Mask Use Detection in Public Places Using the Convolutional Neural Network Algorithm

Mochamad Yoga Wibowo, Hanny Hikmayanti*, Anis Fitri Nur Masruriyah, Elfina Novalia, Nono Heryana

1 Department of Informatics Engineering, Universitas Buana Perjuangan Karawang, 41361, Indonesia
2 Department of Informatics Engineering, Universitas Buana Perjuangan Karawang, 41361, Indonesia
3 Department of Informatics Engineering, Universitas Buana Perjuangan Karawang, 41361, Indonesia
4 Department of Information System, Universitas Singaperbangsa Karawang, 41361, Indonesia
5 Department of Information System, Universitas Singaperbangsa Karawang, 41361, Indonesia

hanny.hikmayanti@ubpkarawang.ac.id

This article contributes to:

1. Introduction

In early 2020, the discovery of a new virus called coronavirus rocked the healthcare community. Almost the whole world is affected by this virus outbreak; In Wuhan, China, COVID-19 was first discovered in December of this year. [1]. SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) and the coronavirus disease 2019 were named by the World Health Organization on 11 February 2020 (COVID—19). WHO has also designated this virus as a world pandemic because it has spread to several countries. Since the publication of the WHO statement, the Indonesian government has issued policies to control the spread of COVID-19 [2]. The policy aims to break the chain of transmission of COVID-19 and is made based on policies from WHO, including washing...
hands, donning masks, keeping a safe distance, avoiding crowds, and limiting movement [3]. Problems occur because many people violate the 5M policy, one of which is wearing a mask. Masks are very important to limit the transmission of COVID-19, because this virus can be transmitted through droplets containing the SARS-CoV-2 virus and enter the body, causing risks. Seeing this spread, COVID-19 also has common characteristics where a person is known for this disease, namely, fever, cough, diarrhoeal exhaustion, and unusual dyspnea. There are documented patients with asymptomatic COVID-19, in addition to the rising number of COVID-19 patients. Patients with asymptomatic COVID-19 frequently dispute having an infection or question the accuracy of the test [4]. So this study, made detection of using masks and not using masks on the face. The goal is to be able to make people more comply with regulations and break the chain of transmission of Covid-19 [5] [6] [7].

Previously, research was conducted by Lattore [8], identifying masks with a dataset labeled in the CNN format. Furthermore, the data is processed using the Convolutional Neural Network (CNN) algorithm, resulting in an accuracy of up to 98% with a mask and 97% without using a mask. Then Chong et al [9], conducted research on android-based mask detection, the initial step in this research was knowledge acquisition which collected types of masks. Then, proceed with the representation of knowledge before it is modeled into a mathematical calculation formula. So, later it will proceed to the final stage, namely the Convolutional Neural Network. The results of the accuracy obtained in this study amounted to 90%. Furthermore, research on the implementation of CNN for detecting the use of masks was carried out by Erdem [10], the research process began with the use of literature and collection of datasets in the form of a group of human faces totaling 1000 images divided into two types namely 500 wearing masks and 500 others not wearing masks. The results obtained were successful in implementing the CNN algorithm in the form of an application to properly classify mask user detectors.

The detection of using masks and not using masks on faces uses the Convolutional Neural Network algorithm with the architectural preprocessing technique used, namely MobileNetV2 [11] [12]. Research in the field of image classification often occurs because this issue is interesting to debate. In this research, several processes were carried out, such as collecting image data sets obtained from the Kaggle website. After the images were collected, preprocessing was carried out so that the images were uniform in size 224 x 224 pixels. Then, the data set resulting from preprocessing is carried out by a split validation process by dividing the data set. 80 training data and 20 validation data, then augmentation is also carried out to increase the number of images such as rotating, zooming, cropping, shifting and flipping randomly but not eliminating the essence or essence of the data [13]. The training is carried out with the tensorflow library and testing is carried out to obtain results from the images that have been trained. Tests on this detection were carried out with several scenarios from several angles to find out whether all angles were detected, the accuracy obtained from model training was 99%, testing was 89% and the confidence value in the bounding box was 100%.

2. Methods

The data set was obtained from the Kaggle website with 1924 images wearing masks and 1933 images without masks. Then, the data that has been collected is stored in a structured folder. Next, the preprocessing stage is carried out on the image by changing the size to 224x224 pixels. At the feature extraction stage, the data is the image of a face wearing a mask, not a show. The application of the Convolutional Neural Network method is used to distinguish faces wearing masks and not wearing masks. The stages of the CNN method can be seen in Figure 1.
The Multi-Layer Perceptron was developed into the Convolutional Neural Network Algorithm with the same processing, namely two-dimensional input. The Convolutional Neural Network technique has a deep network and is often used for image recognition data. Thus, it can provide a high level of accuracy and good results [14]. The core function of the CNN algorithm is that it is used to do direct categorization or classification work through photos, video, text, or sound. The types of CNN layers that are still being developed in the future consist of feature learning and classification layers. Both of these have their respective roles and performance, which are very good for solving machine learning problems [15] [16]. Especially data in the form of images. This was examined by the CNN architecture, which explained that it was actually similar to the pattern of neurons or nerve cells in the human brain. The role is to process various kinds of information in a visual form [17].

