Project Report

On

DETECTION OF MASK IN LIVETREAMS FOR COVID19 PRECAUTION

Submitted to

KIIT Deemed to be University

In partial fulfillment of the requirement of the award

BACHELOR'S DEGREE IN COMPUTER SCIENCE AND ENGINEERING

By
Srishti Verma (1705829)
Shailesh Mishra (1705930)

Under the guidance of

Prof. Anil Kumar Swain

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA-751024

DEC 2020

ABSTRACT

Looking at present scenario where the whole world is suffering due to the novel Corona virus (COVID-19).After the declaration of COVID-19 as a pandemic by World Health Organization (WHO) many countries went into lockdown to stop the spread,which took a great toll on the world economy ,even bringing the small countries to the state of bankruptcy.Hence forcing them to open the months long lockdown even though their is no vaccine present or developed till now.It produces sear concern to the authorities to manage and control the spread. In regards to this, we propose our face mask detection technology that can be installed into different public and private sectors to check whether the peoples are wearing the mask or not which has been described as the minimum and most important measured in COVID-19 prevention by WHO.

Face detection and recognition has been a field of research for quite a long time now. There have been great improvements when it comes to the algorithms for detecting faces. But owing to the current situation, where almost everybody’ s face is covered with masks, the traditional techniques become efficient. Traditional techniques come handy and do the job when it comes to mask detection using feature engineering but it is now possible to train neural network to outperform these traditional methods without any feature engineering Here, comes the need to modify the existing algorithms and train our face detecting models to accurately detect masked faces. It will also calculate the distance between the two person in the image and decide weather they are having a safe distance between them or not that is 1 meter.

KEYWORDS: Covid-19, Masked face detection, Sklearn, Matplotlib, OpenCV, CNN, RetinaFace, Neural Networks , MobileNetV2

INTRODUCTION

The study of face detection has already marked its present in the field of face AI. There are already several techniques to detect faces in images or live streams. But these existing algorithms become inefficient when it comes to masked faces or any other type of occlusion. Since, the present scenario demands everyone to wear masks, the need for a separate algorithm for detecting masked faces becomes inevitable. This project is about a proposed method to efficiently detect whether an image or a person in a live stream has a wore a mask or not.

The World Health Organisation has made masks compulsory for all. The purpose of this project is to detect faces in images or live streams and provide us an output declaring whether the image or the person is wearing a mask or not.

LITERATURE SURVEYS

The model proposed by [1] is based on LLE-CNN. It consists of 3 modules: Proposal module, Embedding module, Verification module. This model reaches an average precision of 76.4% on the training set of the MAFA dataset. It outperforms six benchmarked models in masked face detection (SURF, NPD, ZR, HH, HPM, MT) , provides feature robustness and developed the MAFA dataset that can be further used for more research. However, it was ineffective in detecting masked side faces or images with strong occlusion.

Another Model proposed by [2] is based on CNN. It consists of 4 modules: Proposal module, Classification model, Regression module, Cluster module. This model produces an average precision of 76.8% on the testing dataset. The technique used is quite reliable in detecting masked faces. This framework outperforms 6 famous detection frameworks namely SURF, NPD, ZR, HH, HPM and MTCNN. However, due to having a huge number of non-linear correlations, it is computationally expensive and inefficient with a small dataset.

The model proposed by [3] is based on Faster RCNN with VGG16 backbone. It consists of 4 modules: Mask generator module, Classification module, bounding box regression, Occlusion Segmentation module.The model achieved an average precision of 77.3% on the training set of MAFA dataset. It was detects all kinds of occlusions and is effective on images containing multiple images. Its only demerit was that it couldn’t detect tiny faces.

Although research on face detection has been developing for decades, the framework and algorithm that are for masked face detection are rare.Lin et al. [4] Modified LeNet and designed a detection framework based on a 1000-training data-set sample which is yet less for what we generally have to capture every day .The data-set where single. MTCNN is a face detection framework trained on the Wider-face.But since some of the faces are labels for occlusion in the data-set the MTCNN has weak detection capability.

Zhu [5] and Ramanan also purposed a approach to explicitly model the structure or deformation of faces with DPM.They proposed a tree structured model for faces detection ,which can simultaneously estimate face poses and localize facial landmarks.Mathias et al. trained a DPM-based face detector which achieved an average precision of around 98% on 26000 faces.Chen et al. [6] Then purposed a face detector after observing that aligned face shapes can provide a better features for classification by jointly learning detection and alignment in a unified framework.The result of DPM-based face detectors achieved impressive accuracy but suffered from high computational cost.

Object detection with only image intensities i.e RGB pixel values made the task of feature detection computationally expensive.On 1998 a paper published by Papageorgious et al.[7] discussed the alternate feature set based on Haar wavelets which was later adapted by Paul Viola and Michael Jones[8] to develop Haar-like Features.Haar-like features has nice calculation speed as it used integral images,calculating the haar-like feature of any size can be done in constant time.

