Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Intelligence-Based Medicine 5 (2021) 100037

Contents lists available at ScienceDirect

Intelligence-Based Medicine

journal homepage: www.sciencedirect.com/journal/intelligence-based-medicine

A novel data augmentation approach for mask detection using deep transfer learning

Manas Ranjan Prusty [a,b,*], Vaibhav Tripathi [c], Anmol Dubey [b]

[a] Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, 600127, India
[b] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, 600127, India
[c] School of Electronics Engineering, Vellore Institute of Technology, Chennai, 600127, India

ARTICLE INFO

Keywords:
Mask detection
Safety in public places
COVID-19
YOLOv3
Pre-processing
Object detection
Deep transfer learning

ABSTRACT

At the onset of 2020, the world saw the rise and spread of a global pandemic named COVID-19 which caused numerous deaths and affected millions of people around the world. Due to its highly contagious nature, this disease spread across the world within a short span of time. It forced almost all the nations to implement strict social distancing rules along with use of face masks to reduce the risk of getting infected. While the virus is still on loose, markets and business firms have reopened to keep the economy alive. This calls for modification of existing technological models to cater for the safety of individuals and stop the spread of virus in public places. One such stringent implementation to achieve this safety would be deployment of a mask detection model. The proposed mask detection models can serve as a vital utility in the coming years for ensuring proper enforcement of safety protocols. This research paper explores the use of state of the art YOLOv3 model, a deep transfer learning object detection technique, to develop a mask detection model. Along with the implementation of a standard approach of any object detection algorithm, this paper has proposed the use of a data augmentation approach for mask detection. The proposed model focuses on generating an augmented dataset from the standard dataset with the help of data augmentation done by using image filtering techniques such as grayscale and Gaussian blur. This augmented dataset is used for training the object detection model for mask detection. The mean average precision for the Data augmentation based mask detection model is observed to be 99.8% while training. Finally, a comparison on the model performance is evaluated for the standard and proposed augmented data approach. The experiment conducted showed that the average confidence level for Standard mask detection model was 0.94, 0.93, 0.91 for images of individuals (type A), images with groups of people (type B) and video with the group of people (type C) respectively. The average confidence levels for the Data augmentation based mask detection model for types A, B and C are 0.97, 0.96 0.93 respectively. This paper therefore concludes that the proposed Data augmentation based mask detection model performs better than the Standard mask detection model.

1. Introduction

As mentioned by the World Health Organization (WHO), the coronavirus disease in 2019 otherwise called as COVID-19 declared as a pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This is a highly infectious disease and spreads primarily via small droplets from coughing, sneezing, and talking. The droplets are usually not airborne, however those standing in close proximity have a high probability of getting infected [1,2]. COVID-19 is a highly contagious disease which spreads from person to person within close vicinity. In order to control the spread of this disease, it is advised to keep social distance, wearing a mask and using sanitizers [3,4]. There has been a lot of debate on whether masks should be worn or not. The truth is, no single preventive action holds the golden key in this disease prevention within the context of proper infection control. Each action contributes significantly to the process and complements the other in the containment of COVID-19 virus and ceases its spread.

That being said, the use of face masks is the single greatest preventive measure to save oneself in this pandemic struck world. Although public places have now been opened up for the sake of the economy, the number of cases in every country is very high. According to infectious disease expert and senior scholar at the Johns Hopkins Center for Health

* Corresponding author. Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, 600127, India.
E-mail address: manas.iter144@gmail.com (M.R. Prusty).

https://doi.org/10.1016/j.ibmed.2021.100037
Received 11 October 2020; Received in revised form 8 April 2021; Accepted 16 June 2021
Available online 22 June 2021

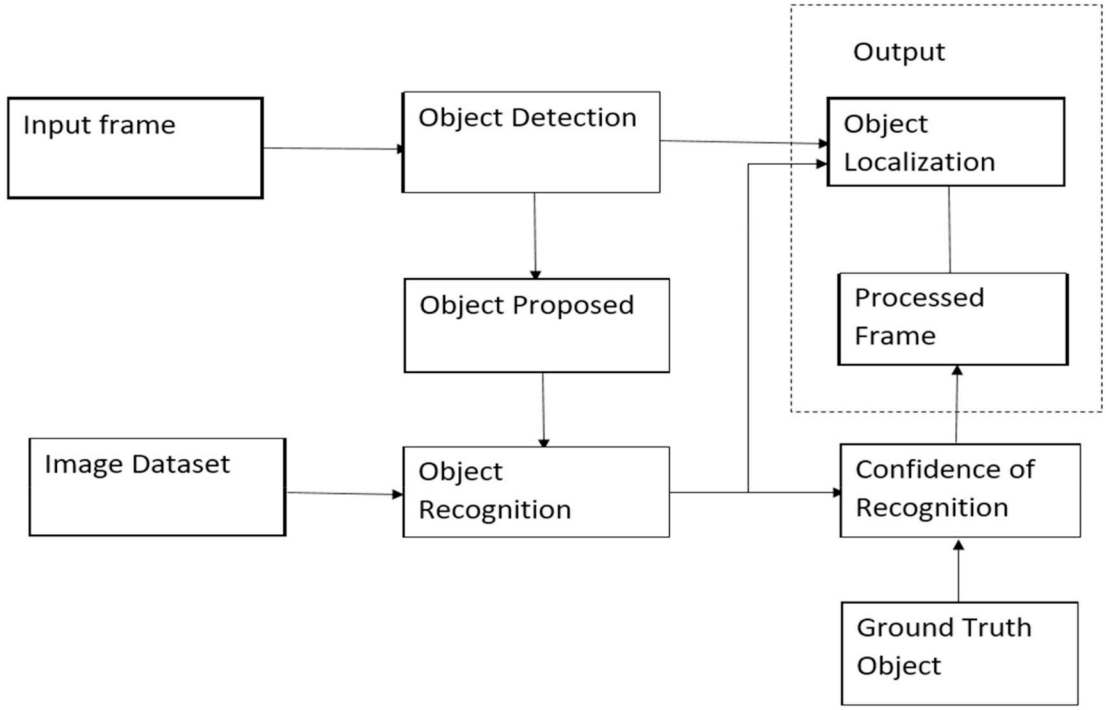

Fig. 1. Object detection model.

