Automatic Face Mask Detection on Gates to Combat The Spread of Covid-19

Musa Dima Genemo
Study Program Computing (Software Engineering)
Gumushane University musa.ju2002@gmail.com

‹β›

Abstract— **The COVID-19 pandemic has spread across the globe, hitting almost every country. To stop the spread of the COVID-19 pandemic, this article introduces face mask detection on a gate to assure the safety of Instructors and students in both class and public places. This work aims to distinguish between faces with masks and without masks. A deep learning algorithm You Only Look Once (YOLO) V5 is used for face mask detection and classification. This algorithm detects the faces with and without masks using the video frames from the surveillance camera. The model trained on over 800 video frames. The sequence of a video frame for face mask detection is fed to the model for feature acquisition. Then the model classifies the frames as faces with a mask and without a mask. We used loss functions like Generalize Intersection of Union for abjectness and classification accuracy. The datasets used to train the model are divided as 80% and 20% for training and testing, respectively. The model has provided a promising result. The result found shows accuracy and precision of 95% and 96%, respectively. Results show that the model performance is a good classifier. The successful findings indicate the suggested work's soundness.**

Keywords: - Accuracy, Computer vision, Classification, Face recognition, Surveillance

1. INTRODUCTION

The COVID-19 pandemic has spread across the globe, hitting almost every country. The epidemic was initially discovered in Wuhan, China, in December 2019. The CODI-19 epidemic has allowed us to pave the road for digital learning to be introduced. To stop the spread of the COVID-19 Pandemic, this article introduces face mask detection at university gates. Face recognition applications employ various deep learning algorithms to look for human faces within larger pictures that often include non-facial items such as landscapes, buildings, and other human body parts such as feet or hands.

The search for human eyes, which is one of the simplest characteristics to identify, is often where face identification algorithms get started. After that, the algorithm could make an effort to recognize the iris, the mouth, the nose, the nostrils, and the eyebrows. After the algorithm has concluded that it has located a facial area, it next conducts further tests to ensure that it has identified a face in the image [1]. The algorithms need to be trained on big data sets that include hundreds of thousands of examples of both face masks and without face mask pictures. This will assist guarantee that the results are accurate. The training helps the algorithms get better at determining whether a picture contains a face mask. Where those faces are located within the image. Knowledge-based, feature-based, template-matching or appearance-based approaches are some of the strategies that may be used in face identification [2]. The deep learning algorithm YOLO is introduced that detects the objects efficiently. The authors do research and improve its version to V5 as shown in figure 1. Each one has positive and negative aspects to consider. The problem occurs that we need an algorithm that accurately recognized the faces. In this paper, we have highlighted the issue that the algorithm accurately recognized the faces with and without masks.

To build a model that detects face masks we used a new deep learning algorithm YOLO V5. This algorithm efficiently resolves the problem and accurately recognizes the faces.
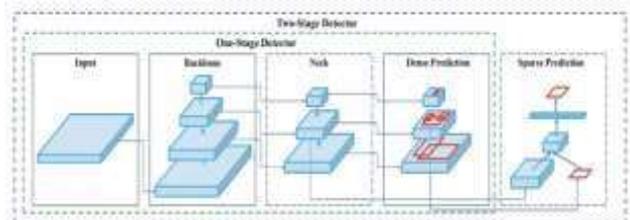

Figure 1: YOLO detection process [4]

The following is the order in which the manuscript is written. The introductory section and the literature review overview section are included in Sections 1 and 2, respectively. The proposed approach is covered in Section 3. Section 4 shows the findings and explanation of the performance evaluation

experiments. The paper's conclusions are presented in Section 5.

2. LITERATURE REVIEW

Knowledge-based approaches, also known as rule-based methods, characterize a face by adhering to certain guidelines. The difficulty of formulating rules that are clearly stated is one of the drawbacks of using this method. Noise and light may have a detrimental effect on facial recognition techniques known as feature invariant methods [3]. These techniques employ distinguishing characteristics of a person's face, such as their eyes or nose, to identify a person's face.