PROPOSED MODEL:

The proposed model includes two integral component- The face detector and the mask detector. The input provided to the face detector can either be an image or a live video steam and the face detector extracts the ROI or the region of interest which in this case are the faces in the input which may be of varying sizes and even overlapping. Then they are send to the second component, which is the mask detector.

There are several face detection algorithms in use right now which can be used for the algorithm. We selected the algorithm on the basis of accuracy and stability of the detection.

1. MTCNN[9] - It uses a deep cascade multitask architecture with three stages of CNN for detecting and localizing faces and facial key points.

2. RetinaFace[10] - It is a single-stage design with pixel-wise face localization on various scales of faces by taking advantages of joint extra-supervised and self-supervised multi-task learning

3. Selective Refinement Network (SRN)[11] - introduces the two-step classification and regression operations selectively into an anchor-based face detector to reduce false positives and improve location accuracy simultaneously.

We decided to go with RetinaFace detector as it was capable of detecting side faces which MTCNN failed on and masked faces which SRN failed on. Comparative study below.

Model	Average Inference Time per image (in seconds), for varying image resolutions		
	480p	720p	1080p
SRN	3.622	8.206	14.654
MTCNN	0.575	0.684	1.260
RetinaFace	0.095	0.113	0.196

The next stage was deciding on the mask detector. There are several algorithms used for mask detection as discussed in the literature survey. All these algorithms use feature engineering or convoluted neural networks to detect masks on face. CNN works more efficiently than Feature engineering.

IMPLEMENTATION

For training the model, we needed a dataset containing a large number of images for both classes: wearing a mask and not wearing a mask. The RMFD[13] includes 5,000 pictures of 525 people wearing masks, and 90,000 images of the same 525 subjects without masks. Sice RMFD mostly contains an asian bias, we also added several 5000 images of random origin. To our knowledge, this is currently the world's largest real-world masked face dataset.

The proposed model involves four steps:

- Data collection and Preprocessing
- Model development and Training
- Model Testing
- Model deployment

Data Collection and Pre-processing

The data is of around 2000 masked and unmasked images from the dataset and created a new dataset with two directories of mask and unmasked images respectively.

The steps of pre-processing are explained as follows:
1.The directory of dataset is loaded to **imutils.paths** to get the list of image paths.
2.The image data is initialized and the list variable is labelled.

3. The image path/ directory is looped over:
   - label from file name is extracted (masked/unmasked)
   - the image is loaded and pre-processed
   - the processed image is inserted in the img_data and img_label list.

4. The img_data and img_label is converted to float numpy array and then the 80-20 training and testing set is generated using sklearn.train_test_split.

5. ImageDataGenerator has been used to augment the data for the training set. Data augmentation is a method that allows practitioners to significantly increase the diversity of the data available for training models, without adding new data to it.

Model Development and Training

MobileNetV2 is a convolutional neural network that is 53 layers deep. It is applied to embedded devices with limited computational capacity to improve efficiency. It is part of TensorFlow-Slim Image Classification Library.

MobileNetV2 is an upgraded version of MobileNetV1, which uses depth wise separable convolution as efficient building blocks, and along with that adds two new features to the architecture:

- linear bottlenecks between the layers
- a shortcut connection between the bottlenecks

