International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

A Survey: Different Techniques of Face Mask Detection

Amey Kalpe[1], Anshul Singh[2], Hrishikesh Kholamkar[3], Pankaj Pathave[4], Vasudha Phaltankar[5], Anil Kumar Gupta[6]

[1,2,3,4]Student, Department of Computer Engineering, Dr D. Y. Patil Institute of Technology, Pune, India
[5]Professor, Department of Computer Engineering, Dr D. Y. Patil Institute of Technology, Pune, India
[6]Senior IEEE Member

Abstract - *The outbreak of COVID-19, also known as Serious Acute Respiratory Syndrome Corona Virus-2 (SARS-CoV2), has caused a significant loss of life after being first discovered in China in the year 2019. World Health Organization has issued several guidelines, wearing a mask being the most important one. Face Mask Detection has seen significant progress in the field of Computer Vision and Image Recognition since the outbreak of the virus. Manual monitoring of the guidelines being followed is time-consuming, which can have serious consequences. Various automated approaches have been put forward to help detect whether masks are being worn correctly or not. The motive of this literature is to put forward a comprehensive survey of these approaches and techniques. Summarization of the models based on the framework used, accuracy rate and additional hardware support has been made.*

Key Words: Machine Learning, Deep Learning, OpenCV, TensorFlow, Keras, MobileNetV2, Raspberry Pi, Computer Vision, YOLO, Convolutional Neural Network.

1.INTRODUCTION

Coronavirus Disease 2019 (SARS-CoV2) has been spreading at an increasingly rapid rate since it has been discovered in the month of December in the year 2019. On the 11th of March 2020, World Health Organization (WHO) announced this crisis as an epidemic affecting people in over 114 countries. The virus spreads through the sneezing or communication of an infected person with other people in the vicinity. Almost all countries have made it a compulsion to wear a mask whenever being outdoors.

Researches have been conducted in the domain of face mask detection with the help of artificial neural networks and deep learning to allow automated detection of people who not wearing face masks. Deep Learning (DL) approaches, as a subset of Machine Learning (ML), are a predominant selection for numerous disease detection. Also, image processing techniques have seen wide adoption in the healthcare field, especially in the detection of cancer [1]. In [1], the author makes an effort to emphasize the DL and image processing techniques to detect COVID-19 and also highlights the challenges associated with the implementations of these technologies. The author highlights one of the major problems in the research as the lack of availability of accurate and adequate data since the performance of DL models depend on large, high-quality datasets. Machine Learning and Deep Learning are discussed in brief in the subsequent subtopics.

1.1 Machine Learning

In Machine Learning, a machine redefines its solution automatically for a particular kind of problem by learning from past experiences, similar to a human approach. Machine learning is the process of teaching a computer to learn and behave in a way similar to humans and to develop its learning abilities over time in an autonomous manner.

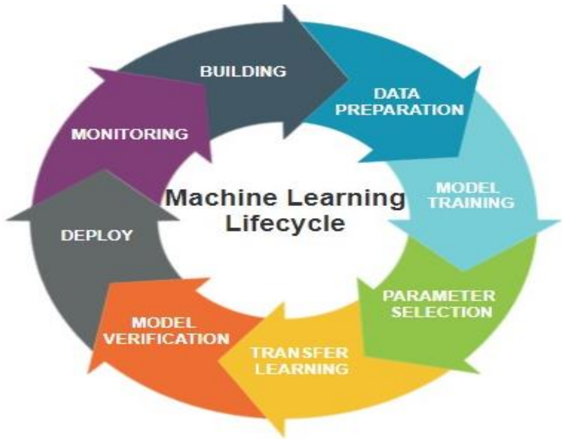

Fig-1. Machine Learning Lifecycle

Essentially, the signal or feedback accessible to the learning system has historically split up the approaches to machine learning into three different categories:

- Supervised Learning: A "teacher" presents the machine with example inputs and ideal outputs, with the aim of learning a rule that is general and maps inputs to outputs.
- Unsupervised Learning: The input data has no labels associated with it; the algorithm has to learn or figure out on its own the labels of the data. Unsupervised algorithms mainly try to find the hidden patterns and try to extract and learn the features using clustering or neural networks.
- Reinforcement Learning: A computer program communicates with a changing/dynamic

environment in order to carry out a specific task (e.g., train robots). The program gets feedback in the form of a reward (positive or negative) as it navigates its problem area, which it aims to maximize [2].

