NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY
[IJIERT] ISSN: 2394-3696
Website: ijiert.org
VOLUME 7, ISSUE 12, Dec. -2020

REAL-TIME FACE MASK DETECTION WITH COMPUTER VISION AND DEEP LEARNING

Chaitanya Krishna Suryadevara
Department of Information Technology
csuryadevara001@my.wilmu.edu

Abstract

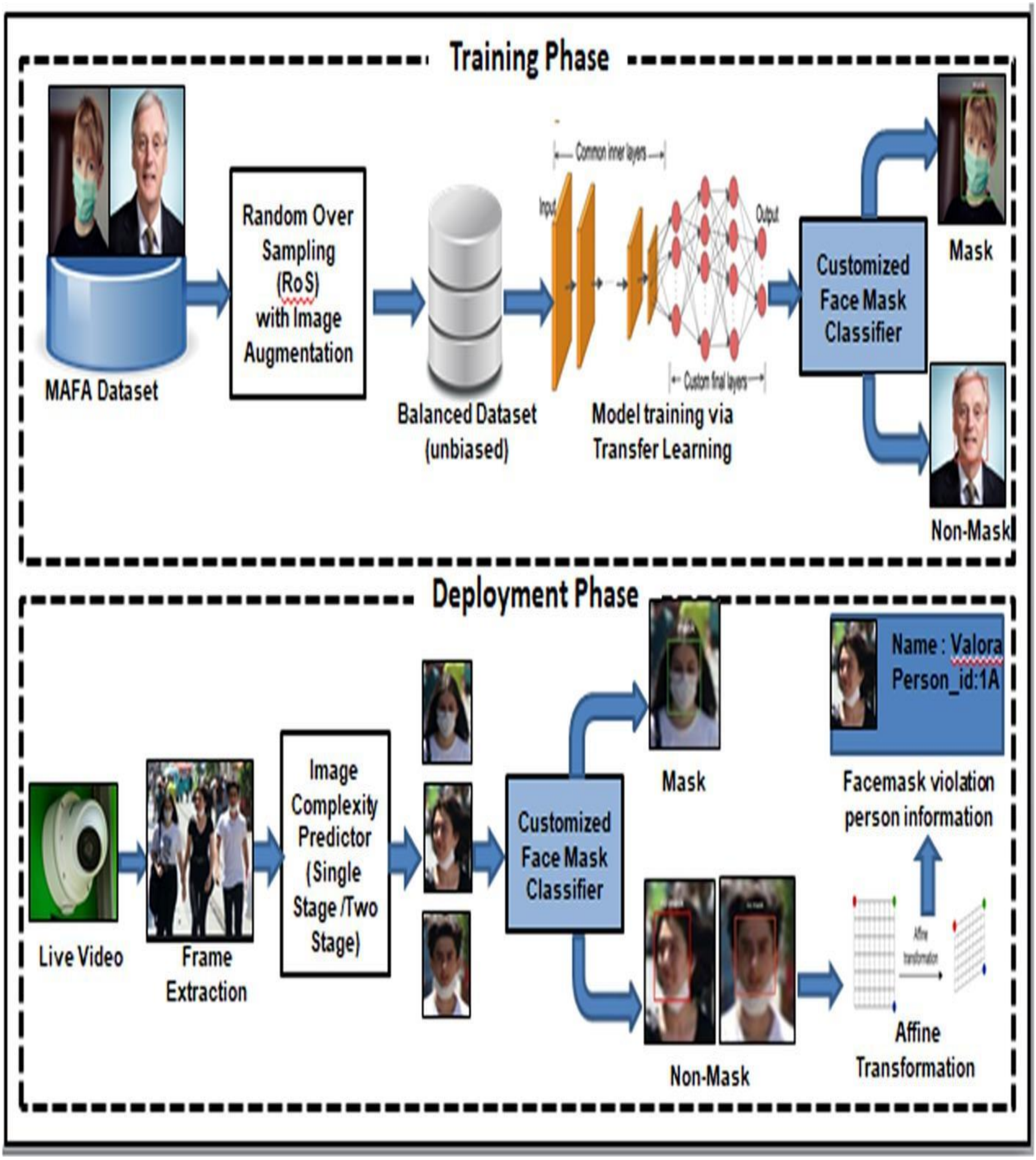
Corona virus sickness has become a big public health issue in 2019. Because of its contact-transparent characteristics, it is rapidly spreading. The use of a face mask is among the most efficient methods for preventing the transmission of the Covid-19 virus. Wearing the face mask alone can cut the chance of catching the virus by over 70%. Consequently, World Health Organization (WHO) advised wearing masks in crowded places as precautionary measures. Because of the incorrect use of facial masks, illnesses have spread rapidly in some locations. To solve this challenge, we needed a reliable mask monitoring system. Numerous government entities are attempting to make wearing a face mask mandatory; this process can be facilitated by using face mask detection software based on AI and image processing techniques. For face detection, helmet detection, and mask detection, the approaches mentioned in the article utilize Machine learning, Deep learning, and many other approaches. It will be simple to distinguish between persons having masks and those who are not having masks using all of these ways. The effectiveness of mask detectors must be improved immediately. In this article, we will explain the techniques for face mask detection with a literature review and drawbacks for each technique.

