iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37726

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Realtime Face-Mask Detection on Raspberry Kit

Kalyan Chakravarthi. M[1], N Anil Chakravarthy[2]
[2](Assistant Professor), [1]Department of Computer Science and Engineering, Raghu Institute Of Technology, Visakhapatnam, AP, India

***Abstract:** Recognition from faces is a popular and significant technology in recent years. Face alterations and the presence of different masks make it too much challenging. In the real-world, when a person is uncooperative with the systems such as in video surveillance then masking is further common scenarios. For these masks, current face recognition performance degrades. Still, difficulties created by masks are usually disregarded. Face recognition is a promising area of applied computer vision . This technique is used to recognize a face or identify a person automatically from given images. In our daily life activates like, in a passport checking, smart door, access control, voter verification, criminal investigation, and many other purposes face recognition is widely used to authenticate a person correctly and automatically. Face recognition has gained much attention as a unique, reliable biometric recognition technology that makes it most popular than any other biometric technique likes password, pin, fingerprint, etc.*

Many of the governments across the world also interested in the face recognition system to secure public places such as parks, airports, bus stations, and railway stations, etc. Face recognition is one of the well-studied real-life problems. Excellent progress has been done against face recognition technology throughout the last years. The primary concern to this work is about facial masks, and especially to enhance the recognition accuracy of different masked faces. A feasible approach has been proposed that consists of first detecting the facial regions. The occluded face detection problem has been approached using Cascaded Convolutional Neural Network (CNN). Besides, its performance has been also evaluated within excessive facial masks and found attractive outcomes. Finally, a correlative study also made here for a better understanding.

I. INTRODUCTION

Rapid advancements in the fields of Science and Technology have led us to a stage where we are capable of achieving feats that seemed improbable a few decades ago. Technologies in fields like Machine Learning and Artificial Intelligence have made our lives easier and provide solutions to several complex problems in various areas.

Face mask detection refers to detect whether a person is wearing a mask or not. In fact, the problem is reverse engineering of face detection where the face is detected using different machine learning algorithms for the purpose of security, authentication and surveillance. Face detection is a key area in the field of Computer Vision and Pattern Recognition. A significant body of research has contributed sophisticated to algorithms for face detection in past. The primary research on face detection was done in 2001 using the design of handcraft feature and application of traditional machine learning algorithms to train effective classifiers for detection and recognition.

The problems encountered with this approach include high complexity in feature design and low detection accuracy. In recent years, face detection methods based on deep convolutional neural networks (CNN) have been widely developed to improve detection performance.

Modern Computer Vision algorithms are approaching human-level performance in visual perception tasks. From image classification to video analytics , Computer Vision has proven to be revolutionary aspect of modern technology. In a world battling against the Novel Corona-virus Disease (COVID-19) pandemic, technology has been a lifesaver. With the aid of technology, ‘work from home’ has substituted our normal work routines and has become a part of our daily lives. However, for some sectors, it is impossible to adapt to this new norm.

We propose a two-stage CNN architecture, where the first stage detects human faces, while the second stage uses a lightweight image classifier to classify the faces detected in the first stage as either ‘Mask’ or ‘No Mask’ faces and draws bounding boxes around them along with the detected class name.

This algorithm was extended to videos as well. The detected faces are then tracked between frames using an object tracking algorithm, which makes the detection robust to the noise . This system can then be integrated with an image or video capturing device like a CCTV camera, to track safety violations, promote the use of face masks, and ensure a safe working environment.

©IJRASET: All Rights are Reserved

II. RELATED WORKS

Initially researchers focused on edge and gray value of face image. It was based on pattern recognition model, having a prior information of the face model. Ada-boost was a good training classifier. The face detection technology got a breakthrough with the famous Viola Jones Detector , which greatly improved real time face detection. Viola Jones detector optimized the features of Haar , but failed to tackle the real world problems and was influenced by various factors like face brightness and face orientation. Viola Jones could only detect frontal well light faces. It failed to work well in dark conditions and with non-frontal images. These issues have made the independent researchers work on developing new face detection models based on deep learning, to have better results for the different facial conditions. We have developed our face detection model using Convolutional Neural Network (CNN), such that it can detect the face in any geometric condition frontal or non-frontal for that matter. Convolutional Neural Network have always been used for image classification tasks.

