Masked Face Recognition via a Combined SIFT and DLBP Features Trained in CNN Model

Nahla Fahad Aljarallah[1] and Diaa Mohammed Uliyan[1*]

nahlafahad8@gmail.com, d.uliyan@uoh.edu.sa

[1] Department of of information and computer science, College of Computer Science and Engineering,

University of Hai'l, Ha'il, Saudi Arabia.

* *Corresponding Author (d.uliyan@uoh.edu.sa)*

Summary

The latest global COVID-19 pandemic has made the use of facial masks an important aspect of our lives. People are advised to cover their faces in public spaces to discourage illness from spreading. Using these face masks posed a significant concern about the exactness of the face identification method used to search and unlock telephones at the school/office. Many companies have already built the requisite data in-house to incorporate such a scheme, using face recognition as an authentication. Unfortunately, veiled faces hinder the detection and acknowledgment of these facial identity schemes and seek to invalidate the internal data collection. Biometric systems that use the face as authentication cause problems with detection or recognition (face or persons). In this research, a novel model has been developed to detect and recognize faces and persons for authentication using scale invariant features (SIFT) for the whole segmented face with an efficient local binary texture features (DLBP) in region of eyes in the masked face. The Fuzzy C means is utilized to segment the image. These mixed features are trained significantly in a convolution neural network (CNN) model. The main advantage of this model is that can detect and recognizing faces by assigning weights to the selected features aimed to grant or provoke permissions with high accuracy.

***Keywords**: Face recognition; masked face; scale invariant features; image segmentation; local binary pattern; CNN model.*

1. Introduction

We are recently witnessing a large-scale spread of the Coronavirus Covid-19, with the number of confirmed cases estimated at more than 84 million cases [1]. The great rapid spread and the damage it caused in terms of loss in economic and human lives led to its declaration as a global pandemic by the World Health Organization.

Coronavirus is an infectious disease that attacks the human respiratory system and is transmitted from one person to another through the spray that comes out of the mouth when sneezing or coughing, and its danger ranges according to the immune strength of the person. The spread of the virus is continuing until this moment. Also, no treatment has been discovered to confront it, so it is necessary to adhere to the imposed prevention instructions of social distancing, wearing face masks and gloves, and using medical alcohol to sterilize consistently [2][3].

One of the most important recommendations recommended by the World Health Organization and imposed by governments on their citizens is wearing a face mask, where the mask greatly contributes to limiting the spread of disease and protects the wearer from infection with the virus, wearing it has become a necessary and usual part of our daily life, so we see employees in various Institutions, students, patients, nurses, vendors, etc. wear masks of all shapes, colors, and sizes.

The mask, in turn, hides a large part of a person's face, as it covers the nose, mouth, and cheeks, including some that conceal the ears as well; from this standpoint, attention has focused on many issues that may be affected due to the wearing masks from a large number of people and hiding a large part of their faces.

Wearing a mask results in difficulty in recognizing the identity of the person who wears it, as this topic has raised concerns in many sectors and areas in which the identity of the person is required, such as the authentication required to unlock the phone or the attendance registration in all kinds of companies as well as in public security systems different in airports, trains, etc. [2].

Resorting to traditional unlocking operations, such as passwords and fingerprints, will increase the spread of the virus because the virus is transmitted by touch. Therefore, face detection and recognition of a person is the best way to

Manuscript received June 5, 2022
Manuscript revised June 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.6.41

unlock or authenticate to reduce the chances of the virus spreading [4].

Scientific efforts are intensifying to confront and respond to the epidemic. Attention is directed to artificial intelligence, which is based on machine and deep learning and computer vision; the field has proven a great development in the recent period in several areas such as object discovery, segmentation, image classification, detection, and recognition, so it was necessary to take advantage of these capabilities in confronting the pandemic.

During this pandemic, AI applications and solutions assisted in predicting the number of illnesses or the spread of the virus, allowing for early warning and suitable preventative actions. We cannot fail to mention here the benefit of A Deep neural networks (DNNs) algorithm in providing a large number of its models that helped in the field of image processing, recognition, and classification related to all areas of the Coronavirus and proved their great ability to give the best and most accurate results, here we refer to the most important ones, which are convolutional neural networks (CNNs) [5][6][7][8]. In addition to the latest models, the graph convolutional networks (GCNs) model.

In the proposed model, the CNN network will be implemented in work, and the method will be based on two main approaches; the first approach is to detect faces and masks, the model can support multiple faces and masks in one image. The image will show each face and the percentage of mask detection.

The second approach is to recognize the face and person's identity to grant the authentication. The recognition will be based on many algorithms starting from feature extraction, identification, and texture information collection. This phase could be challenging due to the limited features and information extracted from the image and the similarity that can be persons look alike.

We will employ these models in our help to address an issue that appeared as a result of the Coronavirus and caused many problems, namely wearing a face mask that hides people's identities. These models enable us to detect whether a person wears the mask or not, which in turn can enable governments to take penalties for non-wearers, as this system has already been implemented in France by integrating artificial intelligence algorithms with live monitoring cameras to monitor people who do not wear the facemask and imposing penalties on them [9][10].

We will also work to exploit the most important capabilities of the before mentioned models, which is considered the trends technology of our time, which is the ability to identify people's identities even with wearing a face mask (face recognition).

From this standpoint, we decided to go into research in this area to help address the problems caused by the Coronavirus pandemic and facilitate the difficulties that result from some of the laws imposed at present.

The field is the focus of researchers' attention at the current time, and here we will review some of the most important related works:

One of the limitations of the field is the lack of data sets available for study, which has prompted some researchers to collect and produce data set as [4] created three types of data sets, the Real-World Masked Face Recognition (RMFRD) data set, the largest real-time mask face data set and the MFDD data set and the Simulated Masked Face Recognition (SMFRD) data set [11]. In contrast, [12] used the same RMFD dataset plus two different SMFD and LFW datasets.

