European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 06, 2020

A Novel Approach To Detect Face Mask To Control Covid Using Deep Learning

T Subhamastan Rao[1], S Anjali Devi[2], P Dileep[3], M Sitha Ram[4]

[1]Associate Professor, CMR Technical Campus, Hyderabad, Telangana, India
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh
[3]Department of Computer Science and Engineering,Malla Reddy College of Engineering and Technology,Kompally,Hyderabad, Telangana, India
[4]Sr.Assistant professor, Department of CSE,Lakireddy Bali Reddy college of engineering, Mylavaram, Vijayawada, Andhra Pradesh, India

Mail: [2]swarnaanjalidevi@gmail.com

ABSTRACT: COVID-19 caused due to corona virus, this virus was first discovered in wuhan on Dec -2019 now it is a pandemic effecting almost every country in the world. This virus transmitting from one person to another person through droplets spawned when a covid patient sneezes, coughs or exhales. Even these droplets reaches ground soon because they are heavy and unable to hang in the air. One of the solution to prevent covid-19 is wearing the face masks, Many governments trying their best to educate citizens to wear masks in public places even they made it mandatory, but majority people are violating this rule.

In current scenario police frequently check for face mask in public places and imposing fine on the people who are not wearing face mask. On other hand some governments introduced technology to detect people with out face mask and send their details to petrol team then they will catch them.

In this paper we are proposing a model which detects public without facemask and thatdata can be used to identify the person who is not wearing the mask using facial detection system then that data is integrated with public identification data base to collect details of that person and fine amount will send to his mobile number and address. using CNN model we have detected persons with mask and without mask.

CNN can able to identify pixel level data when compared to many algorithms available, CNN works more accurately. We implemented a model with two convolution layers with 100 filters in each and applied drop out 0.5% and used Relu, soft max as activation functions at hidden and fully connected layers respectively, Cross entropy used as loss function adam is optimizer and model trained over 1500 images consists of both classes with mask and without mask and cascade classifier is used to classify faces and it is working with 91.21 accuracy.

This AI based mask detection system definitely creates fear in the minds of public and they will start wearing mask in public places so that the spreading of the disease can be controlled that intern useful for wellbeing of the society.

1. INTRODUCTION:

As on 11-Sep2020 there are 28,3,66,652 people worldwide infected with Corona virus and 9,14,501 deaths , 20,3,67,966 people are recovered from covid. All sectors are severely affected because of covid many people lost their jobs throughout the world. More than 50

countries made mandatory to wear mask worldwide.The average growth of the cases for 7 days given below in the fig.1

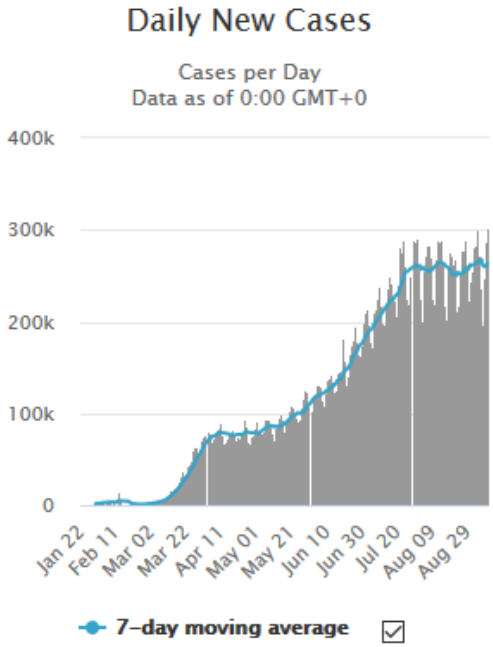

Fig.1 Daily Covid cases growth

As per the WHO, Scientists guidelines wearing face mask and following other kind of precautions such as social distancing , frequent hand wash will help to reduce the spread of corona virus. After these guidelines many countries have framed their own rules regarding face mask to control the spread but there are many people refused to follow government rules. To catch these people police are struggling and they are unable to find every one every where. Face detection and object detection are the some of the computer vision problems will help to identify people who are not wearing masks and assist police to control these people.

Fig2. Sample data set used to train the model

Apart from mask detection there are various applications for object and face detection models in different real time domains such as driver less cars, criminal detection, Vehicle number plate detection etc.

