

# Review on Literature Survey of Human Recognition with Face Mask

Dr. Vandana S. Bhat

Department of Information Science and Engineering  
SDM College of Engineering and Technology  
Dharwad, India

Komal Mainalli

Department of Information Science and Engineering  
SDM College of Engineering and Technology  
Dharwad, India

Arpita Durga Shambavi

Department of Information Science and Engineering  
SDM College of Engineering and Technology  
Dharwad, India

K M Manushree

Department of Information Science and Engineering  
SDM College of Engineering and Technology  
Dharwad, India

Shraddha V Lakamapur

Department of Information Science and Engineering  
SDM College of Engineering and Technology  
Dharwad, India

**Abstract** — The COVID-19 is an unparalleled crisis leading to a huge number of casualties and security problems. To reduce the spread of coronavirus, people often wear masks to protect themselves. This makes face recognition a very difficult task since certain parts of the face are hidden. A primary focus of the researchers during the ongoing coronavirus pandemic is to come up with suggestions to handle this problem through rapid and efficient solutions. This paper aims to present a review of various methods and algorithms used for human recognition with a face mask. Different approaches i.e. Haar cascade, Adaboost, VGG-16 CNN Model, etc. are described in this paper. A comparative analysis is made on these methods to conclude which approach is feasible. With the advancement of technology and time more reliable methods for human recognition with a face mask can be implemented in the future. Finally, it includes some of the applications of face detection. This system has various applications at public places, schools, etc. where people need to be detected with the presence of a face mask and recognize them and help society.

**Keywords** — Viola Jones, Adaboost, Computer Vision, Convolutional Neural Network, MobileNetV2, VGG – 16 Model.

## I. INTRODUCTION

The COVID-19 virus can be spread through contact and contaminated surfaces. There are so many essential equipment's needed to fight against the Corona virus. One of such most essential is Face Mask. Firstly, a face mask was not mandatory for everyone but as the day progresses scientists and Doctors have recommended everyone to wear a face mask. So to detect whether a person is wearing Face Mask or not is an essential process to implement in the society currently which can be used for various applications like at the airport, hospitals, offices, schools, etc. This system can be of great importance at airports to detect travelers whether they are wearing a mask or not and at schools to ensure students are wearing a face mask for their safety. However, wearing the mask face causes the following problems: i) fraudsters and thieves take advantage of the

mask, stealing and committing crimes without being identified. ii) community access control and face authentication have become very difficult tasks when the most part of the face is hidden by a mask. Hence, detecting the face mask and recognizing the person behind the face mask is very important.

In this paper, we will address the different approaches tried to implement the face detection and face recognition system along with the presence of face masks from various papers. To study the existing techniques and analyze which approach is feasible and efficient enough to implement to the current state of the society, we have organized this paper into different sections. Later, by considering the constraints and drawbacks of each approach the best techniques will be concluded.

Recognition has to classify a given face, and there are as many classes as candidates. Consequently, many face detection methods are very similar to face recognition algorithms. Methods are divided into four categories. These categories may overlap, so an algorithm could belong to two or more categories. This classification can be made as follows:

**Knowledge-based methods:** Ruled-based methods that encode our knowledge of human faces. Feature invariant methods. Algorithms that try to find invariant features of a face despite its angle or position. The problem with this approach is; it is difficult to translate human knowledge into well-defined rules. If these rules are strict then they may fail to detect faces that do not pass all the rules. But on the other hand, if the rules are too general then there may be many false detections.

**Template matching methods:** These algorithms compare input images with stored patterns of faces or features. **Appearance-based methods.** A template matching method whose pattern database is learned.

## II. LITERATURE SURVEY

**MAMATA S. KALAS, "REAL TIME FACE DETECTION AND TRACKING USING OPENCV", International Journal of Soft Computing and Artificial Intelligence, ISSN: 2321-404X, Volume-2, Issue-1, May-2014**

### RELATED WORK

Face detection is defined as the procedure that has many applications like face tracking, pose estimation or compression. Face detection is a two-class problem where we have to decide if there is a face or not in a picture. This approach can be seen as a simplified face recognition problem.

