

Transfer Learning Approach for Mask Detection by using Extended VGG16

Neharika Boddakayala, Mentor College, Toronto, Canada

Meysam Effati, Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada

Corresponding Author: Neharika Boddakayala , Email: neharikaboddakayala@gmail.com

Abstract

Wearing masks had been made compulsory in both indoor and outdoor public settings across several countries throughout the duration of the COVID-19 pandemic. According to the World Health Organization (WHO), proper utilization of face masks can practically and effectively lower the potential transmission of the virus through respiratory droplets. In this paper, we introduce an innovative deep learning approach utilizing our enhanced VGG-16 model to identify and categorize instances where face masks are not worn, worn correctly, or worn incorrectly. Our Deep Learning (DL) method incorporates transfer learning by employing pretrained VGG-16 weights. This approach serves to minimize training time while enhancing detection precision. The Real-World Masked Face Dataset (RMFD), Face Mask Detection, and Face Mask Detection datasets are used for training and validation. The results show a classification training accuracy of 97.43% when discerning between the absence of a mask, correctly worn masks, and incorrectly worn masks across various real-world poses.

Keywords: Image Processing, Face Mask Detection, AI Model, Convolutional Neural Networks, VGG16

I. INTRODUCTION

As of July 2023, the global impact of the COVID-19 pandemic has led to more than 6,954,336 deaths and 769,369,823 confirmed cases of the virus [1]. The implementation of universal mask use could have prevented nearly 130,000 of those deaths, in the United States from 22 September 2020 through 28 February 2021 [2].

II. RELATED WORKS

A. MobileNet

For face detection, Ref. [3] uses a Single Shot Multi-box Detector (SSD) and for classification it uses a hybrid approach based on ResNet-50 and MobileNet models called Hybrid ResMobileNet (HResMobileNet). Their parameter optimization is executed using a hybrid optimization algorithm referred to as Adaptive Sailfish Moth Flame Optimization (ASMFO). Their model tested on two classes of mask and no mask images and achieved an accuracy of 95.2%. Ref. [4], uses

MobileNetV2 and FaceNet Model as the feature extractor for the face mask detection of two classes of Mask and no mask. The model was trained using the Real-World Masked Face Dataset (RMFD), which contains 5,000 images of masked faces, 90,000 unmasked faces and 10,000 faces and was tested by detecting faces with masks and without masks. They obtained a detection accuracy of 99.65% in determining if people were wearing or not wearing a mask, 99.52% in the facial recognition of 10 people with masks, and 99.96% for facial recognition without masks. Ref. [5] proposes a model based on the pretrained MobileNet, ResNet50, and AlexNet models. Their model is trained using a custom dataset obtained using random oversampling with data augmentation over the Medical Face Mask dataset (MAFA) which contains 23,858 masked images and 2018 non-masked images. The proposed model achieved an accuracy of 98.2% at mask detection with average inference times of 0.05 seconds per image. Ref. [6] uses a pre-trained MobileNetV2 architecture trained through a custom dataset containing 1914 images of faces with masks and 1918 without. The images were derived from the Real-World Masked Face Dataset (RMFD), Prajna Bhandary's "observations" dataset, which contains 690 images with masks and 686 images without masks, and the Labelled Faces in the Wild (LFW) dataset which consists of 13027 simulated masked faces. The model obtained a 98.7% accuracy and AUC of 0.985. Ref. [7] proposes a Caffe-MobileNetV2 (CMNV2) based on Caffe and MobileNetV2 frameworks and compares its performance with a MobileNetV2 model. The CMNV2 model was trained using 1100 images from Prajna Bhandary's "observations" dataset and the MobileNetV2 model on the ImageNet database. The proposed model achieved an accuracy of 99.64%, precision of 100%, recall of 99.28%, f1-score of 99.64%, and an error rate of 0.36%. Ref. [8] proposes a model based on MobileNetV2 architecture and trains it using a custom dataset, consisting of 3725 images of masked faces and 3,828 unmasked faces, derived from the Face Mask Detection (FMD), Face Mask (FM), and Real-World Mask Face Recognition (RMFR) datasets. The proposed model achieved a precision of 99.98%, a recall of 99.98%, an F1 score of 99.98% and an accuracy of 100%. Ref. [9] uses a MobileNetV2 model trained using 3,832 images 1914 of which are faces with masks and 1918 are those without masks. The model achieved an accuracy of 99.21%. Ref. [10] proposes a hybrid model based on

