COVID-19 Detection of Proper and Improper Wearing of Surgical Face Masks in Regulated Areas

Denise Patricia B. Manalo
College of Engineering, Architecture, and Fine Arts
Batangas State University – Alangilan Campus
Batangas City, Philippines
denisepatricia.manalo@g.batstate-u.edu.ph

Abstract— Coronavirus disease (COVID-19) is the most recent outbreak caused by a recently discovered coronavirus. It is primarily transmitted from person to person via airborne transmission, particularly through close contact. As the number of infected cases and deaths continues to rise, it is becoming increasingly important to limit the spread of the virus to the greatest extent possible. Surgical face masks (also known as medical masks) should be worn by the general public to protect against the coronavirus disease caused by the COVID-19. In order to contribute to the continued improvement of face mask recognition in the fight against COVID-19, the author proposes a COVID-19 detection of proper and improper wearing of surgical face masks in regulated areas. A large dataset of masked faces was used for training deep learning models to distinguish people who are properly wearing masks, not properly wearing masks and those who are not. In order to continue improving face mask recognition in the fight against COVID-19, the author implemented a system using the YOLOv3 algorithm, which utilizes CNN to implement a deep learning technique.

Keywords—deep learning, object detection, yolov3, surgical mask, COVID-19

I. Introduction

Coronavirus disease (COVID-19) is the most recent outbreak caused by a recently discovered coronavirus. COVID-19 is an infectious disease caused by the SARS-CoV-2 virus that affects the respiratory system. It is primarily transmitted from person to person via airborne transmission, particularly through close contact. This occurs with respiratory droplets, which are droplet particles with diameters ranging from 5 to 10 microns. It is most common for droplet transmission to occur when a person is in close quarters (within 1 m) with someone who is experiencing respiratory symptoms and is thus at risk of having his/her mouth, nose, or eyes exposed to potentially infectious respiratory droplets.[1] As the number of infected cases and deaths continues to rise, it is becoming increasingly important to limit the spread of the virus to the greatest extent possible. Surgical face masks (also known as medical masks) should be worn by the general public to protect against the coronavirus disease caused by the COVID-19. Numerous studies have demonstrated that face coverings can contain droplets released by the wearer, which account for the majority of virus transmission. Furthermore, the use of surgical face masks appears to be a feasible method of containing COVID-19 spread.

Past studies have shown that the vast majority of countries have implemented policies requiring people to wear face masks in public places. [2] Prakash et. al [3] states that manual observation of the face mask is a time-consuming task, especially in crowded places like hospitals, airports, train stations, and shopping malls. This prompted researchers to develop an automated system for detecting face masks. Md. Sabbir Ejaz et al. [4] developed a masked and unmasked facial recognition algorithm using the Principal Component Analysis (PCA) technique. PCA is effective at recognizing faces without a mask with a 96.25% accuracy, but its accuracy drops to 68.75% when faced with a mask. Li et al. [5] used YOLOv3 for face detection. YOLOv3 is based on the darknet-19 deep learning network architecture, which was trained using the WIDER FACE and Celebi databases and then evaluated using the FDDB database. This model was 93.9 percent accurate. Rodriguez et al. [6] proposed a system for automatically detecting the presence or absence of the surgical mask that is required in operating rooms. This system is designed to sound an alarm if a member of staff is not wearing a mask. This system achieved a 95% accuracy rate. Whereas, in Hussain et. al [7], the author classified facial emotions into seven categories using a pre-trained VGG16 architecture. They trained the model on the KDEF dataset and achieved an accuracy of 88 percent.

As can be seen from the preceding situation, very few research papers have been published thus far, particularly in the area of face mask detection, but it is clear that existing methods require further improvement. Thus, in order to contribute to the continued improvement of face mask recognition in the fight against COVID-19, the author proposes a COVID-19 detection of proper and improper wearing of surgical face masks in regulated areas.