1.1 Research Problem

Facial recognition solutions identify a person by forming a unique code built on algorithms from multiple points on a person's face, including nose, chin, lips, eyes, and jaw. However, many of these key points are not visible when a person wears a mask—adding challenges to identify the persons based on the lack of facial information for the model. With this research, the researcher will tackle the problem and stress to find the best practice solution for masked face detection and recognition problems.

1.2 Research Questions

1. How to determine the best set of features be used?
2. How to select the best Deep Learning algorithm to be?
3. How to enhance the performance of the best-selected features and Deep Learning algorithms?
4. How to integrate Deep Learning algorithms and feature Extraction for detection and evaluate such integration?
5. What contribution does the classifier add to better-masked face detection and recognition than the currently implemented detection and recognition techniques?

1.3 Limitations and Delimitations of the research

Mask is known to be the most challenging obstacle to facial occlusion since a wide portion of the face, like the nose, is hidden. This dilemma was dealt with in several ways: matching, restoration, and Discard -based.

1. Matching: The goal is to compare images by using a matching method. The face image is usually sampled in a variety of similar-sized patches. Each patch will then be removed from the function. Finally, the method of matching the sensor with the gallery faces is added. The benefit of this strategy is that no overlap between sampled patches prohibits occluded areas from affecting the other information components.
2. Restoration: The occluded areas of the test faces are restored. The occluded regions are identified by setting thresholds for the 3D image's depth map values. The key factor analysis is (PCA).
3. Discard occlusion-based: To prevent a poor restoration method, these methods attempt to detect areas that the image has obscured and fully discard them. The segmentation-based approach is one of the better approaches to first detect the occluded area in the following steps using just the non-occluded portion.

1.4 Significant of the research

The study with the proposed method will support detection for multiple mask types with different mask variations. Also, the support of multiple faces in the image is due to the huge face mask images that the model trains on. The detection and recognition support different angle coverage and wide range faces.

1.5 Research Contribution

This research proposes a hybrid model using Feature selection, deep learning for Masked face detection, and recognition classification. The principal contributions of this work include:

1. Detecting and recognizing masked faces with increasing the accuracy of the model.
2. We introduce a new approach to classify masked face Images. By utilizing many recent developments in deep learning, the model we employ is much deeper than traditional neural network methods.
3. This study is also the first to compare the model's performance using deep learning algorithms, and the study lays out a descriptive visual analysis of the different algorithms.
4. Through a series of experiments, we also explore the effect of the different network structures and parameters on the performance of the deep neural network in consideration.
5. The study also explores the only Python-based deep learning library available to the users scales its capabilities and the functionalities that it provides.

2. Literature Review

This section presents the previous works and studies related to the three approaches in the proposed method, face detection, face mask detection, and face recognition. At the end of each approach, a summary of the previous approaches is presented in addition to our proposed method approach for each problem and how the proposed method solves the problems.

2.1 Face Detection

In [12] proposed a face recognition approach using deep learning and quantization techniques; they extracted features using a CNN algorithm, then Bag-of-Feature's paradigm was applied to quantize the feature maps. Finally, the Multilayer Perceptron (MLP) algorithm is used in the classification process. The results show high recognition accuracy.

In [13] introduced a procedure to recreate 3D face forms with high-faith textures without capturing a broad face texture database from single-view images. The basic concept is to finish the initial texture created with facial information from the input image using a 3D Morphable Models (3DMM) process. The results showed that the approach could outperform cutting-edge methods in qualitative and quantitative comparisons.

In [14] proposed a hybrid system that combines machine learning and deep learning techniques that detects the facemask. The system consists of two parts; the first part extracts features from the three different datasets used: RMFD, SMFD, LFW by Resnet50, while the second part is to classify the facemask using Support Vector Machines (SVM), decision trees, and ensemble algorithm. The result shows that the SVM model achieved a high accuracy rate of 100%.

The previous methods are based on a pre-trained model trained on human faces. This pre-trained model allows the classifiers to understand the human faces and their features without extracting the essential features of the human face from the classifiers.

Our proposed method will train and test human face features and extract faces with different angles and poses.

Also, the model will be able to detect multiple faces in one image.

2.2 Masked Face Detection

In [15] suggested deep learning models, such as the Inception-v3 CNN, using deep learning architectures with training parameters, such as inception-v4, Mask R-CNN, Faster R-CNN, YOLOv3, Xception, and DenseNet, to recognize face masks that detected the face masks with 99.9% accuracy.

In [16] built three types of data sets, Real-world Masked Face Recognition Dataset (RMFRD), which is the largest real-time dataset of the masked faces, Masked Face Detection Dataset (MFDD). They used the Simulated Masked Face Recognition Dataset (SMFRD) in building a face-eye-based multi-granularity model for face recognition. They achieved 95% accuracy.

In [17] built an approach consisting of two stages that detect the facemask and work to remove it using the GAN. The first stage is a map module that detects the person wearing the facemask, and the second stage is an editing module that works to complete the face after removing the mask. The system demonstrated high efficiency in detecting and completing the faces.

In [18] developed a face mask detection framework based mainly on deep learning computer vision techniques as the CNN algorithm was used to detect people who wear a facemask from people who do not wear it. The system was implemented on a camera to track the movement of people in real-time. The system achieved 96% accuracy in detecting masked faces.

In [19] proposed the EyesGAN face recognition system based on the composition of people's faces from their eyes. In the case of wearing a face mask, the system achieved a remarkable accuracy rate of 96.10% by utilizing the perceptual loss and self-attentional mechanism in GANs.

