ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse33942020.pdf
https://doi.org/10.30534/ijatcse/2020/33942020

Real Time Multi-Scale Facial Mask Detection and Classification Using Deep Transfer Learning Techniques

Ssvr Kumar Addagarla[1], **G Kalyan Chakravarthi**[2], **P Anitha**[3]
[1,2,3] Department of Computer Science and Engineering, Gayatri Vidya Parishad College for
Degree and PG Courses (A), Visakhapatnam, India,
[1]ssvrkumar@gvpcdpgc.edu.in
[2]gkalyanchakravarthi@gvpcdpgc.edu.in
[3]anitha501p@gvpcdpgc.edu.in

ABSTRACT

In the era of deep learning, object detection plays an influential role for many industries. Detecting minute things are very much essential without human intervention especially at large scale industries. In this paper we have proposed multiple approaches for Multi-scale facial mask real time detection and classification for the hospital industry, crowd surveillance in the streets and malls are more useful in this COVID-19 Pandemic Situation. In our approach we have implemented two different detection models which are FMY3 using Yolov3 Algorithm and FMNMobile using NASNetMobile and Resnet_SSD300 Algorithms and used two different face mask dataset with 680 and 1400 images respectively. We have analyzed both the models by computing various probabilistic accuracy measures and achieved the 34% Mean Average Precision (mAP) and 91.7% Recall rate on FMY3 Model and achieved the 98% and 99% of accuracy and recall rate on FMNMobile Model. Finally we have shown results of various face mask detections from both the models.

Key words: Object Detection, Face Mask Detection, YoloV3, Resnet_SSD300 Face detection, NASNetMobile, Face Mask Crowd Surveillance.

1. INTRODUCTION

In the present world the need and role of Artificial Intelligence playing crucial role in many industries for the growth and performance of the productivity. Technological advancements in Computer Vision and Deep Learning a subset of the Artificial Intelligence gains more importance in the last decade [1]–[3]. Especially in the field of object detection using Convolutional Neural Networks (CNN) became popular in many fields to address the current societal issues. Many Applications like Employee Attendance, Robotics, Security Surveillance, Image based Content Retrieval, Number plate recognition for the Vehicles, Autonomous Vehicles, Tracking of various objects etc are being implemented [4]. In this paper we presented the Face Mask detection for the utilization in various industries. Moreover in the current Pandemic scenario the need of wearing face mask is essential to prevent the spread of Novel Corona Virus (Covid-19). Especially for the Health care industry face masks are very useful and to track the face masks through Surveillance Cameras.

Many conventional approaches are available for the object detection using Computer Vision but required the manual feature extraction and which is a hectic process. After evolving the usage of the deep learning using CNN based algorithms its became easier compare to the Computer vision based algorithm. Many CNN algorithms are like Spatial, Depth based, Multi-connection, Feature-map, Attention based CNNs are developed to address the various types of problems [5]. Here in our approach we have taken the NASNetMobile[6], Resnet[7], SSD300[8] and YoloV3 [9] are Deep Convolutional based networks for object detection and classification task[10]–[12].

Further we have organized our paper into three main sections namely Materials and methods, Results and discussion, and Conclusion.

2. RELATED WORKS

In this section we have reviewed various object detections techniques for various industrial applications. In [13] proposed a RPN network for object detection using Fast

RCNN. This network model uses VGG16 as the base architecture and achieved the accuracy on 73.2% and tested on benchmark PASCAL VOC datasets. In [14] compared several object detection models and obtained results for evaluation and shown a consistent progress in producing accurate results with better running time. In [15] authors presented a new approach using single stage detection for object detection model YOLO and authors compared with several existing models and produced better mAP. In [16] proposed a Multi Scale CNN for faster and accurate object detection which the network detects the objects at multiple scales in output layers. Object detection performance is computed on KITTI and Caltech datasets. In [17] authors has proposed a new algorithm for facial expression detections for humans called Iterated Local Search and Genetic Algorithms with Back Propagation (ILSGA-BP). For the evaluation of the algorithm authors used FEEDTUM data set for the train and test process. Model achieves higher accuracy compared with back propagation.

