Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8, July, 2021:7282 – 7292

Research Article

A Real Time Prediction and Classification of Face Mask Detection using CNN Model

S. Bavankumar[1], Dr. B. Rajalingam[2], Dr. R. Santhoshkumar[3], Dr. G. JawaherlalNehru[4], P.Deepan[5], N. Balaraman[6], M. Mahashree[7]

Abstract

Current scenario of COVID-19 (Corona Virus Disease) pandemic makes almost everyone to wear a mask in order to effectively prevent the spread of the virus. This almost makes conventional facial recognition technology ineffective in many cases, such as community access control, face access control, facial attendance, facial security checks at train stations, etc. Therefore, it is very urgent to improve the recognition performance of the existing face recognition technology on the masked faces. For that detecting the people with face mask is very essential. In this work, a reliable method based on discard masked region is proposed in order to address the problem of masked face recognition process. The first step is to discard the masked face region and extract the forehead and eyes region. Next, a pre-trained deep Convolutional neural networks (CNN) is applied to extract the best features from the obtained regions. Finally, it show experimental result is achieved an accuracy of more than 98% validation dataset.

Keywords: COVID-19, face recognition, convolutional layer, neural network and diseases.

1. Introduction

The trend of wearing face masks in public is rising due to the COVID- 19 (Corona Virus Disease) pandemic all over the world. Before Covid-19, People used to wear masks to protect their health from air pollution. While other people are self-conscious about their looks, they hide their emotions from the public by hiding their faces. Scientists proved that wearing face masks works on impeding COVID-19 transmission. COVID- 19 is the latest pandemic virus that hit the human health in the last decade. In 2020, the rapid spreading of COVID-19 has forced the World Health Organization to declare COVID- 19 as a global pandemic [1].

More than five million cases were infected by COVID-19 in less than 6 months across 188 countries. The virus spreads through close contact and in crowded and overcrowded areas. The COVID–19 virus can be spread through contact and contaminated surfaces, therefore, the classical biometric systems based on passwords or finger- prints are not anymore safe [2]. Face recognition are safer without any need to touch any device. Recent studies on corona virus have

[1,2,3,4,5,6] Department of Computer Science &Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India
[2,3,4]Associate Professor, [1,5,6 ,7]Assistant Professor
[1]sbavankumar55@gmail.com

proven that wearing a face mask by healthy and infected population reduces considerably the transmission of this virus. However, wearing the mask face causes the following problems:

- Fraudsters and thieves take advantage of the mask, stealing and committing crimes without being identified.
- Community access control and face authentication are become very difficult tasks when a grand part of the face is hidden by a mask.
- Existing face recognition methods are not efficient when wearing a mask which cannot provide the whole face image for description.
- Exposing the nose region is very important in the task of face recognition since it is used for face normalization, pose correction, and face matching. Due to these problems, face masks have significantly challenged existing face recognition methods.

2. Related Works

Face is the natural assertion of identity: We show our face as proof of who we are. Due to this widely accepted cultural convention, face is the most widely accepted biometric modality. Face recognition has been a specialty of human vision: Something humans are so good at that even a days-old baby can track and recognize faces. Computer vision has long strived to imitate the success of human vision and in most cases, has come nowhere near its performance. However, the recent Face Recognition Vendor Test (FRVT06), has shown that automatic algorithms have caught up with the performance of humans in face recognition [4]. This can partly be attributed to the advances in 3D face recognition in the last decade. 3D face recognition has important advantages over 2D; it makes use of shape and texture channels simultaneously, where the texture channel carries 2D image information. However, it is registered with the shape channel, and intensity can now be associated with shape attributes such as the surface normal.

The world is facing a huge health crisis due to the rapid transmission of coronavirus (COVID-19). Several guidelines were issued by the World Health Organization (WHO) for protection against the spread of coronavirus. According to WHO, the most effective preventive measure against COVID-19 is wearing a mask in public places and crowded areas. It is very difficult to monitor people manually in these areas. In this paper, a transfer learning model is proposed to automate the process of identifying the people who are not wearing mask [5]. The proposed model is built by fine-tuning the pre-trained state-of-the-art deep learning model, InceptionV3. The proposed model is trained and tested on the Simulated Masked Face Dataset (SMFD). Image augmentation technique is adopted to address the limited availability of data for better training and testing of the model.
Here in this paper the authors have proposed a mask detection system for the health care personal inside the operation theatre. As the health care personal need to wear a mask in the operation theatre and the proposed system will alert for any personal not wearing the mask. There are two detection system used for face and medical mask wearing. Their system achieved almost 90% recall and less than 5% of false positive rate. They have worked for the medical mask detection from the images that are taken from 5m distance by cameras.