1.2 Deep Learning

Deep learning approaches are used to learn feature hierarchies, which are made up of features obtained from higher-levels of the hierarchy that are produced from lower-level features in the hierarchy. The algorithms in DL make use of these hierarchies to find good representations, with higher-level features resulting from an aggregation of lower-level features.

As described above, this hierarchy helps the machine define complex concepts in terms of multiple simpler concepts. The graph depicting how these definitions are constructed on top of one another is deep, consisting of several layers. As a result, the approach has been coined the name Deep Learning. Deep learning excels at solving problems with analogue inputs (and also outputs), which are not the regular structured format of a table, but text files, image files or maybe audio files.[2].

In the surveyed literature, TensorFlow and Keras API have been widely adopted because of being open-source, easily trainable and the wide community support.

In

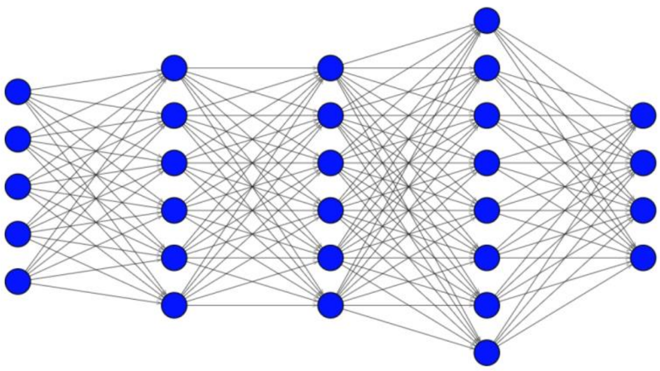

Fig-2. Deep Neural Network

2. Related Work

An approach to detect face masks has been proposed by [3], which uses Image Augmentation to compensate for the less amount of data available at the time of publishing [3]. The authors use the Real-World Masked Face Dataset (RMFD) dataset created by [2]. PyTorch and OpenCV are used for image transformation, resizing the images to a size of 256 x 256. The proposed model uses MobileNetV2 [4] as a classifier to classify the photos from the video or image stream into two classes – "mask", indicated with a green rectangle around the face or "no mask", marked with a red rectangle around the face. MobileNetV2 is trained using PyTorch. [5, 6, 7] use MobileNetV2 in a similar approach. [5] uses image augmentation and a dual of face detection and mask detection in order. Faces from the camera feed are detected in the first step, and then these faces are run through a mask detection model. The model is trained using TensorFlow and Keras. [6] noticed the lack of quality datasets available and created a new dataset which consisted of a combination of various open-source datasets including RMFD, Kaggle's Medical Mask Dataset (MMD) by Mikolaj Witkowski and another one provided by Prajna Bhandary [14] from PyImageSearch. This combination resulted in a dataset consisting of 5521 images having the label "*with_mask*" and 5521 images having the label "*without_mask*", maintaining a complete balance among the two classes. The proposed methodology also uses the Deep Neural Net (DNN) model from the OpenCV library, which includes Single Shot Multibox Detector (SSD), an object detection model; thus, the name, "SSDMobileNetV2" (SSDMNV2). [7], uses the Single Shot Detector (SSD) algorithm, which depends on the specified bounding box against the objects. The author uses the VGG-16 network, a dense network with 16 convolutional layers, for the SSD. The working of the SSD model is identical to the model mentioned in [5]. The author tests three models to find the best fit in the scenario, Xception (Extreme Inception), MobileNetV2 and ResNet50 with accuracies 85%, 100% and 99% respectively. As MobileNetV2 had the best accuracy of 100 percent, the author chose to use it. The reason for the adoption of MobileNetV2 in [5, 6, 7] is the computational efficiency that allows the model usage in embedded devices.