AdaBoost: Adaboost is an algorithm for constructing a "strong" classifier as a linear combination. Adaboost, short for Adaptive Boosting, is a machine learning algorithm. It is a meta-algorithm and can be used in conjunction with many other learning algorithms to improve their performance. Adaboost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. Adaboost generates and calls a new weak classifier in each of a series of rounds. For from a set of training images. This method can be used for both face detection and face locations. In this method, a standard face (such as frontal) can be used. The advantages of this method are that it is very simple to implement the algorithm, and it is easily to determine the face locations such as nose, eyes, mouth, etc. based on the correlation values.

### ALGORITHMS OF FACE DETECTION

Haar-like feature: Haar-like wavelets are binary rectangular representations of 2D waves. A common visual representation is by black (for value minus one) and white (for value plus one) rectangles. The square above the 0-1-interval shows the corresponding Haar-like wavelet in common black-white representation. The rectangular masks used for visual object detection are rectangles tessellated by black and white smaller rectangles. Those masks are designed in correlation to visual recognition tasks to be solved, and known as Haar like feature each call, a distribution of weights is updated that indicates the importance of examples in the data set for the classification. On each round, the weights of each incorrectly classified example are increased, and the weights of each correctly classified example are decreased, so the new classifier focuses on the examples which have so far eluded correct classification.

**Walid Hariri, "Efficient Masked Face Recognition Method during the COVID-19 Pandemic", pp.1-7, July 2020**

The COVID-19 is an unparalleled crisis leading to huge number of casualties and security problems. In order to reduce the spread of coronavirus, people often wear masks to protect themselves. This makes the face recognition a very difficult task since certain parts of the face are hidden. A primary focus of researchers during the ongoing coronavirus

pandemic is to come up with suggestions to handle this problem through rapid and efficient solutions.

### PROPOSED METHOD

**Pre-processing and cropping filter:** The images of the dataset are already cropped around the face, so there is no need of a face detection stage to localize the face from each image. To do so, we detect 68 facial landmarks using Dlib-ml open-source library. According to the eye's location, we apply a 2D rotation to make them horizontal. The next step is to apply a cropping filter in order to extract only the non-masked region. To do so, we firstly normalize all face images into 240 x 240 pixels. Next, we use the partition into blocks. The principle of this technique is to divide the image into 100 fixed-size square blocks (24 x 24 pixels in our case). Then we extract only the blocks including the non-masked region (blocks from number 1 to 50). Finally, we eliminate the rest of the numbers of the blocks.

**Feature extraction layer:** They extract deep features using VGG-16 face CNN descriptor [20] from the 2D images. It is trained on ImageNet dataset which has over 14 million images and 1000 classes. Its name VGG-16 comes from the fact that it has 16 layers. Its layers consist of convolutional layers, Max Pooling layers, Activation layers, Fully connected layers. There are 13 convolutional layers, 5 Max Pooling layers and 3 Dense layers which sums up to 21 layers but only 16 weight layers. In this work, we only consider the feature maps (FMs) at the last convolutional layer, also called channels. These features will be used in the following in the quantization stage.

**Deep bag of features layer:** From the  $i$ th image, we extract feature maps using the feature extraction layer described above. In order to measure the similarity between the extracted feature vectors and the codewords also called term vectors, we applied the RBF kernel as a similarity metric as proposed in. Thus, the first sub layer will be composed of RBF neurons, each neuron is referred to as a codeword. The size of the extracted feature map denotes the number of the feature vectors that will be used in the BoF layer. Here we refer by  $V_i$  to the number of feature vectors extracted from the  $i$ th image.