Keywords: Covid 19, Deep Learning, WHO, Detection, YOLO, Face Mask Detection and Computer Vision.

1. Introduction

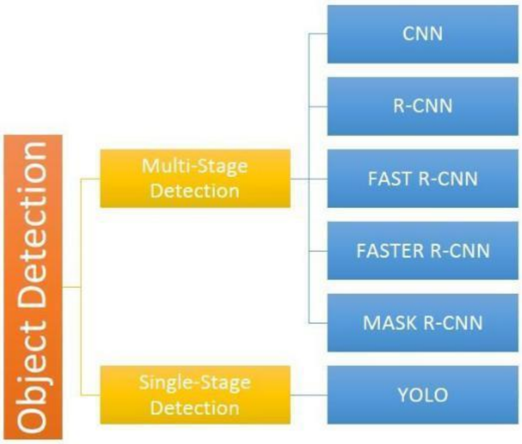
Because of the global pandemic of Covid-19, wearing a face mask in public is becoming increasingly popular. Since the people of Covid-19 refuse secure their health by wearing masks against air pollution. Others, on the other hand, are demure about their appearance and conceal their feelings from broader public by moving their own faces. Putting face masks helps to prevent Covid-19 transmission, according to someone. Covid-19 is most modern epidemic virus to poke human well-being in twentieth centenary. The quick spread of Covid-19 has prompted WHO to announce it an international pandemic in 2020. Covid-19 infected more than 5 million patients in 188 countries in less than six months. Close interaction is how the virus is spreading and in densely populated places. The corona virus outbreak has resulted in unprecedented levels of international scientific cooperation. Machine learning and deep learning powered by computer science will aid in fight for Covid-19 in number of different ways. Machine learning analyses massive amounts of data to estimate Covid-19 dispersion, behaves as a short notification system for prospective pandemics, and categorize sensitive populations. Many countries have regulations requiring people to keep face masks in crowd.

As a result, detecting face masks is a difficult process. Because of the expansion of the corona virus sickness, it has gotten a lot of attention lately since numerous countries have adopted policies such as" N o admission without a mask." Face mask detection is a critical issue in security and the prevention of Covid-19. In the medical industry, a mask minimizes the danger of infection from an infected individual, whether they show symptoms. Face mask detection is employed in a variety of settings, including airports, hospitals, offices, and educational institutions. Face recognition with not a mask is simpler but faces recognition with just a mask is more difficult since masked face feature extraction is more difficult than conventional face feature extraction. Many facial characteristics, such as the nose, lips, and chin, are missing from the covered face. The figure 1 shows the architecture model for Face Mask Detection. There are eight steps are used for the detection of face with mask.