In [20] designed a residual graph Convolutional Network (RCNN) deep facial clustering system comprising more hidden layers. The algorithm k-Nearest Neighbour (kNN) is used to construct its subgraphs for each node. Then the ResNet concept was extended to CNNs, and RCNN was developed to learn how two nodes could be connected. The proposed system is more accurate and has better clustering results than other common facial clustering approaches. Furthermore, the proposed RCNN clustering method automatically detects clusters and expands to large datasets.

In [21] proposed a single camera masked face detection and identification using two approaches: a single-step pre-trained YOLO-face/trained YOLOv3 model on a set of known individuals; and a two-step process using a pre-trained one-stage feature pyramid detector network RetinaFace for localizing masked faces and VGGFace2 for generating facial feature features for efficient mask face verification. RetinaFace and VGGFace2 obtain state-of-the-art outcomes of 92.7 percent overall performance and 98.1 percent face detection in experiments.

In their proposed method, the researchers focused on two main classes, mask and non-mask, without considering covered faces that can mislead the model to predict that the face might or might not have a mask. In the proposed method, the model has been trained on the three major cases mask, non-mask, and covered faces to increase the confidence and certainty that the face or the multiple faces in the image are wearing, covers, or not wearing a facial mask.

2.3 Masked Face Recognition

In [22] proposed a statistical model of the shape and texture of the face using a set of linear basis functions. The model considers the distribution of the input features of the face and generates a high-quality shape and texture using a unified network of Graph CNN and GAN. This system achieved high-quality results for minimal and wilderness images and executed the new methods for 3D facial reconstruction.

In [23] built an MFSR dataset for a face mask to develop a face recognition system. IMAGEN and DCR loss capabilities were used to support image diversity, solve the lack of available data, and support the process of distinguishing features essential to the facial recognition process. The experiments on MFSR have shown the efficiency of the proposed method.

In [24] introduced a framework with three main components: (a) a graph-CNN lightweight, non-parametric decoder equipped to recreate coarse facial geometry from image-centric resent features in a supervised manner, (b) a subnetwork based on CNNs, is highly lightweight (35K parameters) and is trained unregulated to support the first network performance, and (c) a modern mechanism for function sampling and adaptation layer, which injects the first network's sufficient information.

In [25] implemented an ensemble facial recognition framework using a new local graphical descriptor called the Dense Local Graph Structure (D-LGS), its local graphic structure, which uses symmetric LGS. The bilinear interpolation of neighborhood pixels created the extra local graphical structure by locating the pixel dots in the other corner. These corner pixels provide the most stable elements and details related to the local image deformation. Three classifications, K-nearest neighbor, Chi-square, and coefficient of the association, are used in their method.

In [26] provided an image editing method and three types of masked face detection datasets for global masked face detection: the Correctly Masked Face Dataset (CMFD), the Incorrectly Masked Face Dataset (IMFD), and their combination (MaskedFace-Net). MaskedFace-Net was created using the Flickr-Faces-HQ3 (FFHQ) collection of face images publicly available online by NVIDIA Corporation.

In [27] proposed a reliable technique based on occlusion removal and deep learning-based features to remove the masked face region and solve the problem of masked face identification. They utilized three pre-trained deep CNN to extract deep features from the acquired regions: VGG-16, AlexNet, and ResNet-50 (mostly eyes and forehead regions). Finally, the classification procedure uses Multilayer Perceptron (MLP). Compared to other approaches, experimental results on the Real-World-Masked-Face-Dataset indicate good recognition performance.

In [28] proposed a hybrid model combining deep and traditional machine learning for face mask identification. The suggested model is made up of two parts. The first component is made to extract features using Resnet50. The second component uses decision trees, SVM, and an ensemble to classify face masks. Also, three face-masked datasets were used. In RMFD, the SVM classifier has a testing accuracy of 99.64 percent. It scored 99.49 percent testing accuracy in SMFD and 100 percent in LFW.

Face detection models were created by altering attractive images and adding masks to the original images that already recognized the characteristics of the model. The suggested technique understands and extracts the optimal features for each individual using two key features: VGG-16 extraction and CNN features, and was evaluated on the image that was not used in the training and testing procedures. The researchers' face mask detection methods were based on a pre-trained model trained on human faces, and this pre-trained model allows the classifiers to understand human faces and their features based on the pre-trained model.

The suggested technique will train and test human facial traits and extract faces from various perspectives and positions. The model will also recognize several faces in a single shot. The researchers centered their suggested strategy on two fundamental classes: mask and non-mask, excluding covered faces, which might cause the model to predict whether a face has a mask or not incorrectly.

To improve the confidence and assurance that the face or multiple faces in the image are wearing, covering, or not wearing a facial mask, the suggested approach has been trained on the three critical instances mask, non-mask, and covered faces. Face recognition research changed attractive images with masks on the original images that the model comprehended and knew the traits to recognize faces.

The suggested technique understands and extracts the optimal features for each individual using two major feature extractions, VGG-16 and CNN features, and was evaluated on images not included in the training and test processes.

2.4 Literature Review Discussion and Summary

This chapter presented various methods that have been implemented in the area of face mask detection to provide accurate results. Table 1 shows the related work summary. The Summary contains the three main approaches, Face detection, mask detection, and face masked recognition.