In this paper the authors have worked on the masked face detection from the video. The masked person is detected in this presented approach and mainly 4 steps are performed for the detection that are estimation of distance between camera and person, detection of eye line, detection of part of face and detection of eye. They have analyzed their algorithm on various video surveillance systems and achieved a fine accuracy. Li et al. and others [7, 16] used YOLOv3 for

face detection, which is based on deep learning network architecture named darknet-19, where WIDER FACE and Celebrities databases were used for training, and later the evaluation was done using the FDDB database. This model achieved an accuracy of 93.9%.

3. Proposed Works

In this section, we have to develop Face Mask Detection using Convolutional Neural Network model. The first step is to localize the mask region. To do so, a cropping filter is applied in order to obtain only the informative regions of the masked face (i.e. forehead and eyes). Next, the selected regions are described using a deep learning model. This strategy is more suitable in real-world applications comparing to restoration approaches. Recently, some works have applied supervised learning on the missing region to restore them such as this strategy, however, is a difficult and highly time-consuming process. Despite the recent breakthroughs of deep learning architectures in pattern recognition tasks, they need to estimate millions of parameters in the fully connected layers that require powerful hardware with high processing capacity and memory. Thus an efficient quantization based pooling method for face recognition using the VGG-16 pre-trained model is used [12, 13]. At the last convolutional layer (also called channels) feature maps are represented using Bag-of-Features (BoF) paradigm. The Figure 1 shows the block diagram of the proposed Masked Face detection system.

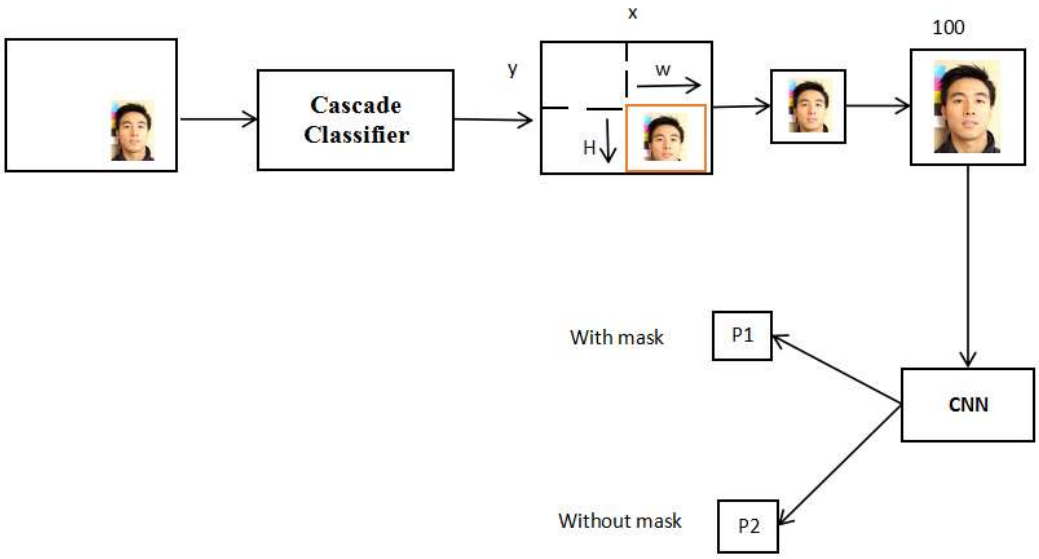

Figure 1 The block diagram of proposed Masked Face detection system

CNNs consist of the following sequential modules (each one may contain more than one layer)