The detection of faces using template-matching approaches involves comparing pictures with conventional face patterns or traits that have been recorded in the past and connecting the two to establish a relationship between the three. The problem with these approaches is that they do not account for differences in position, size, or form. Finding the important attributes of face photos requires statistical analysis and machine learning, both of which are used by appearance-based approaches [4].

This approach, which is also used in the process of feature extraction for face recognition, is broken down into many sub-methods.

The following is a list of some of the more specialized methods that are used in face detection: Taking off the backdrop image. For instance, if a picture has a backdrop that is a single color, or if it has a background that is pre-defined and unchanging, then removing the background from the image may assist show the facial borders. Sometimes the color of the subject's skin may help identify faces in color photographs; however, this is not guaranteed to work with every complexion.

The use of motion to identify faces is yet another possibility. Because a face is generally always moving during the real-time video, users of this approach are required to determine the region of the face that is moving. This approach has a few drawbacks, one of which is the potential for misunderstanding with other moving objects in the background.

A complete technique for detecting faces may be created by combining some of the tactics described in the previous paragraph [5]. Face recognition in photographs can be challenging because of the many variables that can affect the process, differences in camera gain, lighting conditions, and image quality, as well as differences in attitude, emotion, location, orientation, skin color, and pixel values, the existence of spectacles or facial hair, and the photograph's orientation are all factors to consider.

Deep learning, which has been used to make advancements in face identification in recent years, has the benefit of greatly outperforming standard computer vision approaches. These advancements have been made in recent years. Artificial intelligence detects objects. Artificial intelligence required a large amount of data. Then as a solution authors use YOLO, however, a single YOLO version did not detect a large number of objects accurately [6]. In [7] authors, highlight the issue of car number plate detection. Authors use the YOLO algorithm that does not accurately detect all number plates.

The authors in [8], and [9] highlight the issue of face recognition. Many other authors use other ML techniques; however, that techniques fail to capture the facial expressions. The accuracy of the algorithms is not good. The authors resolved the problem of detecting apple plants' pictures. Many authors use deep learning algorithms that fail to recognize that YOLO v4 is used to detect pictures whose accuracy is also low [10-11].

Paul Viola and Michael Jones, computer vision researchers, developed a system in 2001 that could accurately recognize faces in real-time. These developments brought about significant advancements in the face detection approach [1, 12]. The Viola-Jones framework relies on the concept of teaching a model to recognize what constitutes a face and what does not constitute a face.

Once the model has been trained, it will extract certain features, which will then be saved in a file. This will allow the features extracted from fresh photos to be compared with the features that were previously recorded at different phases [13-15]. If the picture being analyzed is successful in passing through each step of the comparison of its features, then a face has been identified and the processes may continue.

Even though it is still widely used, the Viola-Jones framework has certain shortcomings when it comes to the recognition of faces in real-time applications. For instance, the framework may not function correctly if a face is obscured by something like a mask or a scarf. Likewise, if a face is not orientated correctly, the algorithm might not be able to locate it [16].

Other algorithms have been created to help improve processes and eliminate the disadvantages of the Viola-Jones framework, such as the region-based convolutional neural network (R-CNN) and Single Shot Detector (SSD). This has helped to improve the overall quality of the process [3].

In the field of image identification and processing, an artificial neural network known as a convolutional neural network (CNN) is a form of neural network that was developed for the sole purpose of handling pixel input. To localize and categorize the objects seen in pictures, R-CNN will provide region recommendations based on a CNN framework.

SSD requires only one shot to recognize several objects within an image, compared to two shots required by region proposal network-based techniques like R-CNN [27,28]. The first shot is used to generate region proposals, and the second shot is used to detect the object associated with each proposal [2]. As a result, SSD is much quicker than R-CNN.

The COVID-19 outbreak has quickly wreaked havoc on our day-to-day lives, impeding the flow of commerce and travel across international borders. Protecting one's face by using a face mask has emerged as the new standard practice. In [17, 18] the not too distant future, many suppliers of public services will need their customers to wear masks that are suited for the environment to get their services.