Security, Amesh A. Adalja, MD, "Face masks can help protect against many respiratory infections that are spread through the droplet route, and that includes coronavirus and the flu." Speaking to the basis of this recommendation, Dr. Adalja stated that viruses such as the corona virus can spread from an infected person to others through the air by coughing and sneezing or by touching a contaminated surface and then touching your mouth, nose, or eyes prior to hand washing. When a face mask is worn, one can prevent those droplets from coming into contact with one's face or mouth before dropping to the ground [4].

Though people are instructed to wear masks as a protective measure in public places such as malls, workplaces, etc., there are many who still do not abide by the instructions. Most complexes and firms rely on guards to keep a check on such people, which exposes them to a high risk. By automating the process of mask detection, the risk can be reduced. This can be achieved by using object detection algorithms [5] to automate the access to any premise only to people wearing masks. Object detection algorithms have been widely used in the last decade for detecting various objects like Military Gun detection [6] as well as medical purposes for cancerous cell detection [7] and other abnormalities.

This paper focuses on developing an efficient model for mask recognition [8] using an object detection algorithm otherwise called a mask detection model (MDM). The proposed model is trained on a dataset of masked and unmasked people with the usage of a proper pre-processing technique [9]. This model aims at ensuring a more efficient approach for achieving the required automation. This automation would help humans adapt to the new normal and ensure their safety from any such virus in future. It is also another step to automation using AI while keeping in mind the requirements of changing social convention with time and need [10].

This paper will walk through the basic architecture of the chosen objection detection algorithm along with the appropriate classifier in section 2. Section 3 of this paper elucidates on the proposed model for mask detection. The experimental procedure is explained in section 4 of this paper. Section 5 elaborates on the result analysis of the basic model and proposed models and compares their performances. Finally, section 6 draws the conclusion of this paper.

2. Object detection algorithms

Object Detection [5,11] is a deep neural network technique associated with computer vision and image processing that performs the task of detecting instances of certain objects such as a human, vehicle, banner, building, etc. from a digital image or a video. Object detection combined with other advanced technology integrations allows us to perform face detection or pedestrian detection, popularly known as person tracking from a video. This technique is being used in a plethora of areas such as security [12], healthcare [13] and manufacturing [14]. In recent times, object detection algorithms have played a vital role in cancer diagnosis [15], virus detection [16] and other disease diagnosis/analysis which helped doctors and medical science researchers around the world. Research in computer vision focusing on object detection is growing rapidly. Initially, traditional image processing techniques like pixel-by-pixel object matching were used. Later Region-based algorithms were developed like R–CNN, Fast R–CNN, Faster R–CNN and You Only Look Once (YOLO) are a few examples. The algorithm used in this project is YOLO because of its ability to see the entire image during training and testing time hence it gets every detail about the whole image and the object all at once, hence justifying its name. Also YOLO is significantly faster than Faster R–CNN as it is trained to do classification and bounding box regression, which is predicting localization boxes, at the same time [17]. Fig. 1 depicts the general working of an object detection model.

2.1. YOLOv3

You Only Look Once version 3 (YOLOv3) [18] is one of the fastest object detection algorithm. YOLOv3 uses a single convolutional neural network (CNN) [19] to the full image and then divides the images into various regions to predict bounding boxes. This makes the CNN more efficient as multiple predictions can be made at once. YOLOv3 utilizes a regional proposal algorithm which takes an image input and returns the output image with bounding boxes plotted for all patches in an image which are most likely to be the object. This is a faster process as compared to the traditional sliding window algorithm [20], which is an

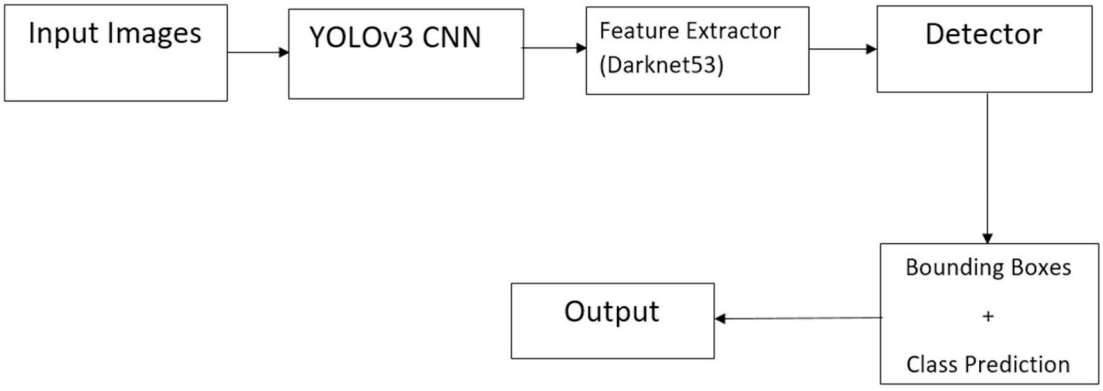

Fig. 2. Standard YOLOv3 object detection model.