1.1 Object detection:

Face recognition is a Artificial intelligence technique in which a given image is processed to find out what is the object inside the image whereas face detection techniques are used to find out where a specific object inside the image. Image classification is labelling objects inside the image.Scale invariant feature transform(SFIT): in this approach we are going to identify patches and apply some mathematical transformations on it to generate feature vector, then that feature vectors between images are compared for matching.

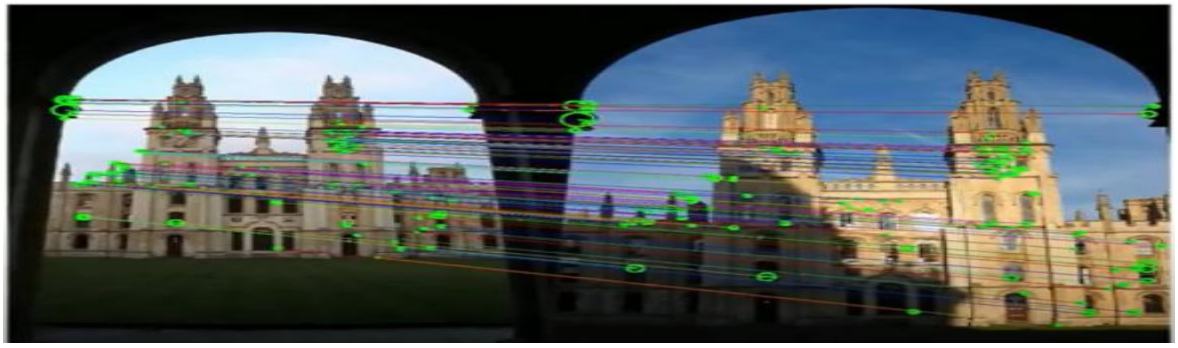

Fig3. Image comparison using SFIT

SIFT algorithm invariant to rotation and scale not only that if we observe above image and compare first image and second background is different in second image some part is covered with shadow even SIFT algorithm invariant to brightness and contrast. So far from images these algorithms are selecting important features and processing them for comparison there is another variant in SIFT which is using all patches in the image for comparison call Dense SIFT.

HOG is one of the feature extraction algorithm used in object detection technique introduced in 2005. The main idea Used in HOG IS detecting intensity of the pixel compared to surrounding pixels it means how darker a pixel is when compared to surrounding pixel(magnitude) . even it identifies the direction of the pixels .

Here the magnitude of the each gradient(Inclination size of dark pixels) calculated then The direction of the inclination is calculated

Once magnitude and direction of all the pixels are calculated then they are normalised and those normalised features are stored in constructed 9 bins with each bin having 20 degree inclination and compare direction of each pixel and check for the corresponding magnitude and store those values in suitable bin some sample pixel magnitude and directions calculated and 9 bins are shown in the below fig.