Techniques for Face Mask Detection

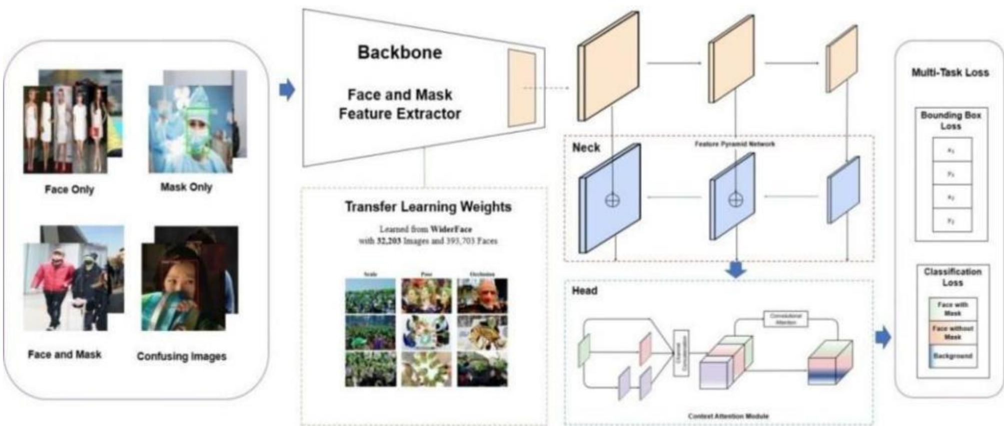
Convolutional Neural Networks (CNNs) are a type of deep neural network motivated by bio- logical phenomena. A CNN is composed of several components, including one with convolutional layer, pooling layer, as well as then fully connected layer, and it learns the spatial patterns of data autonomously and fluidly using the backpropagation method. The CNN kernels are common across entire image positions, making it incredibly parameter efficient. The CNN is a strong option for computer vision problems because of these properties. Because of major advancements in GPU com- punter capability, deep learning technologies have blossomed in recent years. Throughout computer vision, object recognition seems to be a critical task that has attracted a lot of attention. According to the recommended recommendations and tactics for improvement, shows us the current object detection methods:



Object Detection Methods

2.1. Multi-Stage Detectors

The two-stage method uses a heuristic algorithm as CNN to generate a huge number of region recommendations for every image, next classifies and stagnates these eligible provinces. As first object detection technique, deep learning was applied. The figure shows the model architecture of retina face detection. There are many stages in this detection for the detection of masked face is shown in the figure.



Architecture of Retina Facemask

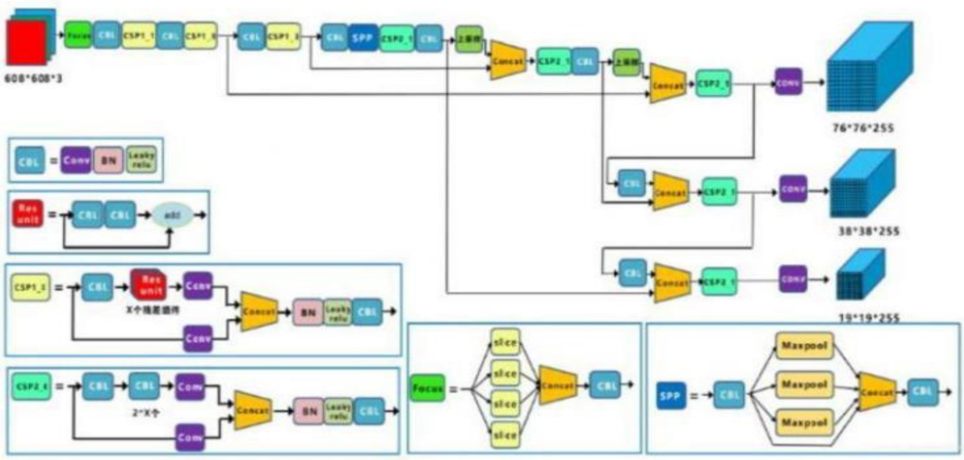
2.2. ***Single Stage Detection***

Classification as well as regression are completed in single shot employing consistent and dense sampling having regard to locations, sizes, and aspect ratios in the one-stage technique:

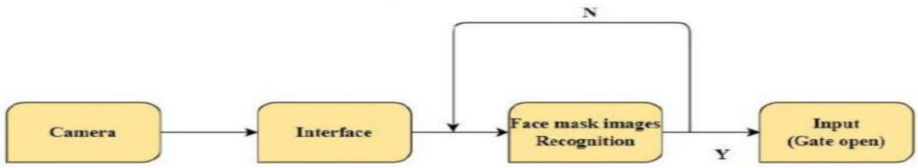
FCOS is a one-stage detector that is both anchor-free as well as proposal- free. In tests, FCOS outperforms common anchor-depending one-stage detectors such as Retina Net, YOLO, and SSD, yet with far less design complication. FCOS entirely skips all anchor box calculation and hyper-parameters, solving object detection in a per-pixel predictions manner, identical to those other dense prediction problems like semantic segmentation. Between one detector, FCOS also reaches state-of-the-art effectiveness. We also demonstrate that FCOS can be employed as RPNs in the two-stage Faster R-CNN detector and surpasses its RPNs by a significant margin.

