Effective approach of face mask position detection and recognition

Om Pradyumana Gupta1,*, Arun Prakash Agrawal2, Om Pal3

1 Department of Computer Science & Engineering, Sharda University, Greater Noida, Uttar Pradesh 201 306, India
2 Department of Computer Science & Application, Sharda University, Greater Noida, Uttar Pradesh 201 306, India
3 Department of Computer Science, University of Delhi, New Delhi 110 007, India

*Corresponding author: Om Pradyumana Gupta, op.gupta@nic.in

ABSTRACT: During recent COVID-19 pandemic across the world, face masks became necessary to stop the spread of infection. This has led to challenges with effective detection and recognition of human faces using the existing face detection systems. This paper proposes a Convolutional Neural Network (CNN) based face mask recognition system, which offers two solutions—recognition of the person wearing face mask and position of face mask i.e., whether the mask is correctly worn or not. The proposed model could play instrumental role of face recognition. In the first stage, with the help of Viola-Jones algorithm, the model detects the position of the face mask. In the second stage, we identify the person with a modified pre-trained face mask recognition DeepMaskNet model facilitates in identifying the person. The proposed model achieves an accuracy of 94% in detecting the face mask position and 99.96% in identifying the masked person. Lastly, a comparison with the existing models is detailed, proving that the proposed model achieves the highest greater performance.

KEYWORDS: DeepMaskNet; Viola-Jones; Convolutional neural network; Pooling layers; Softmax

1. Introduction

People must wear face masks in public places all around the world to reduce the impact of Covid 19. Controlling the spread of COVID-19 has been currently a major problem for WHO policymakers and the entire human race. Wearing a face mask lowers the spread of COVID-19 by lowering the likelihood of respiratory droplets being transmitted, according to most data from the WHO. Droplets can transmit coronavirus. The proper usage of face masks during the pandemic is essential to limiting and preventing the spread of COVID-19 amongst individuals[1,2]. To stop the spread of COVID-19, many nations mandate the use of face masks in public settings. It is necessary to develop automated techniques for spotting the face masks. Additionally, wearing a face mask presents unique difficulties for conventional facial recognition software, which is normally made for faces that are exposed. The failure of facial recognition techniques when used with face masks has created considerable difficulties for application verification and authentication. Additionally, the COVID epidemic has inhibited the adoption of several conventional biometric-based techniques, including fingerprint recognition. Despite the fact that recent research has suggested useful techniques for masked facial recognition, masks still greatly obscure a face[3] and research in detection of proper use of face mask is limited. Although there has been tremendous research in the domain of face mask identification[4-7]. This paper presents a unique approach to not only identify the person wearing the face mask but also inform whether the mask is worn properly or not. One
of the most effective ways to stop transmission is to correctly wear face masks. Such an automatic device may be installed at busy public gates to help monitor and prevent people entering without masks or with the wrong kind of mask on.

The paper is structured as follows. In the next section, Literature Review is discussed. Methodology is detailed in Section 3. The Dataset is described in Section 4. Section 5 is the Experimental Setup, Section 6 is Results, and Section 7 is Conclusion.

2. Literature review

In 2022, Rahman et al.\cite{8} and Venkateswarlu et al.\cite{9} developed face mask position identification system using MobileNet. They have created a system that can identify different face mask placement mistakes. They achieved this by including additional layers, such as dropout, and making more advantageous choice of the number of thick layers, which significantly improved the classification accuracy. In 2021, Ullah et al.\cite{10} is able to both detect face masks and recognize masked faces and suggests a revolutionary DeepMasknet framework that can do both. Because there is no dataset of masked people, they create their own. In 2020, Fan and Jiang\cite{11} developed model based on RetinaNet architecture to check whether a face mask is present. They extracted characteristics from photos using ResNet-50. In 2020, Vijitkunsawat and Chantngarm\cite{12} determines the best model to normalize the face masks, two conventional Machine Learning classifiers one Deep Learning algorithm were discussed. Kumar et al.\cite{13} and Mundial et al.\cite{14} discuss how the tendency to fail in the detection of the methods used by the facial recognition systems have led to many challenges for the verification of true identities in many applications. However, Qin and Li\cite{15} devised a method to automatically tell the placement of the mask on the face. Zhang et al.\cite{16} proves that traditional techniques such as Support Vector Machines (SVM) take longer and consume more space than state-of-the-art models do. Chowdary et al.\cite{17} discusses the robustness and effectiveness of DeepMaskNet model in facial recognition.

3. Proposed methodology

The proposed methodology is divided in to two phases. The System flow chart is presented in Figure 1.