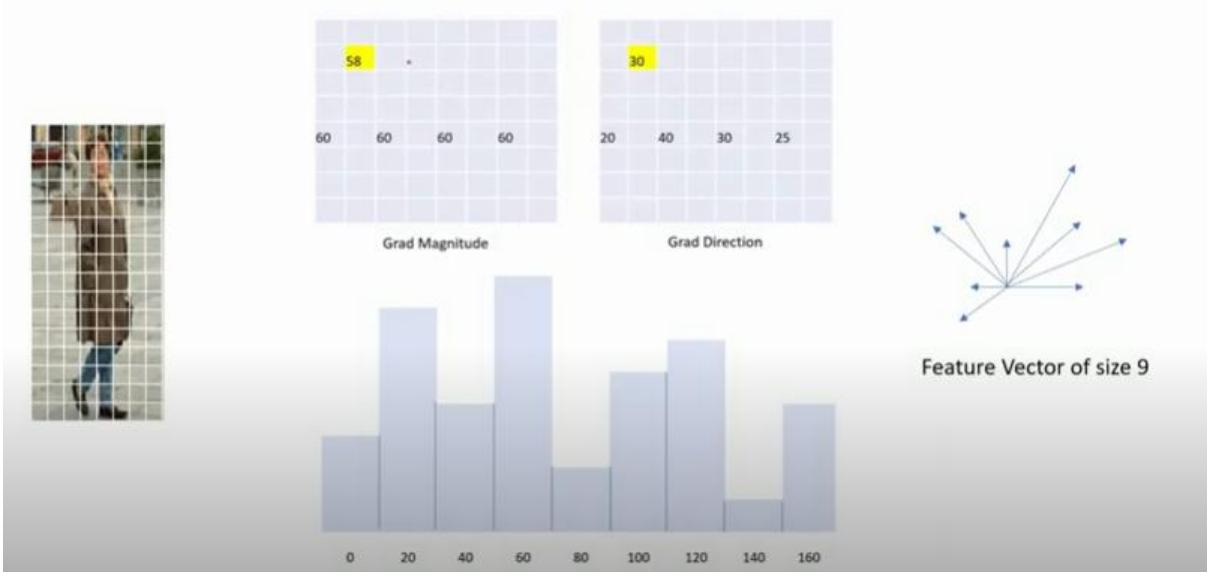

Fig4. Image feature extraction using HOG

These HOG features are used to detect objects by convoluing with corresponding object filter then apply support vector machine classifier on it finally we can detect object using HOG+ SVM classifier an example was shown in the below fig..

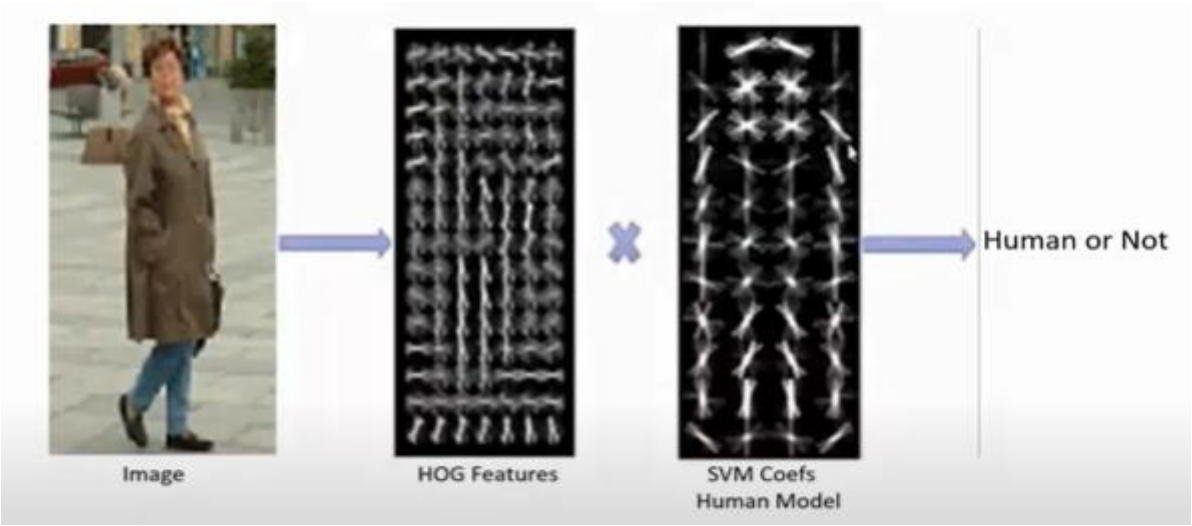

Fig5. Human detection using HOG+SVM

First we need train some images with human in it and some images without human then HOG features are extracted and labeled as human and non humans then they are performed dot product with SVM coefficents finally we are going to get a value weather image cnsists of human or not.

2. LITERATURE SURVEY:

In earlier days face detection models are implemented using edge,line and centre near features and patterns are recognized from those feature. These approaches are used to find binary patterns locally. These approaches are very effective to deal with Gray-scale images and the computation effort required also very less[1]

AdaBoost is a regression based classifier which is going to fit regression function on original data set even some miss classified objects waits also adjusted during back propagation to optimize the results[2]

Viola Jones Detector proposed an real time object model used to detect different classes of objects. It uses 24x24 base window size to evaluate any image with edge ,line and four rectangular features. Harr-like features are like convolutions to check weather given feature is available in the image or not[3]. This model fail to work in when image brightness varies even it exhibits poor performance when images are in different orientations.