The most used automatic algorithm is of course k-means. Let  $F$  the set of all the feature vectors, defined by  $F = \{V_{ij}, i = 1 \dots V, j = 1 \dots V_i\}$  and  $V_k$  is the number of the RBF neurons centers referred by  $c_k$ . Note that these RBF centers are learned afterward to get the final codewords. The quantization is then applied to extract the histogram with a pre-defined number of bins, each bin is referred to as a codeword. RBF layer is then used as a similarity measure, it contains 2 sub layers:

**RBF layer:** Measures the similarity of the input features of the probe faces to the RBF centers. Formally: the  $j$ th RBF neuron  $\phi(X_j)$  is defined by:  $\phi(X_j) = \exp(-\frac{1}{2} \frac{\|x - c_j\|^2}{\sigma_j^2})$ , (1) Where  $x$  is a feature vector and  $c_j$  is the center of the  $j$ th RBF neuron.

**Quantization layer:** The output of all the RBF neurons is collected in this layer that contains the histogram of the

global quantized feature vector that will be used for the classification process. The final histogram is defined by:

$$h_i = \sum_k V_j^N \phi(V_{jk})$$

Where  $\phi(V)$  is the output vector of the RBF layer over the  $ck$  bins.

Once the global histogram is computed, pass the classification stage to assign each test image to its identity. To do so, a Multilayer perceptron classifier (MLP) is applied where each face is represented by a term vector. Deep BoF network can be trained using back-propagation and gradient descent. Note that the 10 cross-validation strategy is applied in our experiments on the RMFRD dataset.  $V = [v_1, \dots, v_k]$  the term vector of each face is noted, where each  $v_i$  refers to the occurrence of the term  $i$  in the given face.  $t$  is the number of attributes, and  $m$  is the number of classes (face identities). Test faces are defined by their codeword  $V$ . MLP uses a set of term occurrences as input values ( $v_i$ ) and associated weights ( $w_i$ ) and a sigmoid function ( $g$ ) that sums the weights and maps the results to an output ( $y$ ).

RMFRD faces were firstly pre-processed as described. Using the normalized 2D faces of sizes 240 x 240 pixels, VGG-16 pretrained model is applied to extract the best features from the last convolutional layer as presented. The quantization is then applied to extract the histogram of 70 bins as presented. Finally, MLP is applied to classify faces. In this experiment, the 10 cross-validation strategy is used to evaluate the recognition performance. The experiments are repeated ten times in the RMFRD dataset, where 9 samples are used as the training set and the remaining sample as the testing set, and the average results are calculated.

**Vinitha.V1, Velantina.V2, "COVID-19 FACEMASK DETECTION WITH DEEP LEARNING AND COMPUTER VISION", International Research Journal of Engineering and Technology (IRJET), Volume: 07, pp.1-6, Aug 2020.**

The mask face detection model that is based on computer vision and deep learning. The model is integration between deep learning and classical machine learning techniques with OpenCV, tensor flow and Keras. We have used deep transfer learning for feature extractions and combined it with three classical machine learning algorithms. We introduced a comparison between them to find the most suitable algorithm that achieved the highest accuracy and consumed the least time in the process of training and detection.

#### PROPOSED SYSTEM

The proposed system focuses on how to identify the person on image/video stream wearing face mask with the help of computer vision and deep learning algorithm by using the OpenCV, Tensor flow, Keras and PyTorch library.

Approach –

1. Train Deep learning model (MobileNetV2)
2. Apply mask detector over images / live video stream

The majority of the images were augmented by OpenCV. The set of images were already labeled "mask" and "no mask".

The images that were present were of different sizes and resolutions, probably extracted from different sources or from machines (cameras) of different resolutions.

Face Mask Detection in webcam stream.

- The flow to identify the person in the webcam wearing the face mask or not. The process is two-fold.
  - a. To identify the faces in the webcam.
  - b. Classify the faces based on the mask.

Identify the Face in the Webcam: To identify the faces a pre-trained model provided by the OpenCV framework was used. The model was trained using web images. OpenCV provides 2 models for this face detector:

1. Floating-point 16 version of the original Caffe implementation.
2. 8 bit quantized version using Tensor flow.

**Yassin Kortli, Maher Jridi, Ayman Al Falou, and Mohamed Atri: "Face Recognition Systems: A Survey", 20(2): 342, pp.1-10, 2020 Jan 7.**

The objective of developing biometric applications, such as facial recognition, has recently become important in smart cities. Besides, many scientists and engineers around the world have focused on establishing increasingly robust and accurate algorithms and methods for these types of systems and their application in everyday life.

