EasyChair Preprint

№ 9188

Real-Time Face Mask Detection by Clever Convolution Neural Networks

Suresh Kumar Kanaparthi, Sai Pratheeka Kaluvala, Vyshnavi Ravula, Jyothika Sai Masimukkula and Sri Indu Dekkapati

October 29, 2022

Real-time Face Mask Detection by Clever Convolution Neural Networks

Suresh Kumar Kanaparthi
Computer Science Engineering – Data Science
Malla Reddy University
Hyderabad, India.
sureshkonline@gmail.com

Sai Pratheeka Kaluvala
Computer Science Engineering
GokarajuRangaraju Institute of Engineering and Technology
Hyderabad, India.
saipratheeka70@gmail.com

Vyshnavi Ravula
Computer Science Engineering
GokarajuRangaraju Institute of Engineering and Technology
Hyderabad, India.
ravulavaishnavi2001@gmail.com

Jyothika Sai Masimukkula
Computer Science Engineering
GokarajuRangaraju Institute of Engineering and Technology
Hyderabad, India.
jyothikasai999@gmail.com

Sri Indu Dekkapati
Computer Science Engineering
GokarajuRangaraju Institute of Engineering and Technology
Hyderabad, India.
sriindu6701@gmail.com

Abstract:

Face Mask Detection is the process of determining if a mask is being worn by a person or not. Our paper's purpose is to determine whether someone is wearing a mask properly or not. When a mask entirely covers a person's mouth and nose, he is wearing it correctly. Conferring to the current works, relatively little study has been done to identify masks over faces. There by, the goal of our work is to develop a technology that can detect masks over the face in public facilities in order to avert the spread of COVID-19 accurately and thus contribute to public welfare. This paper proposes a simpler approach to accomplish this goal utilizing some libraries such as TensorFlow, Keras, etc. We investigate optimum parameter values utilizing the YOLO (You Only Look Once) technique. COVID-19 pandemic has a significant impact on the society, causing global trade and transportation to be disrupted.

Keywords- YOLO, Object detection, Pre-processing, Image segmentation, Data augmentation.

I. INTRODUCTION

COVID-19 pandemic has a significant impact on people's lives, leading global trade and transportation to be affected. During this pandemic, wearing a mask is a necessary part of our lives. This is the most efficient way to avoid COVID-19. Many nations believe that the best way to ensure human safety is to wear a mask. Those who do not follow the rules, however, still exist. When they leave the house, they do not put on a mask. A person whose face is not covered with a mask can be identified using face mask detection.

Face Recognition is a method of identifying an individual depending upon different features of their face by comparing stored representations of each human face in a cluster of people. Face recognition is a natural method for identifying and verify persons. In every sector or institution, a person's safety and authentication are important. As a result, computerized facial recognition employing computers or devices for identity verification 24 hours a day, 7 days a week, and even remotely is gaining popularity in today's society. In pattern recognition and image processing, face identification has emerged as one of the most difficult and fascinating challenges.

Masks have been widely used as one of the most effective ways to protect against the virus. In addition, in many cities across our country, curfews without a mask have been established. Currently, face mask detection systems have become a critical duty, although there has been little research on the subject in the literature.

COVID-19 vaccinations have started to reach the market, although they are not accessible in all parts of the world. So, until this deadly virus is totally eradicated, wearing face masks on a day-to-day basis is a critical practise that will help in the prevention of contagion and keep people safe from infectious germs. When someone coughs, talks, or sneezes, the microbes of the deadly virus are released into the air, which might infect the people around.

Face mask detection in this paper was accomplished using the YOLO object detection algorithm. The newly created YOLO method is employed for face mask identification in this article, which makes it unique. In real-time face mask detection, YOLO delivers better accuracy. While it has been determined in previous research if only people are wearing masks or not, one of the contributions of this paper to the literature is that it was also determined whether someone is wearing a mask properly or not.