In [18] proposed a framework using CNN and RNN for characterize the semantic label dependency and image label relevance. Testing is done on the developed framework and Experimentation is done with public benchmark datasets and demonstrated the proposed architecture yields better results. In [19] proposed a classification framework using spectral spatial features. Experimentation is done on hyper spectral datasets and presented and compared the performance results with commonly used methods for hyper spectral classification. In [20] investigated the classification of the quality of wood boards based on their images using texture feature extraction along with CNN . For the classification compared deep learning techniques with existing traditional algorithms. In [21] proposed a robust deep face detection approach based on R-CNN. Experiments are carried out on FDDB and WIDER FACE-challenging detection benchmarks to explain the efficiency of the approach. In [22] has investigated various algorithms used for the face recognition. authors examined face recognition for supermarket system at the bill desk which automatically detects the customer and based on the number of times he visited the supermarket discounts will be given on the items as a loyalty. The data obtained by the system is used for the business analytics.

In [23] proposed VSFR a spatial resolution algorithm which uses for remotely sensed imagery and tested on rural and urban areas with a satellite sensor dataset and achieved a great performance. In [24]proposed a Deep Coupled ResNet (DCR) model consists of a Trunk network which is trained on three different resolutions and extract the different features of face images. Evaluation of the model shown the DCR model has better performance. In [25]come up with FDNet1.0 a framework model for face detection which is improves Faster RCNN for a better performance. The proposed work is tested on WIDER FACE data set and validated the algorithm. A survey on object detection using deep learning which highlighted the achievements provided a structural taxonomy for the methods according to their roles in detection, datasets and investigated the performance of the various representative methods [4]. In [10] focused on the various CNN applications to image classification and developed deep learning models to the latest ones and analyzed their advantages. In [26] presented a dataset OIDV4 with unified annotations for classification of image, detection of object and detection of visual relationship. For the detection of objects authors provided 15 times more bounding boxes and annotated visual relationship between the objects for detection of visual relationship.

3. RESEARCH METHODS

We have proposed two different multi-scale Face Mask Detections methods 1) Face Mask using YoloV3(FMY3) and 2) Face Mask Using NASNetMobile (FMNMobile). These two approaches rely on the CNN based transfer learning method where as we have taken the pre-trained network models and modified the models according to our problem requirement. In this section we have discussed various algorithms which we have used in our approach. In the Figure 1, we presented our proposed structure for the FMY3 and FMNMobile.

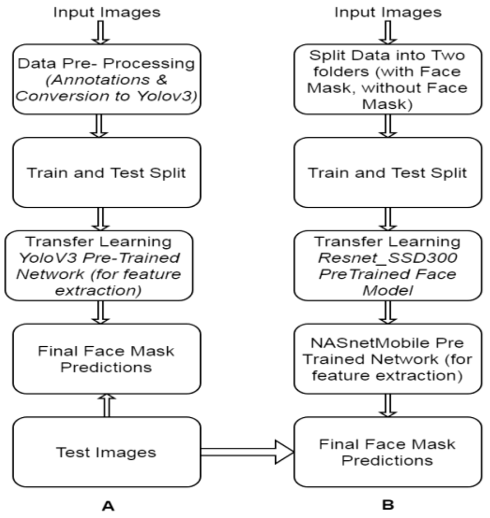
Input Images
Data Pre- Processing (Annotations & Conversion to Yolov3)
Train and Test Split
Transfer Learning YoloV3 Pre-Trained Network (for feature extraction)
Final Face Mask Predictions
Test Images
A
Input Images
Split Data into Two folders (with Face Mask, without Face Mask)
Train and Test Split
Transfer Learning Resnet_SSD300 PreTrained Face Model
NASnetMobile Pre Trained Network (for feature extraction)
Final Face Mask Predictions
B