A. *Single-stage Detectors*

The single-stage detectors treat the detection of region proposals as a simple regression problem by taking the input image and learning the class probabilities and bounding box coordinates. Overfeat and Deep-Multi-Box were early examples. YOLO (You Only Look Once) popularized single-stage approach by demonstrating real-time predictions and achieving remarkable detection speed but suffered from low localization accuracy when compared with two-stage detectors; especially when small objects are taken into consideration. Basically, the YOLO network divides an image into a grid of size GxG, and each grid generates N predictions for bounding boxes. Each bounding box is limited to have only one class during the prediction, which restricts the network from finding smaller objects. Further, YOLO network was improved to YOLOv2 that included batch normalization, high-resolution classifier and anchor boxes. Furthermore, the development of YOLOv3 is built upon YOLOv2 with the addition of an improved backbone classifier, multi-sale prediction and a new network for feature extraction. Although, YOLOv3 is executed faster than Single-Shot Detector (SSD) but does not perform well in terms of classification accuracy. Moreover, YOLOv3 requires a large amount of computational power for inference, making it not suitable for embedded or mobile devices. Next, SSD networks have superior performance than YOLO due to small convolutional filters, multiple feature maps and prediction in multiple scales. The key difference between the two architectures is that YOLO utilizes two fully connected layers, whereas the SSD network uses convolutional layers of varying sizes. Besides, the RetinaNet proposed by Lin is also a single-stage object detector that uses featured image pyramid and focal loss to detect the dense objects in the image across multiple layers and achieves remarkable accuracy as well as speed comparable to two-stage detectors.

B. *Two-stage Detectors*

In contrast to single-stage detectors, two-stage detectors follow a long line of reasoning in computer vision for the prediction and classification of region proposals. They first predict proposals in an image and then apply a classifier to these regions to classify potential detection. Various two-stage region proposal models have been proposed in past by researchers. Region-based convolutional neural network also abbreviated as R-CNN described in 2014 by Ross Girshick et al. It may have been one of the first large-scale applications of CNN to the problem of object localization and recognition. The model was successfully demonstrated on benchmark datasets such as VOC-2012 and ILSVRC-2013 and produced state of art results. Basically, R-CNN applies a selective search algorithm to extract a set of object proposals at an initial stage and applies SVM (Support Vector Machine) classifier for predicting objects and related classes at later stage. Spatial pyramid pooling SPPNet (modifies R-CNN with an SPP layer) collects features from various region proposals and fed into a fully connected layer for classification. The capability of SPNN to compute feature maps of the entire image in a single-shot resulted in significant improvement in object detection speed by the magnitude of nearly 20 folds greater than R-CNN. Next, Fast R-CNN is an extension over R-CNN and SPPNet. It introduces a new layer named Region of Interest (RoI) pooling layer between shared convolutional layers to fine-tune the model. Moreover, it allows to simultaneously train a detector and regressor without altering the network configurations. Although Fast-R-CNN effectively integrates the benefits of R-CNN and SPPNet but still lacks in detection speed compared to single-stage detectors.

Further, Faster R-CNN is an amalgam of fast R-CNN and Region Proposal Network (RPN). It enables nearly cost-free region proposals by gradually integrating individual blocks (e.g. proposal detection, feature extraction and bounding box regression) of the object detection system in a single step. Although this integration leads to the accomplishment of break-through for the speed bottleneck of Fast R-CNN but there exists a computation redundancy at the subsequent detection stage. The Region-based Fully Convolutional Network (R-FCN) is the only model that allows complete backpropagation for training and inference.

©IJRASET: All Rights are Reserved

Feature Pyramid Networks (FPN) can detect non-uniform objects, but least used by researchers due to high computation cost and more memory usage. Furthermore, Mask R-CNN strengthens Faster R-CNN by including the prediction of segmented masks on each RoI. Although two-stage yields high object detection accuracy, but it is limited by low inference speed in real-time for video surveillance.