Travellers who do not wear masks in the airport or onboard risk having their United flying privileges cancelled. The police will impose a fine if a face mask is not worn in malls and supermarkets. Face Mask Detection is an artificial intelligence research. We can tell whether someone is wearing a mask or not. There are two stages to the creation of this work. To train a model to recognizes face masks in photos, you can use convolution or any other pre-trained model. Then, using our trained model, identify faces in video or photos and get an estimation.

Our application helps hospitals keep track of employees who do not wear masks during their shifts. It also keeps track of whether patients who are admitted to the hospital are wearing masks or not. It can also be used in the workplace to see if individuals are wearing masks. Many of these industries are attempting to systematize the detection of such problems, lowering the period of time and manpower required.

This work was built with Keras and TensorFlow, with the model trained for face mask recognition and YOLO Object Detection used to check if an individual's face is covered with a mask or not.

The following is how the rest of the paper is organized: The introduction to our work is provided in the first section. The related work is discussed in Section 2. The methodology, dataset and implementation of our work are detailed in Section 3. Conclusion and future scope of work are discussed in Section 4 and 5 respectively. References are covered in section 6.

II. RELATED WORK

Principal Component Analysis (PCA) is used by the authors of [1] to discern the faces hidden behind the mask. This practice is considered crucial in any security measure. One of a handful of the works centers around distinguishing human faces hidden under masks. The study examines the precision of distinguishing non-masked and masked faces utilizing PCA to recognize an individual. It has been established that non-masked face recognition rate using PCA is comparatively higher than the other one. Be that as it may, when the picture of a masked individual is given as an input, recognition rate is considerably low. Feature Extraction from a face without a mask is far more strenuous than a face with one. The face mask detection rate decreased because of the fact that there are lot more features on an unmasked face. The accuracy of detecting human faces dropped by 70% when a mask was used. At last, PCA which is a standard statistical method is prevalent for traditional face detection. In spite of that, it hasn't produced significant results when masked face detection is performed. Hence, the spotlight from now on will be utilizing progressed ML techniques to enhance the accuracy of masked face recognition.

The authors of [2] have developed a way to determine how a person uses a mask. There are three types of mask wearing conditions which are: Incorrect wearing of a mask, proper wearing of a mask, and not wearing a mask were identified as three types of mask wearing conditions. Finally, SRCNet surpassed standard end-to-end image classification algorithms by more than 1.5% in kappa, achieving an accuracy of 98.70%. Although the proposed SRCNet can achieve high accuracy in recognizing mask-wearing conditions that are important for the prevention of public epidemics such as COVID19, their research still has some limit. First, the medical mask dataset used to determine mask wearing conditions is minimal and cannot cover all poses or situations. In addition, because the dataset lacks video, it is not possible to evaluate the recognition results on the video stream. The algorithm's suggested recognition time for a single frame is a bit long, with an average of 10 frames detected per second, well below the video's standard frame rate of 2 frames per second (fps). The authors of [3] address the challenge of deep face recognition (FR) by using an open protocol in which similar facial features are predicted to have a small maximum intraocular distance. smaller than the minimum distance between classes in a suitably chosen metric space. However, only some existing algorithms can satisfy this condition. To achieve this goal, they proposed SoftMax angular attenuation (A SoftMax), which allows CNN to learn angular discriminant features. The ASoftMax loss can be seen as imposing discriminant restrictions on a super spherical manifold, corresponding to a priori that the faces are also geometrically located on a manifold.