Table 1: Literature Review Summary

Paper title and the Authors	*Main Technique*	*Main Approach*	*Single\ Hybrid*	*Data Set*	*Accuracy %*
Our Proposed Approach	SIFT+DLBP+CNN	Face mask detection	Single	RMFD (Real World Masked Face Dataset) and Celebrities Dataset	99.1
Wu et.al [12]	R-CNN	Face Detection	Single	A human face SCALE SUBSETS	95.36
Lin et.al [13]	R-CNN	Face Detection	Single	FDDB, AFW, and WIDER FACE	91
Mohamed et.al [14]	CNN	Face Detection	Single	Face96	93.33
Hung et.al [15]	HOG-CNN	Face Detection	Hybrid	FEI, LFW, UOF	95.21
Loey et.al [16]	Machine Learning	Face Mask Detection	Hybrid	Masked Face Dataset (SMFD, RMFD)	96.2
Nagrath et.al [17]	DNN	Face Mask Detection	Single	Masked Face Dataset (SMFD, RMFD)	94.64
Chavda et.al [18]	CNN	Face Mask Detection	Single	Masked Face Dataset (SMFD, RMFD)	92.64
Mata et.al [19]	CNN	Face Mask Detection	Single	It consists of about 2000 images	93.05
Sandesara et.al [20]	CNN	Face Mask Detection	Single	RMFD (Real World Masked Face Dataset) and Kaggle Dataset	95.8
Militante et.al [21]	CNN	Face Mask Recognition	Single	Dataset collected contains 25,000 images	95
Mundial et.al [22]	CNN	Face Mask Recognition	Single	Authors collected masked faces	94
Lee et.al [23]	CNN	Face Mask Recognition	Single	Authors collected 186 test images	95.4
Hariri et.al [24]	CNN	Face Mask Recognition	Single	RMFD (Real World Masked Face Dataset) and Kaggle Dataset	91.3

3. Proposed Method

The main steps of our method are introduced as follows: 1) image is acquired from video camera and then Y CbCr color conversion is applied to prepare image for analysis, 2) the masked face is segmented from the image using Y Cb Cr color features, 3) various features in depth are extracted such as SFIT features ROIs: eyes and skin. Furthermore, DLBP texture features are computed and edge detection is applied using Non-maximal suppression of its gradient features .4) the mixed whole features need to reduced using Principle component analysis to speed up time complexity of detection method. 5) CCN will be implemented in work, and the method will be trained [25] as shown in Figure 1.

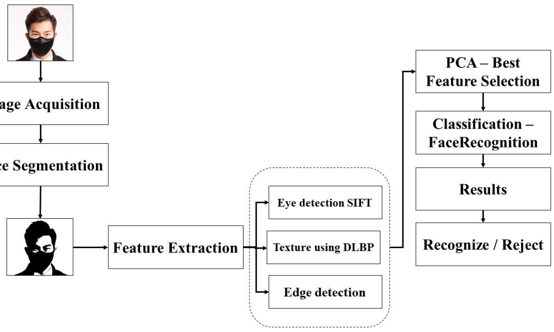
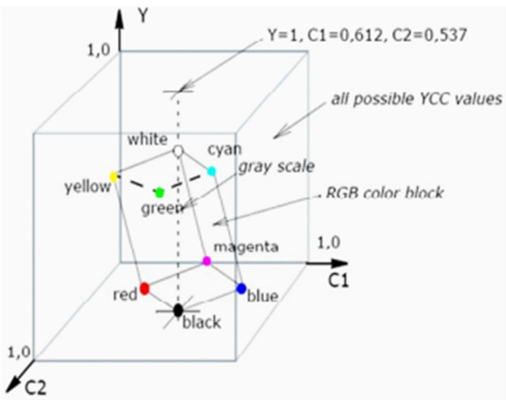


Fig. 1. Flow chart of masked face recognition.

The dataset is The real-world masked face verification dataset [26] contains 4015 face images of 426 people. The dataset is further organized into 7178 masked and non-masked sample pairs, including 3589 pairs of the same identity and 3589 different identities—the images with JPG extension and dimensions of 398*566 with 96 dpi and bit depth of 24.

3.1 Image acquisition



The first step implements RGB color conversion of image to Y Cb Cr channel for analysis. The Y channel of the input image is selected to segment the face from background as shown in Figure 2.

Fig 2. YCbCr color space.

The reason of selecting YCbCr channels is that can introduce color pixels as brightness and two-color difference channels: Cb and Cr, while RGB introduce color as three different channels R(Red), G(Green) and B(Blue). Y channel exploits the human eye perception. It also has the minimum intersection between desire ROI (segmented region) and undesired ROI under illumination variations.

3.2 Face Segmentation using Fuzzy C means (FCM)

Face segmentation contributes to a wide range of applications and has enormous promise for the computer vision paradigm; yet, face segmentation is still far from being solved, particularly in wild and unrestrained environments, and there are several remaining difficulties. There are other success stories, and some compelling work for face segmentation has been documented, particularly in controlled environments. However, the challenge of face segmentation remains unsolved in unconstrained circumstances. Several environmental elements influence the performance of an effective face segmentation system and contribute to robust face segmentation. Occlusions, variations in lighting conditions, noise in various forms, changes in facial expressions and head postures, and so on are examples. Furthermore, the number of datasets accessible for face segmentation is limited [27].

After selection of Y channel, FCM is applied to masked face image. To solve the problem of two or more target class in one target class; clusters, with training data for more than one target cluster, partitioned again using FCM. FCM is basically introduced as follows:

1. Let $S = \{s1, s2, \ldots, sN\}$ is the set of N pixels in the

Image region S and Let $\{\mu 1 \ldots \mu c\}$ is a Fuzzy C set where with $\mu_{ij} = \mu_i(S_j)$ is defined as the association ratio of the jth pixel information in the ith cluster and $\sum_{i=1}^{c} \mu_{ij} = 1$, for all $j = 1, \ldots, N$.