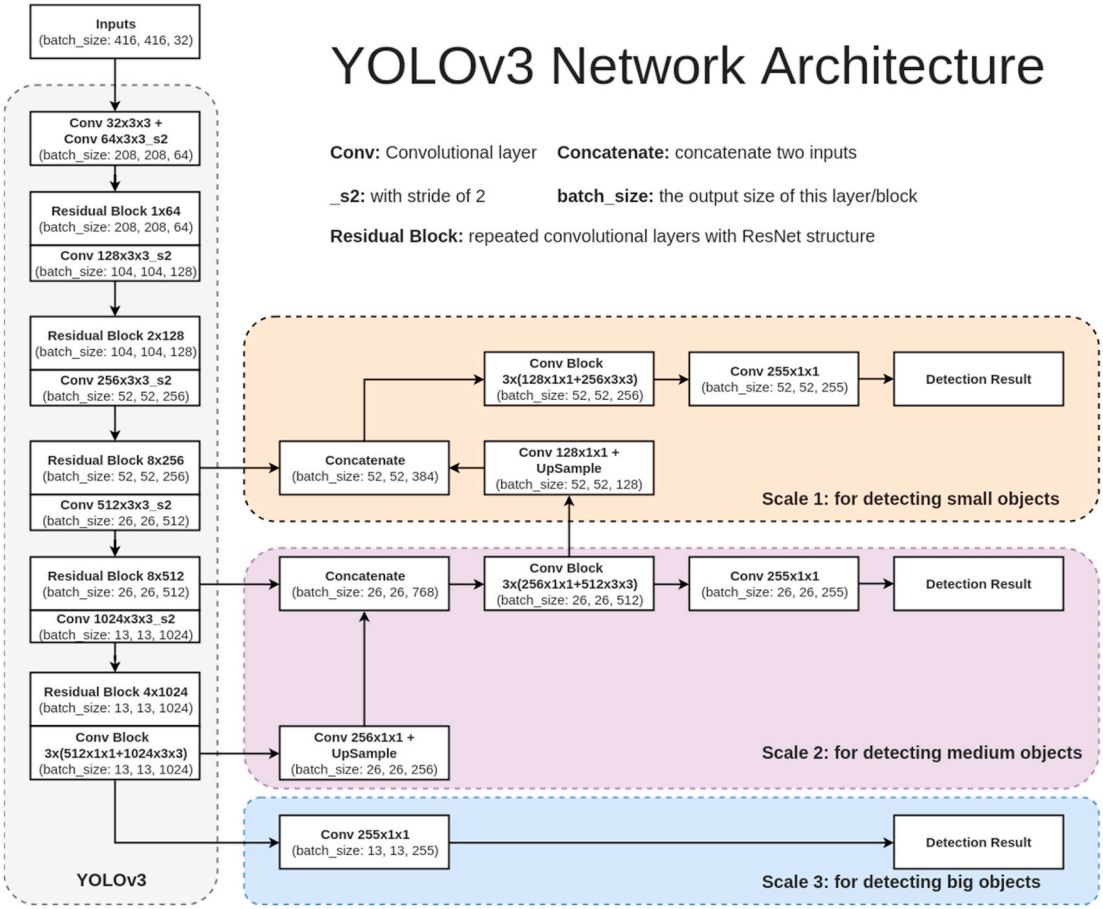

Fig. 3. YOLOv3 network architecture.

exhaustive search for objects as the entire image is searched pixel by pixel. While detecting the object, some regions may be noisy, overlapping or unclear which might be wrongly detected as the object. YOLOv3's regional proposal will counter this issue by returning the region closest to the detected object in the image with the highest probability score. Another feature of this regional proposal method is its high recall rate which enhances the performance of the model. A high recall rate ensures generation of the list of regions containing the object, facilitating higher detection accuracy for multiple objects in a single frame [18].

YOLOv3 divides the image into a 13×13 grid of cells and all the 169 cells are responsible for the bounding boxes that are to be detected in the image. Along with the location of the bounding boxes, a confidence of probability is also predicted. (Confidence level is the probability of detecting the class of an object in an image). Since most of these confidence values fail to reach the threshold, they are eliminated, yet there might exist some discrepancies in the objects detected. In this paper, an attempt has been made to reduce such discrepancies and improve the accuracy by pre-processing the images before feeding it into the YOLOv3 model. Fig. 2 depicts the working of the YOLO model for training the datasets. The input images are convolved using various layers and the loaded Darknet-53 (discussed in detail in section 2.2) extracts the features of the image. After the training is completed using the detector function the output is generated with boundary boxes and class prediction.

Fig. 3 is a diagrammatic representation of YOLOv3 network

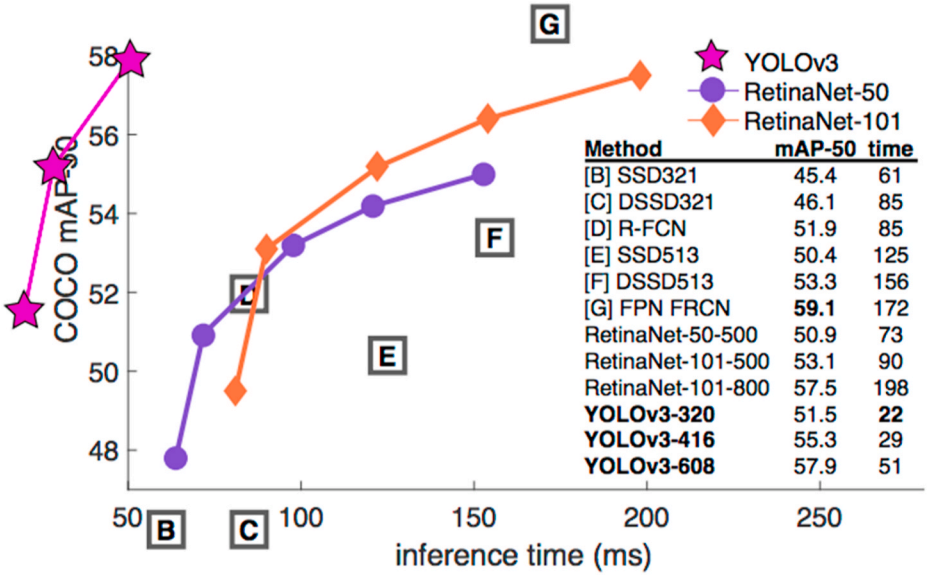

Fig. 4. YOLO vs RetinaNet performance on COCO 50 Benchmark.