![Figure 1. Process flow.](image)
In the first phase, face mask position is detected with the help of Viola-Jones algorithm\cite{18,19}, and an image is classified under i) Nose Uncovered (NU), ii) Nose & Mouth Uncovered (NMU), or iii) No Mask (NM). The classification is done sequentially, and a message is prompted in case the face mask is not properly worn. In case the mask is properly worn, the model enters the second stage in which identification of the person is done. For this purpose, a modified DeepMaskNet CNN is deployed that helps in identification of images of 50 persons.

3.1. Phase 1—Face mask position detection with Viola-Jones algorithm

Ensuring masks are properly worn has been a challenge. The proposed model categorizes images into the following:
1) Nose Uncovered (NU)
2) Nose and Mouth Uncovered (NMU)
3) No mask (NM)

Viola Jones algorithm uses Haar features to detect the presence of a feature in that image. The Haar feature is calculated by running rectangular window (with black and white rectangles) on an image. Haar features are shown in Figure 2. The black rectangle has a value of −1 and white rectangle has a value of +1. The result is calculated by adding the pixels in both black and white rectangles and then subtracting the white rectangle from the black rectangle. This computation can be very slow and around 160,000 features (with 24 × 24 window) and can be reduced with the help of an integral image.

The integral image helps in a faster calculation of applying the Haar features. The number of pixels (on the left & above to the point \((x, y)\)) in the integral image at one location is calculated as follows:

\[
\text{int}_\text{img}(x, y) = \sum_{x', y' \leq (x, y)} \text{img}(x', y')
\]  

(1)

where \(\text{int}_\text{img}(x, y)\) is the integral image and \(\text{img}(x, y)\) is the original image.

\[
r(x, y) = r(x, y - 1) + \text{img}(x, y)
\]  

(2)

\[
\text{int}_\text{img}(x, y) = \text{int}_\text{img}(x-1, y) + r(x, y)
\]  

(3)

where \(r(x, y)\) is the number of total rows.

In Figures 3 and 4, the calculation of the integral image is demonstrated pictographically.

![Figure 2. Different types of Haar features in Viola-Jones algorithm.](image)

![Figure 3. Integral image computation.](image)
Figure 4. Input image (left) to integral image (right).

The cascade classifier algorithm was employed by Viola and Jones. Stratified classification is a feature of this cascade classification technique.

This algorithm is divided into multiple stages. To attain a low false positive rate, the first stage calls for the deployment of more sophisticated classifiers after a basic classifier has rejected the majority of the sub-windows. Because of this, the classifier set is a potent suite capable of classifying faces. As illustrated in Figure 5, the general form of the detection process is a degraded decision tree, sometimes known as a “cascade”. When the first classifier yields a positive result, the second classifier—which has been tuned to have an extremely high detection rate—is evaluated. The third classifier is triggered by a positive result from the second classifier, and so on. If the outcome is negative at any stage, the sub-window is rejected outright.

Figure 5. Cascading classifier.

Using AdaBoost machine learning to train classifiers and modifying the threshold value to reduce false negatives, the cascade stage is constructed. Keep in mind that the goal of this AdaBoost threshold is to minimize error rates in training data. Figure 5 depicts the explanation of the cascade categorization. AdaBoost constructs a strong classifier with a linear combination of weak classifiers as shown in Equation (4).

\[ F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \cdots + \alpha_n f_n(x) \]  

