An Adaptive Neural Network Model for Clinical Face Mask Detection

OLADAPO TOLULOPE IBITOYE[1*], OLUWAFUNSO OLUWOLE OSALONI[1],
SAMUEL OLUFEMI AMUDIPE[2], OLUSOGO JULIUS ADETUNJI[3]
[1]Department of Electrical, Electronics and Computer Engineering,
Afe Babalola University,
Ado Ekiti,
NIGERIA

[2]Department of Mechanical and Mechatronics Engineering,
Afe Babalola University,
Ado Ekiti,
NIGERIA

[3]Department of Computer Engineering,
Bells University of Technology,
Ota,
NIGERIA

Corresponding Author

Abstract: - Neural networks have become prominent and widely engaged in algorithmic-based machine learning networks. They are perfect in solving day-to-day issues to a certain extent. Neural networks are computing systems with several interconnected nodes. One of the numerous areas of application of neural networks is object detection. This area is now very prominent due to the coronavirus disease pandemic and the post-pandemic phases where wearing of clinical face mask is imminent. Wearing a protective face mask in public and a clinical face mask in a hospital environment slows the spread of the virus and any other respiratory-related contagious diseases, according to experts' submission. This calls for the development of a reliable and effective model for detecting face masks on people's faces during compliance checks. The existing neural network models for facemask detection are characterized by their black-box nature and large dataset requirement. The highlighted challenges have compromised the performance of the existing models. The proposed technique utilized the Faster R-CNN model on the Inception V3 backbone to reduce system complexity and dataset requirements. The model was trained and validated with very few datasets and evaluation results show an overall accuracy of 96% regardless of skin tone.

Key-Words: - convolutional neural network, face detection, face mask, masked faces, inception V3, machine learning

Received: July 15, 2022. Revised: September 28, 2023. Accepted: October 8, 2023. Published: October 16, 2023.

1 Introduction

The identification of face masks plays a vital role in security and surveillance systems, especially during the ongoing pandemic caused by the breakout of the coronavirus disease in 2019. The implementation of an effective system for detecting and identifying face masks has become imperative in various domains, such as conducting face mask compliance assessments and enhancing facial security measures. Research conducted on the Corona Virus Disease 2019 (COVID-19) has demonstrated that the utilization of face masks can impede the transmission of this highly contagious pathogen, [1]. As a result, a significant number of organizations have implemented a policy requiring individuals to wear face masks to gain entrance. However, the manual verification and enforcement of face mask usage in public settings remain arduous tasks.

Research on "face mask detection" has recently piqued the interest of the "computer vision community". Research into building automatic face mask identification and recognition of faces covered by masks has led to the development of deep learning applications for digital image processing, [1], [2]. According to, [2], [3], Deep learning refers to a "deep neural network" capacity to absorb new information straight from input data, [4]. A deep learning technique called "Convolutional Neural Network" is mainly employed in object detection and image processing, [2], [5], [6].

Non-occluded datasets that show the primary facial characteristics, such as the eyes, nose, and mouth, were utilized to develop the traditional face recognition systems. Such a system of face recognition is not useful in this era of the pandemic which occasioned the wearing of protective facemasks that occlude human face, [7], [8]. A growing number of research articles containing masked faces datasets have been published, although the effectiveness of such systems on people with dark complexion is relatively poor. This study supports the third Sustainable Development Goal of the United Nations which focuses on good health and wellbeing, [9]. The results of this research will contribute to people's safety and health during a pandemic and afterward.

The rest of the paper is organized as follows. Section two gives an analysis of different techniques used in related works. Section three discusses the methodology. Section four presents the results of the system evaluation. Section five concludes the study with recommendations for future research.

2 Review of Related Works

The development of "masked face detection" systems goes through some stages. Image acquisition is typically the first stage of any object detection system, followed by image pre-processing. Masked face detection is performed at stage three. There are further stages, specifically for systems designed to examine detected masked faces in more detail. The identified stages may include, but are not limited to, mask positioning, gender identification, and identification of masked faces. A typical face mask detection system is shown in Figure 1.

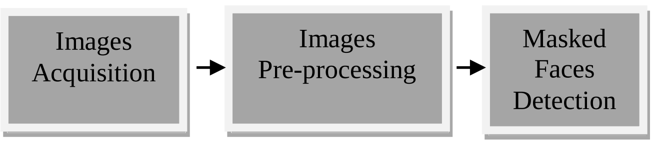

Fig. 1: Typical masked faces detection system, [10].

