input and output layers should be hidden neurons. 2/3 of the input layer plus the output layer should be hidden neurons. Hidden neuron count should be fewer than input layer size.

Number of Epochs: In one Epoch, the neural network passes the entire dataset forward and backward once. Number of epochs specifies how many times the learning algorithm will run over the training dataset. Each training dataset sample contains one epoch to update model parameters. Epochs have many batches.

Learning Rate: Learning rate is a hyper parameter used in neural network training with a tiny positive value between 0.0 and 1.0. Step size or "learning rate" refers to how often weights are increased during training. The model's learning rate determines how rapidly it adapts. Smaller learning rates require more training epochs due to smaller weight changes every update, while bigger learning rates require fewer training epochs.

Activation Functions: The activation functions of neural networks play a significant role in deep learning. The success or failure of a large-scale neural network rests on the activation functions that are used to train the network and produce the desired results in terms of output, accuracy, and computing efficiency.

Optimizers: Optimizers train neural networks without realizing its optimization. Optimizers modify neural network weights and learning rate to minimize losses.

Dropout Rates: The dropout hyper parameter is the likelihood of training a particular node in a layer, where 1.0 implies no dropout and 0.0 means no layer outputs. Hidden layer dropout should be 0.5 to 0.8. Input layers utilize 0.8 dropout rate.

Result Analysis & working procedure

- The proposed method checks social distance and recognizes face masks to ensure public safety. This section describes the solution architecture and explains how the recommended system would prevent coronavirus development. The proposed system uses transfer learning, a deep learning method, and computer vision to automatically watch people in public using a camera attached to a local machine and discover mask wearers.
- Person recognition, safe distance measurement, and face mask detection are the proposed system's main contributions. YOLOV3 (You Only Look Once) beats Faster R-CNN in real-time person detection. The CNN model with 91.2% map using RestNet50 and Open CV. A box will be drawn around every person who is detected. [18] While YOLOV3 is capable of recognizing many objects in a single frame, it can only identify a single person.
- First determine the person's distance from the camera using triangle similarity, then calculate the camera's perceived focal length, assume the person's real height H = 165 cm, then use YOLOV3 person recognition to identify the person's pixel height P

using bounding box coordinates. [19] The focal length of the camera can be estimated using these values using the formula below:

$$F = (P \times D)/H \tag{1}$$

- The person's distance from the camera is then calculated using the real person's height H, the person's pixel height P, and the camera's focal length F. The following formula can be used to calculate the distance from the camera:

$$(H \times F)/P = D \tag{2}$$

- ResNet consists of many residual blocks where residual learning is adopted to every few (usually 2 or 3 layers) stacked layers.

$$y = f(x,\ W) + x \tag{3}$$

  Here W's are the weights and these are learned during training.
- The output of the second convolution is transferred to a regular pooling layer of 5 $\times$ 5. The result is then straightened and gone via a thick layer of 128 neurons using ReLU enactment work. A dropout layer of 0.50 is employed after the thick layer. The outcome after dropout is taken care of into a thick layer of one neuron. The last layer's output is processed in an 8-layer neural network that ultimately produces the parallel grouping.

The architecture can be seen in the above diagram to contain (Fig. 2)

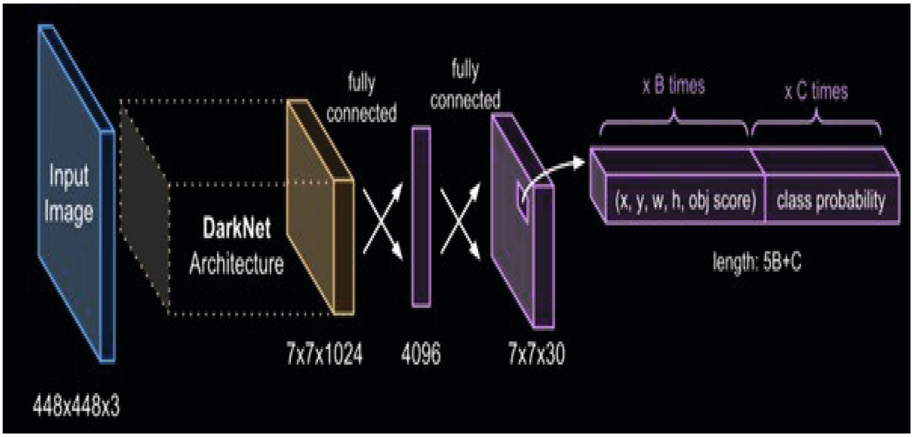

Fig. 2. YOLO Architecture

4 Experimental Results

Four steps are followed in the proposed system, including:

- Data collection,
- preprocessing,
- model building,
- training,
- testing, and implementation

Data Collection and Pre-processing
The recommended technique uses labeled face pictures with varied masks to train our models. We use background subtraction in pre-processing. YOLOV3 detects social distance and masks in real-time. Our face mask detector was trained with 3165 pictures. Before labelling the custom face mask picture dataset, the training and testing sets are divided.

Model building and Training
Our proposed framework will fine-tune the RestNet50 model, a highly efficient architecture that can be deployed to edge devices with minimum computing power [20]. Yolov3 trained 80% of our custom data which can collect on several things with a single frame. In order to train the algorithm, the annotated photos and the custom data set are placed in the project's folder. Image is scaled up to 224224 pixels, transformed to numpy array format, and labels are added to the dataset before we use the YOLOV3 model as input to construct our own model using RestNet50 as the backbone and train our model with the Keras Object Detection API. Before model training, Keras improves data and downloads ImageNet weights. ResNet50 is trained utilizing both pre-trained ImageNet weights and annotated new data set photographs by modifying the head layer weights without changing the base layer weights (Fig. 3).

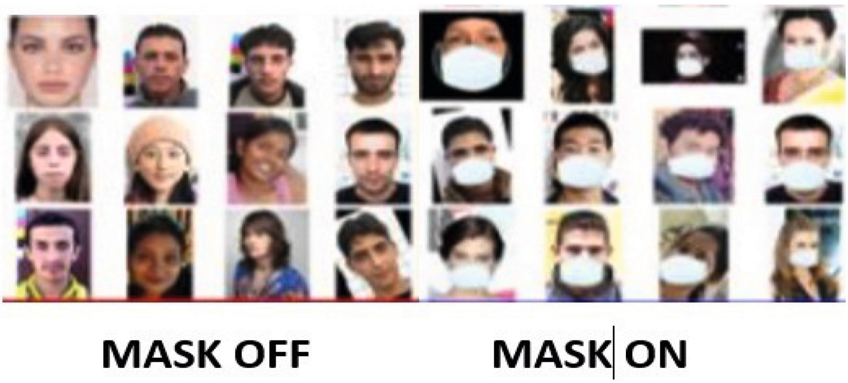

Fig. 3. Data Processing

We trained our model for 1000 steps using the Adam optimization technique, the learning decay rate for updating network weights, and the binary cross entropy for classifying mask types. Initial learning rate (INIT LR) = 1e–4, epochs (EPOCHS) = 20, and batch size (BS) = 32 were the parameters used.

We used a camera for cv2 social distance monitoring, and after an individual has (Fig. 4)

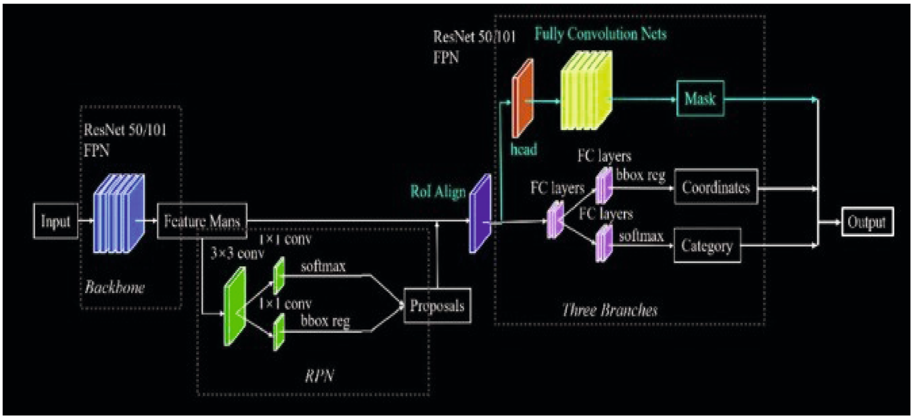

Fig. 4. R-CNN

been recognized, we calculate the center of their image by averaging their left and right extremities. We utilize the Euclidean distance between the points to find out how far apart the people in the picture located (Fig. 5).