III. HARDWARE PLATFORM

The hardware part mainly consists of a digital computer, a Raspberry Pi Kit, R305, 16x2 LCD displays, RS232 and a Buzzer which is being discussed along with their specific functions.

A. Raspberry Pi Kit

The Raspberry Pi is a series of small single- board computers developed by the United Kingdom by the Raspberry Foundation to promote teaching in computer science in schools and developing countries. Peripherals (including key board, mice and cases) are not included with the Raspberry Pi.

It is bundled with on-board wifi, Bluetooth and USB boot capabilities. It has a Broadcom System-On-Chip, which includes an ARM compatible central processing unit (CPU) and an on-chip graphics processing unit (GPU, a video core). CPU speed ranges from 700 to 1.2GHz for Pi 4 and on-board memory ranges from 256MB to 1GB.

Secure digital (SD) cards are used to store the operating system and program memory either in the SDHC or microSDHC sizes. Lower level output is produced by a number of GPIO pins which supports common protocols. It supports Raspbian, a Debian based Linux distribution for download as well as the third party Ubuntu, Windows 10 IT Core, RISC OS and centralized media center distributions.

It promotes Python and Scratch as the main programming languages with support for many other languages. The Raspberry Pi 4 supports 1GB RAM.

B. Liquid Crystal Display (LCD)

Liquid Crystal Display screen is an electronic display module and find a wide range of applications. A 16x2 LCD display is very basic module and is very commonly used in various devices and circuits. LCD stands for liquid crystal display. They come in many sizes 8x1 , 8x2 , 10x2 , 16x1 , 16x2 , 16x4 , 20x2 ,20x4 ,24x2 , 30x2 , 32x2 , 40x2 etc.

These modules are preferred over seven segments and other multi segment LEDs. The reasons being: LCDs are economical; easily programmable; have no limitation of displaying special & even custom characters (unlike in seven segments), animations and so on.

A 16x2 LCD means it can display 16 characters per line and there are 2 such lines. In this LCD each character is displayed in 5x7 pixel matrix.

This LCD has two registers, namely, Command and Data. The command register stores the command instructions given to the LCD. A command is an instruction given to LCD to do a predefined task like initializing it, clearing its screen, setting the cursor position, controlling display etc.

The data register stores the data to be displayed on the LCD. The data is the ASCII value of the character to be displayed on the LCD.

IV. METHODOLOGY

In this model our aim is to detect wheather the person is wearing the mask or not wearing the mask . Face Mask detection has turned up to be an astonishing problem in the domain of image processing and computer vision.

Face detection has various use cases ranging from face recognition to capturing facial motions, where the latter calls for the face to be revealed with very high precision.

Due to the rapid advancement in the domain of machine learning algorithms, the face mask detection technology seem to be well addressed yet. This technology is more relevant today because it is used to detect faces in static images and videos also and follow the steps in figure 1.1