One of the most crucial and challenging tasks in object detection is face detection, [11], [12]. The following are the three categories of face detection. "Boost-based face detection" falls under the first category and makes use of "boosted cascade Haar features and normalized pixels' difference." The second category is based on deformable component models, which replicate the deformation of faces. The third category makes use of CNN, whose features are directly derived from the input images, [13], [14], [15], [16].

The CNN network's several spatial compressions have led to a significant level of system complexity, [17], [18], [19]. Without sacrificing efficiency, a less complicated network will minimize the complexity of the whole system, [20], [21]. The authors in, [22], [23], [24], developed face mask extractors from video clips. The assessment demonstrates great potency with offline images and low potency for real-time operation. Some other basic neural networks have been realized in, [24], [25], [26], [27]. To enhance such a system, a real-time still image extractor from video clips is required.

The majority of the systems proposed in the existing literature have not been implemented in real time. The current detectors also employed a dataset consisting of individuals with fair complexion to train the model. Hence, there is a need for a real-time system that can be trained on a diverse dataset of individuals with varying complexions. Such a system would possess significant value and global relevance.

3 Proposed Methods

The developed system is divided into two phases: model training and implementation. Each phase comprises several tasks that were completed successfully, as indicated in Figure 2. The training process involves validating the model to prevent over-fitting and training the model for best fit. The model is extracted during the implementation phase and then deployed as a full system.

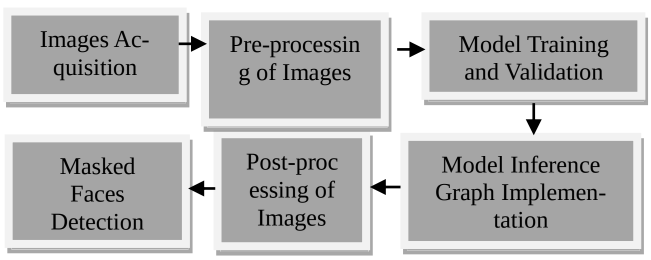

Fig. 2: Proposed system overview

3.1 Image Acquisition and Preprocessing

Acquiring face images is the first step in the model training phase. Next, the images are prepared for additional augmentation, and finally, the model is trained and validated. Acquisition of face images precedes image pre-processing to detect face masks in the model testing phase. After the detection of the face mask, the face images were processed again for effective recognition of the face behind the mask using template matching techniques. The last stage in the implementation phase is the storage of the recognized faces extracted in a database.

One thousand images of dark skin-masked faces and 1,000 images of skin faces without masks were taken. The images were preprocessed by scaling them at a specific ratio to maintain consistency and also by the application of a cropping filter to capture the relevant portions of the masking faces. This speeds up network processing and simplifies computation. The cropping was completed by 240-by-240-pixel normalization of all images.

3.2 Model Training and Validation

A faster region-based convolutional neural network (RCNN) with Inception V3 architecture was used to develop the detection model due to its reduced complexity and ability to learn faster with the limited number of datasets. Convolutions in the original model are more effective in terms of computational complexity because of the employment of clever “factorization techniques”. The Inception V3 model factorizes a convolution of 7×7 and uses an additional classifier to propagate information about labels. The network's performance improved as a result of convolution factorization. For instance, a 3×3 convolution with the same number of filters is computationally $49/9 = 5.44$ times more expensive than a 7×7 convolution over a grid with 'n' filters and 'm' filters. Utilizing a momentum optimizer, the faster-RCNN Inception V3 model was trained. Here, 250 images were utilized to validate the model, and 750 images were utilized to train the model using 15 epochs.

The RPN received its input from the final convolution layer of the CNN. Regression box differences about anchors were predicted by the RPN together with “objectness”. To produce proposals, these offsets were positioned alongside the anchors. The ROI Align layer, followed by the classifier and “box regressor”, received the RPN proposal. The architecture of faster R-CNN is shown in Figure 3. Each feature map channel is designed to undergo independent pooling for extraction. Numerous quantization procedures are required to map the generated proposal to precise indexes during ROI pooling implementation. These quantization operations are capable of introducing misalignments between the ROI and extracted features. This, however, has some negative impact on object detection. To address the misalignment issue, ROI alignment was used in this study to remove all possible quantization operations from the network.