As a result, the identification of face masks has developed into an essential obligation to support international culture. In [19], [20] using several essential Machine Learning technologies such as TensorFlow, Keras, OpenCV, and Scikit-learn, the method presented in this research offers a straightforward approach to accomplishing this goal. The

method that is proposed can correctly recognize the face in the picture or video, and it then decides whether or not the subject is wearing a mask [4]. In addition, it can recognize a face even when it is covered by a mask, both while it is moving and while it is being seen on video [21, 22,30], and [23,29]. The approach attains high precision. To properly determine the presence of masks and avoid producing overfitting, we study the ideal parameter values for the Convolutional Neural Network model (CNN) [33-34].

3. MATERIALS AND METHOD

In this section, we discuss materials and implementation used for facemask detection on gates. Our proposed method consists of four major steps that are: (i) Object Detection (ii) Object Localization (iii) Objectness + box (iv) classification and confidence score. We first use a detector to extract ROI and divide inputs into two parts: the object region, which is crucial, and the context region, which is secondary. The ROI score is used to restore the classification outcome. Finally, we do classification with a confidence score. The detailed proposed flow is shown in figure 2.
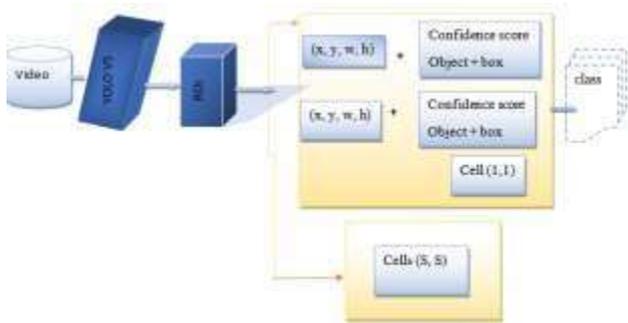

Figure 2: Proposed model for Facemask Detection and classification.

To recognize the faces, we employed the Viola-Jones algorithm. The three strategies employed by [14] are as follows: Cascade, Integral Image (representation), and AdaBoost (Classifier). Several pixel-wise operations are used to compute the integral frame. We used equation 1 to calculate the sum of the left and top pixels of the impacted pixel, and the integral image is calculated.

$$P(x', y') = \sum I(x, y) \; where \, x' \geq x \, and \, y' \geq y$$
$$(1)$$

The interest RIO locations are found using Hessian Matrix Approximation. RIO is determined using equation 3.

$$\beta(I, \alpha) = \begin{bmatrix} Cxx(I, \alpha) & Cxy(I, \alpha) \\ Cxy(I, \alpha) & Cyy(I, \alpha) \end{bmatrix}$$
$$(2)$$

Where $Cxx(I, \alpha)$, $Cxy(I, \alpha)$, $Cyy(I, \alpha)$ are Gaussian convolution of second order derivative.

By forming a rectangle region around the interest locations, the descriptor is extracted. Before computing the Haar wavelet, the region is partitioned into smaller subregions.

We increase the size of the object bounding box by a factor of 1:1 while maintaining the bounding box's center the same. The obtained result illustrates that increasing the bounding box size can increase classification performance.

Because LSVM is very fast, we utilized linear SVM with non-linear X^2-kernel to train the model. To achieve a balance, we use the homogeneous kernel map, which estimates explicit feature mappings to approximate the homogeneous additive kernels, to compute a linear approximation to X^2-kernel [32].