Optimized algorithm to improve efficiency of object detection as the accuracy of the one-stage detector often lags behind of the two -stage detector, the model was trained on VOC 2007 and 2012 train Val with total of 16.551 pictures, for enhancing part the data flipped left and right and random sampling is used. The results show that the one-stage detector gets high accuracy on SSD. By comparing the results, it can draw that first the improved algorithm for objects with similar categories has higher detection accuracy and reduce false detection, second improved algorithm has higher detection confidence for the same object detected.

To effectively apply YOLOV5 revealed that the network structure illustration, the most effective objection detection technique at the moment, in the real world, particularly in the guidance of donning masks in crowded locations, it was suggested to substitute manual inspection with only a deep learning approach and utilize YOLOV5 figure 10 exhibited the network structure illustration, the most influential objection detection technique at the moment. Depicts the system. When visitors enter the mall, they will capture images with the camera, which will then be transferred to interact for face mask identification. The mall gate would be opened and exhibited to pass if face detected within two seconds is indeed a face have a mask; else, it will be reverted to face mask identification until success is achieved. The experimental findings suggest that the suggested algorithm can efficiently recognize face masks and enable staff surveillance.



YOLO V5 Network



When contrasted to Faster R-CNN, YOLO has weaker recall and greater error in localization, fails to detect close items because every grid could only suggest two bounding boxes, and suffers to identify small objects.