Single Shot Multibox Detector (SSD) algorithm, ResNet-10 and MobileNetV2 and compares it to preexisting models including LeNet-5, AlexNet, VGG-16, and ResNet-50. The dataset used to train the model included the 678 images in the MMD dataset and the 1376 images in Prajna Bhandary's "observations" dataset and was able to achieve a training accuracy of 92.64 %. Ref. [11] uses a MultiTask Cascaded Convolutional Neural Network (MTCNN) algorithm to identify facial landmarks and a MobileNetV2 architecture to identify masked regions. It's trained using video footage of masked people from 15 videos on YouTube with an average of 30 FPS. The proposed framework is compared with RetinaFaceMask and Cascaded Framework from Ref. [12] and Ref. [13], respectively. The proposed framework has a lower accuracy of 81.74% than the Cascaded Framework and has a higher precision rate of 94.50% and 84.39% when detecting masks and faces than RetinaFaceMask.

B. YOLOv3 and v4

Ref. [14] uses a web-scraping tool to form a dataset consisting of 300 images of faces with masks and 300 without to train a YOLOv3 model that results in a 96% classification and detection accuracy. Ref. [15] trains a YOLOv3 and TensorFlow model using a custom dataset consisting of 4,095 images of masked and unmasked individuals and obtained a 99% accuracy rate. Ref. [16] uses YOLOv3 and faster R-CNN models trained using the same 7500 images of masked and unmasked individuals, concluding that the faster R-CNN model had a better average precision of 62 compared to YOLOv3 precision of 55 while YOLOv3 has a faster inference time of 0.045s and faster R-CNN time of 0.15s. Ref. [17] uses a YOLO V4 deep learning algorithm whose AP is improved by 10% and FPS by around 12%. than YOLOv3's. Ref. [18] proposes a new deep learning algorithm, called SE-YOLOv3 based on YOLOv3 and trained using a custom dataset, Properly-Wearing-Masked Face Detection Dataset (PWMFD), consisting of 7695 properly masked faces, 10,471 unmasked faces, and 366 incorrectly masked faces. Ref. [19] uses a YOLOv4 algorithm trained by a dataset consisting of 12,133 images with a resolution of 416×416 to achieve a mAP of 69.70% and an AP75 of 73.80%. Ref. [20] uses a YOLOv4 CSP SPP model trained using a custom dataset whose images were derived from the MMD and FMD datasets. The dataset had 1415 images that were split into 3 classes, with 500 instances of a mask worn incorrectly, 4000 where a mask was worn correctly and 500 where none was worn. The model achieved a detection performance of 99.26%.

C. AlexNet

Ref. [21] uses an AlexNet CNN architecture to achieve an accuracy of 98.40%, precision of 99.05% and recall of 83.97% by training it through the RMFD dataset, which includes 5,000 mask faces and 90,000 normal faces of 525 people and the Celeb Faces Attributes (CelebA) dataset which consists of 202,599 of 10,177 unmasked celebrities. Ref. [22] proposes a

Spartan Face Detection and Facial Recognition System that uses AlexNet, VGG16, and FaceNet models. The model is trained using a dataset of 14,676 images sourced from the CASIA WebFace Dataset, which consists of 453,453 images of unmasked faces, the MaskedFace Net dataset, which has 67,193 images of correctly masked faces and 66,900 images of incorrectly masked faces, and by using Web scraping techniques. The model achieved an average accuracy of 97%.