2. FCM compute a fuzzy c cluster (μ^*, v^*) by minimizing an objective function J as follows

$$J = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij} . d_{ij} \quad (1)$$

where d_{ij} is the similarity distance between the jth pixel information and the center of the ith cluster, $Vi = \{Vi1, Vi2, \ldots, Vp\}$, m is the index of fuzziness=1 and d_{ij} is Euclidian distance defined as the similarity measure as follows:

$$d_{ij} = \sum_{k=1}^{p} (s_{ik} - V_{ik})^{\,2} \quad (2)$$

3.3 Feature extraction using

3.3.1 Eye detection using SIFT

The scale-invariant feature transform (SIFT) is used to find feature points. It commonly corresponds to face characteristics like nostrils, although stable spots can also be detected on the user's head strap during trials. Even though there are other SIFT implementations available in Python and other programming languages, the GPU-based solution given by the OpenVIDIA library was chosen. It uses the graphics card's processing capacity to achieve a large speedup (10x) over "conventional" software versions. With an off-the-shelf Nvidia graphics chip that performed feature extraction and matching while relieving the main CPU, speeds of roughly 60 frames per second (640x480 pixels in size) were achieved [28].

3.3.2 Texture descriptor using DLBP

The size and sign of differences between a pixel and its neighbors, As a result, a pixel p(x, y) is denoted by c s and c m. The original LBP approach as shown in Figure 3 is used to code the neighborhood sign: Creating binary values based on the DD difference between the core pixel and each neighbor pixel produces c s from neighboring pixels [29]. Magnitude coding involves giving a binary value to each adjacent pixel based on an absolute depth magnitude T m threshold. The binary sequence for c m is acquired in the same manner as the binary sequence for the sign is obtained [30].

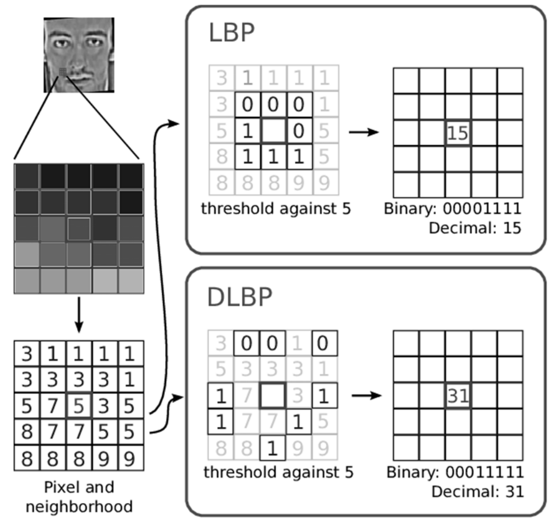


Fig 3. Local binary patterns (LBP) and DLBP.

Assigning a 0 to each neighbor pixel in n that is smaller than the center pixel c, and a 1 to each neighbor greater than c, and interpreting the result as a number in base two might be regarded as an example [31]. As a result, there are two S potential LBP values for a neighborhood of S pixels. By thresholding the value of a 3*3 neighborhood concerning its core pixel and interpreting the result as a binary integer, DLBP is a description of a pixel. In a broader sense, an LBP operator assigns a decimal number to a pair of numbers (c, n).

$$LBPb = \sum_{i=1}^{S} 2^{i-1} \, I(c, n_i) \quad (3)$$

Where c represents a center pixels, n = (n1, . . . nS) corresponds to a set of pixels sampled from the neighborhood of c according to a given pattern in image I [32].

3.3.3 Non-maximal suppression (NMS) Edge detection

NMS is defined as a process of extracting edges by edge thinning. Thinning edges could use computed gradient parameters such as magnitude and directions to extract image contours. This process helps to detect face images by extracting the dominant thin contour edges. Non-maximal suppression algorithm computes the magnitude G of gradient and direction $\boldsymbol{\theta}$ parameters via a 3×3 Sobel image gradient operator as follows:

$$G = \sqrt{G_x^2 + G_y^2} \tag{4}$$

$$G = |G_x| + |G_y| \tag{5}$$

$$\theta = \text{atan}\left(\frac{G_y}{G_x}\right) \tag{6}$$

Where $G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I$ and $G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I$

are computed at each pixel location in the image I(x, y). To perform the contour edges of an input image, read every pixel of the image. For each pixel, two of its neighbors are selected through the direction of the gradient as shown in Figure 4.

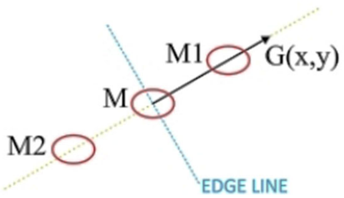


Fig 4. edge line based on gradient parameters G and θ.

As a result, the edge is detected by the condition written in Pseudo code as shown in Figure 5 and 6:

```
BEGIN
IF magnitude G is less than or equal one of the
   both neighbors M1,M2      Then
        Suppress it by making it 0 (black/background).
ELSE
        Keep it unchanged and proceed with the next one.
END
```

Fig 5. Pseudo code of finding thinner edges using NMS algorithm.

Fig. 6. a) input image b) edge detection results c) NMS thin edge detection result.

