INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS

ISSN(print): 2643-9840, ISSN(online): 2643-9875

Volume 05 Issue 11 November 2022

DOI: 10.47191/ijmra/v5-i11-31, Impact Factor: 6.261

Page No. 3214-3218

Facial Mask Detection Using Transfer Learning with Resnet18 and MobileNetV2

Freshta Akhtary[1], Engila Akhtary[2], Qurban Ali Karimi[3]

[1]Lecturer at Parwan University, Parwan, Afghanistan

[2]Lecturer at Kabul University of Medical Sciences Abu Ali Sina

[3]Epidemiologic Research Specialist

ABSTRACT: Identification of individuals using face masks has become a widespread concern as a result of the coronavirus pandemic. One critical method of preventing pandemic spread is for persons to constantly wear masks in public places. To achieve the maximum possible accuracy, new approaches for face mask detection are being developed that utilize convolutional neural networks. These novel techniques attempt to construct a binary face classifier capable of categorizing pixels in an image as either faces or non faces. In this study, we created a model using Resnet18 and MobileNetv2 in order to obtain trustworthy results for picture categorization. Additionally, we constructed our model for detecting live webcam feeds utilizing Transfer Learning on these deep convolutional neural networks, as well as ImageNet weights and OpenCV.

KEYWORDS: Face mask detection, Image classification, MobileNetv2, ResNet18, Semantic segmentation, Transfer learning.

1. INTRODUCTION

These days, face recognition is commonly used as the human face's primary biometric feature. Face detection, identification and differentiation is assurance of several approaches to one of the main areas of study for focused areas like legal applications where offenders or publications can be identified using their live features, actions and characteristics [1] [2] [3]. Face detection is also used for security, like in face locks etc. which pick out specific, distinctive details about a person's face, like distance between eyes etc. and convert it into mathematical representation. Another major use is in the police department where face detection can be used to find missing persons. Thus, face detection and recognition is described as a very important issue in image processing and computer vision [4]. Following the discoveries in face detection, the latest discoveries are for face mask detection where a person's face is detected and classified whether they are wearing a mask or not. Owing to the recent COVID19 times, the need for a face mask detector is of utmost importance, especially in public places. Since there are no available vaccines for COVID19 and our only hope is to wear a mask in public places at all times, the face mask detector, which was previously of no use, has become a necessity to keep people in check. As a result, the motivation for this project stems from the fact that the COVID19 pandemic is spreading and claiming the lives of hundreds of thousands of people every day. So far, there are no reports about any cure or vaccine being available for the disease. WHO is now suggesting that individuals wear face masks to avoid the risk of virus transmission [5]. Furthermore, this project is a step towards ensuring that precautions are taken by people to stop the spread. In this paper, we will be performing detection on live webcam feed with the help of convolutional neural networks and transfer learning to detect faces with masks and generate accurate results. CNNs such as Resnet18 and MobileNetV2 will be trained on various datasets such as the Real-World Face Mask Dataset, Wider Face Dataset, and others to provide us with a model capable of accurately detecting faces with and without masks. Multiple datasets are used to verify the accuracy of our model in all conditions. The rest of the paper is outlined as: Section I gives a reference to the progress currently made in face mask detection and the current research so far on which we plan to expand and build. Section II outlined the related works. Section III suggests the possible methods this paper will experiment with to build upon existing research so as to improve the current face mask detection scenario with reference to our future solutions we plan on implementing. While section IV contains the result of our experiment and discussion showing the accuracy of our model and the prediction it has made on our faces in real time. In section V, the implementation of our paper in the future and the research possibilities we will be exploring will be mentioned alongside our promising results.