Furthermore, the use of masks is critical to containing the virus's spread. Masks act as a simple barrier, preventing respiratory droplets from spreading to others. Masks worn over the nose and mouth have been shown to reduce the spray of droplets. [8] To ensure that masks are as effective as possible, they must be used, stored, and cleaned or disposed of properly. In this context, effective recognition systems are expected to ensure that individuals' faces are concealed in regulated areas.

To accomplish this task, a large dataset of masked faces is required for training deep learning models to distinguish people who are properly wearing masks, not properly wearing masks and those who are not. Also, this work proposes the YOLOv3 method, which has the advantages of high accuracy and low computational complexity. In the sections that follow, we will discuss the architecture of the YOLOv3 algorithm, the preparation and collection of datasets, the technique for detecting proper and improper mask wearing, model training, and testing on a custom data set.

II. Methodology

This section provides details about the methods used to conduct this research. The diagram presented in Fig.1 focuses on the dataset, preprocessing and augmentation techniques, model architecture, and real-time mask detection system. The proposed framework is intended to determine whether or not individuals in a video or image of a public area are wearing face masks properly.

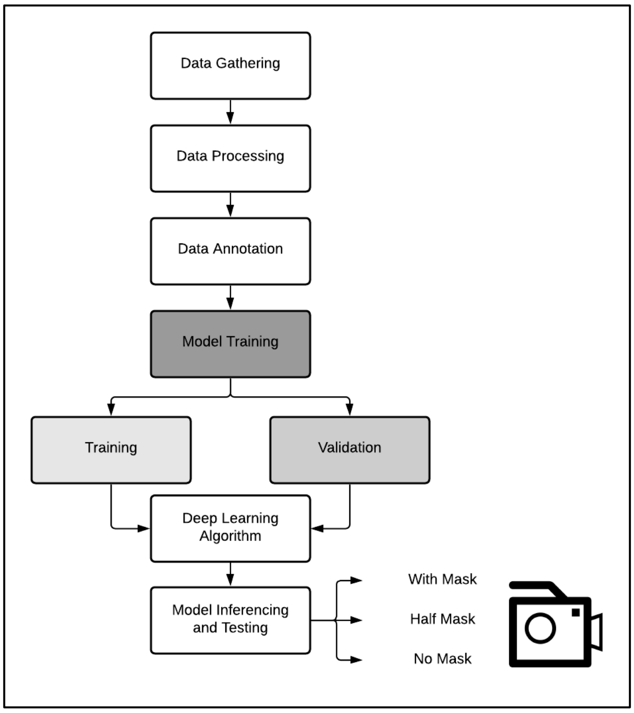

Fig. 1. Block Diagram

A. Data Gathering

Data is critical when it comes to data-driven techniques such as machine learning and deep learning. The more data, the more accurate the result. [9] Additionally, for our purposes of working with YOLO, we require additional data that is properly annotated. The predefined collection of images used in this work comes from MaskedFace-Net [10]. The dataset contains 70,000 high-quality PNG images with a resolution of 1024x1024 and contains a wide range of age, ethnicity, and image background. Additionally, it covers a variety of accessories such as eyeglasses, sunglasses, and hats. Thus, Fig. 2 depicts a sample dataset of the gathered images no masks(a), with masks(b), and half masks(c) with a total number of 300 images.

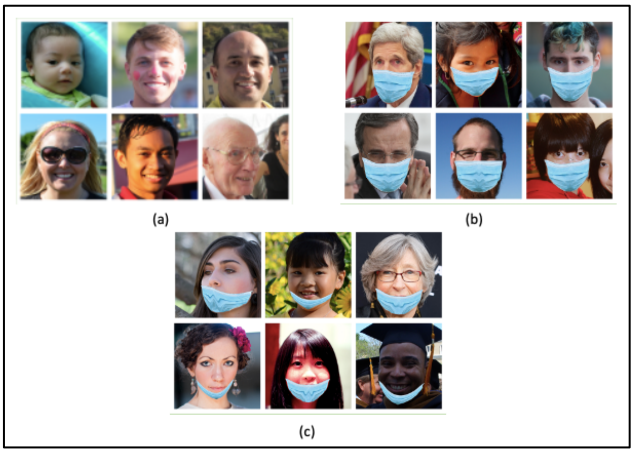

Fig. 2. Sample Dataset for no masks (a), with masks(b), and half masks(c)

B. Data Processing

Following that, our data is unsuitable for feeding into the model. We perform some pre-processing prior to feeding. The dataset contains a few irrelevant images. We removed them and ended up with a dataset of 300 images. A dataset is a collection of instances, and when working with machine learning techniques, we typically require two datasets for distinct purposes.