3.2. Phase 2—Recognition with modified DeepMaskNet

Once the image passes the face mask position check, it is evaluated for face masked person identification. A modified DeepMaskNet CNN model is deployed. Modifications to the Size and Stride were made to enhance the performance of the model. The model has ten layers and can process input images with a resolution of 256 × 256. The distribution of the input to each layer changes during the training. For instance, more surgical masks are present in the dataset as compared to other types of masks that causes parameter training to become extremely time-consuming and requires better initialization. In the DeepMaskNet design, we applied batch normalization to address this covariate shifting and used the LeakyReLU activation function to overcome the dying ReLU problem. The details of the modified DeepMaskNet model are presented in Table 1 and hyperparameter details in Table 2.
Table 1. Layers and architecture of modified DeepMaskNet model.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Layer</th>
<th>Filters</th>
<th>Size</th>
<th>Stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Input</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Convolutional-1</td>
<td>128</td>
<td>3 × 3</td>
<td>1 × 1</td>
</tr>
<tr>
<td>3</td>
<td>Max Pooling</td>
<td></td>
<td>3 × 3</td>
<td>2 × 2</td>
</tr>
<tr>
<td>4</td>
<td>Convolutional-2</td>
<td>512</td>
<td>3 × 3</td>
<td>1 × 1</td>
</tr>
<tr>
<td>5</td>
<td>Max Pooling</td>
<td></td>
<td>3 × 3</td>
<td>2 × 2</td>
</tr>
<tr>
<td>6</td>
<td>Convolutional-3</td>
<td>384</td>
<td>3 × 3</td>
<td>1 × 1</td>
</tr>
<tr>
<td>7</td>
<td>Max Pooling</td>
<td></td>
<td>3 × 3</td>
<td>2 × 2</td>
</tr>
<tr>
<td>8</td>
<td>Convolutional-4</td>
<td>256</td>
<td>3 × 3</td>
<td>1 × 1</td>
</tr>
<tr>
<td>9</td>
<td>Convolutional-5</td>
<td>256</td>
<td>3 × 3</td>
<td>1 × 1</td>
</tr>
<tr>
<td>10</td>
<td>Convolutional-6</td>
<td>256</td>
<td>3 × 3</td>
<td>1 × 1</td>
</tr>
<tr>
<td>11</td>
<td>Max Pooling</td>
<td></td>
<td>3 × 3</td>
<td>2 × 2</td>
</tr>
<tr>
<td>12</td>
<td>Fully Connected + LeakyRelu + Dropout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Fully Connected + LeakyRelu + Dropout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Fully Connected + LeakyRelu + Dropout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Fully Connected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Softmax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Classification</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Hyperparameters of DeepMaskNet Architecture.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Rate</td>
<td>0.01</td>
</tr>
<tr>
<td>epoch</td>
<td>50</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
</tbody>
</table>

The output is passed to a 50-way softmax function for 50 persons.

4. Dataset

We created an internal dataset because there wasn’t a uniform dataset for mask position detection and facial recognition covered by a mask. Different dataset requirements apply to each task. Face mask position requires photographs of several people wearing masks both correctly and improperly, whereas face recognition covered by a mask is trained on multiple images of the same person wearing a mask.

We have divided the data in the ratio of 80:20 for training and test data.

For face-mask position detection, since the Viola-Jones algorithm employed in this study has trained over 10,000 photos of human parts, including the face, nose, mouth, and eye, in a detection phase, training images was not needed\(^{[20]}\).

For masked face recognition, we have developed a database of 10,397 images of 50 people using image augmentation\(^{[21]}\). Table 3 details the image database.
5. Experiment setup

As part of the experimental setup, Viola-Jones algorithm and DeepMaskNet are primarily used. Apart from DeepMaskNet, python libraries such as Imutils, numpy, dlib, random, OpenCV2 etc. are used for face detection, pre-processing and face cropping.

NVIDIA GTX 1070 GPU was used to develop the model.

6. Results

The combination of DeepMaskNet and Skin detection method helps us to reach a higher accuracy. Moreover, Skin detection method is used instead of a traditional face detection method to reduce the amount of computational processing required.

Table 4 and Figure 7 depict the result on test data on our proposed model, DeepMaskNet only, ResNet 18, DenseNet, and DarkNet 53.

### Table 4. Comparison of proposed model with standard models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Model – Viola-Jones + modified DeepMaskNet</td>
<td>99.96</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>DeepMaskNet</td>
<td>98.08</td>
<td>98.51</td>
<td>99.25</td>
<td>98.87</td>
</tr>
<tr>
<td>ResNet 18</td>
<td>99.12</td>
<td>98.33</td>
<td>100</td>
<td>99.15</td>
</tr>
<tr>
<td>DenseNet</td>
<td>99.12</td>
<td>99.83</td>
<td>98.52</td>
<td>99.17</td>
</tr>
<tr>
<td>DarkNet 53</td>
<td>99.51</td>
<td>98.48</td>
<td>97.02</td>
<td>97.74</td>
</tr>
</tbody>
</table>

![Comparison of the performance of the proposed model](image)

**Figure 7.** Confusion matrix for face mask position.
7. Conclusion

The proposed model achieves an accuracy of 94% in detecting the face mask position and 99.96% in identifying the masked person Therefore, we conclude, that with the help of the above model, we can improve the accuracy of the exiting pre-trained models which will help all the face recognition systems deployed. Automated face recognition systems have multiple use cases such as attendance systems, authentication systems, accessibility systems etc. Moreover, the proposed study is completed with the incorporation of face mask position detection that adds to the list of use cases—controlling entry to public places, narrowing the source of spread of virus, identifying defaulters at schools, universities, offices and other areas. The proposed model faces challenges in detecting different types of masks. This can be improved by adding more images of people wearing different types of masks in the training dataset. Secondly, computational heavy but more accurate algorithms such as You-Only-Look-Once (YOLO)[22] can be deployed for face mask position detection.

Author contributions

Conceptualization of the whole research, OPG, APA and OP; methodology, software, OPG; validation part, APA; dataset, OPG, APA and OP; review and editing of the paper, OP. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References