The proposed model efficiently resolves the problem and accurately recognizes the faces. In the detection, the process algorithm collects all features of video frames. In the next step, the layering process started where all features are combined and sent for the prediction. In the last step, the prediction step is taken where all features are gathered from the image and performed the prediction as shown in figure 3.
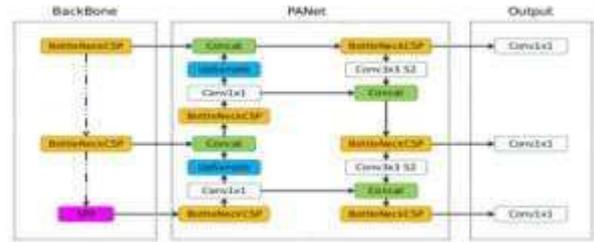

Figure 3: Model Structure of YOLO V5[10]

The model is exploited with fine-tuned parameters. YOLO V5 is used to classify face recognition into with and without masks, as shown in figures 4 [24 - 26]. The dataset is split into 80% training and 20% testing sets and classes are with masks and without a mask. Moreover, the 300 images are from Class 1 which is with a mask, and 450 without a mask after that images are augmented using the technique of rotation, translation, and scaling to get more images and overall images become 1200 for With mask and 1350 without a mask. Furthermore, we used Yolo Fine Tuned algorithm with change weights as well as activation function.

Figure 4: Facemask detection

4. RESULTS AND DISCUSSION:

In this section, we discuss the steps and parameters that were employed during the computation of results. In the facemask detection process, we tested the proposed model. During testing, the proposed model accuracy and error rate were both taken into account while detecting fasemask. The result found are shown in figure 5 and
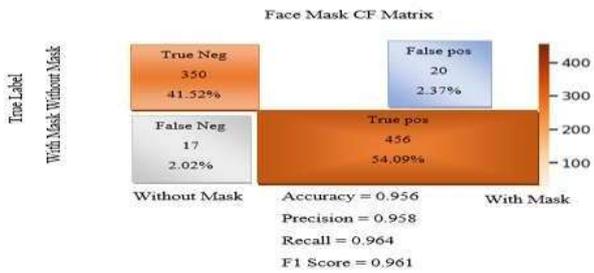
Figure 5: Confusion Matrix

Table 1 Confusion matrix

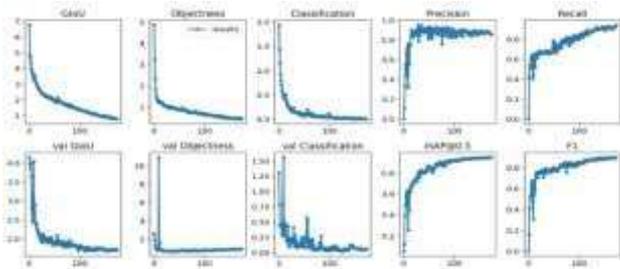| Class	n(Truth)	n(classified)	Accuracy	Precision	Recall	F1 Score
With Mask	367	370	95.61%	0.95	0.95	**0.95**
Without Mask	476	473	95.61%	0.96	0.96	**0.96**

Figure 6: Fasemask detection results

We used the Generalized Intersection of Union (GIoU) loss function for the YOLO V5. The GıoU maximizes the overlap area of actual and predicted boundary boxes. As shown in figure 6, you is initially high at the point of prediction then slowly overlaps with the actual value. This shows that you gives good results that achieve good precision which is 97%. The value you are used for training purpose, this graph shows that this loss function maximizes the overlap area of ground truth and predicted boxes as shown in figure 4. These results show that YOLO V5 accurately classifies face recognition with and without masks as shown in table 1 and figure 5. In the ground reality, there are 367 faces with masks and 476 without.

Moreover, YOLO V5 is used to detect the face accurately. In figure 6, the objectness plot shows how accurately YOLO V5 classifies the faces with and without masks. The objectness loss is due to wrong face recognition in GIoU prediction. The classification plot shows that the classification loss function deviates from predicting 2 for actual classes and 0 for other classes. This shows that YOLO V5 classifies the classes more accurately as shown in the results.

Furthermore, precision and recall plots show that YOLO V5 detects the objects accurately. The high precision shows that YOLO V5 accurately predicts the objects. High precision shows the correctness of prediction. The Recall plot shows how accurately find all the positives by using the YOLO V5 algorithm. The recall is high means our results are good and YOLO V5 detects the faces efficiently.