Figure 1: Proposed Face Mask Detection Model using A) FMY3 B) FMNMobile

A) Yolov3

You only look once (Yolo) which is state-of-the-art single stage object detection technique which uses the Darknet[27] as the backbone architecture. In our process we have considered the Yolov3 initially which uses the Darknet-53 as the backbone network which has 53 convolution layers but for the various task detections it have 53 more layers and a total of 106 convolution layers uses in Yolov3. Yolo algorithm is popularly known for its faster object detection compares to the two stage detector like Fast RCNN[28], Faster RCNN[13] etc. Yolov3 utilizes the input size 416×416 and passed to the network. Yolo feature map kernel is 1×1 and detection kernel shape is 1×1×(B×(5+C) where B indicates the bounding boxes on the feature map and C indicates total number of classes and '5' consist of one object predictor and 4 bounding box attributes. Yolov3 initializes with 9 anchor boxes and which were predicated on our dataset using K-Means clustering algorithm. Yolov3 uses the Non max suppression (NMS) and Generalized Intersection of Union (GIoU) for detecting accurate multi objects in the give image. At the final layers of the network YoloV3 uses the logistic linear regression instead of the Softmax as the predictor. Here logistic linear regression performs well on the multi-scale object detector using minimum confidence threshold set by the user.

B) Single Shot Detector (SSD) and ResNet

SSD is another variant for the one shot detection algorithm which doesn't have the Region Proposal network in its model. SSD achieves the better mAP on benchmark PASCAL VOC Datasets.SSD has also applicable for the multi-class object detector which uses M×N feature map and P number of channels. In the Figure 2 [29], presented SSD-Resent architecture process flow. SSD utilizes the VGG16/19 as backbone architecture for many cases. For our face detection approach we have uses Residual Network (ResNet) as the back bone network for the SSD as the feature extractor. ResNet provides the Skip Connections which works better than the VGG network.

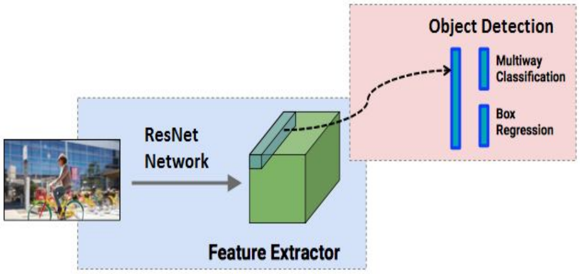

Figure 2: Single Shot Detector Process Flow

In resent layers go deeper up to 152 whereas in the VGG it's up to 19 only. In the Figure 3, presented the ResNet architecture.

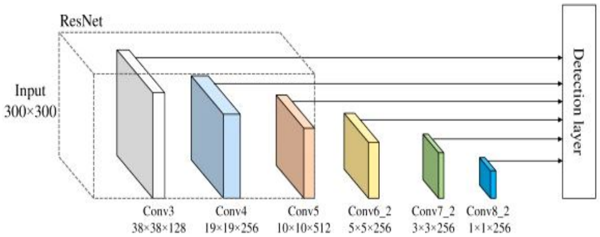

Figure 3: Resent Architecture

C) NASNetMobile

Neural Architecture Search Network (Nasnet) was developed by Google brain team, which uses the two main functionalities are 1) Normal cell 2) Reduction cell which shown in image 4. Initially Nasnet applies its operations on the small dataset and then transfer its block to the large dataset to achieve the higher mAP. A modified droppath called Scheduled droppath for effective regularization is used for improved performance of Nasnet. In the original Nasnet Architecture where the number of cells are not pre-defined and specifically normal and reduction cells are used and architecture show in Figure 4[6]. where as normal cells defines the feature map size and reduction cell returns the reduce feature map in terms of height and width by the factor of two. A Control architecture in Nasnet based on Recurrent Neural network (RNN) is used predicts the entire structure for the network based on the two initial hidden states.