2. RELATED WORKS

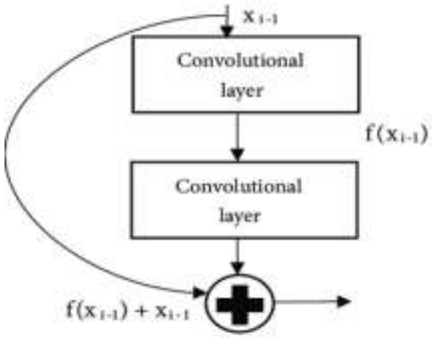
Face mask detection means finding the face of a person and classifying the face of a person as with a mask or without a mask. There has been a lot of development in the field of face detection. For example, the viola jones detector uses concepts of Haar-like features, AdaBoost algorithm, cascade classifier and integral imaging for detection. In the past, when there were no methods to develop feature extractors, the features were learnt using neural networks end to end. Object detection using deep learning is done by two families: the first is the Region CNN [6]. or the RCNN and the second one is the YOLO (You Only Look Once) family. The RCNN family uses two stage object detectors and therefore provides much more accurate results as compare to the YOLO family which uses one stage object detectors. But the YOLO family works faster than the RCNN family and is therefore used for detection in real time. In the same way, face detectors use similar architecture but with more face-related features. However, there is little research about face mask detection instead of just face detection. Current research about facial recognition and facial mask detection includes Deep Dense Face Detector (DDFD). It is a method that can detect faces in various angle ranges even when the faces are covered with obstacles like hands. It does not require pose/landmark annotation and is able to detect faces .It uses a single model based on deep convolutional neural networks [3]. A covered face is detected using skin ratio on each side of the head. To check the accuracy of this method, a human head(covered or uncovered) is taken from any angle and the space between -90 degrees to +90 degrees around the head is checked [7].

Current research about facial mask detection highlights methods like ResNet used as backbone, SSD and FPN as neck, that are able to extract high-level semantic information and then mix this information with the feature maps of previous layers [4]. and in other places where a hybrid model was developed using deep and classical machine learning for face mask detection is presented using Resnet50, decision trees, Support Vector Machine (SVM) [8]., and ensemble algorithm. Given that the testing accuracy for DS1 ranged from 93.44 percent using the decision tree classifier and 99.64 percent using the ensemble classifier, this approach turned out to be highly successful. For DS4, using the decision tree classifier, the testing accuracy ranged between 99.76 percent and 100 percent using the SVM classifier [8]. Other studies have even presented the use of lighter backbones, such as the MobileNetV2 architecture, which provides a significant cost and computational advantage over the standard 2D CNN model.

The process also involves the SSD MultiBox Detector such as ImageNet and PascalVOC which classifies images which are high quality [5], and the predefined training weights of VGG16 architecture are excellent methods to make predictions alongside feature extraction which creates two channels in the data, one for background and other for face. These models were designed in a manner that all the incorrect predictions were not taken into account while showing the result of the detection phase. This was done by first passing using the Median filter and the Closing Operation. The semantic segmentation via FCN [9]. Establishes the facial spatial position with a specific mark. In a single frame, the model can detect multiple facial masks. Post processing gives a significant boost to the precision of the pixel level [10]

3. THE PROPOSED APPROACH

In this paper, we make the use of concepts like convolutional neural networks and attention block modules, which play a bi-fold role in developing a model using Resnet18 and transfer learning to generate accurate results with the help of image classification. This method was found to be better than other models we researched about, like the predefined VGG16 architecture, which is a convolutional neural network model. Even though the VGG16 architecture is distinguished by its simplicity, using only 3x3 convolutional layers with 16 layers of weight stacked in increasing depth on top of one another. Volume size reduction is handled by max pooling. A Softmax classifier is then followed by two fully-connected layers, each with 4,096 nodes [11]. The use of Resnet is preferred due to its ability to solve the problem of vanishing gradient and its ability to find simpler mappings when they exist.



(a)

Residual block	Filter size	Number of filters	Total convolutional layer
1	7×7	64	1
2	3×3	64	4
3	3×3	128	4
4	3×3	256	4
5	3×3	512	4

(b)

Fig. 1. ResNet-18 architecture: (a) basic building block of residual learning, (b) filter information of different convolutional layer [12].