The Map@0.5 plot shows the mean of average precision at 0.5 you. The Map is gradually increasing which shows how accurate the results are and how efficiently the YOLO V5 algorithm detects the faces with and without masks. The F1 is a harmonic mean of precision and recall. As shown in the results, precision and recall are high which means our F1 score should also be high which shows the better performance of our used algorithm YOLO V5 that classifies the faces with and without masks accurately. Our results depict that YOLO V5 detects the faces with and with masks more accurately and works efficiently in classification. We also used the public dataset to compare our results to the current state of the art. The outcome is shown in table 2.

Table 2: State-of-the-art performance evaluation

Dataset	Accuracy
Thermal Cheetah Dataset	0.942
Anki Vector Robot Dataset Dataset	0.985
Drone Gesture Control Dataset	0.971
EgoHands Dataset	0.982
Pascal VOC 2012 Dataset	0.980

5. CONCLUSION

In this paper, the deep learning algorithm YOLO V5 is exploited to accurately classify faces with and without masks. YOLO V5 used the loss functions GIoU, objectness, and classification. That shows how accurately YOLO V5 detects the objects. The dataset is split into 80% training and 20% testing sets and classes are with masks and without a mask. Moreover, the 300 images are from Class 1 which is with a mask, and 450 without a mask after that images are augmented using the technique of rotation, translation, and scaling to get more images and overall images become 1200 for With mask and 1350 without a mask. The used algorithm is evaluated through simulations in which the used algorithm YOLO V5 precision is 96% and accuracy is 95%. The results show that YOLO V5 detects objects more accurately.

REFERENCES

[1]   A. D. Miller, T. B. Murdock, and M. M. Grotewiel, "Addressing Academic Dishonesty Among the Highest Achievers," *Theory Into Practice,* vol. 56, pp. 121-128, 2017.

[2]   Li, E. "Research on Face Detection Methods". 4th International Conference on *Signal Processing and Machine Learning*, 2021.

[3]   Genemo, M.D. "Suspicious activity recognition for monitoring cheating in exams". *Proc.Indian Natl. Sci. Acad.* 88, 1–10 (2022).

[4]   Genemo, M. D. (2022). Suspicious activity recognition for monitoring cheating in exams. Proceedings of the Indian National Science Academy, 1-10.

[5]   Reddy, S., Goel, S., & Nijhawan, R. " Real-time Face Mask Detection Using Machine Learning/ Deep Feature-Based Classifiers For Face Mask

Recognition". *IEEE Bombay Section Signature Conference* (IBSSC), 2021.

[6] Luh, G. "Face detection using a combination of skin color pixel detection and Viola-Jones face detector" *International Conference on Machine Learning and Cybernetics*, 2014.

[7] Vidal, A., Jha, S., Hassler, S., Price, T., & Busso, C. "Face detection and grimace scale prediction of white-furred mice. *Machine Learning with Applications*", 8, 100312.2022.

[8] Alakkari, S., & Collins, J. J. "Eigenfaces for Face Detection" A Novel Study. 12th International Conference on Machine Learning and Applications, 2014.

[9] Jeong, H. J., Park, K. S., & Ha, Y. G. (2018, January). Image preprocessing for efficient training of YOLO deep learning networks. In 2018 *IEEE International Conference on Big Data and Smart Computing* (BigComp) (pp. 635-637).

[10] Garg, D., Goel, P., Pandya, S., Ganatra, A., & Kotecha, K. (2018, November). A deep learning approach for face detection using YOLO. *In 2018 IEEE Punecon* (pp. 1-4). IEEE.

[11] Al-Masni, M. A., Al-Antari, M. A., Park, J. M., Gi, G., Kim, T. Y., Rivera, P., ... & Kim, T. S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Computer methods and programs in biomedicine, 157, 85-94.

[12] Wu, D., Lv, S., Jiang, M., & Song, H. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Computers and Electronics in Agriculture, 178, 105742.

[13] George, J., Skaria, S., & Varun, V. V. (2018, February). Using a YOLO-based deep learning network for real-time detection and localization of lung nodules from low dose CT scans. In Medical Imaging 2018: Computer-Aided Diagnosis (Vol. 10575, p. 105751I). International Society for Optics and Photonics.