Table 7: Summary of reviewed papers

Research Number	Method	Algorithm	Advantages	Disadvantages	Datasets	Results
S.Shivaprasad, 2021 [24]	TenserFlow, OpenCV with Keras modules are utilised to create CNNs.	Algorithm-1: pre-processing and dataset training Algorithm 2: deployment process	Got high accuracy by collecting data from different sources	Issues when applied system with real time	Kaggles mask dataset PrajnaBhandary dataset	The accuracy is 0.96, and the F1-score is 0.92.
Prathmesh Deval, 2021 [6]	proposed system uses CNN and OpenCV	The Tensor Flow and Keras algorithm	1.Proposed system can be implemented in many places. 2. system work in real-time,helps in reducing the transmission rate	need more research to ensure that the person put make on their faces or something else.	Kaggle dataset	Utilizes profound learning strategies in distinctive facial acknowledgment and perceive if the individual is wearing a facemask or not.
Jansi Rani Sella Veluswami 2021 [23]	proposed methodology can be split into two phasesTrainingphaseDeployment Phase	SSD (Single Shot-Multibox Detector) CNN	1. With many datasets testing getting high accuracy 2. Model will help government agencies and health officials fight the global pandemic	Need more research to apply with live video or with security camera	A mixture of 4 datasets was used from kaggle	High accuracy of ~96% and ~99% on two separate testing datasets
Chenchen Zhu 2017 [31]	The Multi-Scaled Region Proposal Network (MS-RPN) was used to produce a set of regional candidates, while CMS-CNN was used to estimate facial region candidates.	Contextual Multi-Scaled Regional-based Convolution Neural Networks (CMS-RCNN)	1.For dealing with small face regions, multi-scaled content is grouped including both regional proposal and RoI detection. 2.Inspired by the intuitive of the human vision system, our suggested network provides intuitive body contextual reasoning.	Further research is needed to enhance fully joint training such that network may be trained from beginning to end.	WIDER FACE FDDB	1.The proposed strategy outperforms powerful baselines on the WIDER FACE Datasets by a significant margin. 2. In comparison to contemporary state-of-the-art face detection algorithms, obtains competitive performance on FDDB.
K Wang,2016 [26]	By using four stages L-Cls-NetL-Cal-NetH-Cls-Net H-Cal-NetFR-Net	deep cascade convolutionalnetwork that usesFast RCNN	1.Higher recall of 91.87% on thechallenging FDDB benchmark 2.outperforming the state-of-the-art methods	By using non-maximum suppressing after L-Cal-Net, you can go very slowly.	FDDB,AFW	Cascade achieves a recall rate of 88.77 percent at 167 errors on FDDB. The recall rate of CNN is 85.67 percent.
Rongqiang Qian, 2016 [18]	Combined MSERs and Edge Boxes	1.Hybridregionproposal method 2.Fast R-CNN	1.A hybrid region suggestion approach that takes into consideration colour and edge complementary information. 2.Rapid R-CNN for classification with bounding box regression at the same time.	Selective searches not adopted in system	Collected road surface traffic sign	The entireAverage Precision (AP) is near about 85:580%.
Qihang Wang 2018 [27]	Three algorithms were employed to find a candidate region of such face in the image.	1.CNN algorithm, 2. The Haar-Adaboost algorithm 3. Candidate search algorithm	1.To retrieve the feature autonomously, this method efficiently use deep convolution network. 2. Eliminates classic face detection model's reliance on manual attributes.	1.The requirement for a high number of samples in order to construct 2. The use of the selective search object concept extraction procedure takes longer. 3. Fast R-CNN cannot be used to create a complete detection method.	Labeled Faces in the Wild LFW	1. a better degree of accuracy in identifying 2. Shorten the time it takes to detect something.
Lin Jiang 2021 [11]	Human faces may be detected automatically using an advanced deep learning algorithm relying on machine vision.	Multiscaled fast RCNN methods based on upper and lower layers (UPL-RCNN)	High precision Time consumption is reduced. There is no correlation mark.	1.Not only this approach effectively recognise human faces, but it can also do it quickly. 2.not yet reached the optimal time consumption	WIDER FACE subway station data set	When compared to the Faster-RCNN model, the UPL-RCNN model exhibits a 16.2%increase.
Huaizu Jiang 2017 [12]	assessments of region ideas, as well as the top face's end-to-end effectiveness detectors.	Faster R-CNN	Designed to detect a wide range of objects. Multiple convolutional layers can be used within an RPN without adding to the computational load.	Additional research is needed into the unique features of human faces.	WIDER FACE (FDDB) IJB-A benchmark	The regional proposal network (RPN) module is responsible for the module's performance.
Wenqi Wu 2018 [28]	method of face detection based on Faster R-CNN	different scales face detector (DSFD)	face proposals of exceptional quality. A three-network Fast R-CNN of the parallel kind is also presented.	-efficiency of parallel-type fast R-CNN -efficiency of parallel-type fast R-CNN -needs further research to increase less weight parameters as well as running speed	FDDB, AFW, PASCAL faces, and WIDER FACE.	The experimental outcomes show that the suggested DSFD approach performs well on common benchmarks.
Mosab Rezaei 2019 [22]	Three different types of image deterioration have been studied: SSD and Faster R-CNN	Faster R-CNN SSD	R-CNN that is quicker is more resistant to Gaussian blur. SSD is much more resistant to JPEG2000 compressed images of lower quality.	The edges are much more noticeable on SSD.	WIDER FACE	For fuzzy pictures, the faster R-CNN is the better option. For JPEG compressed images, SSD is a better structure.
Kaiming He [9]	Faster R-CNN is extended by adding a stream for object mask prediction in parallel with the preexisting stream for bounding box recognition.	Mask R-CNN	Simple to train Adds only a small overhead to Faster R-CNN Running at 5 fps	Need more research for optimization	COCO key point	On each and every task, Mask R-CNN surpasses all other single-model entries, such as the COCO 2016 challenge winners.
Kaihan Lin [13]	Mask R-CNN-based segmentation algorithm with Generalized Intersection over Union (GIoU).	G-Mask	Improve your results in multi-scale face detection challenges. A fresh face dataset includes annotation metadata for segmentation	To increase the model's performance even more, expand the existing dataset. Enhance detection accuracy by optimising the model's architecture.	The 5115 images were randomly chosen from the FDDB and ChokePoint datasets and labeled in a new dataset.	When compared to Faster R-CNN and the classic Mask R-CNN approach, the new G-Mask method exhibits promising face identification performance.
Kaihan Lin [14] R-CNN	Based on enhanced Mask R-CNN, a face detection and segmentation approach has been developed.	G-Mask	Face detection and segmentation are combined into a single framework. To enhance detection performance, GIoU is employed as that of the bounding box loss function.	Improving the speed of the proposed method	FDDB, AFW, and WIDER FACE	Upon that FDDB, AFW, and WIDER FACE benchmarks, the suggested G-Mask technique has shown impressive outcomes.
Zhi Tian 2019 [25]	anchor-free and proposal-free onestage detector FCOS	FCOS	FCOS totally avoids the time-consuming calculations associated with anchor boxes. All hyper-parameters relating to anchor boxes should be avoided. being far less complicated	for every level, restrict the bounding box regression's scope.	MS-COCO dataset	With single-model as well as single-scale tests, 44.7% in AP were achieved.
Hongtao, W. 2020 [10]	Maps output from SSD box to six convolutional layers and the center point will generate two square boxes of different sizes	One-stage detectors and SSD	Good effectiveness for new method. Private parameters are fewer	Not very effective in detecting small targets.	PASCAL VOC	The results show that the module improve detection accuracy while ensuring real-time detection speed.
Guanhao Yang2020 [30]	YOLOV5 is used in this application to recognise faces in the stock, whether they are wearing masks or not, if success, can enter the gate open.	deep learning YOLOV5	1. Recognize persons entering the store, whether they are wearing masks or not. 2. Experiment does have a success percentage of approximately 97.9%. 3. There is no longer a need for human crowd control.	It would not be recognised if the buyer covers part of such mask using his hand.	AIZOOTe team's Face Mask Detection	The suggested algorithm can accurately distinguish face masks and achieve effective staff surveillance, according to the results.
Xing, C 2020 [29]	Modified network based on YOLOv3-Tiny for water conservancy facility inspection	Network combined residual network and YOLOv3 Tiny	Good performance on small objects. Better in five types of chosen targets than YOLOv3-Tiny within lower speed	The proposed network's performance on small targets is quite poor	All images are token by camera mounted on flying drones during the daytime	The proposed network gets 2.9 higher mAP than YOLOv3-Tiny.