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

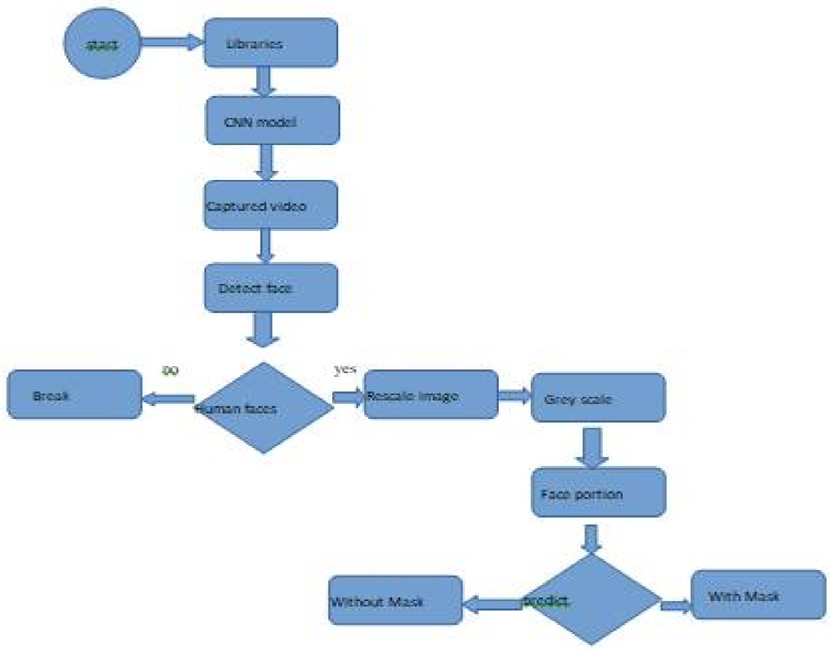


Fig1.1: Working flow chart

1) *Step1:* Initially import all the libraries and packages Raspberry device .
2) *Step2:* Train the data with mask faces and without mask faces , after training is complete . Validate and test the dataset with accuracy , finally completing all these save the file for prediction.
3) *Step3:* Run the video and if it detected the face
4) *Step4:* If it is human face then condition will re-scale the image or if it is not human face the condition will break.
5) *Step5:* After re-scaling the image and check with grey scale values and check to the face recognition values and displays rectangle box in the screen.
6) *Step6:* After face recognition is complete it will predict wheater the face has mask or not.
7) *Step7:* If the person wear mask then in display it will show with a green line rectangle box with the name mask on the screen . If the person wear mask then in display it will show with a red line rectangle box with the name No-mask on the screen .

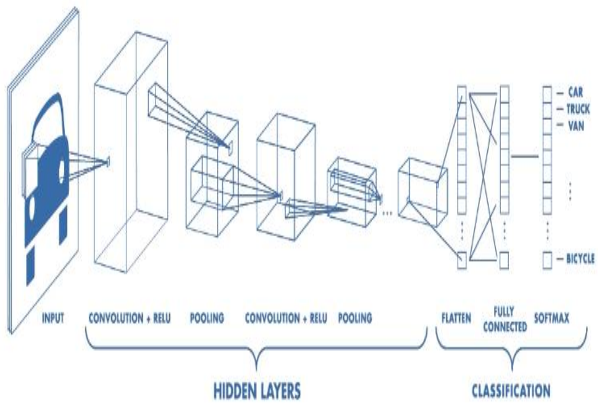
In this way, our proposed method provides Face-Mask detection through raspberry pi camera on real time embedded systems.

A. Convolutional Neural Network(CNN)

Convolutional Neural Network, also known as CNN is a sub field of deep learning which is mostly used for analysis of visual imagery. CNN is a class of deep feed forward (ANN). This Neural Network uses the already supplied data-set to it for training purposes, and predicts the possible future labels to be assigned. Any kind of data This Neural Network uses its strengths against the curse of dimensionality. A portion of the territories where CNNs are broadly utilized are image recognition, image classification, image captioning and object detection etc. The CNNs got immense popularity when Alex discovered it in 2012. In just three years, the engineers have advanced it to an extent that an older 8 layer AlexNet now is converted into 152 layer ResNet. Tasks where recommendation systems, contextual importance or natural language processing (NLP) is considered, CNNs come handy. The key chore of the neural network is to make sure it processes all the layers, and hence detects all the underlying features, automatically. A CNN is a convolution tool that parts the different highlights of the picture for analysis and prediction.

1) Convolutional neural network is also known as Artificial neural network that has so far been most popularly used for analyzing images.
2) CNN has hidden layers called Convolutional layers, and these layers are precisely what makes this CNN.
3) Convolutional layer more precisely able to detect patterns.
4) Through CNN model we insert any object from input it will check convolutional layer and transform through output.

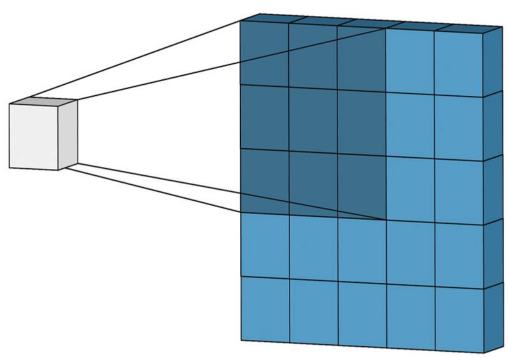
©IJRASET: All Rights are Reserved



B. A CNN Typically has three Layers

1) *Convolution Layer:* The convolution layer is a core building block of the CNN. It carries the main portion of the network's computational load. This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel and the other matrix is the restricted portion of the receptive field. The kernel is spatially smaller than an image but is more in-depth.