D. ResNet-50

Ref. [23] uses a hybrid model that combines a Resnet50 algorithm for feature extraction with decision trees, a Support Vector Machine (SVM), and an ensemble algorithm for classifying face masks. 10 000 images of 5000 masked faces and 5000 unmasked faces, from the RMFD dataset as well as 1570 images, 785 of simulated masked faces and 785 of unmasked faces, from the SMFD dataset were acquired for training, validation, and testing. 13,000 simulated masked faces were additionally acquired from the LFW dataset solely for testing. The SVM algorithm achieved a testing accuracy of 99.64% in the RMFD dataset, 99.49% in SMFD and 100% in LFW. Ref [16] proposes a hybrid model with a ResNet-50 for feature extraction and YOLO v2 for face mask detection. The model is trained using a custom dataset of 1415 images sourced from the Medical Masks Dataset (MMD) which consists of 682 images of masked faces and the Face Mask Dataset (FMD) which consists of 853 images of masked faces and achieved an average precision percentage of 81%. Ref. [24] uses a pretrained ResNet50V2 model trained using the MAFA dataset which contains 35,803 images of masked faces with a minimum size of 32×32 . The model achieves a precision and recall of 97%.

E. Single Shot Multibox Detector

Ref. [25] uses a SSDIV3 network using a Single Shot Multibox Detector and an Inception V3 model trained by the Real-Time Face Mask Dataset (RTFMD), the Real-Time Face Mask Dataset Version 2 (RTFMD-V2), the FMD dataset and 3000 images from the RFMD dataset, 1500 of which are masked faces and 1500 are unmasked. The system obtained a detection accuracy of 96% on the FMD and RFMD datasets and 98.43% on the RTFMD and RTFMD-V2 datasets. Ref. [26] uses a SSDMNV2 hybrid model based on Single Shot Multibox Detector and MobileNetV2 models and trained using the 678 images in the MMD dataset and the 1376 images in Prajna Bhandary's "observations" dataset. The proposed model obtained an accuracy of 99.73% and a precision of 97.82% on images without masks and an accuracy of 99.07% and a precision of 98.13% on images with masks.

F. Other Architectures

Ref. [27] uses a pre-trained Sequential CNN model containing two 2D convolution layers to achieve an accuracy of 95.77% in Prajna Bhandary's "observations"

dataset which consists of 1376 images in which 690 are masked faces and 686 unmasked faces and an accuracy of 94.58% for the 853 images in the Face Mask Detection dataset. Ref. [28] uses a CNN model trained using a dataset containing images 858 images of faces with masks and 681 are those without masks derived from both the FMD dataset and Prajna Bhandary's "observations" dataset. The model attained an accuracy of 98.7%. Ref. [29] uses a histogram-based recurrent neural network (HRNN) with histograms of oriented gradient (HOG) as a feature extractor, and RNN for deep learning and is trained using the Labeled Face in the Wild Stimulated Masked Face Dataset (LFW-SMFD) and RMFD datasets to achieve an accuracy of 99.56%. Ref. [30] proposes a new framework named Out-of-distribution Mask (OOD-Mask) using a YOLOX model and uses a dataset containing (confusion regarding # of images) sourced from the Wider Face dataset and Kaggle (doesn't specify exactly). The model achieved a precision of 83.08% in data where masks are worn correctly, 82.04% where masks are not worn and 53.6% where masks are incorrectly worn. Ref. [31] uses a ResNet-18 model and RetinaFace algorithm and is trained using a custom dataset containing 41,934 images in which 29,532 are correctly worn facemasks, 1528 incorrectly worn and 32,012. The images were sourced from the MAFA and WiderFace datasets. The model achieves a 99.6% accuracy for two classes (Correctly worn masks and No Masks) and 99.2% accuracy for three classes (Correctly worn masks, incorrectly worn masks and No Masks). Ref. [32] proposes a lightweight CNN model trained using a custom dataset containing 3076 no masked faces and 4664 masked faces with sizes of 224×224 sourced from a website, <https://thispersondoesnotexist.com>. The model achieves an average accuracy of 98.47%. Ref. [33] proposes a hybrid model based on DenseNet201 and ResNet101 models and was trained using a total of 2,075 images with 529 incorrectly worn masks, 992 correctly worn masks, and 554 no mask face images. The proposed model achieved a 99.75% accuracy.