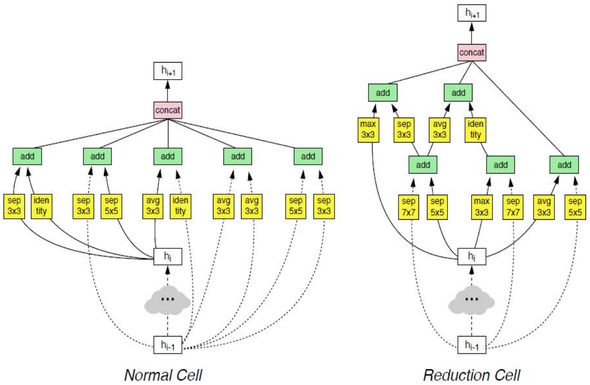

Figure 4: Nasnet Normal and Reduction Cell Architecture

Controller Architecture uses the RNN based LSTM model and Softmax prediction is used for the Convolutional cells prediction and constructed the network motifs recursively. Here in our model NASNetMobile which uses 224×224 as input image size and NASnettLarge model uses 331×331 as the input image size. NASNetMobile uses the Imagenet pre-trained network weights for transfer learning process to detect the face masks.

4. EXPERIMENTATION

In this section, we have trained two different face mask detection with different deep transfer learning approaches. First experiment done using Yolov3 object detection algorithm and second experiment using NASNetMobile model and Resnet_SSD_300 pre-trained model.

A) Face-Mask-Yolov3 (FMY3):

In this process we have used the state of the art Yolov3 object detection algorithm to detect the various states of the face mask for the given input image. To train our face mask detection model we have used the 680 various face mask image dataset. Further we manually annotate the each image using LabelImg tool and generated the annotated XML files for each corresponding image. We categorized objects in the images as 3 class labels namely Good, Bad, None. 'Good' indicates those wear the mask properly, 'Bad' indicates those who haven't wear the mask and 'None' indicates those who haven't properly wear the mask. Then we converted the XML annotated files to the Yolo required format and divided into 440 images into train dataset and 240 images into test dataset (80:20). we have modified some of the Yolov3 configuration parameters in order to achieve the better object detection while training and testing on the face mask image dataset. The following Table 1 shows some of the hyper parameters for the Yolov3 algorithm.

Table 1: Face-Mask Detection Yolov3 Hyper Parameters

Hyper-parameter Name	Value
GIoU Loss gain	3.5
Cls loss gain	37.4
Obj Loss gain	64.3
Lr0 (initial learning rate)	0.005
Image translate	0.05
hsv_h, hsv_s, hsv_v (image augmentation)	0.0138, 0.678, 0.36
Nms-thres	0.5

B) Face-Mask-NASNetMobile(FMNMobile)

In this process we have considered the NASNetMobile as the backbone network architecture to train the face mask images. Here we have gathered 1400 images of with face mask and without face mask. Then we have separated into two different folders. Here we have used the Pre-trained Resnet-SSD100_caffee face detection model to detect the faces in the given input image. For both the experimentations we have used Intel Core I7-8700K processor with 16 GB of RAM and Nvidia RTX 2070 8GB GPU acceleration also been utilized.

5. RESULTS

For both the FMY3 and FMNMobile models we ran 20 epochs to train and validate the models and achieved the considerable detection accuracy. To assess the FMY3 model performance we have computed Precision, Recall , F1 measure, and mean Average Precision (mAP). In FMY3 we have achieved the 45% of mAP and 80% of Recall rate. The following shows various detection results for FMY3 model. In Table 2, presented the various accuracy measures computed using FMY3 Model and shown the sample batch training process in Figure 5, various model parameter plots while training the model in Figure 6 and Figure 7 and sample predictions of the trained model in Figure 8.