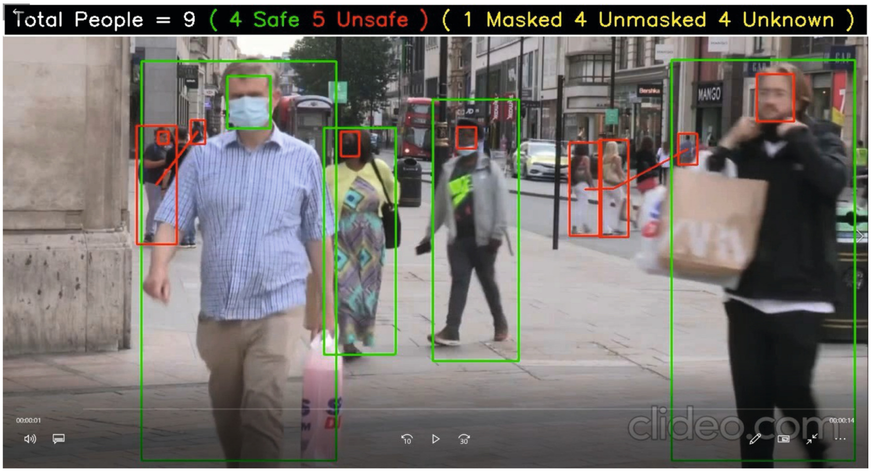

Fig. 5. Detection Result

Model Implementation
The proposed solution to stop the spread of Covid-19 is to use a machine with a camera

that tracks public places in real time. The camera is linked to the custom data set that was used to train the model. This set is installed on the machine.

The camera sends live footage to a computer model, which then monitors public areas in real time, automatically analyzing crowd behavior to determine if individuals are keeping a safe distance from one another and whether or not they are wearing Mask. Our technology has two phases: the first is when an unmasked individual is spotted and his photograph is collected and sent to the State Police Headquarters' control center; the second is when an individual continually detects a social distance violation in certain criteria.

The proposed method would improve public safety by saving time and aiding in the decrease of coronavirus spread; this is because real-time interventions have the potential to dramatically minimize noncompliance (Fig. 6 and Table 1).

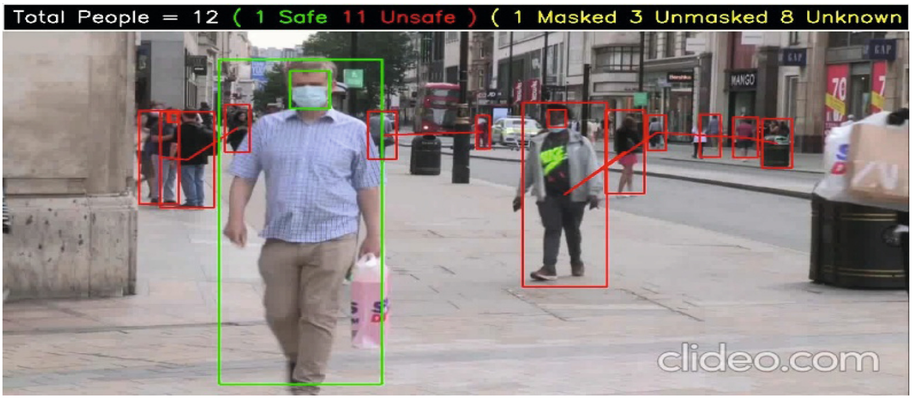

Fig. 6. Detection Result

Table 1. The above figures show the result for the images

Total people	Safe	Unsafe	Masked	Un masked	Un known
9	4	5	1	4	4
12	1	11	3	3	8

We think that by adopting this strategy, we can boost the future effectiveness of plant processes while simultaneously increasing public safety.