G. VGG-16

Ref. [34] uses a pretrained VGG-16 model trained on the Real-world Masked Face Recognition Dataset (RMFRD), which contains 5,000 images of masked faces and 90,000 images of unmasked faces, to achieve an accuracy of 91.3%. Ref. [35] proposes a model based on VGG-16 architecture trained using 5,000 images of masked faces and 5,000 images of unmasked faces. The proposed model achieved a precision of 94%. Ref. [36] uses a VGG-16 model trained through 20,000 images ranging from 800 up to 1200 pixels of masked and unmasked faces sourced from the web using a Python script. The model achieved an accuracy of 97%. Ref. [37] uses 5 pretrained models, specifically VGG-16, MobileNetV2, Inception V3, A2QW-50, and Convolutional Neural Network (CNN) to detect instances of faces with proper masks, faces with

improper masks and faces with no masks. The model was trained using the 12,000 images of unmasked faces from the MAFA dataset, 34,456 images of properly masked faces and 33,106 improperly masked faces from the Masked Face-Net dataset and 4,039 images of unmasked faces from a Bing Images dataset. Of the 5 models VGG-16 achieved the highest accuracy of 99.8%, whereas MobileNetV2 achieved 99.6%, Inception V3 achieved 99.4%, ResNet-50 achieved 99.2% and CNN achieved 99.0%.

In this paper, a new deep learning (DL) and transfer learning method based on VGG-16 is proposed to classify input images into 3 classes. 1) faces with correctly worn masks 2) faces with incorrectly worn masks or partially covered faces and 3) faces without masks. The proposed mask detection method is compared to a standard VGG-16 model to better evaluate its performance.

III. METHODOLOGY

The classification system used in this paper contains two main modules 1) image preprocessing 2) image classification. The preprocessing module resizes input images into uniform square image inputs. Figure 1 illustrates the overall structure of the classification model which is based on VGG-16. The primary modules of the architecture are discussed below.

A. Preprocessing Module

The preprocessing module takes the inputted images and resizes their dimensions to 244×244 pixels, thus changing the varying rectangular figures into uniform square images.

B. Classification Module

The pretrained VGG-16 model (on the ImageNet [38] dataset) consists of 16 layers, 13 of which are convolutional layers used for feature extraction and 3 are fully connected 3×3 layers. The model also has 5 max pooling layers and a SoftMax layer for classification [39].

The proposed model used a pretrained model and connected it to a dense layer with a rectified linear unit (ReLU) activation function. This layer was connected to a dropout layer, another ReLU activated layer and then a dropout layer. A dense layer with a softmax activation function was then added for classification. Table I expresses the output shape and parameters of the layers involved in the proposed architecture.

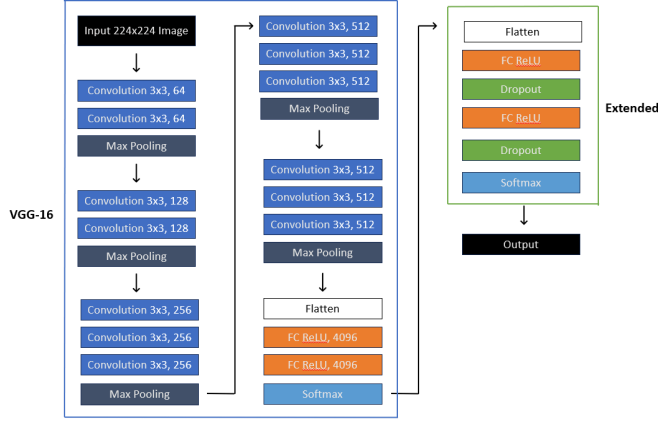


Figure 1. The proposed architecture for the extended VGG-16

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

TABLE I. EXTENDED VGG-16 PARAMETER SUMMARY

Layers	Output Shape	# of parameters
VGG-16	(None, 512)	14714688
Flatten	(None, 512)	0
Dense	(None, 512)	262656
Dropout	(None, 512)	0
Dense	(None, 256)	131328
Dropout	(None, 256)	0
Dense	(None, 3)	771
Total Parameters	15,109,443	
Trainable Parameters	15,109,443	
Non-Trainable Parameters	0	

IV. TRAINING

In this section, the Datasets and the training process are explained.