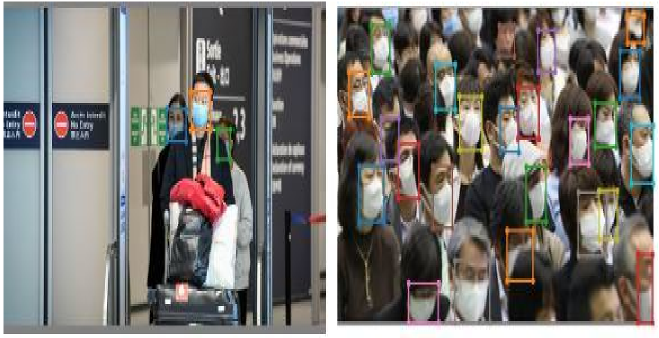

Figure 5: FMY3-Bounding box detection during testing from batch of images

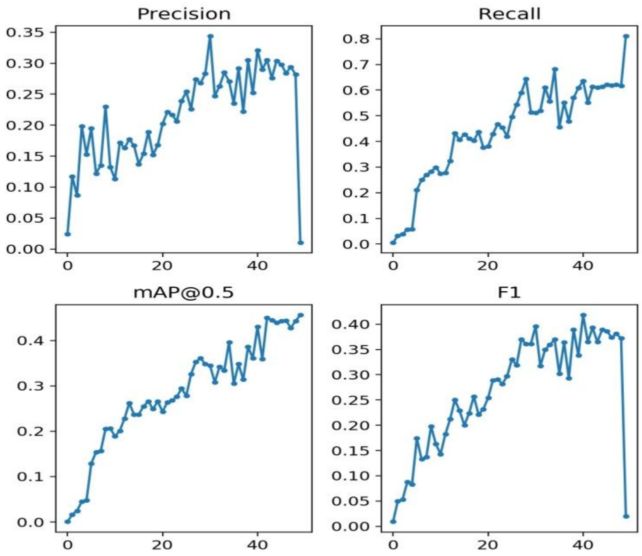

Figure 6: FMY3-Training Results for Accuracy

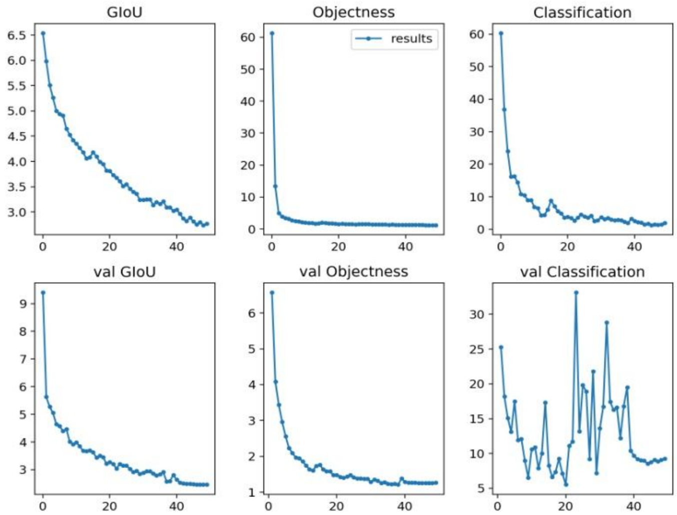

Figure 7: FMY3-Training Results for Generalized IoU, Objecteness, Classification

Table 2: Accuracy Measures For The Fmy3 Model

Class	Recall	MAP@0.5
All	**91.7**	**34.0**
good	93.9	71.7
bad	81.1	12.3
none	100	18.1

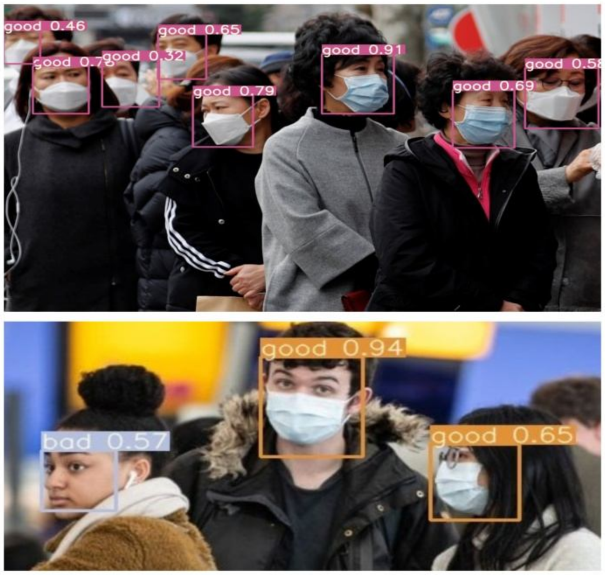

Figure 8: FMY3 Multi-scale Face Mask Prediction and classification

In the Table 3, presented various measures of our trained FMNMobile detection. we have computed standard measures like Precision, Recall, F1-Score, Macro average and weighted average for better understanding of the model. Training and validation loss show in the Figure 9 and sample predictions for the FMNMobile shown in Figure 10.