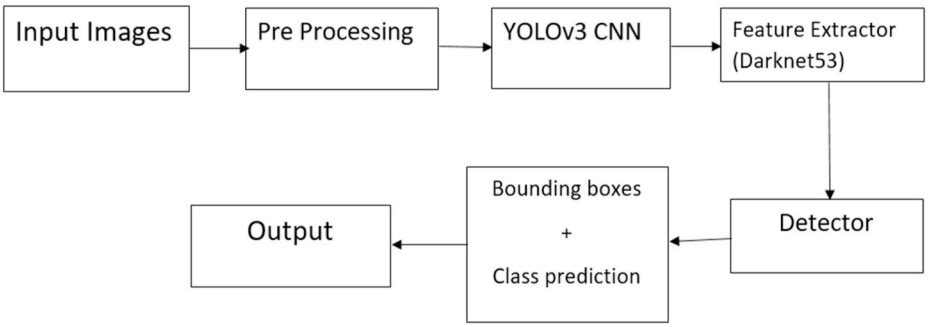

Fig. 5. Proposed YOLOv3 object detection model.

architecture [21]. It represents how different numbers of convolutional layers are used for detecting small, medium and large objects respectively.

Fig. 4 shows how YOLOv3 works better than other state of art detectors like SSD and its variants at COCO mAP 50 benchmark dataset. 50 here denotes 0.5 IoU (Intersection over Union) [22]. An IoU less than 0.5 is classified as a mislocalization of an object and a false positive. It is a simple evaluation metric to evaluate the object detection model. It is calculated as the ratio of area of overlap and area of union of the ground-truth bounding box and predicted bounding box. The ground-truth bounding box refers to the manually drawn bounding box while annotating the test images whereas, the predicted bounding box is the result of prediction from the model. It is evident from the graph in Fig. 4 that YoloV3 (denoted by the pink line) has the least inference time as compared to other models and therefore it is significantly faster.

2.2. Darknet 53

YOLOv3 uses a new network- Darknet-53 [22]. DarkNet-53 is a convolutional neural network that is 53 layers deep. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. Darknet-53 has the fastest floating point operations per second as compared to ResNet-101 and ResNet-152 which have more inefficient layers, making them slower. Thus, Darknet-53 network structure optimally utilizes the GPU and produces results similar to ResNet-152 but with twice the speed [17,23]. Darknet 53 is used as the transfer learning model for training the standard and proposed MDM in this paper.

3. The proposed mask detection model

The authors have proposed the use of pre-processed images for the mask detection model (MDM) as shown in Fig. 5. The pre-processing block includes preparation of an augmented dataset. This augmented dataset includes the standard dataset along with another set created by some filtering techniques. Two such filtering techniques that are performed on the dataset to facilitate enhanced performance of the model are grayscale [24] and Gaussian blur [25].

3.1. Grayscale

A grayscale image is an image with shades of gray only. These images are easy to process for extracting information as their individual pixels carry less information as compared to their coloured counterparts [26]. They have only one colour, i.e., 'gray'. In 'gray' colour all the three red, green and blue components have equal intensity in RGB space, therefore only one intensity value needs to be assigned to all the components, as opposed to a coloured image. There are 256 different shades of grayscale intensities possible ranging from black to white. These intensities are stored as 8-bit integers. Grayscale images are extremely popular as many display and image capture devices can only support 8-bit images. Also, 8-bits are sufficient enough to extract the features for many tasks. Therefore, these are used more when compared to colour images, which

Table 1
Number of masked and unmasked face instances in the dataset.

	Individual	Group	Average (for groups)
Masked	250	725	5
Unmasked	200	650	4

are harder to process.

3.2. Gaussian Blur

Most images are prone to noises and distortions which makes it hard for the object detection model to detect edges. Gaussian blur effect when applied to images reduces the noise in the image and smoothens the edges, making it easier for the model to detect the edges and produce a more accurate region of detection [27]. It uses a Gaussian filter to apply transformation to all the pixels in the image. The formula of a Gaussian function in one dimension is described by equation (1).

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{x^2}{2\sigma^2}}$$

Equation (1). Gaussian function in one dimension.

In two dimensions, it is the product of two such Gaussian functions, one in each dimension as shown in equation (2).

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Equation (2). Gaussian function in two dimension.

Here x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution. Values from this distribution are used to build a convolution matrix which is applied to the original image.

This transformation blurs the original image and all the pixels with object edges become smoother. It makes edge detection easier for the object detection algorithm and hence, more accurate bounding boxes can be plotted.

The authors considered other pre-processing techniques as well such as rotation, contrast change etc. but those techniques produced little increase in the overall accuracy of prediction while increasing redundant data fed to the model. Instead of adding images obtained from such pre-processing techniques, the authors prepared the dataset with images of varying size, resolution, brightness, contrast and object count as discussed in section 4.1 along with the pre-processed images to improve