A. Datasets

In this work, the models were trained using images derived from three public datasets. These include i) Real-World Masked Face Dataset [40], ii) Face Mask Detection Dataset [41] and iii) Face Mask Detection Dataset [42]. In total 15000 images were used, of which 5000 were masked faces, 5000 incorrectly masked faces and 5000 unmasked faces.

The Real-World Masked Face Dataset [40] contains 5,000 images of masked faces, 90,000 unmasked faces and 10,000 faces. The images in this dataset have a resolution of 96 dpi with dimensions of 1024 x 1024 and contain people who are front facing or slightly turned.

2874 images were selected for the Correctly Masked class and 5000 images for the Incorrectly Masked class.



Figure 2. Sample images of faces with correctly and incorrectly worn masks from the RMFD dataset [40].

The Face Mask Detection Dataset [41] contains 4095 images split into two classes with 2165 images of faces with masks and 1930 of faces without masks. These images were sourced from Bing Search API, Kaggle datasets and the RMFD dataset. The images have varying resolutions and dimensions with faces that are front facing or slightly turned. 2126 images were selected for the Correctly Masked class and 1930 images for the No Mask class.



Figure 3. Sample images of faces with correctly worn masks and no masks from the FMD Dataset [41].

The Face Mask Detection Dataset [42] contains 3725 images of faces with masks and 3828 images of faces without masks. 1776 of both images with and without face masks were sourced from Prajna Bhandary's GitHub observations dataset, while the remaining 5777 images were derived from Google search engine. 3070 images were selected for the No Mask class.

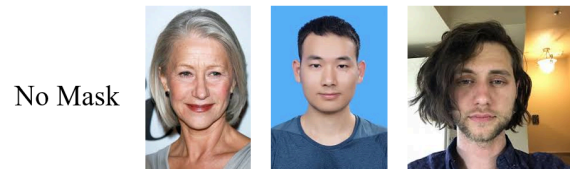


Figure 4. Sample images of faces with no masks from the FMD Dataset [42].

We used a total of 15,000 images for training. There are 10,000 images of faces with masks and 5,000 images of faces with no masks in the dataset. 5,000 correctly

masked faces, 5,000 incorrectly masked faces, and 5,000 unmasked faces. A 70%, 20%, 10% training, validation and validation split was used.

B. Training Procedure

For training, a batch size of 32, learning rate of 0.001 and 15 epochs were used. The training ran on a computer with a 16.0 GB RAM, 64-bit processor, Intel Core (TM) i7-7500U CPU and an Intel HD Graphics 620 graphics card.

C. Training Validation and Results

The model loss and accuracy for both the training and validation are illustrated in Fig. 5. The Extended VGG-16 model specifically achieved a training accuracy of 97.43% and a validation accuracy of 97.62%. As evident from the graphs below, there is a decrease in training and validation losses as well as an increase in training and validation accuracies.

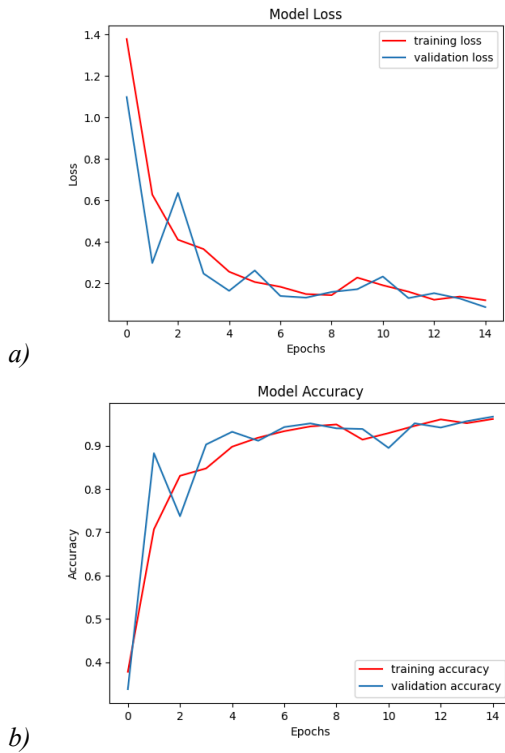


Figure 5. Extended VGG-16 model a) loss b) accuracy

It is noted that the extended VGG-16 model was trained within Google Collab using a T4 GPU.