Researchers from [4] used a cumulative neural network, cross-entropy loss combined with SoftMax is undoubtedly one of the most widely used monitoring (CNN) components. Despite its ease of use, popularity, and high performance, this component does not specifically support learning feature discrimination. They proposed a maximum loss of large amplitude expansion (LSoftmax) in [4], which directly promotes intra-glass compactness and interlayer separation between learned features. . Moreover, LSoftmax can not only adjust the appropriate amplitude, but also avoid over-equipping. They also demonstrate that standard stochastic gradient descent can be used to optimize LSoftmax loss. Extensive testing on four benchmark datasets shows that Lsoftmax's deep learning features that lose points become more discriminatory, significantly improving performance across a wide range of classification and visual verification tasks. Khandelwal et al. proposed a study [5] using computer vision-based object detection models to analyse masked faces and socially distracting behavioural violations using video from security cameras. This solution is designed primarily for industrial environments. For masked face recognition, a two-step method was used. The images were first processed using the MobileNetV2 face recognition model [6]. Then, a binary mask classifier is used to classify the faces as masked or unmasked. Initial data collection is used to train the model. The authors apply SSD to detect class of people because of social distancing. The authors have developed a method to select four points that form a rectangle and perform perspective transformations to measure distances on a single plane. The absolute distance between two points must be provided when comparing these lengths with the threshold. This is possible for a manufacturing plant or limited space, but it would be costly and time consuming for any public place. It is important to emphasize that these two models are distinct and no combined solution has been proposed. "Subject Removal with Interactive Microphones in Facial Imagery" is presented by the authors of [7]. Their goal was to remove the microphone object from the face image and replace it with precise face semantics and fine lines. They demonstrated MRGAN, an interactive technique. On real microphone images, MRGAN outperforms state-of-the-art image-changing algorithms, according to the review.

The authors of [8] have built a masked, no-masked and half-masked model of 95accurately. The authors of [9] analysed bids at their databases and published conclusions. They mentioned the new database used to check the next stage of face recognition algorithms. Researchers of [10] CNN analysed, accused the remaining network (resnnet50) and the Boltzmann machines discriminated, as well as supervised and unattended classifications to improve the effectiveness of Automatic exudate identification. Resety50 with vector support machines have exceeded other networks with accuracy and sensitivity respectively 98% and 0.99, according to conclusions. To diagnose errors, new monks (resnet50) with 51 layers of mute depths are proposed in this study [11]. The authors of [12] presented a remaining learning framework to facilitate the formation of very deep networks that the networks used earlier. They reached a remarkable 28% charm about the Coco object identification data set because of their deeply profound performances. The writers of this article [13] have proposed a system that incorporates extracts of Neuron network functions concluded with standard engine algorithms such as support vectors, gradient and random stimulating machines course. Hybrid models are more effective than standard models when trained from raw pixel values, depending on the results. In this search [14], the characteristics of an image are taken by an erotic nerve cell network and the idea of deep learning. For different purposes, additional classification methods are implemented. This work [15] introduced a dataset to detect occlusal faces with high accuracy. They also recommend LLECNNs for masked face detection based on this dataset. Tests on the MAFA dataset show that the proposed method beats the top six approaches by at least 15.6%.

III. PROPOSED ARCHITECTURE

One of the fascinating parts of Computer vision and profound learning is object recognition. Object location not set in stone in an assortment of computer vision applications, including object following, recovery, video reconnaissance, picture inscribing, picture division, Medical Imagine, and a large number of others. Object identification is a profound learning framework that permits things like individuals, structures, and vehicles to be perceived as items in pictures and films [15]. In image classification, we may just categorize(classify) assuming there is an article in the picture or not as far as the probability, but rather object discovery is to perceive the object with bounding box in the Image (probability). There is a lot of things to examine when it comes to image classification, and you may have implemented it all at least once through a tutorial. Image classification is the challenge of classifying the class that corresponds to an image when it is used as an input.

Redmond et al proposed an algorithm, You Only Look Once (YOLO) in an examination study which is distributed as a paper in the conference at the IEEE/CVF Conference on Pattern Recognition and Computer Vision (CVPR). This has won the OpenCV People's Choice Award. YOLO states the utilization of a start to

finish neural network which gives expectations of class likelihood and bounding boxes at the same time, rather than the technique that is used by object identification algorithms before YOLO, which reused classifiers to perform detection. YOLO delivers cutting edge outcomes by outperforming previous object detection algorithms in real-time by an impressive edge by adopting in a general sense new strategy to object recognition.