In [8], the author proposes a model that integrates InceptionV3's transfer learning with security cameras to recognize people who aren't wearing a mask in public places using the Simulated Masked Face Dataset (SMFD). As this model was proposed before [6], the data was limited. Due to this constraint, to address the limited availability, the images were transformed in several ways like rotation, zoom and flip. Google's InceptionV3 is an architecture consisting of a 48-layer convolutional neural network (CNN). This paper uses a methodology based on transfer learning for classifying people who aren't wearing face masks is proposed, which makes use of the InceptionV3 pre-trained model. Five more layers are added to the network after removing the last layer of InceptionV3.

In [9], The author proposes a model consisting of two components. The primary segment is based on the ResNet-50 deep learning model and is designed for extracting features. The next component is based on You Only Look Once (YOLO) version 2, the successor to YOLO version 1 [10] and is designed to detect medical face masks. Two datasets of medical face masks have been merged into a single dataset for this research. By using the Intersection over Union (IoU) evaluation metric to estimate the best amount of anchor boxes, the proposed system tries to improve the detection performance. The authors conducted experiments using two datasets, MMD and Face Mask Dataset (FMD) from Kaggle. YOLO version 2, an object detection deep network, consists of two networks: feature extraction and detection. ResNet (residual neural network) is a type of deep transfer learning which is based on residual networks. The YOLO version 2 detection network is a CNN with a transform layer (TL), few convolutional layers and an output layer. The TL takes the activations from the convolutional layer and improves the consistency of the DNN. The bounding box forecast is converted by the TL into the goal box's outlines.

The methodology proposed in [11] uses YOLO version 3 to present a new variant of the existing object detection method, Squeeze and Excitation-YOLOv3 (SE-YOLOv3) with the intention to enhance the performance and accuracy of YOLO version 3. The authors try to address the problem of most of the studies not considering the situation of incorrect mask-wearing. A dataset of masked faces from all races was created by the authors, naming it as Properly-Wearing Masked Face Detection Dataset (PWMFD). They included the class of "*incorrect_mask*", in addition to the two previously existing classes. The methodology proposed can detect faces and evaluate whether the mask is worn correctly or not in real-time. The Squeeze & excitation block was added between the convolution layer of Darknet53, which assisted the model in learning the relation between channels. The Generalized Intersection over Union (GIoU) evaluation metric over Mean Squared Error (MSE) was used to improve
the loss function. After these improvements, the analysis revealed that the mean average precision saw an increase of around 6-7 % over the baseline model.

In [15], the author proposes a methodology that uses a CNN to design the classifier. Raspberry Pi is used to implement the project by virtue of OpenCV, TensorFlow and Python. The classification of facial images is done using CNN, and then deep learning models are trained to recognize facial patterns. The author uniquely labels the captured images and creates a knowledge-based dataset. The images are obtained through various optical devices like cellphone cameras or digital cameras. Image Segmentation has been used to identify the Region of Interest (ROI), which would be the area comprising the face. Finally, classification using TensorFlow, OpenCV and VGG-16 CNN model is carried out, and the model is deployed onto the storage device of Raspberry Pi.

[16] uses an approach similar to [15], using the datasets provided by Prajna Bhandary with two classes, *"with_mask"* and *"without_mask"* and another Kaggle Dataset, which consists of the identical two classes, but with different orientations of the faces in the images. OpenCV is used for resizing the images to 100 x 100 and converting images from RGB to grey-scale. The proposed approach uses a pre-trained CNN model consists of two 2d convolution layers linked to a dense neuron layer, TensorFlow as the backend, and Keras for the model's layer implementation.
pt.raj2015@gmail.com

Table-1. Accuracy of baseline models and proposed models in the literature.