Table 3: Accuracy Measures For The FMNMobile Model

Parameter	Precision	Recall	F1-score
With_mask	0.99	0.98	0.99
Without_mask	0.99	0.99	0.99
Accuracy	--	--	0.99
Macro_avg	0.99	0.99	0.99
Weighted_avg	0.99	0.99	0.99

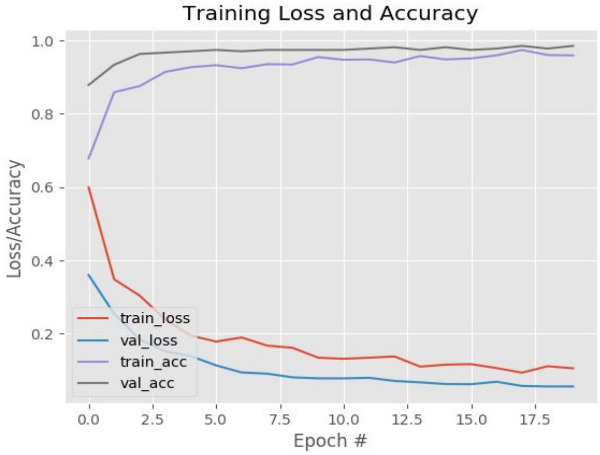

Figure 9: FMNMobile- Training and Validation Curve (loss and Accuracy)

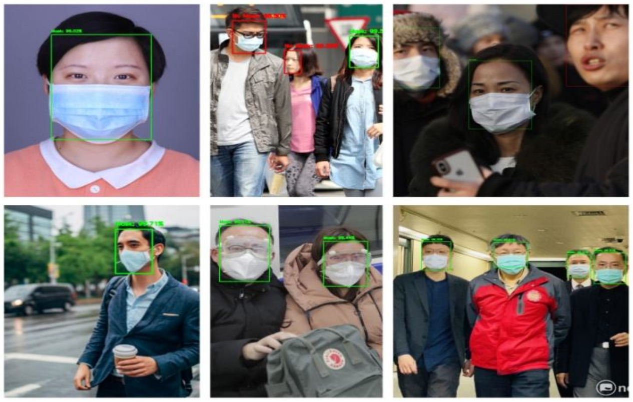

Figure 10: FMNMobile Multi-Scale Mask Predictions and classification

6. CONCLUSION

In both FMY3 and FMNMobile detection models are achieved the good accuracy, where as in the FMY3 full depends on the annotations of the mask in the images. The main reason for choosing another Model (FMNMobile) is initially we detect the human faces using state of the art Resnet_SSD300 model and then NASNetMobile as backbone architecture used to predicts the face masks for classification task. FMNMobile achieves higher accuracy with recall rate of 98% compares to the FMY3. In the Image 8 and Image 10 its clearly projects the slight difference amount bounding box predications and in the case of FMNMobilenet bounding boxes are predicted pretty well. Whereas FMY3 model can be used alone to detect face masks in a crowd in real time video processing and achieved good recall rate of 91.7%. Further FMY3 model accuracy can be improved by considering more face mask images and applying image augmentation techniques.