YOLO is popular because of its high level of precision and ability to run in real-time. Only one step forward is taken using YOLO image processing and YOLO object tracking. The detected objects and bounding boxes are output after non-max suppression, which ensures that the object detection algorithm only recognises each object once Utilizing YOLO, a solitary CNN predicts a few bounding boxes and their class probabilities. By training on full pictures, YOLO enhances detection performance. The YOLO method is used for object recognition and tracking. Unlike prior frameworks that looked at various portions of the image multiple times to detect objects, this one just looks at the same part of the image onceIt just glances at the whole picture once and examines the network once to recognize things. This algorithm has gained a lot of popularity due to its speed and precision. The residual blocks, bounding box regression, and intersection over union (IOU) are the three tactics which YOLO algorithm employs.

Residual blocks: The image is initially partitioned into different grids. The size of every partition is S × S.

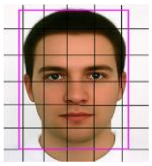


Fig. 1. Image divided into S × S blocks.

Bounding box:

Fig. 2. Picture depicts a bounding box.

$$y = (pc, bx, by, bh, bw, c) \quad (1)$$

Bounding box is a diagram that features an object in a photo. To anticipate the width, height, centre, and class of items, YOLO utilize a solitary bounding box regression [16]. Each bounding box in the picture has the following boundaries: Centre for Bounding Boxes, width *(bw)*, height *(bh)*, and class (for instance, individual, vehicle, face, and so forth) Class is addressed by c.

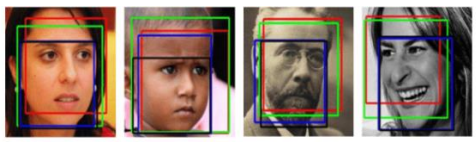


Fig. 3. Different images with multiple bounding boxes.

The idea of intersection over union (IOU) outlines how these boxes cross-over in the object identification. IOU is utilized by YOLO to make a result box that appropriately encompasses the things. The bounding boxes and their confidence scores are anticipated by every matrix cell. Assuming that the expected and real bounding boxes are indistinguishable, the IOU is 1. This approach eliminates bounding boxes that aren't the similar size as the real box [16]. We use Intersection over Union to decide if the expected bounding box is giving us a decent outcome. It computes the intersection of the actual bounding box and the predicted bounding box over their association.

IoU = Area of Yellow Box/Area of Green Box

We can express that the forecast is adequate assuming IoU is greater than 0.5. We picked 0.5 as an arbitrary threshold, yet it tends to be altered to accommodate your individual issue. Instinctively, the higher the threshold, the more precise the prediction become.

YOLO predicts various boxes, each containing a solitary object, using a remarkable neural network that utilizes the properties of the full picture. All of this happens at the same time. To do this the image is segmented into S × S sections. Then, assuming the object's Centre is in one of these areas, that area is responsible for recognizing the object. Every cell in this lattice is responsible for foreseeing 'B' boxes, every one of which contains an object and a score demonstrating the degree of confidence in the object contained in the box. This score should be zero in the event that there are no objects in the cell. Assuming that an item is available in the cell, the score will be equivalent to the intersection over union (IoU) of the predicted box and the image's ground truth.

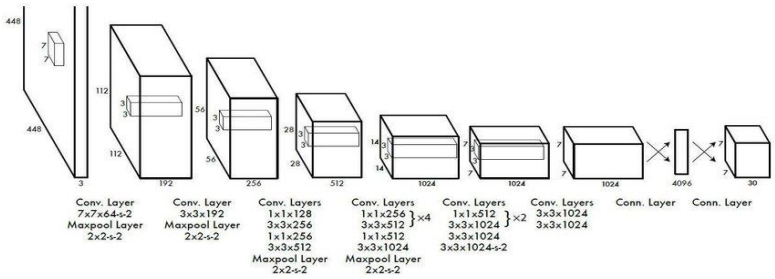


Fig. 4. The above picture depicts the architecture of YOLO.

