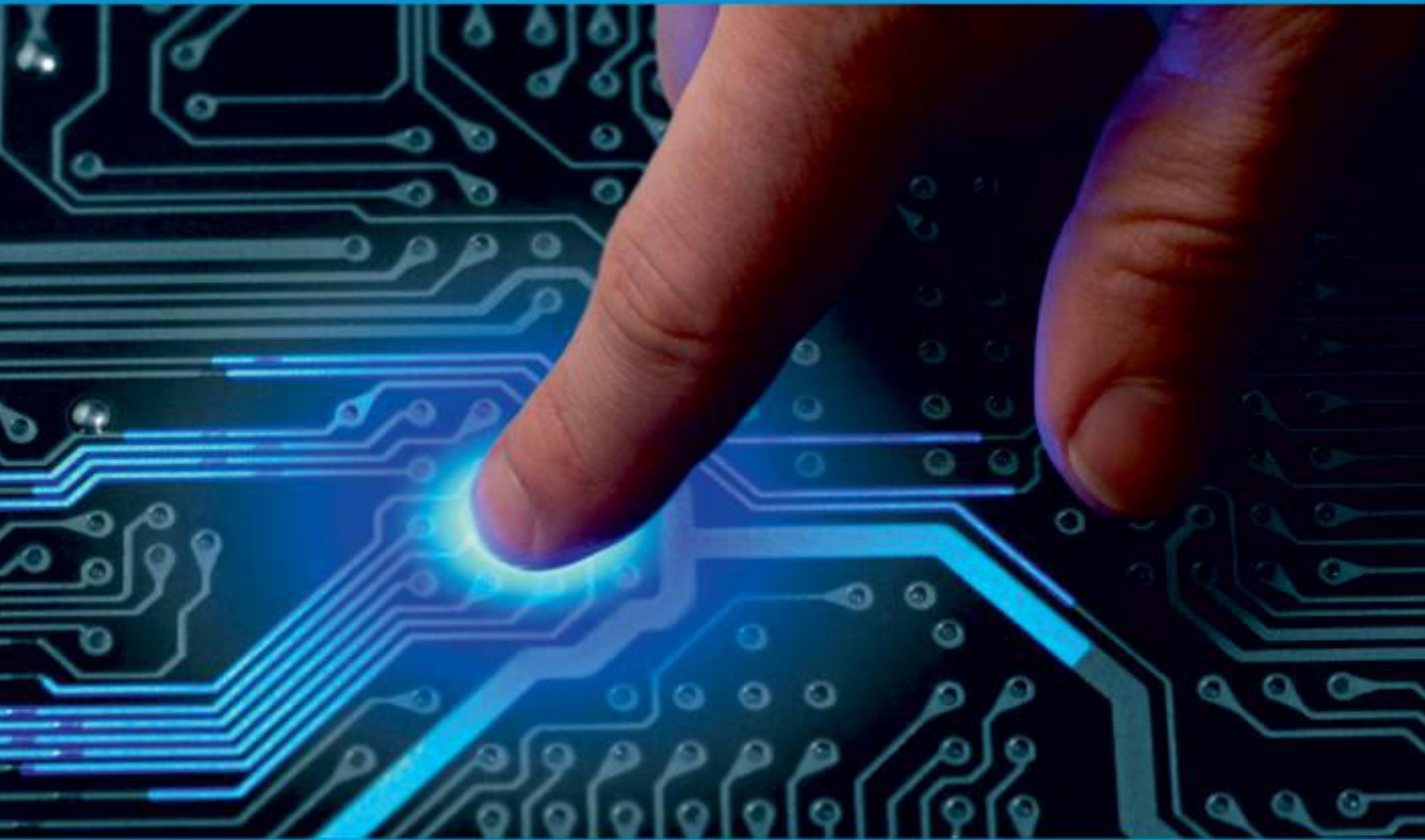
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 10, October 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 10, October 2024| DOI: 10.15680/IJIRCCE.2024.1210049

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Face Mask Detection Using Machine Learning

Mr. Ramakrishna Reddy Badveli, Aditya Patel

Assistant Professor, Department of Computer Science & Applications, The Oxford College of Science, Bangalore, India

MCA Student, Department of Computer Science & Applications, The Oxford College of Science, Bangalore, India