- **Training Dataset**: A dataset that is a collection of data that is fed into our deep learning algorithm in order to train our model.
- **Validation Dataset:** A dataset that we use to validate our model's accuracy but not to train it. This dataset may be referred to as the testing dataset.

From the total number of dataset stated above, 240 images of all three categories (with mask, half mask, and no mask) are used for training and the remaining 60 images are used for validating the model.

C. Data Annotation

Data annotation, also known as image annotation, is the process of accurately labeling data in a variety of formats such as text, video, or images for use in an object detection model. Detection tasks are quite distinct from classification tasks. As a result, data should be thoroughly annotated.

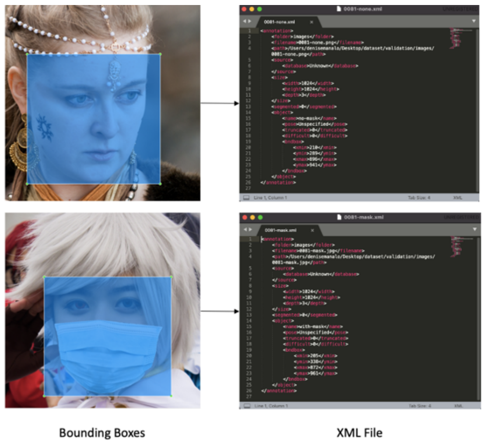

Fig. 3. Dataset Annotation

As illustrated in Fig. 3, this study requires bounding boxes. the researchers used LabelImg to annotate and label the datasets. To annotate the training and validation image datasets, a rectangular bounding box was constructed in the face region of a human using the program called LabelIMG [11]. As a result, an XML file containing the coordinates of the annotated images in PascalVOC format is created.

D. Deep Learning Algorithm

In 2015, Joseph Redmon et al.[12] introduced "You Only Look Once," abbreviated YOLO. This is a unified architecture that is extremely fast. The YOLO model at its core processes images in real time at a rate of 45 frames per second. Later on, YOLOv2 and YOLOv3 were introduced in 2016 and 2018 respectively.

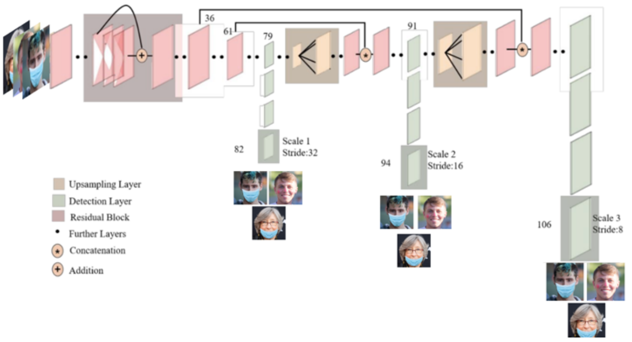

Fig. 4. Yolov3 Architecture

YOLOv3 is the state-of-the-art object detection model, which is followed by other YOLO and YOLOv2 variants. It has achieved remarkable results in terms of object classification and detection. Darknet-19 was used as a feature extractor in previous versions of Yolov2. Yolov3 modifies it slightly and renames it darknet-53. Darknet is a framework for training neural networks written in the C programming language that excels at these tasks.[13]

This study chose YOLOv3, as illustrated in Fig. 4, as the foundation for numerous lightweight networks due to its widespread use in business. Thus, the decision to use YOLOv3 was made based on its fundamental structure and ease of use.

The study makes use of the YOLOv3 model by passing an image through it. This object detector traverses the image in search of the coordinates contained therein. It divides the input data into a grid and then analyzes the target object's features using that grid. The features detected with a high degree of confidence in neighboring cells are combined in one location to produce the model output.