1. Convolution
2. Activation function (Using Relu)
3. Pooling
4. Fully connected layers

3.1.Convolution

Convolution operation is an element-wise matrix multiplication operation. Convolutional layers take the three-dimensional input matrix we mentioned before and they pass a filter (also known as convolutional kernel) over the image, applying it to a small window of pixels at a time (i.e 3x3 pixels) and moving this window until the entire image has been scanned. The convolutional operation calculates the dot product of the pixel values in the current filter window along with the weights defined in the filter. The output of this operation is the final convoluted image. The core of image classification CNNs is that as the model trains what it really does is that it learns the values for the filter matrices that enable it to extract important features (shapes, textures, colored areas, etc) in the image. Each convolutional layer applies one new filter to the convoluted image of the previous layer that can extract one more feature. So, as we stack more filters, the more features the CNN can extract from an image. The results are summed up into one number that represents all the pixels the filter observed. This setting enables the network to learn different features while keeping the number of parameters tractable. Mathematically, the output feature map $y_{i,j}^{(l)}$ at convolutional layer l is calculated as in Equation 1.

$$y_{i,j}^{(l)} = \sigma^{(l)}\left(\sum_{n=1}^{k}\sum_{m=1}^{k} w_{n,m}^{(l)} . x_{i+n,j+m}^{l-1} + b^{(l)}\right) \quad (1)$$

where, the $w_{n,m}^{(l)}$ denoted the convolutional filter with size k x k at layer l, and the $x_{i+n,j+m}^{(l)}$ represent the spatial position of the corresponding feature map at the preceding layer l-1. The algorithm passes the convolutional filter throughout the input feature map using the dot product (.) between them with an addition of a bias unit $b^{(l)}$. Moreover, a non-linear activation function $\sigma^{(l)}$ at layer l is taken outside the dot product to strength the nonlinearity. The Figure 2 shows Operation between image pixel and filter,

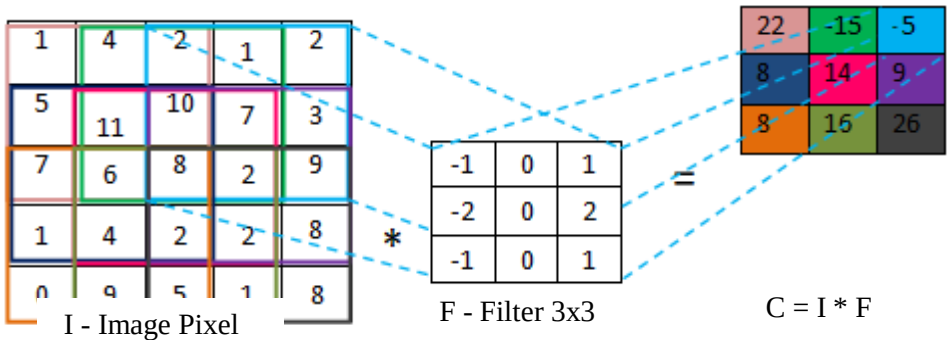

Fig. 2 Convolution Operation between Image Pixel and Filter

3.2.Activation Function

Activation functions are really important for learning and making sense of something really complicated and non-linear dynamic functional mapping between inputs and response variable for an Artificial Neural Network. The convolution layer generates a matrix that is much smaller in size than the original image. This matrix is run through an activation function, which introduces non-linearity to allow the network to train itself via back propagation. The most popular types of activation functions are sigmoid, tanh and ReLU.

ReLU

ReLU activation function given an output x if x is positive and 0 otherwise. ReLU function is the most widely used activation function in neural networks. One of the greatest advantages ReLU has over other activation functions is that it does not activate all neurons at the same time. In practice, ReLU converges six times faster than tanh and sigmoid activation functions. Figure. 3 illustrates the graph of Rectified Linear Unit (ReLU).