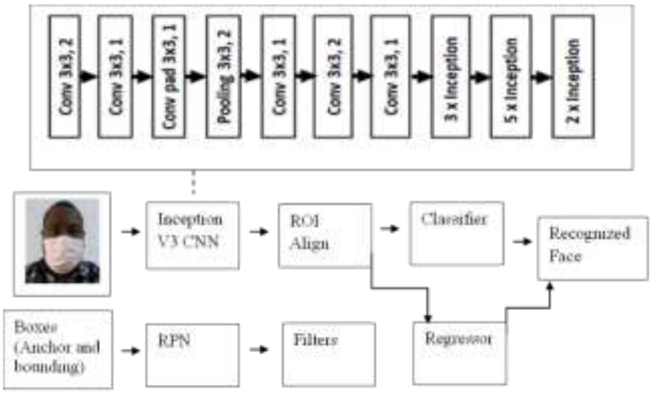

Fig. 3: Faster R-CNN Model on Inception V3 backbone

Upon achieving satisfactory alignment of the ROI, the convolutional layer was further designed to extract distinctive features from the input facial images. The process of convolution involves the utilization of the input image to compute the dot product, resulting in the generation of a feature map with reduced dimensions as the output. The convolution layer's output feature was utilized by the fully connected layer to identify and forecast the bounding box score for the given facial picture input. The optimizer was supported by the utilization of the Adaptive Gradient Algorithm, which facilitated the adaptation of the learning rate throughout the training process of the model. Additionally, the technique of early stopping was employed to halt the training process when there was a lack of discernible improvement. This approach proved beneficial in mitigating the potential issue of model over-fitting and minimizing the duration of the training period. Upon successful training, the inference graph of the model was generated.

3.3 System Implementation

The inference graph generated after a successful model training was implemented on Jupyter Notebook, an open-source web application that permits the creation and implementation of documents that contain codes, algorithms, visualizations, and narrative texts. To achieve good results from the detector, the extracted face mask images were further subjected to image processing. Scale uniformity through rescaling of the extracted images has been performed using a suitable equation. Image binarization was also performed using a suitable equation to remove a certain number of unwanted details from the extracted face images.

3.4 System Evaluation

The accuracy of training and validation processes was obtained from the accuracy-epochs curves generated by the model. These were engaged in the evaluation of the trained model. Training and validation losses were also computed by the model. These losses amount to the trained model classification loss, which is a measure of the predictive inaccuracy of the model. The overall loss function of the model is obtained from the model classification loss. After a successful training procedure, the system was validated with 250 images (positive and negative).

After validating the model, the entire system was tested in real-time with 50 random faces with masks and 50 random faces without masks. The system was evaluated using specificity and accuracy as defined in Equation 1 and Equation 2.

$$Specificity = 100\frac{TN}{TN + FP} \quad (1)$$

$$Accuracy = 100\frac{(TP + TN)}{(TP + FP + TN + FP)} \quad (2)$$

where TN is “true negative”, FP is “false positive” and TP is “true positive”.

In this study, TP is defined as the number of masked faces correctly detected with masks; TN is defined as the number of faces correctly detected without masks; FP is defined as the number of faces wrongly detected as having masks.

4 Results

To demonstrate how the model responded to the training and validation datasets, the training accuracy, training loss, validation accuracy, and validation loss per epoch curves were automatically computed using the model prediction operation. The results of the 15 training and validation epochs are shown in Table 1 and Table 2.

Table 1. Accuracy and loss results for model training

Epochs	Training Accuracy	Training Loss
1	0.7580	0.8510
2	0.9250	0.2500
3	0.9450	0.2000
4	0.9500	0.1800
5	0.9550	0.1500
6	0.9600	0.1350
7	0.6200	0.1000
8	0.9640	0.0900
9	0.9680	0.0800
10	0.9700	0.0700
11	0.9720	0.0650
12	0.9750	0.0670
13	0.9770	0.0690
14	0.9790	0.0500
15	0.9800	0.0400

Table 2. Accuracy and loss results for model validation

Epochs	Validation Accuracy	Validation Loss
1	0.9215	0.3000
2	0.9450	0.1800
3	0.9625	0.1500
4	0.9645	0.1350
5	0.9670	0.1000
6	0.9680	0.1400
7	0.9690	0.1000
8	0.9695	0.1000
9	0.9700	0.1400
10	0.9710	0.1000
11	0.9720	0.1000
12	0.9750	0.1000
13	0.9770	0.1002
14	0.9750	0.1003
15	0.9720	0.1005

Figure 4 illustrates a graphical plot of the relationship between accuracy and epoch. The data presented in the plot demonstrates that the maximum level of accuracy, specifically 0.9800, was achieved during the 15th epoch. The 13th epoch yielded a validation accuracy of 0.9770, which represents the maximum accuracy achieved for the best fit. It is worth mentioning that the validation accuracy experienced a decline after the 13th epoch, which suggests the occurrence of overfitting during the 14th and 15th epochs.