The YOLO architecture was influenced by Google Net's picture classification model. There are 24 convolutional layers in the YOLO network, trailed by two completely connected layers. It also has 1 × 1 convolutional layers that alternate, reducing the feature spaces from previous layers. The convolution layers used in YOLO are taken from the ImageNet task's pre-trained model, sampled at half the resolution (244 × 244) and then doubled. All of the layers in YOLO are leaky ReLu, with the last levels using a linear activation function.

The architecture shown below gives a decent understanding of how the implementation works. Other library functions in this work, such as OpenCV, Keras, Imutils, and TensorFlow, are also relevant. The architecture is a complex model that hides the source code beneath. The figure below aids us in determining the complexity of the application and its outcomes.

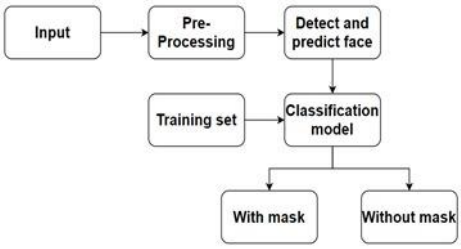
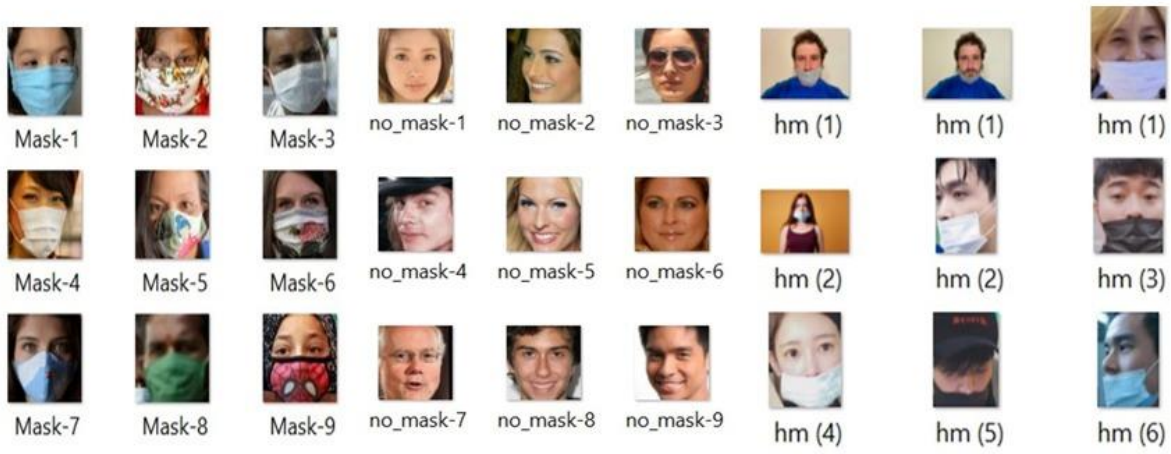


Fig.5. Proposed architecture.

3.1 Dataset

Our dataset consists of two categories of images: people with masks and people without masks. We have collected our dataset from the internet. Some of the images of our dataset are as follows:



3.2 Experimental results

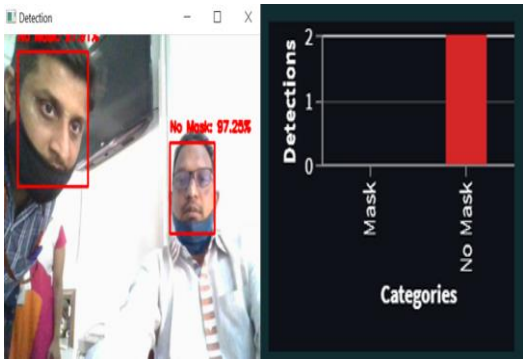


Fig. 9. The above picture consists of a frame where there are two people wearing a mask, but not entirely covering their nose and mouth region, hence the model identifies that the two people detected in the frame are not wearing the mask properly, hence it is displayed as no mask, and a corresponding graph is displayed.