Figure 3 Rectified Linear Unit (ReLU)

3.3.Pooling

The pooling layer can generalize the convolved features through down-sampling and thereby reduce the computational complexity during the training process. A filter is passed over the results of the previous layer and selects one number out of each group of values (typically the maximum, this is called max pooling). This allows the network to train much faster, focusing on the most important information in each feature of the image. Given a pooling/subsampling layer *q*, the feature output F^q can be derived from the preceding layer $f^{(q-1)}$ through the Equation 3.2.

$$F_{i,j}^{q} = \max(f_{1+p(i-1),1+p(j-1)}^{q-1},..., f_{pi,l+p(i-1)}^{q-1},..., f_{1+p(i-1),pj}^{q-1},..., f_{pi,pj}^{q-1} \quad (3.2)$$

where p x p is the size of the local spatial region, and $1 \le i, j \le (m-n+1)/p$, here *m* refers to the size of input feature map, while n corresponds to the size of the filter. Then it simply summarizes the input features within local spatial region using the maximum value. On two-dimensional feature maps, pooling is typically applied in 2×2 patches of the feature map with a stride of 1 or 2. Figure.4 demonstrates the max, min and average pooling layer with 2x2 filters of stride 2.

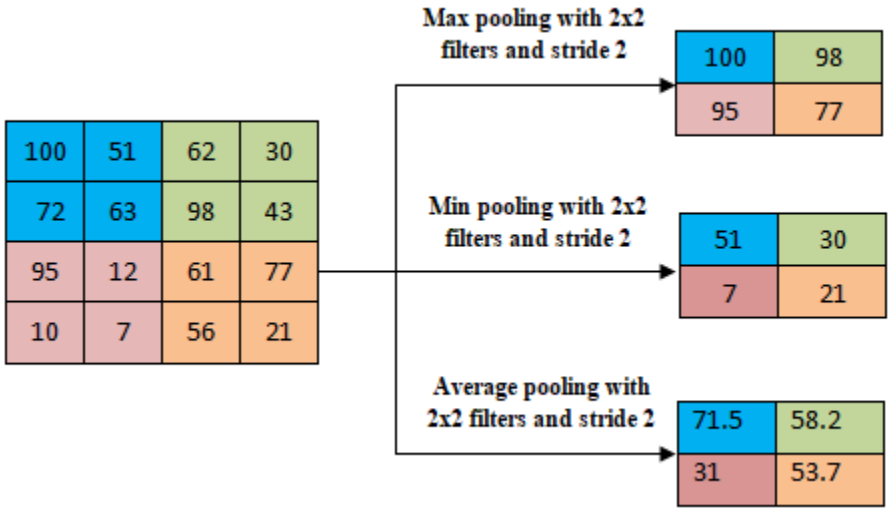

Figure 4 Max, Min and Average Pooling Layer with 2x2 Filters and Stride 2

- ***Max pooling*** is a pooling operation that calculates the maximum, or largest, value in each patch of each feature map.
- ***Min pooling*** is a pooling operation that calculates the minimum or smallest value in each patch of each feature map.
- ***Average pooling*** involves calculating the average for each patch of the feature map. This means that each 2×2 square of the feature map is down sampled to the average value in the square. Once the higher level features are extracted, the output feature maps are flattened into a one-dimensional vector, followed by a fully connected output layer.

3.4.Fully Connected Layers

After pooling, there is always one or more fully connected layers. These layers perform the classification based on the features extracted from the image by the previously mentioned convolution processes. The last fully connected layer is the output layer which applies a softmax function to the output of the previous fully connected layer and returns a probability for each class.

4. Experimental Results and Analysis

4.1.Dataset Description

The two face mask classifier models were trained in the dataset. The dataset images for masked and unmasked faces were collected from image datasets available in the public domain, along with some data scraped from the Internet. Masked images were obtained from the Real-world Masked Face Recognition Dataset (RMFRD) (Wang, Z. et al., 2020) and Face Mask Detection dataset by Larxel on Kaggle (Larxel, 2020). RMFRD images were biased towards Asian faces. Thus, masked images from the Larxel (Kaggle) were added to the dataset to eliminate this bias. RMFRD contains images for unmasked faces as well. However, as mentioned before, they were heavily biased towards Asian faces.