3.4 PCA – Best Feature Selection

We apply the Principal Component Analysis (PCA) technique to reduce the dimensionality of the pre-extracted multidimensional feature groups and minimize the number of features to increase data processing speed and timeliness of harmful application identification. In our method we have extracted 3 sets of features from the masked face after face segmentation task. The set of selected features are: 1) eyes features using SIFT, 2) edge features suing NMS and 3) face texture features using DLBP. These features are saved in feature vector foe each face image to represent a rich descriptor information to be used in CNN model later. The spirit of using PCA is how to best fit high-dimensional image features Si into low dimensional features in d space using a linear transformation method while preserving the main and rich descriptor information about the image as much as possible [35-37]. The theoretical concept of PCA is introduced as follows:

1. Suppose how to make a d vector $\bar{s}_0$ represent n feature of image a high dimension space, and we need to compute $\varepsilon_0(\bar{s}_0)$ sum square of distance

$$E_0(\bar{s}_0) = \sum_{i=1}^{n} \|\bar{s}_0 - \bar{s}_i\|^2 \quad (7)$$

2. Mean of n features in the image S defined by $\bar{\omega}$, that is, $\bar{\omega} = \frac{1}{n}\sum_{i=1}^{n} \bar{s}_i$, and $E_0(\bar{s}_0)$ of $\bar{s}_0$ is computed.

$$\begin{aligned} \mathrm{E}_0(\bar{s}_0) &= \sum_{i=1}^{n} \|(\bar{s}_0 - \bar{\omega}) - (\bar{s}_i - \bar{\omega})\|^2 \\ &= \sum_{i=1}^{n} \|(\bar{s}_0 - \omega)\|^2 + \sum_{i=1}^{n} \|(\bar{s}_i - \omega)\|^2 \end{aligned} \quad (8)$$

3. All features need to fit into the line direction $\bar{e}$ whose the mean value is $\bar{\omega}$ in the linear equation

$$\bar{s} = \bar{\omega} + \alpha\bar{e} \quad (9)$$

4. The projection α_i of $\bar{s}_i$ on line $\bar{e}$ is the expression of $\bar{s}_i$ in one dimentional d space, where $\bar{s}_i = \bar{\omega} + \alpha_i\bar{e}$ and θ_i is the direction between features vector $|\bar{s}_i - \bar{\omega}|$ and vector $\bar{e}$, $|\bar{e}| = 1$. It is defined as follows

$$\alpha_i = |\bar{s}_i - \bar{\omega}| \cdot \mathrm{Cosine}\ \theta_i = e^{-T}(\bar{s}_i - \bar{\omega}) \quad (10)$$

5. To find the best angle of line $\bar{e}$, we need to minimize $E_0(\bar{e})$ as follows:

$$E_0(\bar{e}) = \sum_{i=1}^{n} \alpha_i^2 \parallel \bar{e} \parallel^2 - 2\sum_{i=1}^{n} \alpha_i \bar{e}^T(\bar{s}_i - \bar{\omega}) + \sum_{i=1}^{n} \|(\bar{s}_i - \bar{\omega})\|^2 \quad (11)$$

6. updating the equation of α_i by

$$E_1(\bar{e}) = -\bar{e}^T \varepsilon\, \bar{e} + \sum_{i=1}^{n} \|(\bar{s}_i - \bar{\omega})\|^2 \quad (12)$$

where $\varepsilon = \sum_{i=1}^{n}(\bar{s}_i - \bar{\omega})(\bar{s}_i - \bar{\omega})^T$ is the divergence matrix. It can be noticed that the above equation that if $E_0(\bar{e})$ is reduced.

3.5 Convolution Neural network (CNN) training

CNN model is defined as a type of deep learning algorithms which is applied in image analysis such as face recognition, object detection and pattern recognition. It is simply having a set of convolutional layers. These layers are set of kernels which plays an important role in building a feature map of input image. Theses layers can be stacked to extract complex feature map. In our paper, we adopted a CNN model which includes three basic components: 1) Convolutional layer, 2) Pooling layers, and 3) Fully connected layer which are stacked modules with identity loops. The main task for the convolutional layers is map its' correlated inputs into the next Convolution layer based on various filters inside the layers. The outputs of the filters is computed via a Rectified Linear Unit ReLU activation function [38]. The Max Pooling function is utilized to identify the SIFT features and to reduce the size of selected features in the masked face that reduce the computational cost of the CNN model. The parameters of CNN listed in Table 2.

Table 2: The parameters of CNN.

Symbol	*Descriptions*	*Output*	*Configurations*
I	input	32×32	Segmented Image
Conv1	Convolution-1	30×30	$128, 3 \times 3$, ReLU
Conv2	Convolution-2	28×28	$128, 3 \times 3$, ReLU
MaxP1	Max-pooling-1	14×14	$128, 2 \times 2$, stride 2
Conv3	Convolution-3	12×12	$256, 3 \times 3$, ReLU
Conv4	Convolution-4	8×8	$256, 3 \times 3$, ReLU
MaxP2	Max-pooling-2	5×5	$256, 2 \times 2$, stride 2
Conv5	Convolution-5	3×3	$512, 3 \times 3$, ReLU
Conv6	Convolution-6	3×3	$512, 91 \times 1$, ReLU
AP	Average Pooling	3×3	512 filters
O	SoftMax	1×1	n classes

It used four convolution layers and Average Pooling (AP) is employed as classifier. For more details about CNN, the basic architecture of CNN is presented in Figure 7.

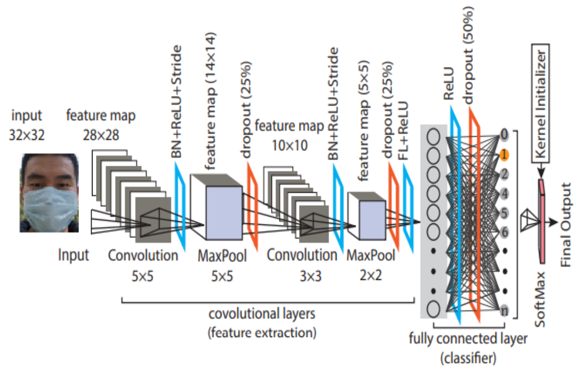


Fig 7. CNN architecture model [39].