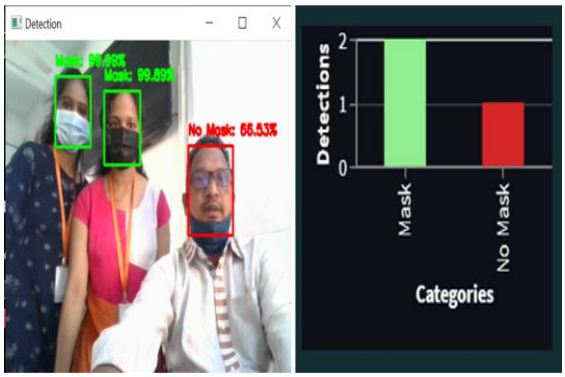


Fig. 10. The above picture consists of a frame which consists of three people, out of which two people are wearing a mask, while one person not wearing the mask properly, hence the model identifies that the two out of three people detected in the frame are wearing a mask, and the corresponding graphs shows that there are two people in the frame with a mask and one person who is not wearing a mask (not wearing the mask properly).

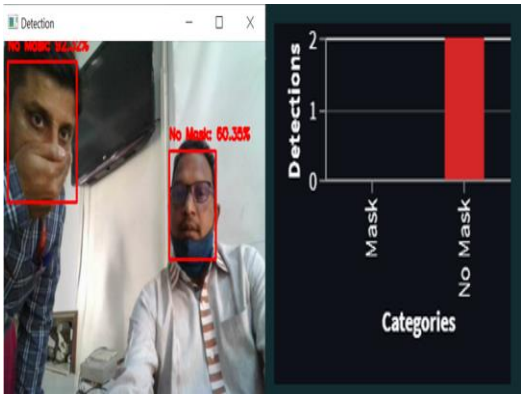


Fig. 11. The above picture consists of a frame where a person covers his face with a hand, while the other person not covering the entire nose and mouth region, hence the model detects that the people present in the frame are not wearing a mask, and a corresponding graph is displayed.

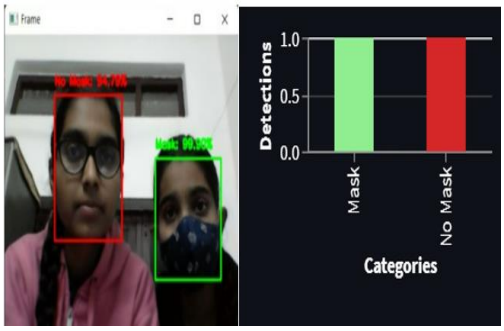


Fig. 13. The above picture depicts the result where the model identifies that one of the persons in the frame is wearing a mask, while the other doesn't. Hence the corresponding result is displayed over the faces and a graph of the result is also displayed.

IV. CONCLUSION

In our work, face mask detection was presented which was able to detect whether a mask is worn by a person or not. Precautions should be taken to avert the spread of COVID-19. The foremost measure being wearing a mask. Face masks have lately been mandatory in more than fifty nations throughout the world. In public places such as supermarkets, public transportation, offices, and stores, people must conceal their faces. Retailers frequently utilize software to track the number of individuals who enter their businesses. Our work focuses on this detection. This detector can be used in various locations, which includes airports, supermarkets, and other crowded areas, to supervise the public and prevent the spread of the virus by keeping an eye on who is following basic norms and who is not. Hence, the model which we've trained can effectively classify people into two categories: those whose face is covered with a mask and those whose doesn't and display the results accordingly.

V. FUTURE SCOPE

To begin with, the proposed method can be implemented in any high-resolution video surveillance system. The proposed method can be successfully used to perform face mask detection, according to the results of the experiments. It is a real-time software application that can be used in high-resolution video surveillance system, public places such as supermarkets, airports, and other places where wearing a mask is mandatory. The software may be extended to work with additional IOT devices to deny entry for the people who doesn't follow the norm. Furthermore, we emphasize that our work works on devices with minimal computational capability and can process photos and video streams in real time, making our solution practical in the real world. We can also assign hardware such as speakers which can produce an alarm if a mask is not worn, and alerts the person.