V. COMPARISON

A comparison study for mask classification and detection was conducted between the proposed Extended VGG-16 model and the standard VGG-16 model. The ImageNet

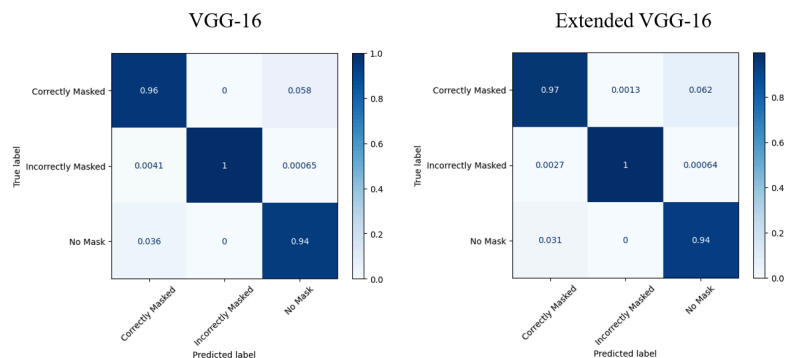
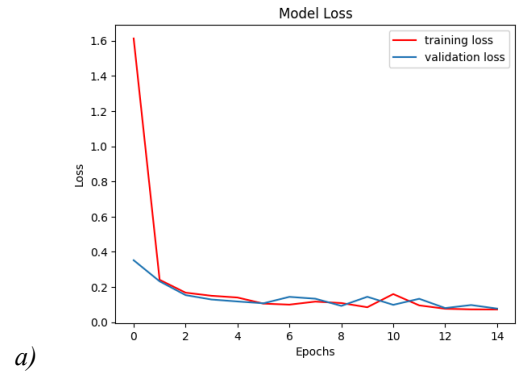
database was used to train, test, and validate the standard VGG-16 model. As illustrated in Fig. 6, the standard VGG-16 model achieved a training accuracy of 96.15% and a validation accuracy of 97.62 %.

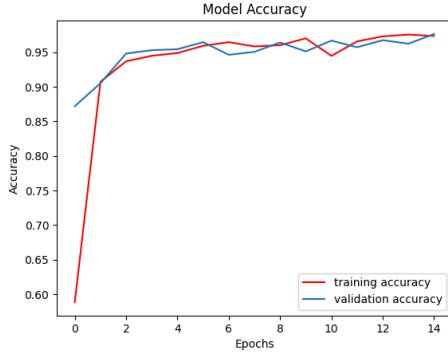
Demonstrated in Table II, the extended VGG-16 achieved a greater accuracy for validation and training and performed approximately the same compared to the VGG-16.

TABLE II. ACCURACY COMPARISON FOR VGG-16 & EXTENDED VGG-16

Method	Training Accuracy	Validation Accuracy	Testing Accuracy
VGG-16	96.15%	96.69%	97%
Extended VGG-16	97.43%	97.62%	96.67%

As illustrated in Fig. 7, the proposed extended VGG-16 model performed consistently to the standard VGG-16 model in correctly classifying images into the Incorrectly Masked and Unmasked classes. The extended VGG-16 model performed better in classifying images into the Correctly Masked class, compared to the standard model.





b)

Figure 6. VGG-16 model a) loss b) accuracy

Figure 7. Confusion matrix comparison between the standard VGG-16 model, and the Extended VGG-16 model.

VI. CONCLUSION

In this paper, we present an innovative deep learning mask detection and classification system. The deep learning structure is based on our unique extended VGG-16 structure for real-time classification of three classes: faces with no mask, faces with a correctly worn mask, and faces with an incorrectly worn mask/face covering. We obtained an overall training, validation, and testing accuracy of 97.43%, 97.62% and 96.67% respectively. Our future work will consist of adopting similar deep learning structures with other architectures such as MobileNet-V2, Single Shot Multibox Detector (SSD), or AlexNet.