4. Results

4.1 Data Set

Because of the recent epidemic of the COVID-19 virus around the world, people across the country wear masks, and there appear a large number of masked face samples. [25] created the world's largest masked face dataset to accumulate data resources for possible intelligent management and control of similar public safety events in the future.

Based on the masked face dataset Figure 8, corresponding masked face detection and recognition algorithms are designed to help people in and out of the community when the community is closed. Also, the upgrade of face recognition gates, facial attendance machines, and facial security checks at train stations is adapted to the application environment of pedestrians wearing masks.

Fig. 8. Dataset Samples

The dataset contains:

1. Real-world masked face recognition dataset: We crawled the samples from the website. After cleaning and labeling, it contains 5,000 masked faces of 525 people and 90,000 normal faces.
2. Simulated masked face recognition datasets: We put masks on the faces in the public face datasets and obtained the simulated masked face dataset of 500,000 faces of 10,000 subjects.
3. The real-world masked face verification dataset contains 4015 face images of 426 people. The dataset is further organized into 7178 masked and non-masked sample pairs, including 3589 pairs of the same identity and 3589 different identities.

4.2 The Implementation

The images will be loaded from the data set for the training and processing phase, and the training will be both on masked and non-masked images with multiple images for each person. The classifier with feature extraction support detects and recognizes the test images with accuracy rate and certainty. The classifier evaluation will be based on the metrics (Accuracy, Specifity, and recall). Figure 9 shows the six initial results from the proposed method. The first three results detected faces from the original image and masks with the lowest accuracy rate (98.84%). The last three images faced faces with no masks but different poses in the original image. The model detected the faces, and there is no mask been in the images with the lowest accuracy rate reached (98.26%).

The proposed model will continue to develop and reach a better accuracy rate in recognition and authentication users.

Fig. 9. Model accuracy (Face and mask detection)

4.3 The comparison

In this part, we compared the proposed model's performance with other related approaches in Table 3 and 4.

To compare the performance of these approaches, we used the common metric: Accuracy.

Table 3: Face Detection Results

Authors	*Main Technique*	*Single\ Hybrid*	*Data Set*	*Accuracy %*
The Proposed method	(SIFT + DLPB) + CNN	Single	Masked Face Dataset (SMFD, RMFD)	99.02
Wu [12]	R-CNN	Single	A human face SCALE SUBSETS	95.36
Lin [13]	R-CNN	Single	FDDB, AFW, and WIDER FACE	91
Mohamed [14]	CNN	Single	two public face datasets	91.6
Hung [15]	CNN	Single	Face96	93.33

Table4: Face Mask Detection Accuracy results

Authors	Main Technique	Single\ Hybrid	Data Set	Accuracy %
The Proposed method	(SIFT + DLPB) + CNN	Single	Masked Face Dataset (SMFD, RMFD)	96.2
Nagrath et.al.[17]	Machine Learning	Hybrid	Masked Face Dataset (SMFD, RMFD)	94.64
Chavda et.al.[[18]	DNN	Single	Masked Face Dataset (SMFD, RMFD)	92.64
Mata et.al.[[19]	CNN	Single	Masked Face Dataset (SMFD, RMFD)	93.05
Sandesara et.al.[[20]	CNN	Single	consists of about 2000 images	95.8
Militante et.al.[[21]	CNN	Single	RMFD (Real World Masked Face Dataset) and Kaggle Dataset	96
Mundial et.al.[[22]	CNN	Single	Authors private dataset	94
Lee [23]	CNN	Single	Authors collected 186 test images	93.4
Hariri [24]	CNN	Single	authors collected masked faces	91.3

5. Conclusion

The main task of our method is to detect masked face based on combined features extracted from the segmented image and deep learning CNN model. The selected features represent a rich information helps to detect face. We consider the points in the image detected by Scale invariant features such as eyes. Then, we extract Local Binary Pattern (DLBP) texture features and used it into a novel model to detect the masked face. Later, the proposed method employed edge detection and saved these features into a feature vector combined with a deep learning models for masked face recognition. Significant experiments are implemented for verifying the proposed method. Evaluation results have established which the proposed method outperformes several state-of-the-art masked face detection methods.

6. References

[1] Al_azrak, F. M., Elsharkawy, Z. F., Elkorany, A. S., el Banby, G. M., Dessowky, M. I., El-Samie, A., & Fathi, E. (2020). Copy-move forgery detection based on discrete and SURF transforms. Wireless Personal Communications, 110(1), 503–530.

[2] Arefi, M. F., & Poursadeqiyan, M. (2020). A review of studies on the COVID-19 epidemic crisis disease with a preventive approach. Work, 66(4), 717–729.

[3] Aswal, V., Tupe, O., Shaikh, S., & Charniya, N. N. (2020). Single Camera Masked Face Identification. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 57–60.

[4] Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., Chapman, B., Amrhein, T., Mong, D., Rubin, D. L., & others. (2019). Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. Artificial Intelligence in Medicine, 97, 79–88.

[5] Bianchi, F. M., Grattarola, D., Livi, L., & Alippi, C. (2021). Graph neural networks with convolutional arma filters. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[6] Cabani, A., Hammoudi, K., Benhabiles, H., & Melkemi, M. (2021). MaskedFace-Net–A dataset of correctly/incorrectly masked face images in the context of COVID-19. Smart Health, 19, 100144.

[7] Chen, J., Hachem, E., & Viquerat, J. (2021). Graph neural networks for laminar flow prediction around random two-dimensional shapes. Physics of Fluids, 33(12), 123607.

[8] Cheng, S., Tzimiropoulos, G., Shen, J., & Pantic, M. (2020). Faster, better and more detailed: 3d face reconstruction with graph convolutional networks. Proceedings of the Asian Conference on Computer Vision.