ABSTRACT: The primary mode of COVID-19 transmission is through respiratory droplets that are produced when an infected person talks, coughs, or sneezes. To avoid the fast spread of the virus, the WHO has instructed people to use face masks in crowded and public areas. This paper proposes the rapid real-time face mask detection system or RRFMDS, an automated computer aided system to detect a violation of a face mask in real-time video. In the proposed system, single-shot multi-box detector is utilized for face detection, while fne-tuned MobileNetV2 is used for face mask classification. The system is lightweight (low resource requirement) and can be merged with pre-installed CCTV cameras to detect face mask violation. The system is trained on a custom dataset which consists of 14,535 images, of which 5000 belong to incorrect masks, 4789 to with masks, and 4746 to without masks. The primary purpose of creating such a dataset was to develop a face mask detection system that can detect almost all types of face masks with different orientations. The system can detect all three classes (incorrect masks, with mask and without mask faces) with an average accuracy of 99.15% and 97.81%, respectively, on training and testing data. The system, on average, takes 0.14201142 s to process a single frame, including detecting the faces from the video, processing a frame and classification.

KEYWORDS:- Face mask detection,COVID-19,Machine learning, face detection.

1. INTRODUCTION

COVID-19, also known as SARS-CoV-2, is highly contagious. The COVID-19 pandemic has impacted nearly every country, wreaking havoc on available healthcare facilities and treatment systems. Direct contact with infected respiratory droplets spreads the virus (produced via sneezing and coughing). Anyone who comes into contact with virus-infected surfaces and subsequently touches their face may also become sick and experience symptoms such as shortness of breath, cough and fever. Although COVID-19 vaccinations have been developed and their widespread distribution began in early December 2020, they do not eradicate the virus; rather, they minimize complications and morbidity associated with COVID-19. The WHO emphasizes wearing a face mask in crowded and public places, as it prevents virus transmission via the nose or oral passages [1–3]. In COVID-19-afected countries, the government has instituted laws mandating the wearing a face mask. Face masks (e.g., cotton, surgical, N-95) offer 50–95% protection against the COVID-19 virus [4]. In the circumstances described, it is an excellent practice to always use a face mask to prevent exposure to COVID-19. In this context, determining whether or not a person in a public gathering or an organization is wearing a mask has been the subject of a significant amount of research. Conventional procedures for checking a face mask violation are not always feasible and are error prone. Conventional procedures include human force monitoring people for not wearing face masks manually. Therefore, there is a need for a system that can automatically do this task. Moreover, the human force could be saved and deployed to other essential tasks. This article is part of the topical collection "Computer Aided Methods to Combat COVID-19 Pandemic" guest edited by David Clifton, Matthew Brown, Yuan-Ting Zhang and Tapabrata Chakraborty. * Burhan ulhaque Sheikh sbuhaque@myamu.ac.in Aasim Zafar azafar.cs@amu.ac.in 1 Department of Computer Science, Aligarh Muslim University, Aligarh 202002, Uttar Pradesh, India 288 Page 2 of 19 SN Computer Science (2023) 4:288 SN Computer Science During the last 10 years, significant strides have been made in the field of computer-aided deep neural network (DNN) approaches, which have demonstrated promising results in classification tasks [5, 6], pattern recognition [7], and other areas. DNN models can detect minute variations between images, ultimately allowing them to precisely recognize an object in an image. In face mask detection systems, strategies based on deep learning (DL) and machine learning (ML) have been used to develop reliable and accurate systems that are both quick and effcient. A large number of researchers have developed a variety of DNN architectures for the purpose of detecting face mask violations. Most

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

of these researchers have chosen to base their methods on transfer learning and hybrid approaches (combination of deep learning and machine learning). DL models have shown high sensitivity and specificity when detecting face mask violations, as shown in Table 6. This paper proposes a system named RRFMDS (Rapid Real-Time Face Mask Detection System) for effective and accurate face mask detection. During the entirety of the development process, deep learning strategies, such as convolution neural networks (CNN), were applied throughout the process. The system consists of a face detector, intermediate block, and face mask detection modules. To construct a face identification module, we employed a single-shot multidetector, based on Resnet-10 [8, 9]. Furthermore, to identify face masks, we used the transfer learning of the state-of-theart model, MobileNetV2 [10].