The plot of loss against epoch is depicted in Figure 5. The plot demonstrates that the 15th epoch exhibited the lowest training loss, with a recorded value of 0.0400, indicating its suitability for achieving optimal fit. Additionally, it was noted that the lowest validation loss for optimal fitting remained steady for epochs 10, 11, and 12. It is worth mentioning that the validation loss started increasing at the 13th epoch, this is again evidence of overfitting at the 13th epoch.

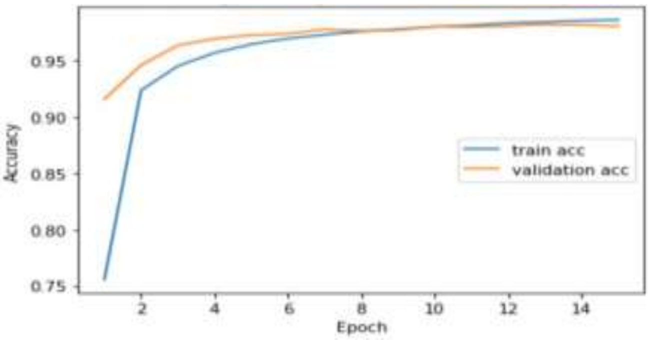

Fig 4: Computed plot of training accuracy and validation accuracy

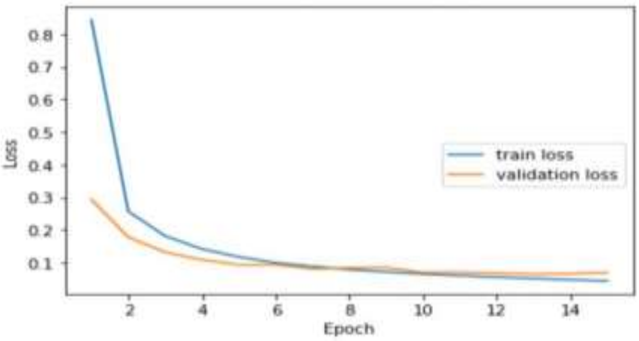

Fig. 5: Computed plot of training loss and validation loss

The findings of this study indicate the efficacy of the proposed system, despite its utilization of a limited dataset. A total of fifty individuals, consisting of twenty-five individuals wearing masks and twenty-five individuals not wearing masks, were selected to undergo facial capture utilizing the developed technique. The true positive (TP) value is determined to be 24, whereas the true negative (TN) value is also 24 and the false positive (FP) value is 1. The model's specificity and accuracy were determined to be 96% each, based on the calculations using Equation 1 and Equation 2.

5 Conclusion

In this study, a system of masked face detection was developed using a Faster-Region-based Convolutional Neural Network with Inception V3 architecture. The system leverages the unique features of Region of Interest Align to resolve the issues of misalignments caused by the use of Region of Interest Pooling engaged in the traditional Faster-RCNN. The techniques and the developed system were implemented using a Python-based integrated development environment called “Anaconda Navigator”. Regardless of skin tone or gender, the developed masked faces detector achieved an accuracy of 96% during the evaluation of the system in real time. A robust system with the capacity to capture and process a wide range of areas at a time may be included in future research and development on masked face detection systems.

References:

[1] M. F. Ali and M. . S. Al-Tamimia, "Face mask detection methods and techniques: A review," *Int. J. Nonlinear Anal. Appl.,* vol. 13, no. 1, pp. 3811-3823, 2022.

[2] N. Ullah, A. Javed, M. A. Ghazanfar, A. Alsufyani and S. Bourouis, "A novel DeepMaskNet model for face mask detection and masked facial recognition," *Journal of King Saud University–Computer and Information Sciences,* vol. 34, no. 2022, pp. 9905-9914, 2022.

[3] M. A. Firas Amer and M. S. Al-Tamimi, "Face mask detection methods and techniques: A review," *Int. J. Nonlinear Anal. Appl.,* vol. 13, no. 1, pp. 3811-3823, 2022.

[4] K. K. Archana, R. Abishek, S. Archana and V. Jagadeeshwaran, "Face Mask Detection System," *International Journal of Research and Analytical Reviews,* vol. 9, no. 2, pp. 63-67, 2022.

[5] Y. Hu, Y. Xu, H. Zhuang, Z. Weng and Z. Lin, "Machine Learning Techniques and Systems for Mask-Face Detection—Survey and a New OOD-Mask Approach," *Applied Sciences MDPI,* vol. 12, no. 9171, pp. 1-37, 2022.