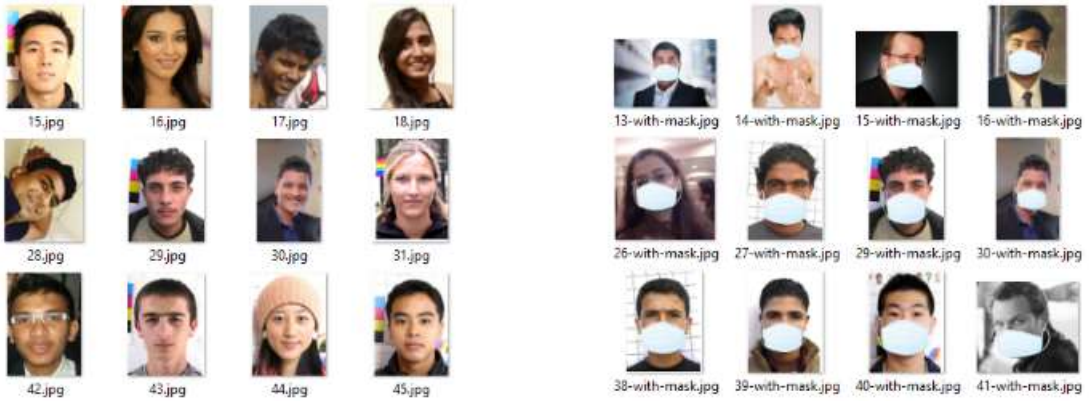

Figure 5 Sample dataset of with out and with masks

Table 1. Datasets

Class name	Description	No. of images
Mask	Faces with masks correctly used	690
Without mask	Faces with no masks or masks incorrectly used	686

Face Mask Classifier Model Training:

For the second stage, two CNN classifiers were trained for classifying images as masked or unmasked. The models were trained using the Keras framework. The dataset was split into train and test sets in a ratio of 75:25. That is partition the data into training and testing splits using 75% of the data for training and the remaining 25% for testing.

Data augmentation was performed using the ImageDataGenerator class in Keras. The input image size was set as 224 x 224. To selected an initial learning rate of 0.001. Besides this, the training process included checkpointing the weights for best loss, reducing the learning rate on plateau, and early stopping. Each model was trained for 20 epochs and the weights from the epoch with the lowest validation loss were selected. Based on a comparative analysis of performance.

4.3. Result Analysis

In this section, Tensorflow and Keras to create the CNN that classifies the images as with or without mask. Randomly split the dataset in separate train / test sets. Then call it twice (one for the images that contain a mask and one fot the images that do not) with a train / test split of 80% (80% used for training and 20% for test). The model consists of 10 layers in total. The first 6 layers form 3 sequential Convolution - ReLu - Pooling groups. Then, a flatten layer is applied to reshape the output of the CNN to a single dimension. After the flatten layer, a dropout layer is applied. This layer randomly drops 30% (rate = 0.3) of the tensors in order to avoid overfitting. In the end, a fully connected (dense) layer is applied that classifies the images based on the features extracted in the previous layers of the CNN and the final layer outputs the probability of each class label.

Train the model with the following function. First, open 2 training streams ("flows") from the 2 directories of train and test (validation) images. We also save checkpoints during training in separate directories for each checkpoint. Finally, it call the fit_generator function of the model and training begins. During the process, we keep track of training and validation accuracy and loss (we will use the values later to plot learning curves). Then label the outputs of the CNN ans apply colors to the results (red for without mask, green with mask).

The OpenCV framework to implement live face detection using the default webcam of the computer. It used the very common Haar Feature-based Cascade Classifiers for detecting the features of the face. This cascade classifier is designed by OpenCV to detect the frontal face by training thousands of images. The learning curves of the model for 75 epochs of training are shown as graph in Figure 6.