REFERENCES

1. Y. LeCun, Y. Bengio, and G. Hinton, "**Deep learning,**" *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. https://doi.org/10.1038/nature14539
2. C. A. Llorente and E. P. Dadios, "**Development and characterization of a computer vision system for human body detection and tracking under low-light condition,**" *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 2, pp. 251–254, 2019. https://doi.org/10.30534/ijatcse/2019/24822019
3. M. Akour, O. Al Qasem, H. Alsghaier, and K. Al-Radaideh, "**The effectiveness of using deep learning algorithms in predicting daily activities,**" *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2231–2235, 2019. https://doi.org/10.30534/ijatcse/2019/57852019
4. L. Liu *et al.*, "**Deep learning for generic object detection: A survey,**" *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, 2020. https://doi.org/10.1007/s11263-019-01247-4
5. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
6. B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "**Learning Transferable Architectures for Scalable Image Recognition,**" *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8697–8710, 2018.
7. K. He, X. Zhang, S. Ren, and J. Sun, "**Deep residual learning for image recognition,**" in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
8. W. Liu *et al.*, "**Ssd: Single shot multibox detector,**" in *European conference on computer vision*, 2016, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
9. J. Redmon and A. Farhadi, "**Yolov3: An incremental improvement,**" *arXiv Prepr. arXiv1804.02767*, 2018.
10. H. Salman, J. Grover, and T. Shankar, "**Hierarchical Reinforcement Learning for Sequencing Behaviors,**" vol. 2733, pp. 2709–2733, 2018.
11. V. Bharat and D. Malik, "**Study of Detection of Various types of Cancers by using Deep Learning: A Survey,**" *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 4, pp. 1228–1233, 2019. https://doi.org/10.30534/ijatcse/2019/31842019

\12. A. C. Calvin Ng, "**Training of a deep learning algorithm for quadcopter gesture recognition,**" *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 211–216, 2020. https://doi.org/10.30534/ijatcse/2020/32912020

13. S. Ren, K. He, R. Girshick, and J. Sun, "**Faster r-cnn: Towards real-time object detection with region proposal networks,**" in *Advances in neural information processing systems*, 2015, pp. 91–99.
14. A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "**Salient object detection: A benchmark,**" *IEEE Trans. image Process.*, vol. 24, no. 12, pp. 5706–5722, 2015.
15. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "**You only look once: Unified, real-time object detection,**" in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
16. and N. V. 1 Zhaowei Cai1(B), Quanfu Fan2, Rogerio S. Feris2, "**A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection Zhaowei,**" *Springer*, vol. 9914 LNCS, no. 17, p. V, 2016.
17. M. Alsmadi, "**Facial recognition under expression variations.,**" *Int. Arab J. Inf. Technol.*, vol. 13, no. 1A, pp. 133–141, 2016.
18. J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "**Cnn-rnn: A unified framework for multi-label image classification,**" in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2285–2294.
19. W. Zhao and S. Du, "**Spectral--spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach,**" *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, 2016. https://doi.org/10.1109/TGRS.2016.2543748
20. C. Affonso, A. L. D. Rossi, F. H. A. Vieira, A. C. P. de Leon Ferreira, and others, "**Deep learning for biological image classification,**" *Expert Syst. Appl.*, vol. 85, pp. 114–122, 2017.
21. H. Wang, Z. Li, X. Ji, and Y. Wang, "**Face r-cnn,**" *arXiv Prepr. arXiv1706.01061*, 2017.
22. P. Sanmoy, S. Acharya, and K. Bhuva, "**A Review on Facial Recognition Algorithm; Their Application in Retail Stores,**" *Acad. Mark. Stud. J.*, vol. 22, no. 3, pp. 1–8, 2018.
23. C. Zhang *et al.*, "**A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification,**" *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 133–144, 2018.

https://doi.org/10.1016/j.isprsjprs.2017.07.014

24. Z. Lu, X. Jiang, and A. Kot, "**Deep coupled resnet for low-resolution face recognition,**" *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 526–530, 2018.
25. C. Zhang, X. Xu, and D. Tu, "**Face detection using improved faster rcnn,**" *arXiv Prepr. arXiv1802.02142*, 2018.
26. A. Kuznetsova *et al.*, "**The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale,**" *arXiv Prepr. arXiv1811.00982*, 2018.
27. J. Redmon and A. Farhadi, "**YOLOv3: An Incremental Improvement,**" 2018.
28. R. Girshick, "**Fast R-CNN,**" *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1440–1448, 2015. https://doi.org/10.1109/ICCV.2015.169
29. J. Huang *et al.*, "**Speed/accuracy trade-offs for modern convolutional object detectors,**" *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3296–3305, 2017. https://doi.org/10.1109/CVPR.2017.351