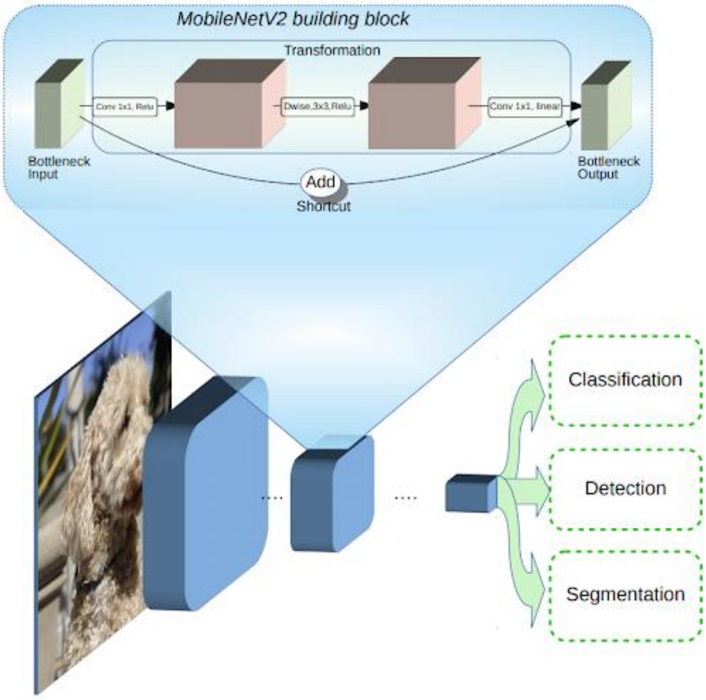
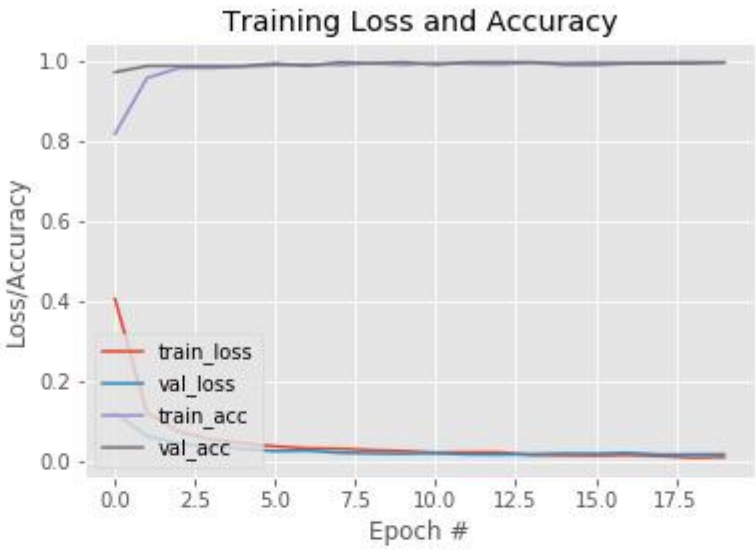


Fig 5.1 MobileNetV2 architecture

The training the model is done using the MobileNetV2 network to create the base model. The head of the model is constructed and placed on top of the base model. This will become the actual model we will train. The model is compiled and finally the head is trained over 20 epochs.



Plot for training loss and accuracy

Model Testing

The testing phase includes:

- Predicting the network and then printing the classification report

	precision	recall	f1-score	support
with_mask	0.99	1.00	1.00	181
without_mask	1.00	0.99	1.00	179
micro avg	1.00	1.00	1.00	360
macro avg	1.00	1.00	1.00	360
weighted avg	1.00	1.00	1.00	360

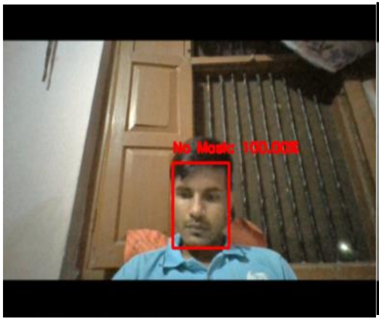
Fig 5.6. classification report

After the testing phase we visualise the prediction accuracy and loss on the trained set and save the mode created.
The steps are as follows:

- The video stream is first initialized and the camera sensor is aloe\wed to warm up.
- The frames are looped over:
  - The frames are resized
  - Detect faces in the frame and determine whether it is having mask or not

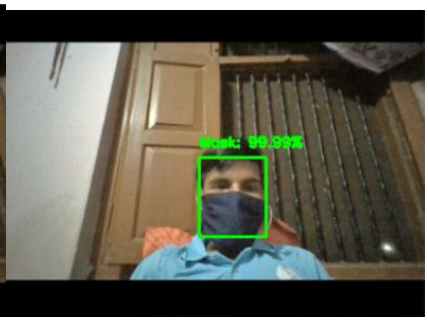
- Displaying the output:
  - The detected face locations are looped over
  - The bounding boxes and predictions are unpacked
  - The class label and the colour of the bounding box is determined and the probability is included
  - The bounding box along with the label is displayed on the output frame

The output produced by our proposed model.



no mask output

Light colour mask output



Dark colour mask output

The proposed model gives a precision of 100% on a non-masked face, 99.93% accuracy on a face covered with a light-coloured mask and 99.99% accuracy on a face covered with a dark coloured mask for one test case.
The average precision thus stands at 99.97% for the particular test case.

CONCLUSION AND FUTURE SCOPE

Face detection has been researched and has more regions that need to covered out of which face occlusion is one major challenge.

Conclusion

In this paper, we have created a real-time face mask detection model which gives an average precision of 99.97%. The paper consists of research, planning and implementation of the model. Even though the model produces an incredible accuracy, it has a drawback. The model is dataset dependant. If the dataset does not include varieties, the model fails in such situations.