REFERENCES

- [1] W. H. Organization. "WHO Coronavirus (COVID-19) Dashboard." <https://covid19.who.int/> (accessed August 14, 2023, 2023).
- [2] "Modeling COVID-19 scenarios for the United States," *Nature medicine*, vol. 27, no. 1, pp. 94-105, 2021.
- [3] R. A. S. Naseri, A. Kurnaz, and H. M. Farhan, "Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach," *Applied Soft Computing*, vol. 134, p. 109933, 2023.
- [4] J. S. Talahua, J. Buele, P. Calvopiña, and J. Varela-Aldás, "Facial recognition system for people with and without face mask in times of the covid-19 pandemic," *Sustainability*, vol. 13, no. 12, p. 6900, 2021.
- [5] S. Sethi, M. Kathuria, and T. Kaushik, "Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread," *Journal of biomedical informatics*, vol. 120, p. 103848, 2021.
- [6] A. H. I. Al-Rammahi, "Face mask recognition system using MobileNetV2 with optimization function," *Applied Artificial Intelligence*, vol. 36, no. 1, p. 2145638, 2022.
- [7] B. A. Kumar and M. Bansal, "Face mask detection on photo and real-time video images using Caffe-MobileNetV2 transfer learning," *Applied Sciences*, vol. 13, no. 2, p. 935, 2023.
- [8] S. Habib, M. Alsanea, M. Aloraini, H. S. Al-Rawashdeh, M. Islam, and S. Khan, "An Efficient and Effective Deep Learning-Based Model for Real-Time Face Mask Detection," *Sensors*, vol. 22, no. 7, p. 2602, 2022.
- [9] X. Su, M. Gao, J. Ren, Y. Li, M. Dong, and X. Liu, "Face mask detection and classification via deep transfer learning," *Multimedia Tools and Applications*, pp. 1-20, 2022.
- [10] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," *Sustainable cities and society*, vol. 66, p. 102692, 2021.
- [11] A. S. Joshi, S. S. Joshi, G. Kanahasabai, R. Kapil, and S. Gupta, "Deep learning framework to detect face masks from video footage," in *2020 12th international conference on computational intelligence and communication networks (CICN)*, 2020: IEEE, pp. 435-440.
- [12] M. Jiang and X. Fan, "Retinamask: A face mask detector," 2020.
- [13] W. Bu, J. Xiao, C. Zhou, M. Yang, and C. Peng, "A cascade framework for masked face detection," in *2017 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM)*, 2017: IEEE, pp. 458-462.
- [14] 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," in *2020 11th international conference on computing, communication and networking technologies (ICCCNT)*, 2020: IEEE, pp. 1-5.
- [15] A. A. Abed, A. Al-Ibadi, and I. A. Abed, "Real-time multiple face mask and fever detection using YOLOv3 and TensorFlow lite platforms," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 922-929, 2023.
- [16] S. Singh, U. Ahuja, M. Kumar, K. Kumar, and M. Sachdeva, "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," *Multimedia Tools and Applications*, vol. 80, pp. 19753-19768, 2021.
- [17] S. Susanto, F. A. Putra, R. Analia, and I.K.L.I.M. Suciningtyas, "The face mask detection for preventing the spread of COVID-19 at Politeknik Negeri Batam," in *2020 3rd International Conference on Applied Engineering (ICAE)*, 2020: IEEE, pp. 1-5.
- [18] X. Jiang, T. Gao, Z. Zhu, and Y. Zhao, "Real-time face mask detection method based on YOLOv3," *Electronics*, vol. 10, no. 7, p. 837, 2021.
- [19] J. Yu, X. Zhang, T. Wu, H. Pan, and W. Zhang, "A Face Detection and Standardized Mask-Wearing Recognition Algorithm," *Sensors*, vol. 23, no. 10, p. 4612, 2023.
- [20] C. Dewi and R.-C. Chen, "Automatic medical face mask detection based on cross-stage partial network to Combat covid-19," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 106, 2022.
- [21] S. Gupta, S. Sreenivasu, K. Chouhan, A. Shrivastava, B. Sahu, and R. M. Potdar, "Novel face mask detection technique using machine learning to control COVID-19 pandemic," *Materials Today: Proceedings*, vol. 80, pp. 3714-3718, 2023.