All types of security systems must protect all personal data. The most commonly used type for recognition is the password. However, through the development of information technologies and security algorithms, many systems are beginning to use many biometric factors for the recognition task.

These biometric factors make it possible to identify people's identities by their physiological or behavioral characteristics. They also provide several advantages, for example, the presence of a person in front of the sensor is sufficient, and there is no more need to remember several passwords or confidential codes anymore.

#### FACE RECOGNITION

Three basic steps are used to develop a robust face recognition system:

The face recognition system begins first with the localization of the human faces in a particular image. The purpose of this step is to determine if the input image contains human faces or not. The variations of illumination and facial expression can prevent proper face detection. To facilitate the design of a further face recognition system and make it more robust, pre-processing steps are performed. Many techniques are used to detect and locate the human face image, for example, Viola-Jones detector, histogram of oriented gradient (HOG), and principal component analysis (PCA). Also, the face detection step can be used for video and image classification, object detection, region-of-interest detection, and so on.

Feature Extraction:

The main function of this step is to extract the features of the face images detected in the detection step. This step

represents a face with a set of features vector called a “signature” that describes the prominent features of the face image such as mouth, nose, and eyes with their geometry distribution. Each face is characterized by its structure, size, and shape, which allow it to be identified. Several techniques involve extracting the shape of the mouth, eyes, or nose to identify the face using the size and distance. HOG, Eigen face, independent component analysis, linear discriminant analysis (LDA), scale invariant feature transform (SIFT), Gabor filter, local phase quantization (LPQ), Haar wavelets, Fourier transforms, and local binary pattern (LBP) techniques are widely used to extract the face features.

#### Face Recognition:

This step considers the features extracted from the background during the feature extraction step and compares it with known faces stored in a specific database. There are two general applications of face recognition, one is called identification and another one is called verification. During the identification step, a test face is compared with a set of faces aiming to find the most likely match. During the identification step, a test face is compared with a known face in the database in order to make the acceptance or rejection decision. Correlation filters (CFs), convolutional neural network (CNN), and also k-nearest neighbor (K-NN) are known to effectively address this task.

**C.Jagadeeswari, M. Uday Theja, “Performance Evaluation of Intelligent Face Mask Detection System with various Deep Learning Classifiers”, International Journal of Advanced Science and Technology, Vol. 29, No. 11s, pp.3074-30780, (2020).**

Corona virus disease (COVID-19) is an airborne infectious disease caused by a newly discovered corona virus. The best way to prevent or slow down the transmissions to have knowledge on the COVID-19 virus, the disease it can cause and how it passes. There are many steps suggested by WHO (World Health Organization) to prevent the spread. One of which is wearing medical masks which is highly desirable even after the lockdown period until a vaccine /Medicine is invented.

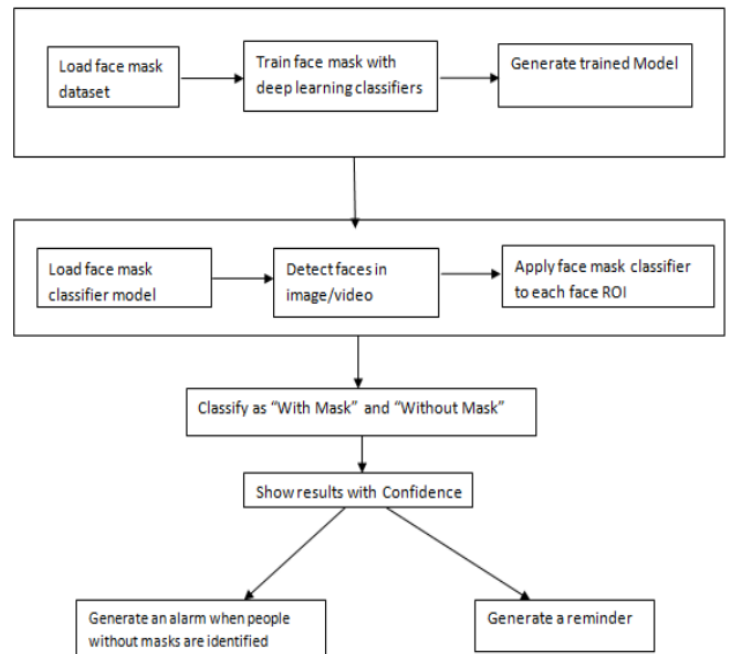
This system aims at classifying whether a person is wearing a mask or not by taking input from Images and Real time streaming Videos.