During the forward pass, the kernel slides across the height and width of the image-producing the image representation of that receptive region. This produces a two-dimensional representation of the image known as an activation map that gives the response of the kernel at each spatial position of the image. The sliding size of the kernel is called a stride.

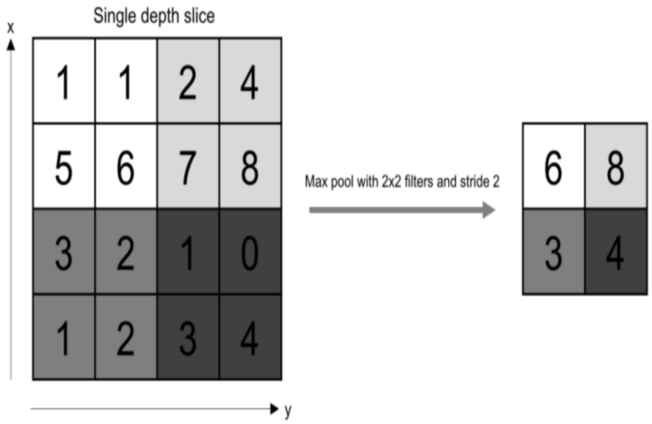


If we have an input of size W *W *D and Dout number of kernels with a spatial size of F with stride S and amount of padding P, then the size of output volume can be determined by the following formula:

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

2) *Pooling Layer:* The pooling layer replaces the output of the network at certain locations by deriving a summary statistics of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually.



If we have an activation map of size*W*W*D*, a pooling kernel of spatial size *F*, and stride *S*, then the size of output volume can be determined by the following formula:

$$W_{out} = \frac{W - F}{S} + 1$$

This will yield an output volume of size *Wout*Wout*D*.

3) *Fully Connected Layer:* Neurons in this layer have full connectivity with all neurons in the preceding and succeding layer as seen in regular FCNN. This is why it can be computed as usual by a matrix multiplication folowed by a bias effect.

The FC layer helps to map the representation between the input and and output.

.Designing a Convolutional Neural Netwrok
Our convolutional neural network has architecture as follows:

[INPUT]

→[CONV 1] → [BATCH NORM] → [ReLU] → [POOL 1]

→ [CONV 2] → [BATCH NORM] → [ReLU] → [POOL 2]

→ [FC LAYER] → [RESULT]

V. EXPERIMENTAL EVALUATION

A. *Control Overfitting*

To address RQ5 and avoid the problem of overfitting, two major steps are taken. First, we performed data augmentation . Second, the model accuracy is critically observed over 60 epochs both for the training and testing phase. The observations are reported inFig. 2.1 and 2,2.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

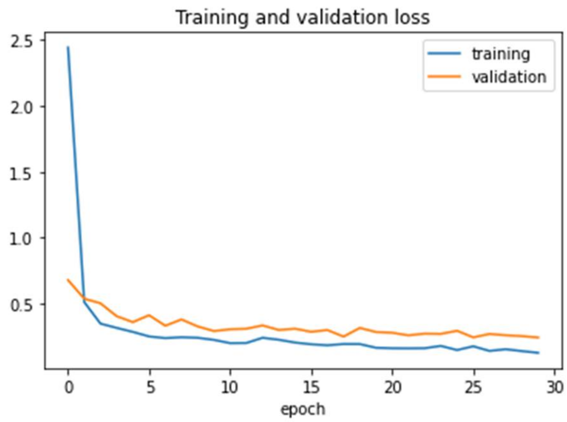


Figure 2.1 : Training and validation loss

Figure 2.2 : Training and validation accuracy

It is further observed that model accuracy keeps on increasing in different epochs and get stable after epoch=3 as depicted graphically in Fig. 2.2 above. To summarize the experimental results, we can say that the proposed model achieves high accuracy in face and mask detection with less inference time and less memory consumption as compared to recent techniques. Significant efforts had been put to resolve the data imbalance problem in the existing MAFA dataset, resulting in a new unbiased dataset which is highly suitable for COVID related mask detection tasks. The newly created dataset, optimal face detection approach, localizing the person identity and avoidance of overfitting resulted in an overall system that can be easily installed in an embedded device at public places to curtail the spread of Coronavirus.