[6] P. Gupta, V. Sharma and S. Varma, "A novel algorithm for mask detection and recognizing actions of human," *Elsevier,* vol. 198, no. 2022, pp. 1-10, 2022.

[7] N. Mheidl, M. Fares, H. Zalzale and J. Fares, "Effects of Face Masks on Interpersonal Relatioshionships during COVID 19 Pandemic," *Frontiers in Public Health,* vol. 8, pp. 1-6, 2020.

[8] Y. Said, "Pynq-YOLO-Net: An Embedded

Quantized Convolutional Neural Network for Face Mask Detection in COVID-19 Pandemic Era," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 11, no. 9, pp. 100-106, 2020.

[9] S. Morton, D. Pencheon and N. Squires, "Sustainable Development Goals (SDGs), and their implementation: A national global framework for health, development and equity needs a systems approach at every level," *British Medical Bulletin,* vol. 124, pp. 81-90, 2017.

[10] O. Ibitoye, "A Brief Review of Convolutional Neural Network Techniques for Masked Face Recognition," in *2021 IEEE Concurrent Processes Architectures and Architectures and Embedded Systems Virtual Conference (COPA),* 2021.

[11] W. Hariri, "Efficient Masked Face Recognition Method During the Covid-19 Pandemic," *Preprint,* pp. 1-8, 2020.

[12] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra and S. Agarwal, "Face Mask Detection using Transfer Learning of Inception V3," *Preprint,* pp. 1-10, 2020.

[13] P. Shitala, Y. Li, D. Lin and D. Sheng, "maskedFaceNet: A Progressive Semi-Supervised Masked Face Detector," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021.

[14] S. Shete, K. Tingre, A. Panchal, V. Tapse and . B. Vyas, "Mask Detection and Tracing System," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* vol. 7, no. 2, pp. 406-412, 2021.

[15] B. Qin and D. Li, "Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19," *Sensors, MDPI,* vol. 20, no. 18, pp. 1-23, 2020.

[16] N. U. Din, K. Javed, S. Bae and J. Yi, "A Novel GAN-Based Network for Unmasking of Masked Face,”," *IEEE Access,* vol. 8, p. 44276–44287, 2020.

[17] E. Ryumina, D. Ryumin, D. Ivanko and A. Karpov, "A Novel Method for Protective Face Mask Detection using Convolutional Neural Networks and Image Histograms," in *4th Int. Worksh. on “Photogrammetric & computer vision techniques for video surveillance, biometrics and biomedicine,* 2021.

[18] M. Ngan, P. Grother and K. Hanaoka, "Face recognition accuracy with masks using pre-COVID-19 algorithms:," NISTIR 8311, United States of America, 2020.

[19] A. Mahore and M. Tripathi, "Detection of 3D Mask in 2D Face Recognition System Using DWT and LBP," in *IEEE 3rd International Conference on Communication and Information System.*, 2018.

[20] Y. Li, K. Guo, Y. Lu and L. Liu, "Cropping and attention based approach for masked face recognition," *Applied Intelligence,* vol. 51, pp. 3012-3025, 2021.

[21] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria and J. Hemanth, "A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," *Sustainable Cities and Society,* vol. 66, pp. 1-11, 2021.

[22] I. Madhura and N. Mehendale, "Real-Time Face Mask Identification Using Facemasknet Deep Learning Network," *Preprint,* pp. 1-7, 2020.

[23] S. Y. Wang, B. Luo and J. Shen, "Face Masks Extraction in Video," *Springer,* vol. 127, pp. 625-641, 2019.

[24] M. Wang and W. Deng, "Deep Face Recognition: A Survey," *Neurocomputing - Preprint,* pp. 1-75, 2020.

[25] D. Ekberjan and A. S. Albert, "Continuous Real-Time Vehicle Driver Authentication Using Convolutional Neural Network Based Face Recognition," *IEEE,* 2018.

[26] B. Batagelj, P. Peer, V. Štruc and S. Dobrisek, "How to Correctly Detect Face-Masks for COVID-19 from Visual Information," *Apllied Science MDPI,* vol. 11, no. 2070, pp. 1-24, 2021.

[27] M. McDonald and Y. Cen, "COVID-19 Face Mask Detection Alert System," *Computer Engineering and Intelligent Systems,* vol. 13, no. 2, 2022.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Oladapo Tolulope Ibitoye conceptualized the research idea, supervised the entire experimental process of the research, wrote the original draft, reviewed and edited the final draft.

All other authors equally contributed in the experimental process.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The study was supported by the Afe Babalola University, Nigeria.

Conflict of Interest

The authors have no conflict of interest to declare.