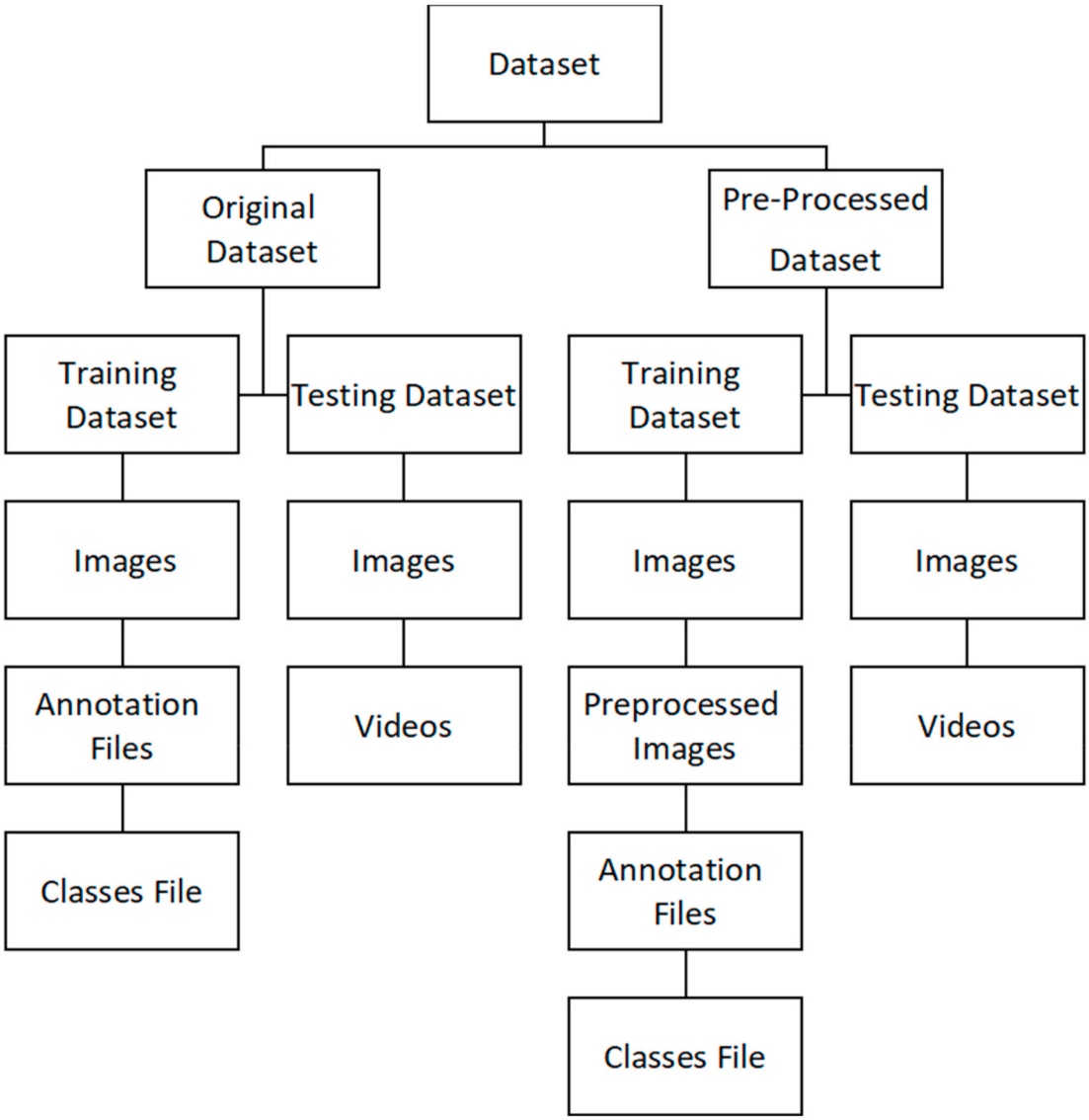

Fig. 6. Dataset classification.

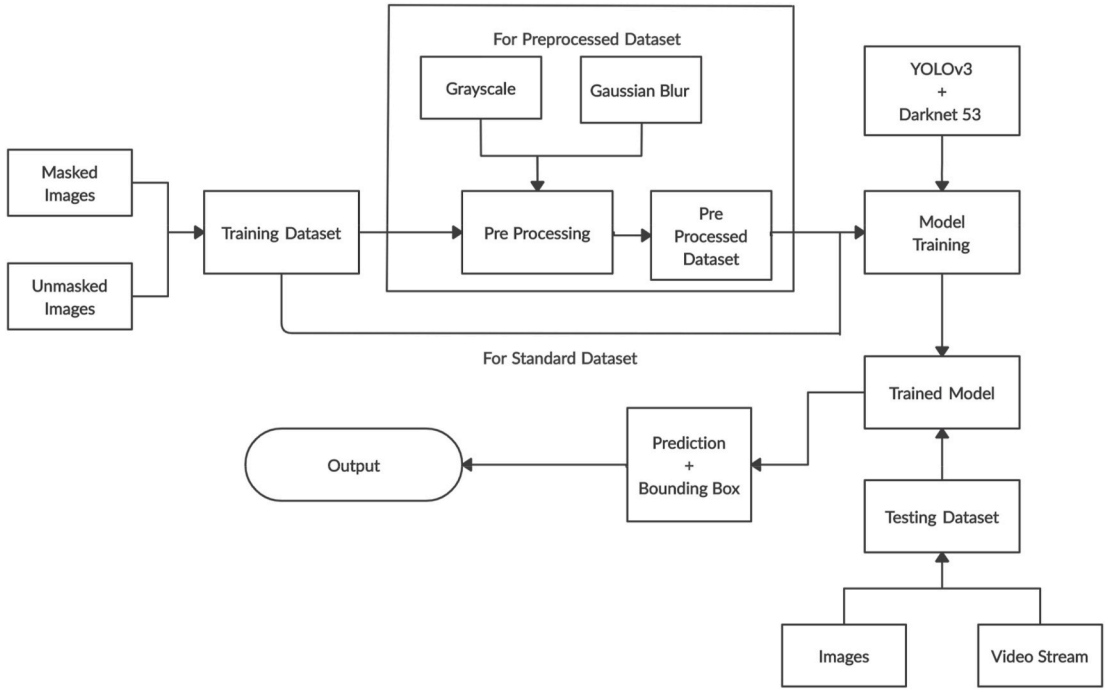

Fig. 7. Workflow of the MDM

performance of the proposed MDM.

4. Experiment

4.1. Dataset preparation

The dataset used in the experiment is taken from Kaggle[1] [28] and contains the images of masked as well as unmasked people of various ethnicities and in various places. For building a model which can perform well regardless of the location where it is implemented, the authors prepared the dataset by collecting a wide range of images with people hailing from different regions of the world in different backgrounds. Images of varying contrast, brightness and size are included as well as to improve the quality of the dataset. The dataset used is of 350 MB size containing 800 images with a 90/10 training and testing split. This dataset contains images of masked individuals, unmasked individuals and images with mixed groups, i.e., containing both masked and unmasked individuals. After collecting the dataset, the images are manually annotated and classified into two classes called Mask and No Mask. Hence, the paper focuses on binary classification. The annotation files for every image contained the bounding box indices of masked or unmasked faces and the respective class index.

Table 1 describes the detailed breakdown of masked and unmasked face instances utilized in the dataset for training the Data augmentation based MDM. A total of 975 masked face instances and 850 unmasked faces instances were utilized, with the group photos consisting of an average of 5 masked faces and 4 unmasked faces. Instead of taking images with only masked or unmasked individuals particularly per image, the dataset is prepared by taking images with single individuals as well as multiple people (in a group or crowd), with some wearing masks while others unmasked and annotated accordingly. This would give our model a better training dataset enabling it to refer and detect multiple faces in a group and predict whether masks are being worn or not by each individual in the frame.