5 Conclusion and Future Work

To prevent the spread of the COVID-19 virus and to help authorities in reducing their physical surveillance work in containment zones and public areas where surveillance is required, this paper proposes an approach that uses computer vision and ResNet50 classifiers to monitor public places automatically.

This proposed solution will therefore be handy for automatically keeping tabs on public spaces now that the lockout is ending. We have discussed in detail how to keep user data safe while also keeping an eye on people's physical and social distances and identifying masks created to protect people's health.

By putting the model into operation on my personal workstation, I was able to test the approach in a live environment (Computer). Therefore, the retail, healthcare, and business sectors will rely primarily on the face mask detection and social distancing system as their primary digital solution.

Learn how digital tools can improve your ability to serve your citizens. In addition to medical and administrative settings, this technology can be used in other places like airports, construction sites, factories, schools, and public buildings.

If used appropriately, the COVID-19 mask detector we're designing today can safeguard you and others. Real-time measures can reduce violations, therefore the proposed method would save time and prevent coronavirus spread. We believe using this strategy in the future will boost public safety and plant efficiency.

References

1. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: RestNet50: inverted residues and linear bottlenecks. IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
2. Fu, C., Liu, W., Ranga, A., Tyagi, A., Berg, A.: DYOLOV3: deconvolutional yolo model. arXiv preprint arXiv:1701.06659, (2017)
3. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Type pyramid networks for object detection. IEEE Conference Proceedings on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
4. Goodfellow, I, Bengio, Y., Courville, A.: Deep learning process, Chapter 6 (2016)
5. Farfade, S.S., Saberian, M.J., Li, L.: Multi-view face recognition using deep convolutional neural networks. In: ACM ICMR, pp. 643–650 (2015)
6. Masita, K.L, Hasan, A.N., Satyakama, P.: Pedestrian identification by mean of R-CNN object detector. IEEE Latin American Conference on Computational Intelligence (2018)
7. Hu, G., et al.: When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. IEEE (2015)
8. Girshick, R.: "Fast R-CNN". In: Proceedings of the IEEE International Conference Computer Vision, pp. 1440–1448 (2015)
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with area proposal networks. In: Proceedings of the Advance Neural Information Processing Systems, pp. 91–99 (2015)
10. Liu, H., Chen, Z., Li, Z., Hu, W.: An powerful method of pedestrian detection based on YOLOv2. Math. Eng. Issues **2018**, 3518959 (2018)
11. Viola, P., Jones, M.: Fast object detection using an enhanced cascade of simple features.In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. I. CVPR 2001, Kauai, HI, USA (2001)
12. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark position in the wild. In: IEEE CVPR, pp. 2879–28862012 (2012)
13. Howard, A.G.: Effective convolutional neural networks for mobile vision applications (2017)
14. Zhu, X.: Semi-supervised study of learning literature survey. Computer Science, University of Wisconsin-Madison **2**(3), 4 (2006)

15. Oro, D., Ferna´ndez, C., Saeta, J.R., Martorel, X., Hernando, J.: Real-time GPU-based face detection in HD video streams. IEEE 2011 IEEE International Conference on Computer Vision (2011)
16. Glass, R.J., Glass, L.M., Beyeler, W.E., Min, H.J.: Targeted social distancing architecture for pandemic influenza. Emerg. Infect. Dis. **12**, 1671–1681 (2006)
17. Ge, S., Li, J., Ye, Q., Luo.: Detection of masked faces in the wild with LLE-CNNs. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 426–434. Honolulu, HI (2017)
18. Pan, S.J., Yang, Q.: a survey of transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (Oct.2010)
19. Mohamed, S.S., Tahir, N.M., Adnan, R.: Background modelling and background subtraction efficiency for object detection. In: 6th International Colloquium on Signal Processing its Applications. Mallaca City **2010**, 1–6 (2010). https://doi.org/10.1109/CSPA.2010.5545291
20. Piccardi, M.: Background subtraction techniques: a analysis. IEEE Int. Conf. Syst. Man Cyber. **4**, 3099–3104 (2004)(IEEE Cat. No.04CH37583). The Hague (2004). https://doi.org/10.1109/ICSMC.2004.1400815