A. Neural Network Architecture

The model being built contains a ResNet18 [12]. Which acts as the backbone of this structure. Resnet is a short form of a residual network which is different from a normal convolutional neural network. In general, in a deep convolutional neural network, several layers are placed one after the other to train for classification. At the end of each layer, the network learns some features. In residual learning, the layers fit the residual mapping instead of learning the underlying mapping. Residual can be simply understood as subtraction of features learned from the input of that layer. ResNet does this using shortcut connections, i.e., jumping over some layers. It is effective as compared to a normal deep CNN due to skipping and also solves the problem of vanishing gradient.

Convolutional Neural Network (CNN) [7], uses convolution kernels for extracting higher-level features to converge with the initial images or function maps. We will also be experimenting with MobileNets [13], deep neural networks which are lightweight and have a streamlined structure. They are so streamlined because they use depth-wise separable convolutions. MobileNets contain two simple global hyper-parameters, which let the researcher select the right sized model for their model. Seeing particular constraints of the problem, these hyper-parameters efficiently trade off latency and accuracy. These showcase detailed experiments on resource and accuracy tradeoffs and support strong performance when compared to other popular ImageNet classification-based models. The effectiveness of MobileNets is demonstrated across a wide range of applications such as large-scale geo-localization, object detection, finegrain classification, and face attributes. MobileNets serve as a lighter backbone and it will be easier to implement OpenCV on such models [14]. These will inherit the parameters of the backbone and neck from pretrained Wider Face dataset and ImageNet so as to increase the efficiency of our model and help train our model better due to its limited dataset size.

B. Transfer Learning

Transfer learning is a machine learning method where a model made for a certain task is used as the base model for some other problem. It is a method that stores knowledge gained from one problem and uses it for some other problems [15]. This methodology we have applied is famous in deep learning. Pre-trained models are the place where we start when working with natural language processing and computer vision gives the immense computational and time resources needed to build neural network models on these issues and the enormous skill leaps they provide on related issues.

In this model, the approach for transfer learning that we use is a pre-trained model approach where we take a model we use pre-trained weights and reuse that model by tuning it to our dataset. Example: In our face mask detector, we take pre- trained imagenet weights from mobilenetv2 and then fine tune our model to our dataset.

Fig. 2. Real time detection of face mask (with mask)

Fig. 3. Real time detection of face mask (multiple people)

4. EXPERIMENTAL RESULT AND DISCUSSION

The current research has led us to the most efficient answers to our problem and the development of these models along with the training gave us the following results. We also connected each of these to OpenCV and have implemented a real-time working model for the same. Read Table 1 for the accuracies and losses of the two models we tested on, where we discussed the training accuracy and loss as well as the accuracy and loss on our validation dataset.

Table 1.

Model	Accuracy		Loss	
	train_accuracy	**val_accuracy**	**train_loss**	**val_loss**
MobileNetv2	0.9915	1.00	0.0707	0.0208
Resnet18	0.9920	0.9990	0.0000	0.0024