E. Model Evaluation

To ensure that the best-trained model was chosen for model inference detection, the trained models were compared using the mAP (mean Average Precision). It will generate files for use during the testing procedure. The model's detection accuracy increases as the mAP value increases; the mAP value is the average of the AP's values (mean average precision).

Average Precision (AP): The average precision of ranked retrieval results is a measure that combines recall and precision. [14] For a single information need, the average precision is calculated as the mean of the precision scores obtained after retrieving each relevant document.

$$Average\ Precision = \frac{\sum_r P_r}{R} \qquad (1)$$

where r denotes the rank of each pertinent document, R denotes the total number of pertinent documents, and P_r denotes the precision of the top-r retrieved documents. [15]

Mean Average Precision (mAP): The term mAP (mean average precision) refers to the average of the two terms AP. To begin calculating the mAP, determine the AP for each class. The mAP is the average of all the APs for all classes. The following equation calculates the mAP of the object detection model using the APs of two classes:

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \qquad (2)$$

where AP_k denotes the Average Precision of class k and n is the number of classes. The threshold is the intersection over union (IoU) score for detected objects in an object detection model. Thus, the mAP is calculated after the AP is determined for each class in the dataset. [16]

F. Model Inferencing and Testing

The study developed a graphical user interface (GUI) using the Anaconda Integrated Development Environment (IDE), Qt Designer, PyQt5, and the ImageAI detection library. The graphical user interface is responsible for three critical functions: image detection, live feed detection, and video detection. The chosen deep learning model's h5 file and accompanying JSON configuration file were used for model inference.

The researchers conducted the testing procedure using a new collection of photographs and video files. These images were used to avoid biases in testing accuracy findings because they were not among the 300 photos used for training and validation.

$$Accuracy = \frac{Number\ of\ Detected\ Object}{Total\ Number\ of\ Objects} x\ 100 \qquad (3)$$

III. Results And Discussions

This section contains the findings and discussions from the study's training, validation, and testing results.

A. Training and Validation Results

Fig. 5 depicts the training and validation results for the dataset. The blue line pattern represents the validation loss, while, the green line pattern represents the training loss. As indicated in the graph, the experimental training level is set to forty-two (42) epochs with its corresponding loss value that indicates the performance of the model.

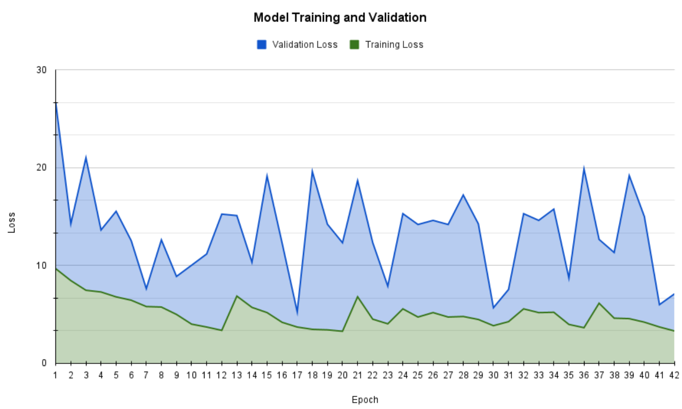

Fig. 5. Model Training and Validation Results

As illustrated in Fig. 5, the training and validation loss score for training epoch 42 was the smallest with 3.31% and 3.80%, respectively. While, the training and validation loss score for training epoch 1 was the largest with 9.68% and 16.96%, respectively. It was evident that the lower