References

[1] M. S. Ejaz, M. R. Islam, M. Sifatullah and A. Sarker. "Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition.", (May2019):1-5.https://doi.org/10.1109/ICASERT.2019.8934543.

[2]BoshengQinandDongxiaoLi.“Identifying Facemask-wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID- 19.” (May 2020), [online] Available: https://doi.org/10.21203/rs.3.rs-28668/v1.

[3] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. “Sphereface: Deep hypersphere embedding for face recognition.” (July 2017):13.https://arxiv.org/abs/1704.08063v4.

[4] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-margin softmax loss for convolutional neural networks.” (November 2016).https://arxiv.org/abs/1612.02295v4.

[5] A. T. Tran, T. Hassner, I. Masi, and G. Medioni, “Regressing Robust and Discriminative 3D Morphable Models with a Very Deep Neural Network.” (July 2017).https://arxiv.org/abs/1612.04904v1.

[6] Prateek Khandelwal, Anuj Khandelwal, Snigdha Agarwal, Deep Thomas, Naveen Xavier and Arun Raghuraman. “Using Computer Vision to enhance Safety of Workforce in Manufacturing in a Post COVID World.”, (May 2020):6.https://arxiv.org/abs/2005.05287.

[7] M.K.J. Khan, N. Ud Din, S. Bae, J. Yi/“Interactive removal of microphone object in facial images.” (October 2019):1115. http://dx.doi.org/10.3390/electronics8101115

[8] Z. Wang, et al. “Masked face recognition dataset and application.”(March 2020).https://doi.org/10.48550/arXiv.2003.09093

[9]Erik Learned-Miller, Gary B. Huang, AruniRoyChowdhury, Haoxiang Li, G. Hua. “Labeled faces in the wild: a survey.”, M. Kawulok, M.E. Celebi, B. Smolka (Eds.), Advances in Face Detection and Facial Image Analysis, Springer International Publishing, Cham (April 2016):189-248. http://dx.doi.org/10.1007/978-3-319-25958-1_8.

[10]Khojasteh,Parham; Aparecido,Leandro;Passos Júnior, Leandro; Carvalho, Tiago; Rezende, Edmar; Aliahmad, Behzad; Papa, João; Kumar, Dinesh. “Exudate Detection in Fundus Images Using Deeply-learnable Features.”(October 2018). https://www.researchgate.net/publication/328575488_Exudate_Detection_in_Fundus_Images_Using_Deeply-learnable_Features.

[11]Long Wen, Xinyu Li, LiangGao.“A transfer convolutional neural network for fault diagnosis based on ResNet-50.”, Neural Comput. Appl., 32 (10) (February 2019): 6111-6124.https://doi.org/10.1007/s00521-019-04097-w.

[12]Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. “Deep residual learning for image recognition.” 2016 EEE Conference on Computer Vision and Pattern Recognition, (June 2016): 770- 778. https://doi.org/10.1109/CVPR.2016.90

[13]AykutÇayir, IşilYenidoğan, HasanDağ.“Feature extraction based on deep learning for some traditional machine learning methods”,3rd International Conference on Computer Science and Engineering (UBMK), (September 2018):494–497. https://doi.org/10.1109/UBMK.2018.8566383

[14] M. Jogin, Mohana, M.S. Madhulika, G.D. Divya, R.K. Meghana, S. Apoorva.“Feature Extraction using Convolution Neural Networks (CNN) and Deep learning.”,3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), (May 2018): 2319–2323, https://doi.org/10.1109/RTEICT42901.2018.9012507

[15] Shiming Ge, Jia Li, Qiting Ye and Zhao Luo. "Detecting Masked Faces in the Wild with LLE-CNNs", (July 2017): 426-434. https://doi.org/10.1109/CVPR.2017.53.

[16] C. Li, R. Wang, J. Li, L. Fei.“Face detection based on YOLOv3.” in Recent Trends in Intelligent Computing, Communication and Devices, Singapore, (January 2020): 277– 284. http://dx.doi.org/10.1007/978-981-13-9406-5_34.