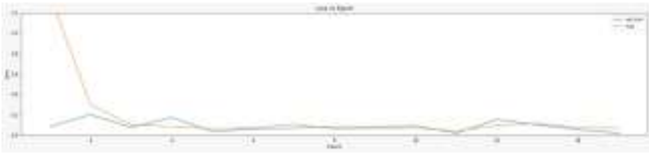


Fig. 4. Training and validation loss during fine tuning for MobileNetv2

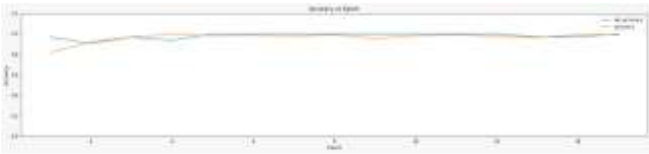


Fig. 5. Training and validation accuracy during fine tuning for MobileNetv2

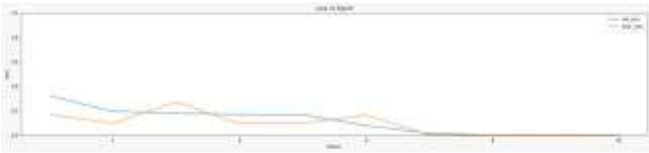


Fig. 6. Training and validation loss during fine tuning for Resnet18

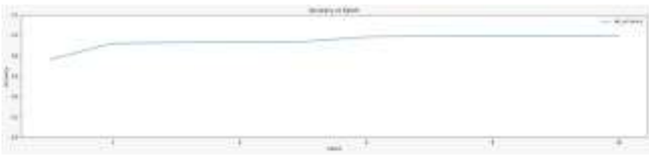


Fig. 7. Training accuracy during fine tuning for Resnet18

5. CONCLUSION AND FUTURE WORKS

Our research work has led us to the conclusion that there are several effective models that can detect the presence of face masks, and this creates the possibility of various different technological solutions to our COVID-19 problem. This research resulted in the realization that using Resnet18 with transfer learning is an effective method to detect face masks due to the lack of related data to train the models and this creates very effective results and using models like mobile net leads to better results in real time when there is low computational power involved and hence performs better when the input is a live video. Further experimentation continues after this paper as well. The following will be improved in future work, and the aim of this paper is to share our findings and present the best methodologies available in a brief paper so as to present the latest research on face mask detection and compile the best methodologies found to conveniently develop a better model in the future. OpenCV implemented on the models presents great opportunities for us to detect the violation of not wearing a face mask in real time, even with low image quality and helps authorities use technology for the greater good of mankind.

6. ACKNOWLEDGMENTS

We wish to extend our appreciation to the Department of Information Technology, Faculty of Computer Sciences, Parwan & KMUs University's for their generous support.

REFERENCES

1) H. Najma, H. Ansafb, and O. A. Hassen, “AN EFFECTIVE IMPLEMENTATION OF FACE RECOGNITION USING DEEP CONVOLUTIONAL NETWORK,” J. SOUTHWEST JIAOTONG Univ. Vol.54, vol. Vol.54, no. No.5, pp. 1–9, 2019, doi: 10.35741/issn.0258-2724.54.5.29.
2) O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,” no. Section 3, pp. 41.1-41.12, 2015, doi: 10.5244/c.29.41.
3) S. S. Farfade, “Multi-view Face Detection Using Deep Convolutional Neural Networks.”
4) M. Jiang and X. Fan, “Retinamask: A Face Mask Detector,” arXiv, 2020.
5) S. Yadav, “Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence July 2020,” no. August, 2020, doi: 10.22214/ijraset.2020.30560.
6) B. Novak, V. Ilić, and B. Pavković, “YOLOv3 Algorithm with additional convolutional neural network trained for traffic sign recognition,” pp. 165–168, 2020.
7) M. Loey, G. Manogaran, M. Hamed, N. Taha, N. Eldeen, and 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, no. May 2020, p. 108288, 2021, doi: 10.1016/j.measurement.2020.108288.
8) V. Jakkula, “Tutorial on Support Vector Machine (SVM).”
9) J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation.”
10) T. Meenpal, “Facial Mask Detection using Semantic Segmentation,” 2019 4th Int. Conf. Comput. Commun. Secur., no. April 2020, pp. 1–5, 2019, doi: 10.1109/CCCS.2019.8888092.
11) A. Rosebrock, “ImageNet: VGGNet, ResNet, Inception, and Xception with Keras,” Pyimagesearch, 2017. https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/.
12) K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
13) A. G. H. And and W. Wang, “‘Applications,’” 2012.
14) A. Mordvintsev, “OpenCV-Python Tutorials Documentation,” 2017.
15) J. West, D. Ventura, and S. Warnick, “Spring Research Presentation: A Theoretical Foundation for Inductive Transfer,” Brigham Young Univ. Coll. Phys. Math. Sci., 2007.