Convolution networks are mainly used for classification problems there are various kinds of CNN architectures such as VGG-16 this architecture consists of 2 convolution layers with input size224 kernel (64,3x3) followed by max pool with size 2x2 then again two convolution layers followed by max pool then three convolution layers with max pool again three convolution layers and max pool and three fully connected layers final FC is soft max this architecture works fine when compared to AlexNet[4,7].

Google Net architecture fundamentally using inception method by constructing small convolution layers to reduce number of parameters it having around 22 layers with convolution and max pooling etc it can able to work effectively over Alexnet it can able to bring down 60 million features in Alex net to 4 million features[8].

In this paper Deepnueral networks which adopts residual learning to train the models more deeper around 152 layers are used in this which is 8 times more than VGGnet with minimum complexity. This approach achieved relatively better performance in object detection over COCO data set[9]

In this paper UNet and SEnet are used to perform segmentation of heart ventricular segmentation . This model is arrange the weights in such a way like more weights are given to useful features and less weights are given to unimportant features.[10]Support vector machines are used to perform classification on objects which is going to build an equation for constructing line and classifies the objects based on the values mapping to this line.

Semantic segmentation method was used to detect facial mask in this paper they have used VGG net for training and FCN is used to semantically segment the faces available in the image[11] performed experiments on multi parsing human data sets and achieved higher accuracy.

In this paper medical image processing was done. They have taken human brain images and are trained by using FCN to identify tumours very effectively in this paper rather than using 2D segmentation for detecting tumour we have used 3D segmentation[12]. Tumuluru,Lakshmi Ramaniet. al.[13], used CNN model to detect human face which is used efficiently in security related applications. In this paper they have collected various facial features such as mouth, nose, eyes stored as facial template and used it for detecting difference between faces.

Malathi, J. et. al., [14] mainly focused on identifying forgery images used in different places like in social media, and other publicity required places. In this paper various techniques are proposed to find out features of a forgery image like image spicing,copy move attack which can be handled by using correlation analysis to find duplicate features.Patelet. al., [15]proposed a model to find out the quality of the ironore by extracting the features from sample material in the mining industry. It is very important to asses the quality of the ore . SVR support vector regressor used for online measure of the quality of the ore. In this process they have extracted 280 features extracted for object identification, SFFS was model was developed using SVR.

Object detection become the important area in the field of image investigation there are various techniques are there for image analysis in [15]. In this paper author introduced a

wavelet based nueral network for feature extraction and learning which is working efficiently in object detection[.

Satapathy, Sandeep Kumar, et al.,[16]proposed a model to detect number plate which is very important problem helping police to chase many criminal cases. Authors used OCR based approach to detect characters in the number plate and they are stored and processed to client server based model for collecting the details of the owner.Pathaket.al.,[17] proposed multi dimension biometric authentication system which will work effectively in low lightening conditions here accuracy was improved by using entropy based CNN.

Medical plant detection becoming very important problem which will help ordinary people to detect spices [18].In this paper authors proposed a model using CNN. It was trained with medical leaf images and it can able to detect medical plants more accurately.

Human pose detection is one of the important research are nowadays drawing lot of attention [19].In this paper author proposed a model which is going to detect traditional dance based on human pose. To achieve this, they have used CNN and various steps in traditional dance was trained and model learned from it and effectively able to detect traditional dance.

Ravi, Sunitha, et al. [20], sine language detection was implemented by training a CNN model which can able to detect signs in the real world video. Which is very much useful in the driver less cars also. Even it is useful in sign language in machine translation.

Joint angular displacement approach was used through CNN for further enhanceing the capabilities of CNN to capture 3D motion sign language in real time which can be applied to many real time applications in these days[21].Patel, Ashok Kumar et. al.,[22]proposed a model to find out the quality of the ironore by extracting the features from sample material in the mining industry. It is very important to asses the quality of the ore . SVR support vector regressor used for online measure of the quality of the ore. In this process they have extracted 280 features extracted for object identification, SFFS was model was developed using SVR.