Title	Average Precision
LLE-CNN model (Literature Survey-1)	76.4%
MTCNN+VGG model (Literature Survey-2)	76.8%
Faster RCNN+VGG model (Literature Survey-3)	77.3%
RetinaFace + CNN Model (Proposed model)	99.97%

Fig 7.1.Comparison table

Future Scope

This masked face detection model can used in:

- **Masked face recognition** – a model that can detect masked faces and recognise the person which can be used in attendance systems in schools, colleges and workplaces.
- **No-Mask-No-Entry software**- This software can be installed at the entry points of any commercial places (stores, shopping complexes etc) or public transport systems (buses, metros etc.) which will detect whether a person is wearing a mask or not and accordingly produce an alarm if mask is not found thus preventing that person from entering.

REFERENCES

[1]. Detecting Masked Faces in the Wild with LLE-CNNs; Shiming Ge , Jia Li , Qiting Ye, Zhao Luo; 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[2]. Masked Face Detection Via a Novel Framework; Qiting Ye; 2nd International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018)

[3]. Masquer Hunter: Adversarial Occlusion-aware Face Detection; Yujia Chen, Lingxiao Song, Ran He; arXiv:1709.05188v1 [cs.CV] 15 Sep 2017

[4]. C.Zhu, Y. Zheng, K. Luu, and M. Savvides. CMS-RCNN:contextual multi-scale region-based CNN for unconstrainedface detection.CoRR, abs/1606.05413, 2016

[5]. E.Y. Li, B. Sun, T. Wu, and Y. Wang. Face detection with end-to-end integration of a convnet and a 3d model. InECCV,pages 122–138, 2016.

[6]. D.B. Klare, B. Klein, E. Taborsky, andet al.. Pushing the fron-tiers of unconstrained face detection and recognition iarpajanus benchmark a. InIEEE CVPR, pages 1931–1939, 2015.

[7]. F.P. Viola and M. Jones. Rapid object detection using a boostedcascade of simple features. InIEEE CVPR, pages 511–518,2001

[8]. Viola and Jones,"rapid object detection using a boosted cascade of simple features",computer cool vision and pattern recognition,2001

[9]. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks: Zhang, Kaipeng et al. 2016

[10]. RetinaFace: Single-stage Dense Face Localisation in the Wild: Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, Stefanos Zafeiriou 2019: arXiv:1905.00641 [cs.CV]

[11]. Improved Selective Refined Network for Face Detection; Shifeng Zhang, Rui Zhu, Xiaobo wang, Hailin Shi, Tianyu Fu, Shuo Wang, Tao Mei, Stan Z. Li; arXiv:1901.06651v3 [cs.CV] 30 Jan 2019

[12]. Papageorgiou,Oren and Poggio,"A geeral framework for object dection",International Conference on Computer Vision,1998.

[13]. Masked Face Recognition Dataset and Application; Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, Heling Chen, Yu Miao, Zhibing Huang, and Jinbi Liang; arXiv:2003.09093v2 [cs.CV] 23 Mar 2020

[14]. Messom,C.H. and Barczak,A.L.C,"Fast and Efficeient Rotated Haar-like Features using rotated Integral Images",Australian Conference on Robotics and automation ,2006

INDIVIDUAL CONTRIBUTION REPORT

MASKED FACE DETECTION

SRISHTI VERMA
1705829

Abstract: This paper shows how we have researched about face detection and masked face detection and implemented a face detection algorithm that is able to detect faces and tell us whether the person is wearing a mask or not with an accuracy of 98.8%.
The work has been done keeping in mind the present scenario of the pandemic and is a small step towards helping the society in preventing the spread of the corona virus.

Individual contribution and findings: I studied about the various methods that can be used for masked face detection and came up with the proposal of using RetinaFace and CNN. I looked after the webcam integration module of the project.

Individual contribution to project report preparations: I edited and prepared the entire report . I also took care of the langusage of the report.

Individual contribution for presentations and demonstration: I prepared the Introduction slides of the presentation and discussed about face detection, methods of face detection and masked face detection.

Prof. Anil Kumar Swain

Srishti Verma (1705829)

INDIVIDUAL CONTRIBUTION REPORT

MASKED FACE DETECTION

SHAILESH MISHRA
1705930

Abstract: This paper shows how we have researched about face detection and masked face detection and implemented a face detection algorithm that is able to detect faces and tell us whether the person is wearing a mask or not with an accuracy of 98.8%.
The work has been done keeping in mind the present scenario of the pandemic and is a small step towards helping the society in preventing the spread of the corona virus.

Individual contribution and findings: I was completely into the implementation part of the project. I used my knowledge about CNN and opencv to pre-process the images and then train the model. The main focus was to find the Bounding Box coordinates of the detected face and to successfully extract it. Thus, my contribution to the project forms the basis of the Masked Face Detection.

Individual contribution to report preparations: I researched and added the literature survey part of the report along with the abstract.

Individual contribution for presentation and demonstration: I prepared the slides on Implementation and discussed thoroughly how we have used each library and each classifier to prepare the proposed model.

Prof Anil Kumar Swain	Shailesh Mishra (1705930)