In order to implement and demonstrate the system developed theoretically, we created a prototype that represents the system. Thus the whole system that is being developed is given below, (Fig. No: 2.3).

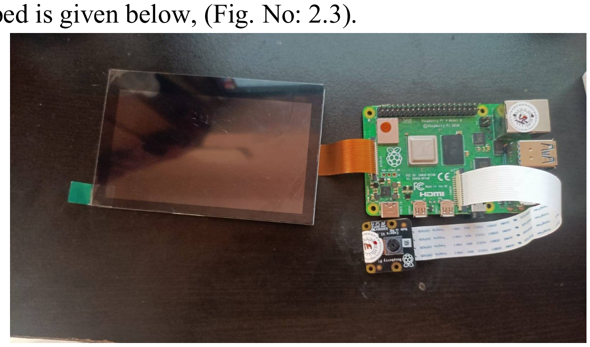


Fig.1.3: Prototype

The Realtime face-mask Detection system provides good solution . A novel architecture for an economic Face-Mask detection technology is proposed and implemented in this paper.

REFERENCES

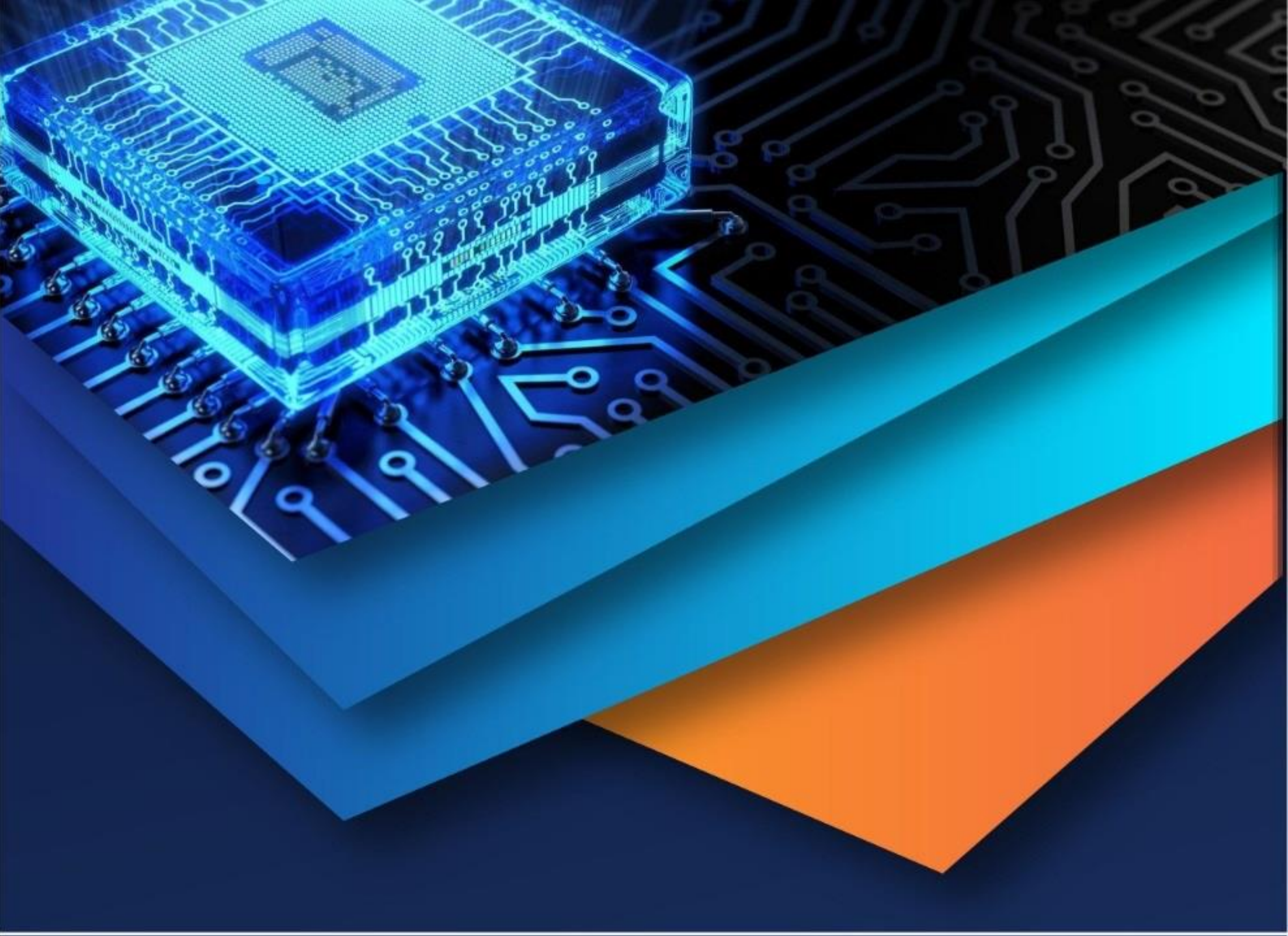
[1] Roomi, Mansoor, Beham, M.Parisa, "A Review Of Face Recognition Methods," in International Journal of Pattern Recognition and Artificial Intelligence, 2013, 27(04), p.1356005
[2] A. S. Syed navaz, t. Dhevi sri, Pratap mazumder, " Face recognition using principal component analysis and neural network," in International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC), vol. 3, pp. 245-256, Mar. 2013.
[3] Turk, Matthew A., and Alex P. Pentland. "Face recognition using eigenfaces," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586-591.
[4] H. Li, Z. Lin, X. Shen, J. Brandit, and G. Hua, "A convolutional neural network cascade for face detection," in IEEE CVPR, 2015, pp.5325-5334.
[5] Wei Bu, Jiangjian Xiao, Chuanhong Zhou, Minmin Yang, Chengbin Peng, "A Cascade Framework for Masked Face Detection," in IEEE International Conference on CIS & RAM, Ningbo, China, 2017, pp.458-462.
[6] Shiming Ge, Jia Li, Qiting Ye, Zhao Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," in IEEE Conference on Computer Vision and Pattern Recognition, China, 2017, pp. 2682--2690.
[7] M. Opitz, G. Waltner, G. Poier, and et al, "Grid Loss: Detecting Occluded Faces," in ECCV, 2016, pp. 386-402.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