3. METHODOLOGY:

In this paper we have proposed a new CNN architecture called as M-CNN to detect weather a person wearing mask or not. The proposed architecture was shown in the below Fig.6.This architecture accepts a input size of 150X150x3 and first convolution layer convolved for feature extraction with 3X3 kernel of size 100. Second layer is Max Pooling layer of size 2X2. Third layer takes the input from previous layer and again convolved with 3X3 kernel of size 100 then max pooling of size 2X2 then followed by flatten layer next a drop out layer finally 2 dense layers first layer consists of Relu as activation function and second layer soft max as activation function.

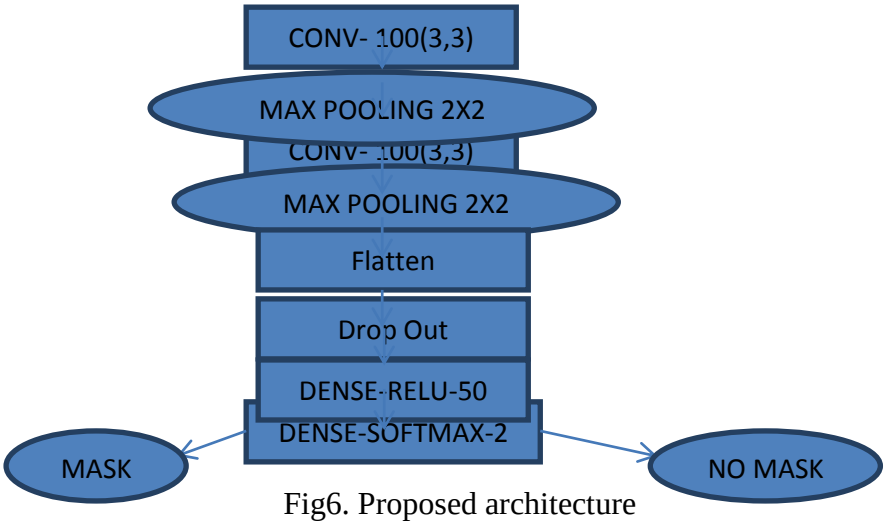

Fig6. Proposed architecture

Convolution layer:

In convolution process multiplication is performed between input and set of weights also called as filters or kernel. It is smaller than input and dot product is performed between patch of the input with filter and finally a single value is produced like this process is repeated for all the patches in the input image.

Out put layer = (WD-FL+2PA)/ST+1 --- Eq.1

WD = Input image size, FL = Filter size, PA = Padding, ST = Stribe --- Eq.2

Max pooling layer:

Pooling layers applied on the feature map generated from through convolution process, it down sample the patches that contain input image. The common methods in pooling are Max pooling and Avg pooling in this methodology we have used Max pooling.The following formula used to calculate max feature.

Max pool = $(n_{hight} - fl + 1)/str X (n_{width} - fl + 1)/str \times c_{hannel}$ --- Eq.3

n_{hight} = height of feature map

n_{width} = width of feature map

fl= size of filter, str=stribe length

c_{hannel} =number of channels in the feature map

In this model we have taken 2X2 max pooling to down sample feature map

Flatten layer:Flatten layer is used to convert two dimensional feature map into a single vector with all features of the image.

Drop out probability applied in this model is 0.5 which is used to reduce the over fitting of data by applying regularization. From the flatten layer drop out layer drops some nodes randomly after applying drop out network performance improved when it is exposed to testing data.

Fully connected layer/dense is a linear operation on input vector in this model we have used two dense layers first layer is used Relu as activation function second dense layer used soft max as activation function.

In the first dense layer we have used ReLu to normalize the values:

Gf(k) = max(0,k) where in this always negative values are assigned to zero any value greater than zero assigned to that value.

In second dense layer we have used softmax activation function which takes input vector and divide it by two classes on with mask other one is with out mask.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K \text{ and } \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K$$

---- Eq.4

We have used Adam optimization function to update learning rate and to update weights based on the loss back propagated .

We have used binary cross entropy loss function to find the loss

$$CE = -\sum_{i=1}^{C'=2} t_i log(s_i) = -t_1 log(s_1) - (1 - t_1) log(1 - s_1)$$

---Eq.5

4. RESULTS AND DISCUSSION:

We have used around 1500 images they are divided into training set and testing set these images consists of with mask and without mask. Training testing split is 80%,20% respectively. Then these images are trained to the M-CNN model. We have used different convolution, max pool and flatten , dropout, fully connected layers. We have used adam as optimization function and cross entropy as loss function network was trained in 100 epochs and achieved 91.5 accuracy . This model can able to detect single person in an image or multiple persons in a single image into two classes with mask and with out mask effectively around 91.5 accurately.