II. RELATED WORK

Pattern learning and object recognition are the inherent tasks that a computer vision (CV) technique must deal with. Object recognition encompasses both image classification and object detection [12]. The task of recognizing the mask over the face in the pubic area can be achieved by deploying an efficient object recognition algorithm through surveillance devices. The object recognition pipeline consists of generating the region proposals followed by classification of each proposal into related class [13]. We review the recent development in region proposal techniques using single-stage and two-stage detectors, general technique for improving detection of region proposals and pre-trained models based on these techniques.

2.1 Single-stage detectors The single-stage detectors treat the detection of region proposals as a simple regression problem by taking the input image and learning the class probabilities and bounding box coordinates. OverFeat [8] and Deep Multi Box [9] were early examples. YOLO (You Only Look Once) popularized single-stage approach by demonstrating real-time predictions and achieving remarkable detection speed but suffered from low localization accuracy when compared with two-stage detectors; especially when small objects are taken into consideration [10]. Basically, the YOLO network divides an image into a grid of size GxG, and each grid generates N predictions for bounding boxes. Each bounding box is limited to have only one class during the prediction, which restricts the network from finding smaller objects. Further, YOLO network was improved to YOLOv2 that included batch normalization, high resolution classifier and anchor boxes. Furthermore, the development of YOLOv3 is built upon YOLOv2 with the addition of an improved backbone classifier, multi-sale prediction and a new network for feature extraction. Although, YOLOv3 is executed faster than Single-Shot Detector (SSD) but does not perform well in terms of classification accuracy [14,15]. Moreover, YOLOv3 requires a large amount of computational power for inference, making it not suitable for embedded or mobile devices. Next, SSD networks have superior performance than YOLO due to small convolutional filters, multiple feature maps and prediction in multiple scales. The key difference between the two architectures is that YOLO utilizes two fully connected layers, whereas the SSD network uses convolutional layers of varying sizes. Besides, the Retina Net [11] proposed by Lin is also a single-stage object detector that uses featured image pyramid and focal loss to detect the dense objects in the image across multiple layers and achieves remarkable accuracy as well as speed comparable to two-stage detectors.

2.2 Two-stage detectors In contrast to single-stage detectors, two-stage detectors follow a long line of reasoning in computer vision for the prediction and classification of region proposals. They first predict proposals in an image and then apply a classifier to these regions to classify potential detection. Various two-stage region proposal models have been proposed in past by researchers. Region-based convolutional neural network also abbreviated as R-CNN [16] described in 2014 by Ross Girshick et al. It may have been one of the first large-scale applications of CNN to the problem of object localization and recognition. The model was successfully demonstrated on benchmark datasets such as VOC-2012 and ILSVRC2013 and produced state of art results. Basically, R-CNN applies a selective search algorithm to extract a set of object proposals at an initial stage and applies SVM (Support Vector Machine) classifier for predicting objects and related classes at later stage. Spatial pyramid pooling SPPNet [17] (modifies R-CNN with an SPP layer) collects features from various region proposals and fed into a fully connected layer for classification. The capability of SPNN to compute feature maps of the entire image in a single-shot resulted in significant improvement in object detection speed by the magnitude of nearly 20 folds greater than R-CNN. Next, Fast RCNN is an extension over R-CNN and SPPNet [18,12]. It introduces a new layer named Region of Interest (RoI) pooling layer between shared convolutional layers to fine-tune the model. Moreover, it allows to simultaneously train a detector and regressor without altering the network configurations. Although Fast-R-CNN effectively integrates the benefits of R-CNN and SPPNet but still lacks in detection speed compared to single-stage detectors [19]. Further, Faster R-CNN is an amalgam

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

of fast R-CNN and Region Proposal Network (RPN). It enables nearly cost-free region proposals by gradually integrating individual blocks (e.g. proposal detection, feature extraction and bounding box regression) of the object detection system in a single step [20,21]. Although this integration leads to the accomplishment of break-through for the speed bottleneck of Fast R-CNN but there exists a computation redundancy at the subsequent detection stage. The Region-based Fully Convolutional Network (R-FCN) is the only model that allows complete backpropagation for training and inference [22,23]. Feature Pyramid Networks (FPN) can detect nonuniform objects, but least used by researchers due to high computation cost and more memory usage [24]. Furthermore, Mask R-CNN strengthens Faster R-CNN by including the prediction of segmented masks on each RoI [25]. Although two-stage yields high object detection accuracy, but it is limited by low inference speed in real-time for video surveillance [14 Blitz Net improved SSD by adding semantic segmentation layer to achieve high detection accuracy. The number of datasets with diverse features pertaining to human faces with and without mask are given in Table 1. An extensive study conducted on available face-related datasets reveal that there exist principally two kinds of datasets. These are: i) masked face and ii) face masked datasets. The masked face datasets are more concentrated on including the face images with a variant degree of facial expression and landmarks whereas face mask centric datasets

Table 1
Different Categories of Datasets.