For experimenting on the proposed model, this dataset is modified and two sets of pre-processed images are added to the original dataset. One set of these images are grayscale images and the other set are Gaussian blurred images. Manual annotation of these pre-processed images is carried out. Thereby, the new augmented dataset for the proposed model is prepared. The flowchart representation of the dataset preparation is shown in Fig. 6. The model is then trained on this augmented dataset.

4.2. Training procedure

Fig. 7 briefly explains the workflow of the experimental procedure. The experiment is conducted in two phases. In the first phase, the standard dataset is used for the training of the MDM and in the second phase, the augmented dataset is used for training the MDM. Google Colaboratory (Colab), a cloud based python programming platform that provides 12 GB of graphics processing units (GPU), is used for the experiment.

The procedure used for conducting the experiment is explained as follows.

1. As shown in Fig. 7, in this paper, each dataset follows a specified procedure before the training. All the images in the datasets are annotated using labelImg[2] [29]. A classes.txt file contains two classes “Not Wearing mask” (class number: 0) and “Wearing Mask ’’ (class number: 1) to classify the images accordingly. A bounding box is created around the objects and labelled accordingly in each image. A unique.txt file is created for every image that stores the details like class number and the coordinates of the bounding boxes for every object present.
2. The final datasets now have all the images, their corresponding annotated text files and a file with the list of classes.
3. The authors used a python framework for implementing the above mentioned algorithms. The authors made a copy of the standard

[1] https://www.kaggle.com/.

[2] Labeling is a graphical image annotation tool written in python. Annotations are saved in XML files for pascal VOC format and.txt files for YOLO format.

Table 2
Average confidence level for different sections for MDM.

	Images of individuals	Images of group of people	Video with a group of people
Standard MDM	0.94	0.93	0.91
Data augmentation based MDM	0.97	0.96	0.93

Table 3
Minimum and maximum confidence level.

	Standard MDM (minimum)	Data augmentation based MDM (minimum)	Standard MDM (maximum)	Data augmentation based MDM (maximum)
Individual's Images	0.64	0.76	0.99	0.99
Group Images	0.55	0.82	0.99	0.99
Video	0.56	0.68	0.96	0.98

dataset with the pre-processed images and their corresponding annotated text files. The datasets are fed to the training model.

4. The training is run for both the datasets subsequently, firstly for the standard dataset and next for the augmented dataset till the models converge to minimum MSE. The weights are saved upon completion of the training procedure for both the models.
5. After the training the next step is to validate the performance of the model. The test dataset is given in both image and video format to both the trained MDMs for testing purposes.
6. The validation function is so designed that if the objects in the video frames or the image matched any of the classes with a confidence of more than 20%, the object will be detected with a bounding box along with the confidence level. The output generated for each image and video stream in the testing phase is then recorded in the console for both models and further analysed.

5. Results and discussion

The testing of the MDM is performed on different inputs to analyse output response for possible scenarios in the real world environment where it can be implemented. Outputs for images and video streams are then recorded and analysed. The result analysis of the models is done using the following sections for better understanding.

- ❖ Images of individuals (Masked and unmasked)
- ❖ Images of a group of people (Masked and unmasked)
- ❖ Video with a group of people (Masked and unmasked)

The following results are obtained by averaging the confidence level of the testing dataset. The testing dataset consists of 20 images of individuals, 20 images of groups and 20 frames per video for 5 videos. The testing is done on both models to determine the average confidence level in order to compare the two models.

From Table 2, it is evident that MDM using augmented dataset produces a better average confidence level for all the three sections compared to MDM using standard dataset.

From Table 3, it can also be observed that the model trained on Data augmentation based MDM is producing better results with higher accuracy for all inputs as compared to the standard MDM trained on original images only. This brings into light that using images with an augmented dataset increases the confidence of the model.

Some of the sample output from both models and their comparison is shown below.

5.1. Result analysis of images of individuals (masked and unmasked)

As the image contains only a single individual, it can be observed that the bounding box is plotted accurately around the face with high confidence level. The sample testing unmasked image produces a confidence level of 0.88 using standard MDM as shown in Fig. 8(a) while 0.92 using Data augmentation based MDM as shown in Fig. 8(b). For the sample testing masked image, confidence level is 0.96 using standard MDM as shown in Fig. 9(a) while 0.99 using Data augmentation based MDM as shown in Fig. 9(b). Hence, it can be observed that in both the cases, the confidence level is better in Data augmentation based MDM compared to standard MDM.

5.2. Result analysis of images of groups of people

The image contains a group of people wearing masks as well as unmasked. It can be observed that the bounding box is plotted around each face present in the frame with correct prediction of classes. The overall confidence levels in both Fig. 10 and Fig. 11 for Data augmentation based MDM is better than the standard MDM. In Fig. 10(a), the maximum confidence level for masked people is 0.99 and 0.98 for unmasked people, whereas in Fig. 10(b) the maximum confidence level is 0.99 for both the classes. In Fig. 11, it is observed that, though the standard MDM gives a better confidence level in Fig. 11(a) compared to Fig. 11(b), the average confidence level is better in case of Data augmentation based MDM.

5.3. Result analysis of video with groups of people (masked and unmasked)

The video contains people wearing and not wearing masks, it can be observed that the bounding box is plotted around each face present in frame at that particular instant of time. While some bounding boxes seem to overlap due to close faces in frame and the accuracy seems to be slightly lower than that of individual face mask detection, the

Fig. 8. Output of individual unmasked image using (a)Standard MDM (b)Data augmentation based MDM.

Fig. 9. Output of individual masked image using (a) Standard MDM (b) Data augmentation based MDM.