We have used cap = cv2.VideoCapture(0) to capture videos from web cam and this and cascade classifier for object detection from the live web cam.

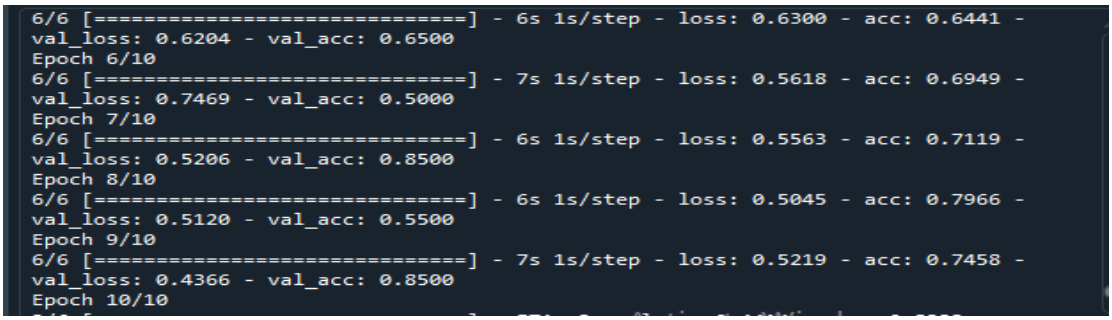

Fig.7 learning image data in different epoch

Fig.8 Detecting single person without mask

Fig.9 Detecting multiple persons without mask

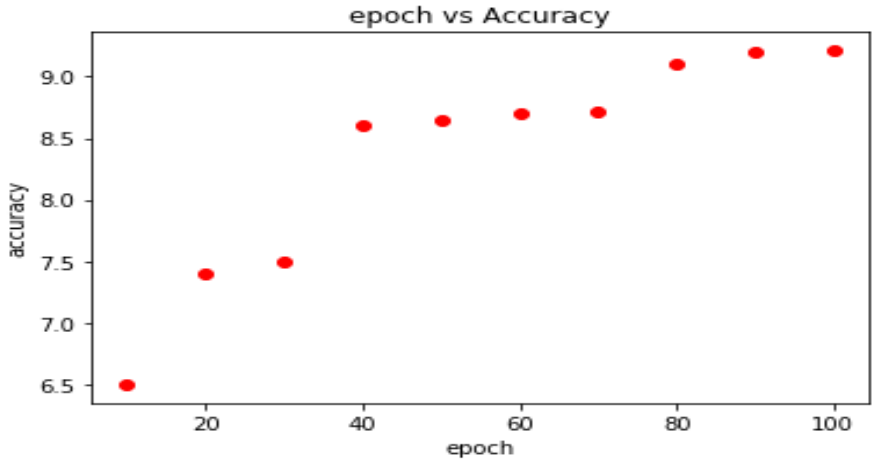

Fig.10 Epoch Vs Accuracy

5. CONCLUSION:

In this pape,r we have implemented a M-CNN model to detect weather a person or group of persons in a image wearing mask or not. We have trained the model in such away that it will work in varying dimensions and light conditions very effectively. This model can be used by any regulating authorities to control spreading of covid through proper monitoring and punishing the public who are not wearing masks. It can be further extended by integrating with app and recognizing the person and integrating his address to send fine to the citizens who are violating rules of wearing mask.

6. REFERENCES:

[1] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scaleand rotation invariant texture classification with local binary patterns,"*IEEE Transactions on Pattern Analysis and Machine Intelligence*,vol. 24, no. 7, pp. 971–987, July 2002.

[2] T.-H. Kim, D.-C. Park, D.-M.Woo, T. Jeong, and S.-Y. Min, "Multi-classclassifier-based adaboost algorithm," in *Proceedings of the Second Sinoforeign-interchange Conference on Intelligent Science and IntelligentData Engineering*, ser. IScIDE'11. Berlin, Heidelberg: Springer-Verlag,2012, pp. 122–127.