Type of Datasets	Dataset	Scale	#Faces	#masked face images	Occlusion
Masked face detection Datasets	FDDB [31]	2845	5171	–	–
	MALF [32]	5250	11931	–	✓
	calebA [33]	200000	202599	–	–
	WIDERFACE [34]	32203	194000	–	✓
Face masked datasets	MAFA [35]	30811	37824	35806	✓
	RMFRD [36]	95000	9200	5000	✓
	SMFRD [36]	85000	5000	5000	✓
	MFDD [36]	500000	500000	24771	✓

include those images of faces that are mainly characterized by occlusions and their positional coordinates near the nose and mouth area. Table 1 summarizes these two kinds of prevalent datasets. The following shortcomings are identified after critically observing the available literature: 1. Although there exist several open-source models that are pre-trained on benchmark datasets, but a few models are currently capable of handling COVID related face masked datasets. 2. The available face masked datasets are scarce and need to strengthen with varying degrees of occlusions and semantics around different kinds of masks. 3. Although there exist two major types of state of art object detectors: single-stage detectors and two-stage detectors. But none of them truly meets the requirement of real-time video surveillance devices. These devices are limited by less computational power and memory. So, they require optimized object detection models that can perform surveillance in real-time with less memory consumption and without a notable reduction in accuracy. Single-stage detectors are good for real-time surveillance but limited by low accuracy, whereas two-stage detectors can easily produce accurate results for complex inputs but at the cost of computational time. All these factors necessitate to develop an integrated model for surveillance devices which can produce benefits in terms of computational time as well as accuracy. To solve these problems, a deep-learning model based on transfer learning which is trained on a highly tuned customized face mask dataset and compatible with video surveillance is being proposed and discussed in detail in the next section.

III. PROPOSED ARCHITECTURE

1. **Data Collection**:
   - Collect a diverse dataset of images containing individuals both wearing masks and not wearing masks.
   - Sources can include publicly available datasets, such as the "Face Mask Detection" dataset on Kaggle, or data collection via web scraping and manual curation.
2. **Data Preprocessing**:
   - **Image Augmentation**: To improve model robustness, apply transformations like rotation, flipping, scaling, and color adjustments.
   - **Normalization**: Scale pixel values to a range of 0-1 for better convergence during training.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- o **Labeling**: Ensure images are labeled correctly (e.g., "mask" vs. "no mask").

3. **Model Selection**:
- o Choose a suitable model architecture. Options include:
- ▪ **Convolutional Neural Networks (CNN)**: For image classification tasks.
- ▪ **Transfer Learning**: Utilize pre-trained models like MobileNet, VGG16, or ResNet50, which can be fine-tuned on the mask dataset to improve performance.

4. **Model Training**:
- o Split the dataset into training, validation, and testing sets (commonly 70/15/15).
- o Use a suitable loss function (e.g., binary cross-entropy for binary classification).
- o Optimize using algorithms like Adam or SGD.
- o Monitor performance metrics such as accuracy, precision, recall, and F1-score.

5. **User Interface**:
- o Develop a simple interface that displays real-time feedback (e.g., alerts for individuals without masks).
- o Integrate with existing surveillance or monitoring systems if applicable.

6. **Monitoring and Maintenance**:
- o Continuously monitor the model's performance and update the dataset with new images to account for changing scenarios (e.g., new mask types, angles).
- o Implement retraining pipelines to periodically update the model with new data.

Tools and Technologies
- **Frameworks**: TensorFlow, Keras, PyTorch for model training.
- **Deployment**: Flask or FastAPI for creating REST APIs, OpenCV for image processing.
- **Cloud Services**: AWS, Google Cloud, or Azure for scalable deployment options.