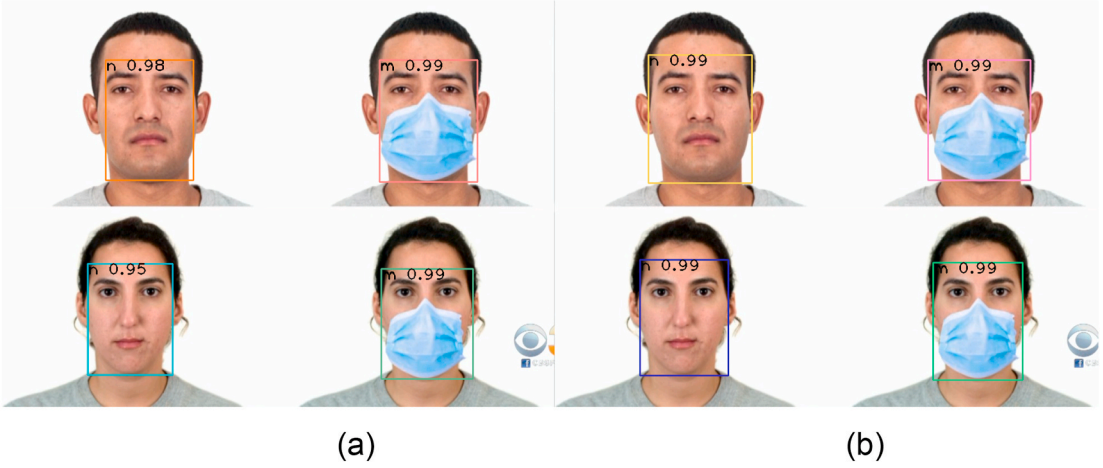

Fig. 10. Output of groups of masked and unmasked people image using (a) Standard MDM (b) Data augmentation based MDM.

Fig. 11. Output of groups of masked and unmasked people image using (a) Standard MDM (b) Data augmentation based MDM.

Fig. 12. Video Frame Prediction from model trained on (a) Standard MDM (b) Data augmentation based MDM.

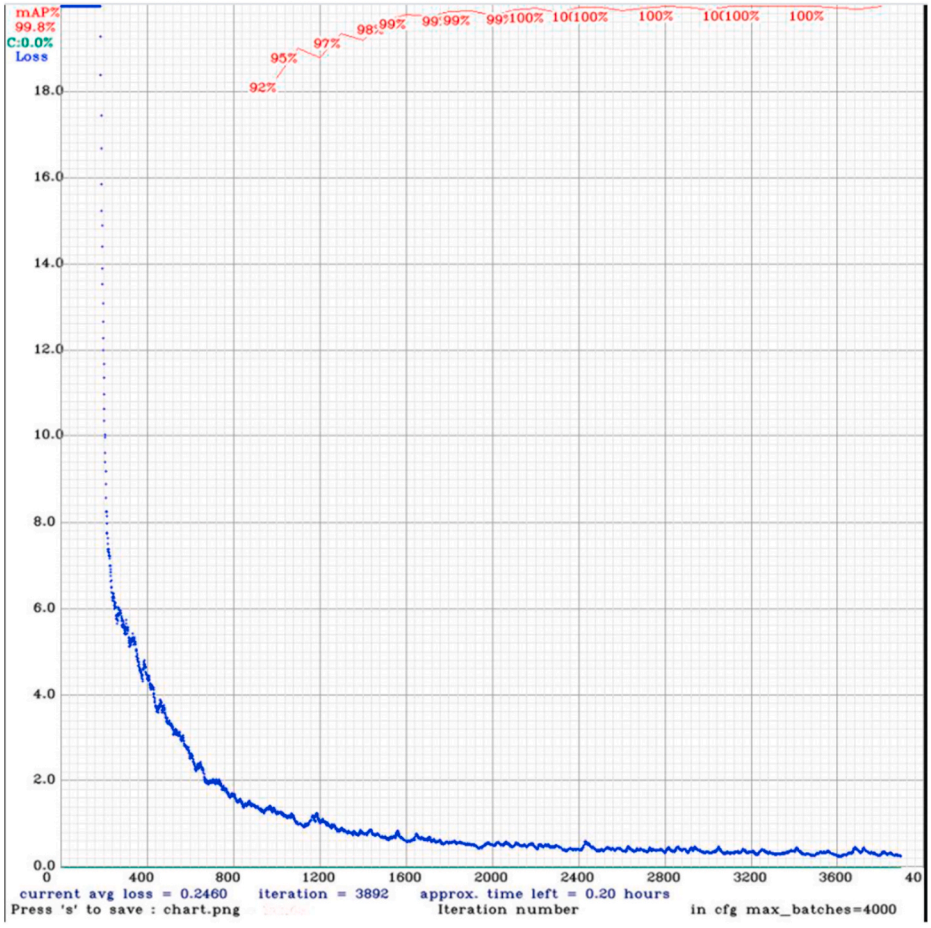

Fig. 13. mAP and loss versus number of iteration graph.

confidence level is still high and the boxes are plotted with the correct class names for each face in frame. It can be observed that if any person in video removes his/her mask, the accuracy for wearing masks decreases till it starts to show a new bounding box with "No Mask" and increasing accuracy. For the video frames, maximum confidence level is 0.97 in Fig. 12(a) while 0.99 in Fig. 12(b). As it can be observed, the confidence level is higher in both cases for the Data augmentation based MDM as compared to the standard MDM.

Fig. 13 is a mAP (mean average precision) and loss versus epoch graph plotted in real time while training the Data augmentation based MDM. It gives values for important evaluation metrics like mAP that comes out to be 99.8% while the number of iterations completed was 3892. The average loss at this point is around 25%.

6. Conclusion

The mask detection model finds its application in places where there is a need to detect masks. One such scenario being the present pandemic due to COVID-19 where there is a need to monitor people wearing masks in public places. This automated system would avoid the spread of the disease to the security personnel or people given the mask monitoring job. In order to produce such an automated system, this paper introduces a novel approach for mask detection using augmented dataset. It uses a filtering technique for data augmentation which enhances the model performance compared to standard MDM. This technique of mask detection can serve as a prototype for future development.