3. Conclusion

In this paper, we have discussed some research papers about facial mask detection. As we know nowadays mask detection is a very challenging task. The applications of Facial Mask Detection are used especially for the prevention of spreading Corona Virus, tracking and identifying criminals and anti-spoofing etc. 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. 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 YOLO algorithm give better results with more accuracy and are more successful in real life.

References

1. M.S.H. Al-Tamimi, Combining convolutional neural networks and slantlet transform for an effective image re- trieval scheme, Int. J. Electr. Comput. Eng. 9(5) (2019) 4382–4395.
2. M.S.H. Al-Tamimi, I.A. Abdulmunem and S.S. Sulaiman, Improved merging multi convolutional neural networks framework of image indexing and retrieval, Int. J. Adv. Sci. Tech. 29(8) (2020) 1884–1901.
3. M. Dasgupta, O. Bandyopadhyay and S. Chatterji, Automated helmet detection for multiple motorcycle riders using CNN, IEEE Conf. Inf. Commun. Technol. Allahabad, India, 2019, p. 1–4.
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) 196–200.
5. M.S. Ejaz, M.R. Islam, M. Sifatullah and A. Sarker, Implementation of principle component analysis on masked and non-masked face recognition, IEEE 1st Int. Conf. Adv. Sci. Engin. Robotics Technol. 2019, pp.1–5.
6. R. Girshick, Fast R-CNN, Proc. IEEE Int. Conf. Comput. Vision 2015, pp. 1440–1448.
7. K. He, G. Gkioxari, P. Doll´ar and R. Girshick, Mask R-CNN, Proc. IEEE Int. Conf. Comput. Vision 2017, pp. 2961–2969.
8. W. Hongtao and Y. Xi, Object detection method based on improved one-stage detector, 5th Int. Conf. Smart Grid Electric. Autom. (ICSGEA), IEEE, 2020, pp. 209–212.
9. Z. Liang, J. Shao, D. Zhang, L. Gao, Small object detection using deep feature pyramid networks, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 11166 LNCS, pp. 554–564, doi: 10.1007/978-3-030-00764-5_51.
10. Chen D, Hua G, Wen F, Sun J (2016) Supervised transformer network for efficient face detection. In: European conference on computer vision. Springer, pp 122–138.
11. Ejaz MS, Islam MR, Sifatullah M, Sarker A (2019) Implementation of principal component analysis on masked and non-masked face recognition. In: 2019 1St international conference on advances in science, engineering and robotics technology (ICASERT). IEEE, pp 1–5.
12. Karen S, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition.
13. K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, in: Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.
14. J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, ImageNet: A large-scale hierarchical image database, 2010, doi: 10.1109/cvpr.2009.5206848.
15. Khan N, Yaqoob I, Hashem IAT, Inayat Z, Ali WKM, Alam M, Shiraz M, Gani A (2014) Big data: survey, technologies, opportunities, and challenges. Scientif World J, 2014.