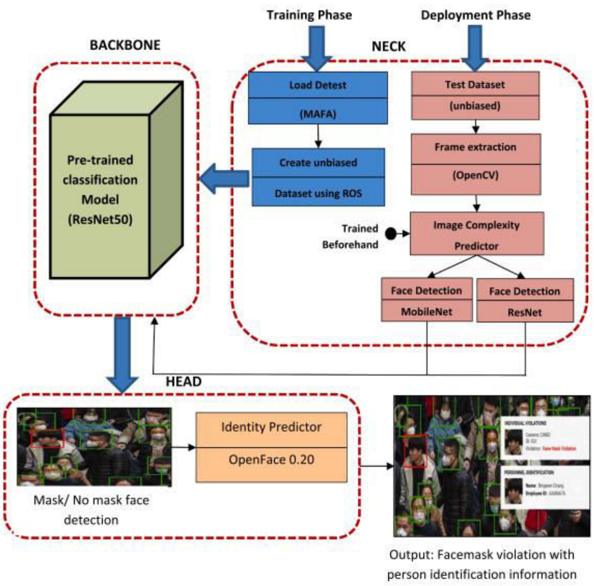


Fig. 2. Proposed Architecture.

Head Fig. 2. Proposed Architecture. S. Sethi et al. Journal of Biomedical Informatics 120 (2021) 103848 5 stands for identity detector or predictor that can achieve the desired objective of deep-learning neural network. In the proposed architecture, the trained facemask classifier obtained after transfer learning is applied to detect mask and no mask faces. The ultimate objective of enforcement of wearing of face mask in public area will only be achieved after retrieving the personal identification of faces, violating the mask norms. The action can further, be taken as per government/ office policy. Since there may exist differences in face size and orientation in cropped ROI, affine transformation is applied to identify facial using OpenFace 0.20. The detailed description of each task in the proposed architecture is given in the following subsections.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A. Creation of unbiased facemask dataset A facemask-centric dataset, MAFA with a total of 25,876 images, categorized over two classes namely masked and non-masked was initially considered. The number of masked images in MAFA are 23,858 whereas non-masked images are only 2018. It is observed that MAFA is put up with an extrinsic class imbalanced problem that may introduce a bias towards the majority class. So, an ablation study is conducted to analyze the performance of the image classifier once with the original MAFA set (biased) and then with the proposed dataset (unbiased). 3.1.1. Supervised pre-training We discriminatively pre-trained the CNN on the original biased MAFA dataset. The pre-training was performed using the open-source Caffe python library [7]. In short, our CNN model nearly matches the performance of Madhura et al. [11], obtaining a top-1 error rate 1.8% higher on MAFA validation set. This discrepancy may cause due to simplified training approach

Fig. 4. Variety of Occlusions Present in Dataset.

B. Image complexity predictor for face detection To address problem 3 identified in Section 2, various face images are analyzed in terms of processing complexity. It is observed that dataset, we consider primarily, contains two major classes that is, mask and nonmask class but the mask class further, contains an inherent variety of

Table 2
Comparison between MobileNet-SSD, ResNet50 and Their Various Combinations based on Random vs. Hard/Soft Complexity of Test Data.

Comparison Parameters	MobileNet-SSD to ResNet50 (Left to Right)				
	100–0%	75–25%	50–50%	25–75%	0–100%
Random split (mAP)	0.8868	0.9095	0.9331	0.9650	0.9899
Soft/hard split (mAP)	0.8868	0.9224	0.9631	0.9892	0.9899
Image complexity	–	0.05	0.05	0.05	–
prediction time (ms)	0.05	1.92	3.08	5.07	6.02
Mask detection time (ms)					
Total Computation Time (ms)	0.05	1.97	3.13	5.12	6.02

occlusions other than surgical/cloth facemask, for example, occlusion of ROI by other objects like a person, hand, hair or some food item as shown in Fig. 4. These occlusions are found to impact the performance of face and mask detection. Thus, obtaining an optimal trade-off between accuracy and computational time for face detection is not a trivial task. So, an image complexity predictor is being proposed here. Its purpose is to split data into soft versus hard images at the initial level followed by a mask and non-mask classification at a later level through a facemask classifier. The important question that we need to answer is how to determine whether an image is soft or hard. The answer to this question is given by the "Semi-supervised object classification strategy" proposed by Lonescu et al. . The image is also