- [22] Z. Song, K. Nguyen, T. Nguyen, C. Cho, and J. Gao, "Spartan face mask detection and facial recognition system," in *Healthcare*, 2022, vol. 10, no. 1: MDPI, p. 87.
- [23] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," *Measurement*, vol. 167, p. 108288, 2021.
- [24] M. A. Ai *et al.*, "Real-time facemask detection for preventing COVID-19 spread using transfer learning based deep neural network," *Electronics*, vol. 11, no. 14, p. 2250, 2022.
- [25] R. Jayaswal and M. Dixit, "AI-based face mask detection system: a straightforward proposition to fight with Covid-19 situation," *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 13241-13273, 2023.
- [26] T. A. Kumar, R. Rajmohan, M. Pavithra, S. A. Ajagbe, R. Hodhod, and T. Gaber, "Automatic face mask detection system in public transportation in smart cities using IoT and deep learning," *Electronics*, vol. 11, no. 6, p. 904, 2022.
- [27] A. Das, M. W. Ansari, and R. Basak, "Covid-19 face mask detection using TensorFlow, Keras and OpenCV," in *2020 IEEE 17th India Council International Conference (INDICON)*, 2020: IEEE, pp. 1-5.
- [28] M. M. Rahman, M. M. H. Manik, M. M. Islam, S. Mahmud, and J.-H. Kim, "An automated system to limit COVID-19 using facial mask detection in smart city network," in *2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2020: IEEE, pp. 1-5.
- [29] W.-J. L. Chong, S.-C. Chong, and T.-S. Ong, "Masked Face Recognition Using Histogram-Based Recurrent Neural Network," *Journal of Imaging*, vol. 9, no. 2, p. 38, 2023.
- [30] 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*, vol. 12, no. 18, p. 9171, 2022.
- [31] F. Crespo, A. Crespo, L. M. Sierra-Martínez, D. H. Peluffo-Ordóñez, and M. E. Morcho-Cayamcela, "A computer vision model to identify the incorrect use of face masks for COVID-19 awareness," *Applied Sciences*, vol. 12, no. 14, p. 6924, 2022.
- [32] H. Farman, T. Khan, Z. Khan, S. Habib, M. Islam, and A. Ammar, "Real-time face mask detection to ensure COVID-19 precautionary measures in the developing countries," *Applied Sciences*, vol. 12, no. 8, p. 3879, 2022.
- [33] E. Aydemir *et al.*, "Hybrid deep feature generation for appropriate face mask use detection," *International journal of environmental research and public health*, vol. 19, no. 4, p. 1939, 2022.
- [34] W. Hariri, "Efficient masked face recognition method during the covid-19 pandemic," *Signal, image and video processing*, vol. 16, no. 3, pp. 605-612, 2022.
- [35] C. Vimal and N. Shirivastava, "Face and Face-mask Detection System using VGG-16 Architecture based on Convolutional Neural Network," *International Journal of Computer Applications*, vol. 975, p. 8887.
- [36] S. V. Militante and N. V. Dionisio, "Deep learning implementation of facemask and physical distancing detection with alarm systems," in *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 2020: IEEE, pp. 1-5.
- [37] S. Hussain *et al.*, "IoT and deep learning based approach for rapid screening and face mask detection for infection spread

control of COVID-19," *Applied Sciences*, vol. 11, no. 8, p. 3495, 2021.

- [38] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, pp. 248-255, 2009.
- [39] L. Ali, F. Alnajjar, H. A. Jassmi, M. Gocho, W. Khan, and M. A. Serhani, "Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures," *Sensors*, vol. 21, no. 5, p. 1688, 2021.
- [40] X-zhangyang, "Real-World-Masked-Face-Dataset", *GitHub*, 2021,
<https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset> (accessed August 14, 2023).
- [41] chandrikadeb7, "Face-Mask-Detection", *GitHub*, 2022,
<https://github.com/chandrikadeb7/Face-Mask-Detection> (accessed August 14, 2023).
- [42] Larxel, "Face Mask Detection", *Kaggle*, 2020,
<https://www.kaggle.com/datasets/andrewmvd/face-mask-detection> (accessed August 14, 2023).