References

[1] Liu Jiaye, et al. Community transmission of severe acute respiratory syndrome coronavirus 2, shenzhen, China. Emerg Infect Dis 2020;26(6):1320–3. https://doi.org/10.3201/eid2606.200239. 2020.

[2] Thaper R. Transmission of SARS-CoV-2 through the air. Curr Med Res Pract 2020; 10(4):196–7. https://doi.org/10.1016/j.cmrp.2020.07.005.

[3] Humphreys J. The importance of wearing masks in curtailing the COVID-19 pandemic. J Fam Med Prim Care Jun. 2020;9(6):2606–7. https://doi.org/10.4103/jfmpc.jfmpc_578_20.

[4] Esposito S, Principi N, Leung CC, Migliori GB. Universal use of face masks for success against COVID-19: evidence and implications for prevention policies. Eur Respir J Jun. 2020;55(6). https://doi.org/10.1183/13993003.01260-2020.

[5] Zhao Z-Q, Zheng P, Xu S-T, Wu X. Object detection with deep learning: a review. IEEE Transactions on Neural Networks and Learning Systems Nov. 2019;30(11): 3212–32. https://doi.org/10.1109/TNNLS.2018.2876865.

[6] Verma GK, Dhillon A. A handheld Gun detection using faster R-CNN deep learning. In: Proceedings of the 7th international conference on computer and communication technology, New York, NY, USA; Nov. 2017. p. 84–8. https://doi.org/10.1145/3154979.3154988.

[7] Fong S, Yang X-S. White learning: a white-box data fusion machine learning framework for extreme and Fast automated cancer diagnosis. IT Professional Sep. 2019;21(5):71–7. https://doi.org/10.1109/MITP.2019.2931415.

[8] Wang Zhongyuan, et al. Masked face recognition dataset and application. arXiv: 2003.09093 [cs], Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.09093.

[9] Alasadi Suad A, Bhaya Wesam S. Review of data preprocessing techniques in data mining. J Eng Appl Sci 2017;12(16):4102–7. https://doi.org/10.36478/jeasci.2017.4102.4107.

[10] Evans G. Solving home automation problems using artificial intelligence techniques. IEEE Trans Consum Electron Aug. 1991;37(3):395–400. https://doi.org/10.1109/30.85542.

[11] Sharma KU, Thakur NV. A review and an approach for object detection in images. Int J Comput Vis Robot Jan. 2017;7(1/2):196–237. https://doi.org/10.1504/IJCVR.2017.081234.

[12] Dhlamini SM, Kachienga MO, Marwala T. Artificial intelligence as an aide in management of security technology. in AFRICON 2007 Sep. 2007:1–5. https://doi.org/10.1109/AFRCON.2007.4401470.

[13] Gao J, Yang Y, Lin P, Park DS. Computer vision in healthcare applications. Journal of Healthcare Engineering 2018:5157020. https://doi.org/10.1155/2018/5157020. 2018.

[14] Fuchs K, Grundmann T, Fleisch E. Towards identification of packaged products via computer vision: convolutional neural networks for object detection and image classification in retail environments. In: Proceedings of the 9th international conference on the internet of things, New York, NY, USA; Oct. 2019. p. 1–8. https://doi.org/10.1145/3365871.3365899.

[15] Saba T. Recent advancement in cancer detection using machine learning: systematic survey of decades, comparisons and challenges. Journal of Infection and Public Health Sep. 2020;13(9):1274–89. https://doi.org/10.1016/j.jiph.2020.06.033.

[16] Bzhalava Z, Tampuu A, Bała P, Vicente R, Dillner J. Machine Learning for detection of viral sequences in human metagenomic datasets. BMC Bioinf Sep. 2018;19(1): 336. https://doi.org/10.1186/s12859-018-2340-x.

[17] Huang Jonathan, et al. Speed/accuracy trade-offs for modern convolutional object detectors. arXiv:1611.10012 [cs], Apr. 2017, Accessed: Oct. 08, 2020. [Online]. Available, http://arxiv.org/abs/1611.10012.

[18] Redmon J, Farhadi A. YOLOv3: an incremental improvement. arXiv:1804.02767 [cs], Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767.

[19] Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: 2017 international conference on engineering and technology (ICET); Aug. 2017. p. 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186.

[20] Nguyen N-D, Do T, Ngo TD, Le D-D. An evaluation of deep learning methods for small object detection. Journal of Electrical and Computer Engineering Apr. 2020; 2020:3189691. https://doi.org/10.1155/2020/3189691.

[21] Rao A. One stop for object detectors. Medium Jun. 13, 2020. https://medium.com/swlh/one-stop-for-object-detectors-2c99daa08c50. [Accessed 12 December 2020].

[22] Joseph Redmon. Darknet: open source neural networks in C. 2016 2013, [Online]. Available: http://pjreddie.com/darknet/.

[23] Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. arXiv:1612.03144 [cs], Apr. 2017, [Online]. Available, http://arxiv.org/abs/1612.03144.

[24] Roy S, Mitra A, Setua SK. Color grayscale image representation using multivector. In: Proceedings of the 2015 third international conference on computer, communication, control and information technology (C3IT); Feb. 2015. p. 1–6. https://doi.org/10.1109/C3IT.2015.7060218.

[25] Singhal P, Verma A, Garg A. A study in finding effectiveness of Gaussian blur filter over bilateral filter in natural scenes for graph based image segmentation. In: 2017 4th international conference on advanced computing and communication systems (ICACCS); Jan. 2017. p. 1–6. https://doi.org/10.1109/ICACCS.2017.8014612.

[26] Kanan C, Cottrell GW. Color-to-Grayscale: does the method matter in image recognition? PloS One Jan. 2012;7(1):e29740. https://doi.org/10.1371/journal.pone.0029740.

[27] Gedraite ES, Hadad M. Investigation on the effect of a Gaussian Blur in image filtering and segmentation. In: Proceedings ELMAR-2011; Sep. 2011. p. 393–6.

[28] Kaggle “. Your machine learning and data science community. https://www.kaggle.com/. [Accessed 28 September 2020].

[29] darrenl. tzutalin/labelImg 2020.