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

flipped horizontally and the same pyramid is applied over it. The 4096 features extracted from each bin are combined to obtain a single feature vector followed by normalization using L2-norm. The obtained normalized feature vector is further, used to regress the image complexity score. Thus, the model automatically predicts image complexity for each image in T. Having identified the hardness of the test images using an image complexity predictor, a soft image is proposed to process through a fast single-stage detector while the hard image is accurately processed by two-stage detector. We employ MobileNet-SSD model for predicting the class of soft images and faster RCNN based on ResNet50 for predicting hard images. The algorithm for image complexity predictor is outlined below: Algorithm. Image_Complexity_Predictor () 1. Input: 2. Image ← input image 3. Dfast ← single-stage detector Table 2 Comparison between MobileNet-SSD, ResNet50 and Their Various Combinations based on Random vs. Hard/Soft Complexity of Test Data. Comparison Parameters MobileNet-SSD to ResNet50 (Left to Right) 100–0% 75–25% 50–50% 25–75% 0–100% Random split (mAP) Soft/hard split (mAP) 0.8868 0.8868 0.9095 0.9224 0.9331 0.9631 0.9650 0.9892 0.9899 0.9899 Image complexity prediction time (ms) Mask detection time (ms) – 0.05 0.05 1.92 0.05 3.08 0.05 5.07 – 6.02 Total Computation Time (ms) 0.05 1.97 3.13 5.12 6.02 S. Sethi et al. Journal of Biomedical Informatics 120 (2021) 103848 7 4. Dslow ← two-stage detector 5. C ← Image complexity 6. Computation: 7. If (C = Soft) R ← Dslow(Image) 8. else R ← Dfast(Image) 9. Output: 10. R ← set of region proposals Table 2 summarizes mAP score and Computation time for various combinations of MobileNet and ResNet50 over test dataset. The various combinations are made by splitting the test dataset into different proportion of images processed by each detector starting from pure MobileNet (100–0%) to three intermediate splits (75–25%, 50–50%, 25–75%) to pure ResNet50 (0–100%). Here, the test data is partitioned based on random split or soft versus hard spilt given by Image Complexity Predictor. To reduce the bias, the average mAp over 5 runs is recorded for random spilt. The elapsed time is measured onInter I5, 2.5 GHZ CPU with 8 GB RAM.

AlexNet

	Mask	No Mask
Mask	TP: 4351	FP:103
No Mask	FN:227	TN:4518

MobileNet

	Mask	No Mask
Mask	TP: 4669	FP:48
No Mask	FN:104	TN:4378

ResNet50

	Mask	No Mask
Mask	TP: 4657	FP:51
No Mask	FN:83	TN:4403

Fig. 6. Confusion Matrix Obtained for Various Pre-trained Models.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Google colaboratory was used to train a model. It is a platform that enables us to write Python scripts for machine learning, deep learning, and data analysis and then execute those programs while having unrestricted access to the cloud's resources. Whenever a new Colab session is started, the computer will automatically assign a random GPU, CPU, and amount of disc storage. There is a wide variety of graphic processing units (GPU) that are readily accessible, including T4s, P4s, Nvidia K80s, and P100s.

In this paper, Keras, Tensorflow, Sklearn, Matplotlib, and Numpy APIs were used. Keras and Tensorflow are advanced neural network packages used to design a classifier.

- *Keras and Tensorflow:* Used to design a classifier, MobileNetV2.
- *SKlearn:* For data analysis, such as computing the metrics of the model.
- *OpenCV::* To perform image processing operations and also used to load the face detector model.
- *Imutils:* For video streaming.
- *Matplotlib:* To plot the learning curves of accuracy and loss.
- *Time:* To compute the average processing of the frame.