[9] Cho, S. W., Baek, N. R., Kim, M. C., Koo, J. H., Kim, J. H., & Park, K. R. (2018). Face detection in nighttime images using visible-light camera sensors with two-step faster region-based convolutional neural network. Sensors, 18(9), 2995.

[10] Christlein, V., Spranger, L., Seuret, M., Nicolaou, A., Král, P., & Maier, A. (2019). Deep generalized max pooling. 2019 International Conference on Document Analysis and Recognition (ICDAR), 1090–1096.

[11] Lee, G.-H., & Lee, S.-W. (2020). Uncertainty-aware mesh decoder for high fidelity 3d face reconstruction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6100–6109.

[12] Wu, W., Yin, Y., Wang, X., & Xu, D. (2018). Face detection with different scales based on faster R-CNN. IEEE Transactions on Cybernetics, 49(11), 4017–4028.

[13] Lin, J., Yuan, Y., Shao, T., & Zhou, K. (2020). Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5891–5900.

[14] Mohamed, S. S., Mohamed, W. A., Khalil, A. T., & Mohra, A. S. (2020). Deep learning face detection and recognition. International Journal of Advanced Science and Technology, 29(2), 1–6.

[15] Hung, B. T. (2021). Face recognition using hybrid HOG-CNN approach. In Research in Intelligent and Computing in Engineering (pp. 715-723). Springer, Singapore.

[16] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021a). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement, 167, 108288.

[17] 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.

[18] Chavda, J. Dsouza, S. Badgujar and A. Damani, "Multi-Stage CNN Architecture for Face Mask Detection," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-8, doi: 10.1109/I2CT51068.2021.9418207.

[19] Mata, B. U., & others. (2021). Face Mask Detection Using Convolutional Neural Network. Journal of Natural Remedies, 21(12 (1)), 14–19.

[20] Sandesara, A. G., Joshi, D. D., & Joshi, S. D. (2020). Facial Mask Detection Using Stacked CNN Model. Int. J. Sci. Res. Comput. Sci. Eng. Inform. Technol.

[21] Militante, S. v, & Dionisio, N. v. (2020). Real-time facemask recognition with alarm system using deep learning. 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), 106–110.

[22] Mundial, I. Q., Hassan, M. S. U., Tiwana, M. I., Qureshi, W. S., & Alanazi, E. (2020). Towards facial recognition problem in COVID-19 pandemic. 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), 210–214.

[23] Lee, S. H. (2020). Deep learning based face mask recognition for access control. Journal of the Korea Academia-Industrial Cooperation Society, 21(8), 395–400.

[24] Hariri, W. (2021). Efficient masked face recognition method during the covid-19 pandemic. ArXiv Preprint ArXiv:2105.03026.

[25] Din, N. U., Javed, K., Bae, S., & Yi, J. (2020). A novel GAN-based network for unmasking of masked face. IEEE Access, 8, 44276–44287.

[26] Geng, L., Zhang, S., Tong, J., & Xiao, Z. (2019). Lung segmentation method with dilated convolution based on VGG-16 network. Computer Assisted Surgery, 24(sup2), 27–33.

[27] Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., & Grieve, B. (2018). Towards on-farm pig face recognition using convolutional neural networks. Computers in Industry, 98, 145–152.

[28] Kolar, Z., Chen, H., & Luo, X. (2018). Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. Automation in Construction, 89, 58–70.

[29] Kumar, D., Garain, J., Kisku, D. R., Sing, J. K., & Gupta, P. (2020). Unconstrained and constrained face recognition using dense local descriptor with ensemble framework. Neurocomputing, 408, 273–284.

[30] Lin, K., Zhao, H., Lv, J., Li, C., Liu, X., Chen, R., & Zhao, R. (2020). Face detection and segmentation based on improved mask R-CNN. Discrete Dynamics in Nature and Society, 2020.

[31] Liu, K., Kang, G., Zhang, N., & Hou, B. (2018). Breast cancer classification based on fully-connected layer first convolutional neural networks. IEEE Access, 6, 23722–23732.

[32] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021b). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement, 167, 108288.

[33] Mukherjee, S., Boral, S., Siddiqi, H., Mishra, A., & Meikap, B. C. (2021). Present cum future of SARS-CoV-2 virus and its associated control of virus-laden air pollutants leading to potential environmental threat–a global review. Journal of Environmental Chemical Engineering, 9(2), 104973.

[34] Pan, Y., & Zhang, L. (2021). Dual attention deep learning network for automatic steel surface defect segmentation. Computer-Aided Civil and Infrastructure Engineering.

[35] Qi, C., Zhang, J., Jia, H., Mao, Q., Wang, L., & Song, H. (2021). Deep face clustering using residual graph convolutional network. Knowledge-Based Systems, 211, 106561.

[36] Ren, Y., Huang, J., Hong, Z., Lu, W., Yin, J., Zou, L., & Shen, X. (2020). Image-based concrete crack detection in tunnels using deep fully convolutional networks. Construction and Building Materials, 234, 117367.

[37] Ryumina, E., Ryumin, D., Ivanko, D., & Karpov, A. (2021). A NOVEL METHOD FOR PROTECTIVE FACE MASK DETECTION USING CONVOLUTIONAL NEURAL NETWORKS AND IMAGE HISTOGRAMS. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

[38] Saadat, S., Rawtani, D., & Hussain, C. M. (2020). Environmental perspective of COVID-19. Science of the Total Environment, 728, 138870.

[39] Sanz, H., Valim, C., Vegas, E., Oller, J. M., & Reverter, F. (2018). SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinformatics, 19(1), 1–18