The classification of the images is done by training the model in 2 phases:

Phase 1: Face mask dataset is loaded into the system. Different classifiers like MobileNetV2, ResNet50, and VGG16 are used to generate a trained model.

Phase2: Load the face mask classifier model.

Detect faces in the images/video stream. Apply the classifier to each face RoI. Classify the images to be “With Mask” and “Without Mask” with Confidence.



This system may then be interfaced with

Case 1: Existing access control system so that violators can be restricted.

Case 2: There could be some scenarios in work places where people may forget or just put off the mask when it becomes uneasy for them to get accustomed to the new face masks. In such cases, alarm by the system may be disturbing other workers. Hence the concerned authorities can take proper measures to alert the user so that they can wear the mask again.

Classifier	Epochs	Train/test size	Optimizer	Train loss	Train Accuracy	Test Loss	Test accuracy
Mobilenet V2	20	90/10	ADAM	0.0090	0.9981	0.0071	1.0000
			ADAGRAD	0.2454	0.9148	0.1811	0.9819
			SGD	0.1549	0.9502	0.0216	0.9855

From the Table 1 it is observed that performance of ADAM optimizer is good in both training and testing when compared with other two optimizers ADAGRAD and SGD. Table 2: Results of the proposed system with Resnet50 classifier:



Classifier	Epochs	Train/test size	Optimizer	Train loss	Train Accuracy	Test Loss	Test accuracy
Resnet50	20	90/10	ADAM	0.0068	0.9975	0.0557	0.9856
			ADAGRAD	0.1087	0.9693	0.0019	1.0000
			SGD	0.1114	0.9693	0.0100	1.0000

From the Table 2 it is observed that performance of ADAM optimizer is good in both training and testing when compared with other two optimizers ADAGRAD and SGD and all test accuracies are good.

### III. COMPARISON ANALYSIS OF DIFFERENT APPROACHES

TABLE I

Methodology	Computer Vision	Convolutional Neural Network
Approach	When a computer looks at an image with a specific goal, the irrelevant information is not taken into account. This helps reduce the types of bias that humans might introduce to a process, whether intentionally or unintentionally.	Accuracy in image recognition problems. This helps us to get the results accurate and differentiate between "mask" and "no mask".
	When the device fails because of a virus or other software issues, it is highly probable that Computer Vision and image processing will fail.	CNN automatically detects the important features without any human supervision.
	Time and error rate are reduced in the process of Computer Imaging. It reduces the cost of hire and train special staff (human force) to do the activities that computers does.	If there is no good GPU they are quite slow to train (for complex tasks). They use to need a lot of training data.
		It is computationally very expensive and time consuming to train with traditional CPUs.
		Once we train the system, the predictions are pretty fast.