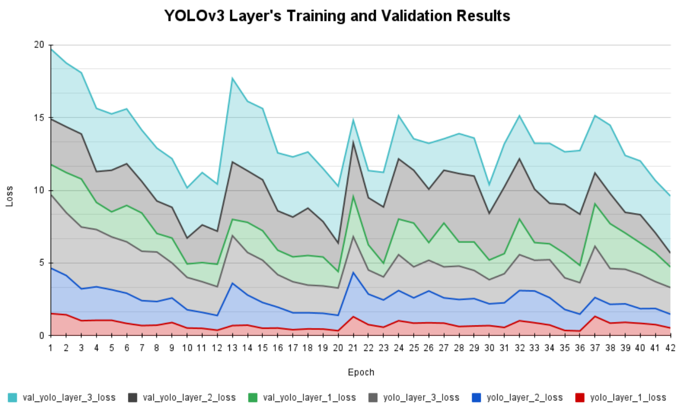

Fig. 6. Model Layers' Training and Validation Results

Whereas, it was illustrated in Fig. 6, that as the training duration increases, the loss decreases. Since the loss value indicates the performance of the model, a lower score indicates a better result. On the other hand, during the training process, it gained knowledge from the datasets provided.

B. *Evaluation Results*

The mAP number which is expressed as a percentage, indicates the precision of the dataset validation. The validity number is close to its maximum possible value, as the mAP is equal to one (1) or 100%.

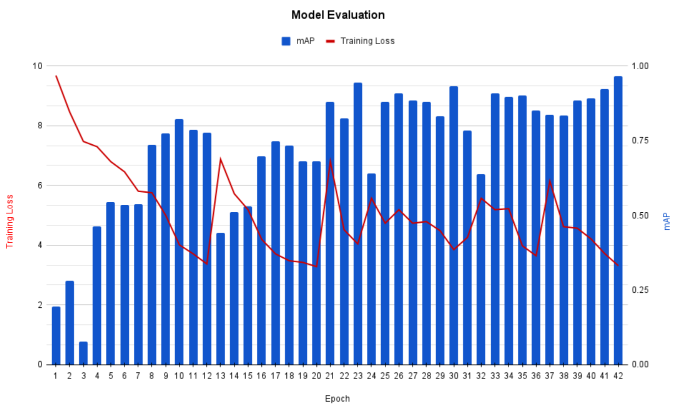

Fig. 7. Model Evaluation Result

Fig. 7 illustrates the mAP obtained during the model evaluation. Model 3 has the lowest performance in this graph, with a mAP of 0.0777. While, Model 42 achieved the best results, with a mAP of 0.9668 (96.68%) and an equal training loss of 3.31. Thus, the model with the greatest mAP, model 42, was therefore employed in testing and inferencing.

C. *Inference and Testing of Model*

Each frame obtained by the capturing device was subjected to the most performing model in the training period, model 42. Fig. 8 illustrates the graphical user interface created to monitor the system's performance and model 42 inference. Also, three options for implementation were used: import picture, live feed, or import video. The detection system exported the results of the detected proper and improper wearing of face mask using Live Feed as a CSV file presented in Fig. 9.

Furthermore, this study produced a video clip and downloaded a video online [17] to assess model 42's inference ability. The video dataset for the deployment and testing sites was created using a camera video. The video clip contains 900 frames and runs for 40 seconds. The output demonstrates how detection sensitivity varies. Because it detects the item classes correctly, the per-frame video testing demonstrates a detection accuracy of 99%, as illustrated in Fig. 9 with a total frame count of 120.