Models/Architecture	**Dataset**	**Environment**	**Additional Hardware Support**	**Image Augmentation Used**	**Accuracy**
MobileNet	SMFD	Real Time	Yes	No	92.89 %
MobileNet V2 ([5], [6], [7])	Local Dataset, RMFD, [14]	Real Time	Yes	Yes	98 %, 93 %, 100 %
Xception (Extreme Inception) [7]	ImageNet [13], JFT	Real Time	NA	No	85 %
ResNet 50	ImageNet [13]	Real Time	NA	No	99 %
VGG16	ImageNet [13]	Real Time	Yes	No	98.5 %
Inception V3 [8]	SMFD	Real Time	Yes	No	100 %
YOLO	ImageNet [13]	Real Time	Yes	No	76.5 %
YOLOv2 with ResNet50 [9]	MMD, FMD	Real Time	NA	Yes	81 %
YOLOv3	Common Objects in Context (COCO) [12]	Real Time	Yes	No	93.8 %
SE-YOLOv3 [11]	PWMFD	Real Time	Yes	Yes	96.2 %

3. CONCLUSIONS

The survey tries to reflect the comparison of different approaches to face mask detection, demonstrating how adding a Squeeze and Excitation block to YOLO version 3 improves the accuracy by adding parameters to each channel at negligible additional computational cost, how using MobileNetV2 makes it possible to implement mask detection in embedded devices for real-time detection, and how using CNNs to classify the images automates the feature detection. In the surveyed literature, the authors have different perspectives on the problem domain, as some have tried to improve upon the existing models while some implemented their own methodologies. On the basis of this survey, it can be concluded that the discipline of Deep Learning has the capability to impede the spread of the virus, using energy-efficient hardware and fast and efficient processing software in companion.

REFERENCES

1) Sweta Bhattacharya, Praveen Kumar Reddy Maddikunta, Quoc-Viet Pham, Thippa Reddy Gadekallu, Siva Rama Krishnan S, Chiranji Lal Chowdhary, Mamoun Alazab, Md. Jalil Piran, “Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey, Sustainable Cities and Society”, Volume 65, 2021, 102589, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2020.102589.
2) Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., ... & Liang, J. (2020). Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093.
3) Vinitha.V, Velantina.V (2020, August). Covid-19 Facemask Detection With Deep Learning And Computer Vision. International Research Journal of Engineering and Technology, 07(08), 3127-3132.
4) Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).
5) Saiyam Jain, M. G. (2021, January). Face Mask Detection using Machine Learning and Deep Learning. International Research Journal of Engineering and Technology, 08(01), 433-437.
6) Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., & Hemanth, J. (2021). SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustainable Cities and Society, 66, 102692. doi:10.1016/j.scs.2020.102692R.
7) Lippert, Prof & Bergner, Benjamin & Ahmed, Akhyar & Ali, Raza & Adeel, Saniya & Shahriar, Md Hasan & Mojumder, Md. (2020). Face Mask Detector. 10.13140/RG.2.2.32147.50725.
8) Chowdary, G. J., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020, December). Face mask detection using transfer learning of inceptionv3. In International Conference on Big Data Analytics (pp. 81-90). Springer, Cham.
9) Loey, M., Manogaran, G., Taha, M. H., & Khalifa, N. E. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society, 65, 102600. doi:10.1016/j.scs.2020.102600
10) Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
11) Jiang, X., Gao, T., Zhu, Z., & Zhao, Y. (2021). Real-Time Face Mask Detection Method Based on YOLOv3. Electronics, 10(7), 837.
12) Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.
13) Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). IEEE.
14) “prajnasb/observations”, GitHub, 2020. [Online]. Available:https://github.com/prajnasb/observations/tree/master/experiements/data. 2020.
15) Militante, S. V., & Dionisio, N. V. (2020, August). Real-Time Facemask Recognition with Alarm System using Deep Learning. In 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC) (pp. 106-110). IEEE.
16) Das, A., Ansari, M. W., & Basak, R. (2020, December). Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV. In 2020 IEEE 17th India Council International Conference (INDICON) (pp. 1-5). IEEE.