V. METHODOLOGY

The methodology for developing a face mask detection system involves several key steps: defining clear objectives to enhance public safety, collecting a diverse dataset of images depicting individuals with and without masks, and preprocessing the data through cleaning, labeling, and augmentation. The next phase includes selecting appropriate model architectures, such as CNNs or pre-trained models using transfer learning, followed by training the model with

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

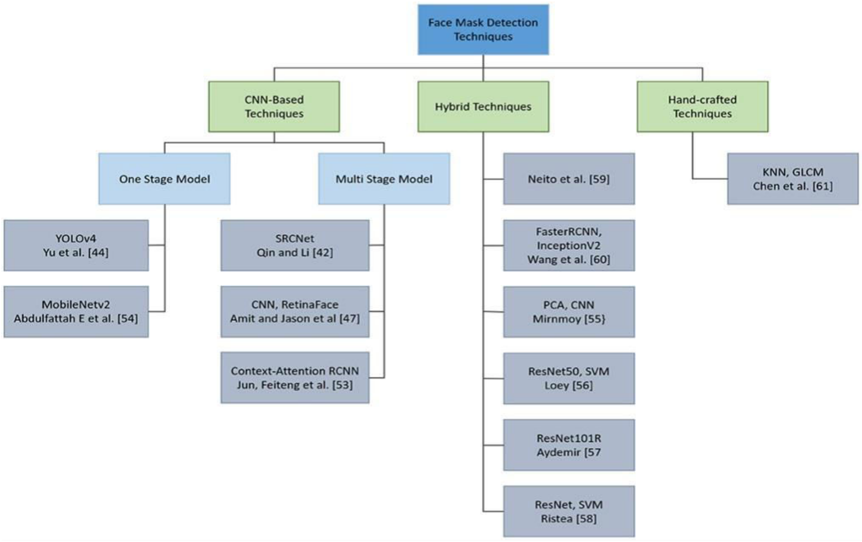
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

careful monitoring of performance metrics to avoid overfitting. After evaluation using accuracy, precision, recall, and F1 score, the model is deployed in real-time applications, integrated with a user-friendly interface, and continuously monitored for improvements and updates based on real-world performance.

1. Convolution layer: This is the main building block of a CNN. It comprises a set of kernels (or filters), parameters of which are to be learned during training. The size of the filters is usually smaller than the actual image and the filters scan the image and create an activation map.
2. Rectified Linear Unit (ReLU): This is an activation layer that helps prevent exponential growth in the computation required to operate the neural network and it sets the negative input value to zero.
3. Maxpooling layer: This is a pooling operation that selects the maximum element from the region of the feature map covered by the filter, thereby extracting the most prominent features from the previous feature map to a new feature map.
4. Batch normalization: This layer re-centers and re-scales input layers so as to normalize them, thereby making training faster and more stable.
5. Fully connected layer: This layer multiplies the input by a weight matric and then adds the bias vector, which gives the probability values for classification into various groups.
6. Loss function: This is a way of evaluating how well a CNN algorithm models the dataset used for training. It does this by applying a soft-max layer to the input data sample.

In this section, existing approaches to face mask detection are reviewed, analyzed, and their performance and limitations are discussed. Due to the superior performance of CNN-based algorithms in feature extraction, most face mask detection models utilize CNN-based techniques. CNN-based algorithms outperform other techniques in feature extraction, which is why they are the most used method for face mask detection models. Some studies considered an approach that aggregates classical ML techniques and CNN deep-learning ones, i.e., a hybrid approach. A few studies used solely classical ML, hand-crafted techniques.



VI. CONCLUSION AND FUTURE SCOPE

In conclusion, the development of a face mask detection system using machine learning is a structured process that emphasizes data collection, preprocessing, model selection, training, evaluation, and deployment. By leveraging diverse datasets and employing robust algorithms, such as convolutional neural networks and transfer learning, the system can achieve high accuracy and reliability in real-world scenarios. Continuous monitoring and updates are vital to adapt to changing conditions and improve performance over time. Ultimately, this methodology not only contributes to public health initiatives but also showcases the potential of machine learning in addressing urgent societal challenges.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The future scope for face mask detection systems is promising and multifaceted. As technology advances, there is potential for integrating more sophisticated algorithms, such as deep learning techniques that leverage larger and more diverse datasets for improved accuracy. Enhanced capabilities could include real-time detection in varied environments, even in crowded settings, and the ability to identify different types of face coverings, such as various mask styles or facial coverings.

Moreover, the incorporation of additional functionalities, such as temperature checks or social distancing monitoring, can create comprehensive health monitoring systems that contribute to public safety. Integration with smart surveillance systems, mobile applications, and IoT devices can facilitate seamless operation in public spaces, workplaces, and transportation hubs.

Furthermore, ongoing research into ethical AI and bias mitigation will ensure that these systems operate fairly across diverse populations. With the growing importance of health security, the development of adaptive models that can evolve based on emerging health guidelines and trends will be essential.