[14] Magalhães, S. A., Castro, L., Moreira, G., Dos Santos, F. N., Cunha, M., Dias, J., & Moreira, A. P. (2021). Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse. Sensors, 21(10), 3569.

[15] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable cities and society, 65, 102600.

[16] Zhuang, Zhemin, Guobao Liu, Wanli Ding, Alex Noel Joseph Raj, Shunmin Qiu, Jingfeng Guo, and Ye Yuan. "Cardiac VFM visualization and analysis based on YOLO deep learning model and modified 2D continuity equation." Computerized Medical Imaging and Graphics 82 (2020): 101732.

[17] Ardhianto, P., Subiakto, R. B. R., Lin, C. Y., Jan, Y. K., Liau, B. Y., Tsai, J. Y., ... & Lung, C. W. (2022). A Deep Learning Method for Foot Progression Angle Detection in Plantar Pressure Images. Sensors, 22(7), 2786.

[18] Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. ACM Computing Surveys (CSUR), 51(5), 1-36.

[19] Guo, Yanming, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew. "Deep learning for visual understanding: A review." Neurocomputing 187 (2016): 27-48.

[12] Zhou, Xinyi, Wei Gong, WenLong Fu, and Fengtong Du. "Application of deep learning in object detection." In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), pp. 631-634. IEEE, 2017.

[21] Xiao, Youzi, Zhiqiang Tian, Jiachen Yu, Yinshu Zhang, Shuai Liu, Shaoyi Du, and Xuguang Lan. "A review of object detection based on deep learning." Multimedia Tools and Applications 79, no. 33 (2020): 23729-23791.

[22] Mittal, P., Singh, R., & Sharma, A. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. Image and Vision computing, 104, 104046.

[23] Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. Neurocomputing, 396, 39-64.

[24] Srivastava, S., Narayan, S., & Mittal, S. (2021). A survey of deep learning techniques for vehicle detection from UAV images. Journal of Systems Architecture, 117, 102152.

[25] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., & Garcia-Rodriguez, J. (2018). A survey on deep learning techniques for image and video semantic segmentation. Applied Soft Computing, 70, 41-65.

[26] Zujovic, J., Gandy, L., Friedman, S., Pardo, B., & Pappas, T. N. (2009, October). Classifying paintings by artistic genre: An analysis of features & classifiers. In 2009 IEEE International Workshop on Multimedia Signal Processing (pp. 1-5). IEEE.

[27] Kobylin, O. A., Gorokhovatskyi, V. O., Tvoroshenko, I. S., & Peredrii, O. O. (2020). The application of non-parametric statistics methods in image classifiers based on structural description components. Telecommunications and Radio Engineering, 79(10).

[28] Dredze, M., Gevaryahu, R., & Elias-Bachrach, A. (2007, August). Learning fast classifiers for image spam. In CEAS (pp. 2007-487).

[29] Kobylin, O. A., Gorokhovatskyi, V. O., Tvoroshenko, I. S., & Peredrii, O. O. (2020). The application of non-parametric statistics methods

in image classifiers based on structural description components. Telecommunications and Radio Engineering, 79(10).

[30] Wenger, J., Kjellström, H., & Triebel, R. (2020, June). Non-parametric calibration for classification. In International Conference on Artificial Intelligence and Statistics (pp. 178-190). PMLR.

[31] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," *IEEE TPAMI* , 2009.

[32] A. Vedaldi and A. Zisserman, *"Efficient additive kernels via explicit feature maps*," in *CVPR*, 2010

[33] Wu, Z., Xiong, Y., Yu, S., & Lin, D. (2018). Unsupervised feature learning via non-parametric instance-level discrimination. arXiv preprint arXiv:1805.01978.

[34] Gorokhovatskyi, V. O., Tvoroshenko, I. S., & Vlasenko, N. V. (2020). Using fuzzy clustering in structural methods of image classification. Telecommunications and Radio Engineering, 79(9).

[35] Chen, R. C. (2019). Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. *Image and Vision Computing*, 87, 47-56.