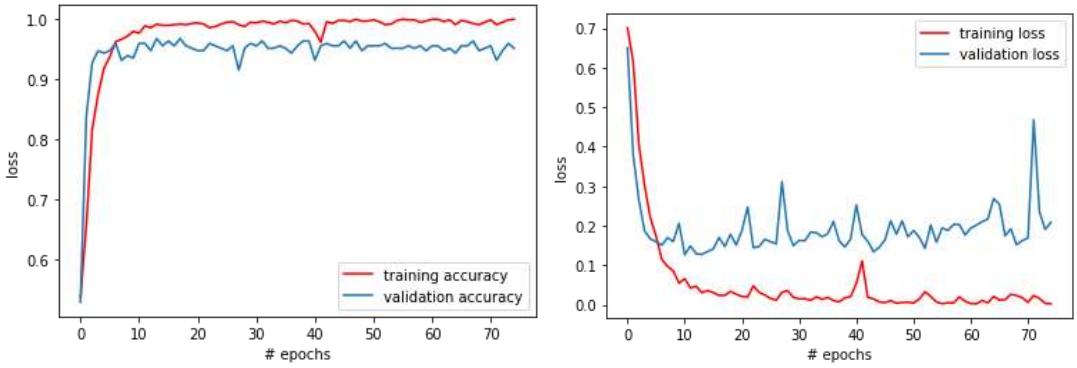

Figure 6 Training and validation accuracy and loss

From the Table 3, it can be seen that the proposed Masked Face detection system performed efficiently with more than 98% accuracy.

Figure 7 Screenshots of Mask and No Mask Detection with single person using CNN

Figure 8 Screenshots of No Mask, Partial Mask and Full Mask Detection with two persons using CNN

Figure 9 Screenshots of No Mask Detection with Partial Face (turned face) and Full Face using CNN

Evaluation Metrics

Table 2 Accuracy of the proposed Face Mask Detection system

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
10	0.9871	0.9435	0.0413	0.1689
20	0.9971	0.9556	0.0142	0.1621
30	0.9999	0.9677	0.0051	0.1579
40	0.9999	0.9677	0.0021	0.1649
50	0.9999	0.9718	0.0086	0.1503
60	1.0000	0.9723	0.0076	0.1550
70	0.9999	0.9865	0.0054	0.1435
75	1.0000	0.9923	0.0210	0.1123

Table 3 Comparison of proposed methods for Facial Mask Detection

Methods	Accuracy	
	Single Person	Multiple Person
CNN	95.0 %	94.7 %
VGG16	98.7 %	96.2 %

Conclusion

In real-world scenarios (i.e. unconstrained environments), human faces might be occluded by other objects such as facial mask. This makes the face recognition process a very challenging task. Consequently, current face detection methods will easily fail to make an efficient recognition. The proposed method improves the generalization of face detection process in the presence of the mask. To accomplish this task, a deep learning based method and quantization based technique is proposed to deal with the detection of the masked faces. The proposed method

can also be extended to richer applications such as violence video retrieval and video surveillance. The proposed method achieved a high detection performance. It is worth stating that this study is not limited to this pandemic period since a lot of people are self-aware constantly, they take care of their health and wear masks to protect themselves against pollution and to reduce other pathogens transmission. The future work aims at improving the masked face detection and also to recognize people with face mask in them.

References

1. Alyuz, B. Gokberk, and L. Akarun. 3-d face recognition under occlusion using masked projection. IEEE Transactions on Information Forensics and Security, 8(5):789–802, 2013. Bagchi,
2. D. Bhattacharjee, and M. Nasipuri. Robust 3d face recognition in presence of pose and partial occlusions or missing parts. arXiv preprint arXiv:1408.3709, 2014.
3. U. Din, K. Javed, S. Bae, and J. Yi. A novel gan-based network for unmasking of masked face. IEEE Access, 8:44276–44287, 2020.
4. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3d face recognition under expressions, occlusions, and pose variations. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(9):2270–2283, 2013.
5. Duan, J. Lu, J. Feng, and J. Zhou. Topology preserving structural matching for automatic partial face recognition. IEEE Transactions on Information Forensics and Security, 13(7):1823–1837, 2018.
6. P.Deepan and L.R. Sudha, "Comparative Analysis of Remote Sensing Images using Various Convolutional Neural Network", EAI End. Transaction on Cognitive Communications, 2021. ISSN: 2313-4534, doi: 10.4108/eai.11-2-2021.168714.
7. S. Gawali and R. R. Deshmukh. 3d face recognition using geodesic facial curves to handle expression, occlusion and pose variations. International Journal of Computer Science and Information Technologies, 5(3):4284–4287, 2014.
8. He, H. Li, Q. Zhang, and Z. Sun. Dynamic feature matching for partial face recognition.IEEE Transactions on Image Processing, 28(2):791–802, 2018.
9. E. King. Dlib-ml: A machine learning toolkit. The Journal of Machine Learning Research, 10:1755– 1758, 2009.
10. P.Deepan and L.R. Sudha, "Deep Learning and its Applications related to IoT and Computer Vision", Artificial Intelligence and IoT: Smart Convergence for Eco-friendly Topography, Springer Nature, pp. 223-244, 2021, https://doi.org/10.1007/978-981-33-6400-4_11.
11. L. Koudelka, M. W. Koch, and T. D. Russ. A prescreener for 3d face recognition using radial symmetry and the hausdorff fraction. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, pages 168–168.
12. P.Deepan and L.R. Sudha, "Object Classification of Remote Sensing Image Using Deep Convolutional Neural Network", The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems, pp.107-120, 2020. https://doi.org/10.1016/B978-0-12-816385-6.00008-8.
13. P.Deepan and L.R. Sudha, "Remote Sensing Image Scene Classification using Dilated Convolutional Neural Networks", International Journal of Emerging Trends in Engineering Research, Vol. 8, No.7, pp.3622-3630, 2020, ISSN: 2347-3983.

14. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classi¦cation with deep convolutional neural networks. In Advances in neural information processing systems, IEEE, pp. 1097–1105, 2012.
15. C. Lian, Z. Li, B.-L. Lu, and L. Zhang. Max-margin dictionary learning for multiclass image categorization. In European Conference on Computer Vision, pp. 157–170. Springer, 2010.
16. P. Deepan and L.R. Sudha, Effective utilization of YOLOv3 model for aircraft detection in Remotely Sensed Images, Materials Today: Proceedings, Elsevier Publisher, 20121, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.02.831
17. R. Santhoshkumar, M. Kalaiselvi Geetha, J. Arunnehru, 'SVM-KNN based Emotion Recognition of Human in Video using HOG feature and KLT Tracking Algorithm, *International Journal of Pure and Applied Mathematics*, vol. 117, No. 15, 2017, pp.621-624, ISSN: 1314-3395.
18. R. Santhoshkumar, M. Kalaiselvi Geetha, 'Deep Learning Approach: Emotion Recognition from Human Body Movements', *Journal of Mechanics of Continua and Mathematical Sciences (JMCMS)*, Vol.14, No.3, June 2019, pp.182-195, ISSN: 2454-7190.
19. R. Santhoshkumar, M. Kalaiselvi Geetha, 'Vision based Human Emotion Recognition using HOG-KLT feature*' Advances in Intelligent System and Computing, Lecture Notes in Networks and Systems*, Vol.121, pp.261-272, ISSN: 2194-5357, Springer https://doi.org/10.1007/978-981-15-3369-3_20
20. R. Santhoshkumar, M. Kalaiselvi Geetha, 'Human Emotion Prediction Using Body Expressive Feature', *Microservices in Big Data Analytics, IETE Springer Series*, ISSN 2524-5740, 2019, (Springer), https://doi.org/10.1007/978-981-15-0128-9_13
21. R. Santhoshkumar, M. Kalaiselvi Geetha, 'Emotion Recognition System for Autism Children Using Non-verbal Communication', *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol.8, No.8, June 2019, pp.159-165, ISSN: 2278-3075.
22. B.Rajalingam, R. Santhoshkumar (2020) "Intelligent Multimodal Medical Image Fusion with Deep Guided Filtering", Multimedia Systems, Springer-Verlag GmbH Germany, part of Springer Nature 2020
23. K.P. Sanal Kumar, S Anu H Nair, Deepsubhra Guha Roy, B. Rajalingam, R. Santhosh Kumar " Security and privacy-aware Artificial Intrusion Detection System using Federated Machine Learning" Computers & Electrical Engineering, Volume 96, Part A, December 2021, 107440
24. Dr. B. Rajalingam, Dr. R.Santhoshkumar, Dr. G. Govinda Rajulu, Dr. R. Vasanthselvakumar, Dr. G. JawaherlalNehru, Dr. P. Santosh Kumar Patra "Survey On Automatic Water Controlling System For Garden Using Internet Of Things (Iot)" The George Washington International Law Review, Vol.- 07 Issue -01 April-June 2021.