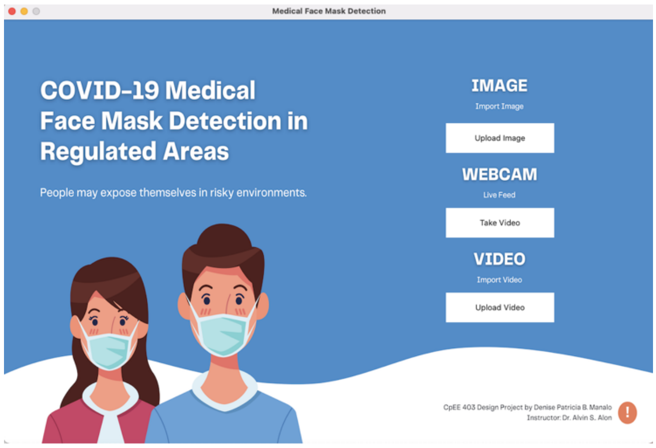

Fig. 8. Model Inference's GUI

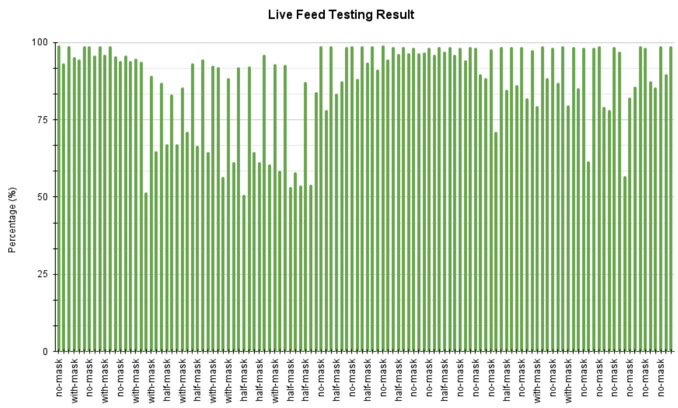

Fig. 9. Live Feed Testing Results

Fig. 10. Live Feed and Video Frames Testing Results

Also, as demonstrated by the per-frame video testing, the performance of proper and improper wearing of face mask is quite impressive. As illustrated in Fig.10, for the live feed video frames(a), the no mask, half-mask, and with mask outcome results were 98.853, 94.461 and 94.603, respectively. Whereas, video frames(b) testing results shows the highest percentage result of 98.8022 for no mask, and 92.5712 for with mask.

TABLE I. COMPARISON OF VARIOUS FACE MASK DETECTION ALGORITHMS

METHOD	NETWORK	YEAR	mAP
LLE-CNN [17]	VGG	2017	76.4%
RetinaNet [18]	ResNet-50-FPN	2017	81.4%
Faster R-CNN [19]	ResNet-50-FPN	2017	82.8%
Cascade R-CNN [20]	ResNet-50	2018	83.2%
Yolov3 [21]	Haar Cascade	2020	90.1%
Yolov3 Weighted Loss	DarkNet-19	2021	91.9%
Faster R-CNN [22]	DarkNet-19	2021	93.2%
Ours	DarkNet-53	2021	96.7%

Table 1 shows the performance of various face mask detection algorithms and demonstrates how the work of this study outperformed previous algorithms for identifying proper and improper wearing of face mask. It demonstrates that, when compared to previous studies conducted in the specified years above, the current study achieved the highest mAP percentage of 96.7 percent.

IV. CONCLUSION

The most recent outbreak of coronavirus disease (COVID-19) was caused by a newly discovered coronavirus. As the number of infected cases and deaths continues to rise, it is becoming increasingly important to limit the spread of the virus to the greatest extent possible. Furthermore, the use of surgical face masks appears to be a feasible method of containing COVID-19 spread. Past studies have shown that the vast majority of countries have implemented policies requiring people to wear face masks in public places.

Thus, in order to continue improving face mask recognition in the fight against COVID-19, the author implemented a system using the YOLOv3 algorithm, which utilizes CNN to implement a deep learning technique. The investigation yielded a mAP of 0.9668 for model 42. This indicates that, even with a small number of datasets, YOLOv3 is an effective technique for identifying proper and improper face mask wear in regulated areas. When put to the test, the system achieved an excellent overall accuracy score of 98.85%.

V. RECOMMENDATION

The future work of the study will include adding more data in order to obtain a more precise result in detection. Additionally, it will include mask variants, such as a variety of colors and shapes, as a result of the variety of masks available today. Also, training and evaluating more models can result in a more accurate result. Finally, a new version of YOLOv4 has been released that can be used to compare which algorithm performs better with these models.

ACKNOWLEDGEMENT

The author wishes to express heartfelt appreciation to the College, Department, and University with which she is affiliated. Additionally, to the Course Instructor, Dr. Alvin S. Alon for the immense knowledge and guidance. Most especially to the author's parents for the moral encouragement and support.