TABLE II

Methodology	Haar like features and Adaboost	VGG- 16 CNN Model	MobileNetV2
Approach	The wavelet template have ability to capture high-level knowledge about the object class (structural information expressed as a set of constraints on the wavelet coefficients) and incorporate it into the low-level process of interpreting image intensities.	In order to avoid a bad reconstruction process, these approaches aim to detect regions found to be occluded in the face image, and discard them completely from the feature extraction and classification process.	MobileNetV2 is a state of the art for mobile visual recognition including classification, object detection and semantic segmentation.
	Due to the non-invariant nature of the normal Haar-like features, classifiers trained with this method are often incapable of finding rotated objects.	VGG16 significantly outperforms the previous generation of models in the ILSVRC-2012 and ILSVRC-2013 competitions.	This classifier uses Depth wise Separable Convolution which is introduced to dramatically reduce the complexity cost and model size of the network, and hence is suitable to Mobile devices.
	Calculates the coefficients of wavelets by the average intensities of the pixels of a region may increase learning time.	The size of VGG-16 trained ImageNet weights is 528 MB. So, it takes quite a lot of disk space and bandwidth that makes it inefficient.	In MobileNetV2, another best module that is introduced is inverted residual structure.
	Haar-like features are more robust to illumination changes than color histogram. The Integral Image allows the sum of pixel responses within a given sub-rectangle of an image to be computed quickly.	In VGG16 instead of having a large number of hyper-parameters they focused on having convolution layers of 3x3 filter with a stride 1 and always used same padding and maxpool layer of 2x2 filter of stride 2.	Non-linearity in narrow layers is deleted. Keeping MobileNetV2 as backbone for feature extraction, best performances for object detection.

#### IV. FUTURE WORK

Human recognition with face mask has various applications in different domains. The various methodologies discussed in this paper can be based on the particular demands of the application. As every approach has its very own pros and cons we need to determine the best approach according to the necessity. Face detection is gaining the interest of marketers. It can be used at various domains like airports where this system can be of great importance at airports to detect travellers whether they are wearing mask or not. Travellers' data can be captured as videos in the system at the entrance. Hospitals - This system can be integrated with CCTV cameras and that data may be administered to see if their staff is wearing mask or not. Offices - This system can help in maintaining safety standards to prevent the spread of Covid-19, to detect whether the person is wearing mask or not. The scope of this system extends to security systems of wide range right from Malls, hospitals, IT companies and in many such public areas.

#### V. CONCLUSION

Different methods and approaches of face mask detection and recognition have been reviewed in this paper. In comparison, Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detector. The key advantage of a Haar-like feature over most other features is its calculation speed. Adaboost can be less susceptible to the over fitting problem than most learning algorithms. Bad feature of adaptive boosting is its sensitivity to noisy data and outliers. In real-world scenarios human faces might be occluded by other objects such as facial mask. This makes the face recognition process a very challenging task. Deep learning-based method and quantization-based technique achieves a high recognition performance. MobileNetV2 is a very effective feature extractor for object detection and segmentation. MobileNetV2 provides a very efficient mobile-oriented model that can be used as a base for many visual recognition tasks. For the best of our knowledge, this work addresses the problem of masked face recognition and different approaches during COVID-19 pandemic. It is worth stating that this study is not limited to this pandemic period since a lot of people are self-aware constantly, they take care of their health and wear masks to protect themselves against pollution and to reduce other pathogens transmission

#### REFERENCES

- [1] Mamata s. Kalas, "REAL TIME FACE DETECTION AND TRACKING USING OPENCV" , International Journal of Soft Computing and Artificial Intelligence, ISSN: 2321-404X, Volume-2, Issue-1, May-2014
- [2] Walid Hariri, "Efficient Masked Face Recognition Method during the COVID-19 Pandemic", pp.1-7, July 2020
- [3] Viniitha.V1, Velantina.V2, "Covid-19 facemask detection with deep learning and computer vision" International Research Journal of Engineering and Technology (IRJET), Volume: 07, pp.1-6, Aug 2020.
- [4] Yassin Kortli, Maher Jridi, Ayman Al Falou, and Mohamed Atri "face recognition systems: a survey", 20(2): 342, pp.1-10, 2020 Jan 7.
- [5] Viniitha.V1, Velantina.V2, "COVID-19 FACEMASK DETECTION WITH DEEP LEARNING AND COMPUTER VISION" International Research Journal of Engineering and Technology (IRJET), Volume: 07, pp.1-6, Aug 2020.