Overall, the future of face mask detection systems is not just limited to compliance but extends into broader public health strategies, making them invaluable tools in managing health crises and enhancing community safety.

REFERENCES

[1]. Ellis R. WHO changes stance, says public should wear masks. World Health Organization; 2020.
[2]. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. Lancet Respir Med. 2020;8(5):434–6.
[3]. World Health Organization. Advice on the use of masks in the context of COVID-19: interim guidance, 6 April 2020 (No. WHO/2019-nCov/IPC_Masks/2020.3). World Health Organization; 2020.
[4]. "How mask antiviral coatings may limit covid- 19 transmission". 2020 [Online]. Available: https://www.optometrytimes.com/view/ .
[5]. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems. ACM; 2012. p. 1097–105.
[6]. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations. 2015.
[7]. Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell. 2000;22(1):4–37.
[8]. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. Ssd: single shot multibox detector. In: European conference on computer vision. Cham: Springer; 2016. p. 21–37.
[9]. Anisimov D, Khanova T. Towards lightweight convolutional neural networks for object detection. In: 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE; 2017. p. 1–8.
10. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Adam, H. Mobilenets: effcient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. 2017. 11. https://github.com/chandrikadeb7/Face-Mask-Detection [12] M. Inamdar, N. Mehendale, Real-Time Face Mask Identification Using Facemask net Deep Learning Network, SSRN Electron. J. (2020), https://doi.org/10.2139/ssrn.3663305.
[13] S. Qiao, C. Liu, W. Shen, A. Yuille, Few-Shot Image Recognition by Predicting Parameters from Activations, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018, https://doi.org/10.1109/CVPR.2018.00755.
[14] A. Kumar, Z.J. Zhang, H. Lyu, Object detection in real time based on improved single shot multi-box detector algorithm, J. Wireless Com. Netw. 2020 (2020) 204,https://doi.org/10.1186/s13638-020-01826-x.
[15] A. ´ Morera, A. ´ Sanchez, ´ A.B. Moreno, A.D. ´ Sappa, J.F. V´elez, SSD vs. YOLO for detection of outdoor urban advertising panels under multiple variabilities, Sensors(Switzerland) (2020), https://doi.org/10.3390/s20164587.
[16] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based Convolutional Networks for Accurate Object Detection and Segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) (2015) 142–158, https://doi.org/10.1109/TPAMI.2015.2437384.
[17] K. He, X. Zhang, S. Ren, J. Sun, Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2015),https://doi.org/10.1109/TPAMI.2015.2389824.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[18] R. Girshick, Fast R-CNN, in: Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter,2015, pp. 1440–1448,doi: 10.1109/ICCV.2015.169.
[19] N.D. Nguyen, T. Do, T.D. Ngo, D.D. Le, An Evaluation of Deep Learning Methodsfor Small Object Detection, J. Electr. Comput. Eng. 2020 (2020), https://doi.org/10.1155/2020/3189691.
[20] Z. Cai, Q. Fan, R.S. Feris, N. Vasconcelos, A unified multi-scale deep convolutional neural network for fast object detection, Lect. Notes Comput. Sci. (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2016), https://doi.org/10.1007/978-3-319-46493-0_22.
[21] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, A.C. Berg, DSSD : Deconvolutional Single ShotDetector, 2017, arXiv preprint arXiv:1701.06659 (2017).
[22] A. Shrivastava, R. Sukthankar, J. Malik, A. Gupta, Beyond Skip Connections: TopDown Modulation for Object Detection, 2016, arXiv preprint arXiv:1612.06851(2016).
[23] N. Dvornik, K. Shmelkov, J. Mairal, C. Schmid, BlitzNet: A Real-Time DeepNetwork for Scene Understanding, in: Proceedings of the IEEE InternationalConference on Computer Vision, 2017, doi: 10.1109/ICCV.2017.447.
[24] Z. Liang, J. Shao, D. Zhang, L. Gao, Small object detection using deep feature pyramid networks, in: Lecture Notes in Computer Science (including subseriesLecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018,vol. 11166 LNCS, pp. 554–564, doi: 10.1007/978-3-030-00764-5_51.
[25] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, in: Proc. IEEE Int. Conf.Comput. Vis., vol. 20