REFERENCES

[1] World Health Organization (WHO) "Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations" https://www.who.int/newsroom/commentaries/detail/modes-of-transmission-of-virus-causingcovid-19-implications-for-ipc-precaution-recommendations [Online; accessed 08 Dec 2021]

[2] X. Liu and S. Zhang, "COVID-19: Face masks and human-to-human transmission," *Influenza Other Respi. Viruses*, vol. 14, no. 4, pp. 472–473, Jul. 2020, doi: 10.1111/IRV.12740.

[3] I. B. Venkateswarlu, J. Kakarla, and S. Prakash, "Face mask detection using MobileNet and global pooling block," *4th IEEE Conf. Inf. Commun. Technol. CICT 2020*, Dec. 2020, doi: 10.1109/CICT51604.2020.9312083.

[4] Ejaz, M.S., Islam, M.R., Sifatullah, M., Sarker, A.: Implementation of principal component analysis on masked and non-masked face recognition. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). pp. 1–5 (2019)

[5] Li, C., Wang, R., Li, J., Fei, L.: Face detection based on yolov3. In: Recent Trends in Intelligent Computing, Communication and Devices, pp. 277–284. Springer (2020)

[6] Nieto-Rodr´ıguez A., Mucientes M., B.V.: System for medical mask detection in the operating room through facial attributes. vol. 9117, pp. 138–145. Springer (2015)

[7] S. A. Hussain and A. S. A. Al Balushi, "A real time face emotion classification and recognition using deep learning model," in Journal of Physics: Conference Series, 2020, vol. 1432, no. 1, p. 012087: IOP Publishing.

[8] Center for disease control and prevention (CDC) "Considerations for Wearing Masks" https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html [Online; accessed 08 Dec 2021]

[9] M. R. Bhuiyan, S. A. Khushbu, and M. S. Islam, "A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3," *2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020*, Jul. 2020, doi: 10.1109/ICCCNT49239.2020.9225384.

[10] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, "MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19," *Smart Heal.*, vol. 19, Mar. 2021, doi: 10.1016/J.SMHL.2020.100144.

[11] "tzutalin/labelImg: LabelImg is a graphical image annotation tool and label object bounding boxes in images." https://github.com/tzutalin/labelImg (accessed Dec. 15, 2021).

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Accessed: Dec. 15, 2021. [Online]. Available: http://pjreddie.com/yolo/.

[13] "Custom Object Detection using Darknet | Towards Data Science." https://towardsdatascience.com/custom-object-detection-using-darknet-9779170faca2 (accessed Dec. 15, 2021).

[14] C. S. Jensen and R. T. Snodgrass, "Absolute Time," *Encycl. Database Syst.*, pp. 1–1, 2020, doi: 10.1007/978-0-387-39940-9_1407.

[15] N. Craswell and S. Robertson, "Average Precision at n," *Encycl. Database Syst.*, pp. 193–194, 2019, doi: 10.1007/978-0-387-39940-9_487.

[16] "Mean Average Precision (mAP) Explained | Paperspace Blog." https://blog.paperspace.com/mean-average-precision/ (accessed Dec. 15, 2021).

[17] "How to Wear (Don) & Take Off (Doff) Surgical Face Mask Tutorial PPE - YouTube." https://www.youtube.com/watch?v=etZK-GrUYgM (accessed Dec. 15, 2021).

[18] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," *undefined*, vol. 2017-January, pp. 426–434, Nov. 2017, doi: 10.1109/CVPR.2017.53.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, ''Focal loss for dense object detection,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2999–3007.

[20] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards realtime object detection with region proposal networks,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[21] Z. Cai and N. Vasconcelos, ''Cascade R-CNN: Delving into high quality object detection,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6154–6162

[22] T. Q. Vinh and N. T. N. Anh, "Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier," *Proc. - 2020 Int. Conf. Adv. Comput. Appl. ACOMP 2020*, pp. 146–149, Nov. 2020, doi: 10.1109/ACOMP50827.2020.00029.

[23] R. Liu and Z. Ren, "Application of Yolo on Mask Detection Task," Feb. 2021, Accessed: Dec. 15, 2021. [Online]. Available: https://arxiv.org/abs/2102.05402v1.