[3] P. Viola and M. J. Jones, “Robust real-time face detection,” *Int. J.Comput. Vision*, vol. 57, no. 2, pp. 137–154, May 2004.
[4] P. Viola and M. Jones, “Rapid object detection using a boosted cascadeof simple features,” in *Proceedings of the 2001 IEEE Computer SocietyConference on Computer Vision and Pattern Recognition. CVPR 2001*,vol. 1, Dec 2001, pp. I–I.
[5] J. Li, J. Zhao, Y. Wei, C. Lang, Y. Li, and J. Feng, “Towards realworld human parsing: Multiple-human parsing in the wild,” *CoRR*, vol.abs/1705.07206.
A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classificationwith deep convolutional neural networks,” in *Advances in Neural InformationProcessing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, andK. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
[6] K. Simonyan and A. Zisserman, “Very deep convolutional networks forlarge-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.
[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,”2015.
[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for imagerecognition,” *2016 IEEE Conference on Computer Vision and PatternRecognition (CVPR)*, pp. 770–778, 2016.
[9] K. Li, G. Ding, and H. Wang, “L-fcn: A lightweight fully convolutionalnetwork for biomedical semantic segmentation,” in *2018 IEEE InternationalConference on Bioinformatics and Biomedicine (BIBM)*, Dec2018, pp. 2363–2367.
[10] X. Fu and H. Qu, “Research on semantic segmentation of high-resolutionremote sensing image based on full convolutional neural network,” in*2018 12th International Symposium on Antennas, Propagation and EMTheory (ISAPE)*, Dec 2018, pp. 1–4.
[11] S. Kumar, A. Negi, J. N. Singh, and H. Verma, “A deep learningfor brain tumormri images semantic segmentation using fcn,” in*2018 4th International Conference on Computing Communication andAutomation (ICCCA)*, Dec 2018, pp. 1–4.
[12] Tumuluru,Lakshmi Ramani, B., Saibaba, C.M.H., Venkateswarlu, B., Ravinder, N. OpenCV algorithms for facial recognition,*International Journal of Innovative Technology and Exploring Engineering*,Volume 8, Issue 8, 1 June 2019, Pages 927-933
[13] Malathi, J., T. Satya Nagamani, and KNVSK Vijaya Lakshmi. "Survey: Image forgery and its detection techniques." *Journal of Physics: Conference Series*. Vol. 1228. No. 1. IOP Publishing, 2019.
[14] Krishnaveni, G., B. Lalitha Bhavani, and NVSK Vijaya Lakshmi. "An enhanced approach for object detection using wavelet based neural network." *Journal of Physics: Conference Series.* Vol. 1228. No. 1. IOP Publishing, 2019.
[15] Satapathy, Sandeep Kumar, et al. "Deep learning based image recognition for vehicle number information." *International Journal of Innovative Technology and Exploring Engineering* 8 (2019): 52-55.
[16] Pathak, Mrunal, Vinayak Bairagi, and N. Srinivasu. "Multimodal Eye Biometric System Based on Contour Based E-CNN and Multi Algorithmic Feature Extraction Using SVBF Matching."*International Journal of Innovative Technology and Exploring Engineering*
[17] Volume-8 Issue-9, July 2019.
[18] .Dudi, B., and Dr V. Rajesh. "Medicinal Plant Recognition Based On CNN And Machine Learning." *International Journal Of Advanced Trends In Computer Science And Engineering*, ISSN (2019): 2278-3091.
[19] Kishore, P. V. V., et al. "Indian classical dance action identification and classification with convolutional neural networks." Advances in Multimedia 2018 (2018).

[20] 20Ravi, Sunitha, et al. "Multi modal spatio temporal co-trained CNNs with single modal testing on RGB–D based sign language gesture recognition." *Journal of Computer Languages* 52 (2019): 88-102.
[21] Kumar, E. Kiran, et al. "Training CNNs for 3-D sign language recognition with color texture coded joint angular displacement maps." *IEEE Signal Processing Letters* 25.5 (2018): 645-649.
[22] Patel, Ashok Kumar, Snehamoy Chatterjee, and Amit Kumar Gorai. "Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades." Earth Science Informatics 12.2 (2019): 197-210.