[8] X. Zhu and D. Ramanan, “Face Detection, pose estimation and landmark localization in the wild,” in IEEE CVPR, 2012, pp.2879-2886.
[9] Florian Schroff, Dmitry Kalenichenko, James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823
[10] Ankan Bansal Rajeev Ranjan Carlos D. Castillo Rama Chellappa, “Deep Features for Recognizing Disguised Faces in the Wild,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2018.
[11] Maneet Singh, Richa Singh, Mayank Vatsa, Nalini K. Ratha, and Rama Chellappa. “Recognizing disguised faces in the wild,” in IEEE Transactions on Biometrics, Behavior, and Identity Science 1, no. 2, 2019, pp. 97-108.
[12] K. Yan, S. Huang, Y. Song, W. Liu and N. Fan, “Face recognition based on convolution neural network,” in 36th Chinese Control Conference (CCC), Dalian, 2017, pp. 4077- 4081.
[13] Amarjot Singh, Devendra Patil, G Meghana Reddy, Sn Omkar, “Disguised Face identification (DFI) with facial Key Points using Spatial Fusion Convolutional Network,” in IEEE International Conference on Computer Vision Workshops (ICCVW),Venice, Italy, 2017, pp. 1648 - 1655.
[14] Chandra Churh Chatterjee, “Towards Data Science,” [Online]. Available:https://towardsdatascience.com/basics-of-the-classiccnn- a3dce1225add. [Accessed 1 September 2019].

©IJRASET: All Rights are Reserved



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)