In Figure 1, the stages of the CNN method are divided into two main layers, namely feature learning and classification [18]. At the feature learning stage, it has a feature extraction layer which consists of a convolution layer and a pooling layer. The convolution layer will carry out the process of moving a convolution filter of a certain size into an image. This convolution process is an important part of CNN, the result of a convolution is used as input to produce a feature representation. The next step is to apply the ReLU (Rectifier Linear Unit) activation function to get the output from the convolution and make the output non-linear. Furthermore, this pooling layer process helps to reduce the spatial size of the convolution features, thus minimizing the calculations required in order to speed up computation and during the training process is faster [19] [20].

A popular pooling technique called max pooling seeks to identify the highest value at each filter shift. In this research, the average value was calculated using average pooling. Each neuron that was obtained and extracted during the feature learning stage is then classified during the next step, classification. There are three stages in the classification process: flatten, completely connected, and softmax. In the flatten step, the feature map is converted into a vector so that it can be used as input for fully connected. Furthermore, at the fully connected layer stage where all activated neurons from the previous layer will be connected to all neurons in the next layer. The last step, namely softmax, calculates the possible value of each class which helps to determine the class of the given input. The range of values in probability is 0 to 1.

The data set resulting from the preprocessing will be used for the split validation process using the skit-learn library. The goal is that the resulting model provides maximum and accurate results. Then, augmentation or data manipulation techniques are also carried out without losing the essence or essence of the data. For example, the dataset is rotated, zoomed, cropped, shifted and flipped randomly. The purpose of this process is to reproduce the dataset as well as increase the accuracy of the model. The results of this process will be used as model training using the tensorflow library. The test was carried out in several scenarios from different angles, at an angle of 45 degrees with the face turned upwards; the detection of this mask could not be detected. Whereas at other angles all are detected. In the testing process the convolution matrix method is used, which is a method used to measure the performance of the classification method. There is an equation to find accuracy and precision in the convolution matrix method, using the following Equation (1):

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%
\]

The confusion matrix is a technique for evaluating the performance of a classification model in machine learning and deep learning. The confusion matrix is represented by a table containing 4 values, namely, true positive (TP), true negative (TN), false positive (FP) and false negative (FN). True Positive (TP) represents a condition where the model can correctly predict the positive class. True Negative (TN) is the same as True Positive (TP) except that what is predicted is the negative class. In contrast to True Positive (TP) and True Negative (TN), False Positive (FP) is a condition in which the model justifies the prediction of an object when in fact it is not, while False Negative (FN) is a condition in which the model gives predictions that are incorrect even though the actual true value. The level of accuracy is one performance that can be measured using a confusion matrix shown in Figure 2.

---

*Figure 2. Confusion Matrix*
3. Results and Discussion

The data set obtained from the Kaggle website is made into a folder with a mask and without a cover; this folder will later be used at the preprocessing stage so that the size of all images is uniform to 224×224 pixels. Furthermore, it is converted into an array form, because computers do not have human-like capabilities that are able to recognize an object. The computer can only interpret the data included in the image. Preprocessing results are shown in Figure 3.

![Preprocessing Results](image3)

The results obtained accurately in the training process using data from split validation and augmentation can be seen in Figure 4. The image has accuracy which is the ratio of True (positive and negative) predictions to the entire data, then recall or the ratio of successful positive predictions to the total number of positive observations correct. Then, the f1-score is a comparison of the average precision and recall weights.

<table>
<thead>
<tr>
<th></th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>dengan_masker</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>386</td>
</tr>
<tr>
<td>tanpa_masker</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>387</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>773</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>773</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>773</td>
</tr>
</tbody>
</table>

![Model Training Accuracy Results](image4)

After the training process is carried out, of course it will produce the best CNN model, then the model will be tested. The goal is to determine the performance of the model. In the test process, a test was carried out on video data (real time) via a webcam. The prediction results from video data in real-time using a webcam can be seen in Figure 5, the detected face is not wearing a mask and Figure 6 is detected using a mask.

![Predictive Results Without Masks](image5)

![Predictive Results With Masks](image6)
Furthermore, tests were carried out from various angles to see the results of detecting masks and not using masks, whether they were successful or not. In the testing process with several scenarios, the results are obtained:

- TP (True Positive) = 16
- TN (True Negative) = 0
- FP (False Positive) = 2
- FN (False Negative) = 0

In the test that has been carried out 16 times. So the percentage of testing accuracy in identifying wearing and not wearing masks is 89%. The results are as follows.

Accuracy = \(\frac{TP + TN}{TP + FP + TN + FN}\) × 100%

= \(\frac{16 + 0}{16 + 2 + 0 + 0}\) × 100%

= \(\frac{16}{18}\) × 100%

= 89%

This test was carried out with 16 angle scenarios, including angles of 15 degrees and 45 degrees with the face looking up, right side, left side, down and front. At an angle of 45 degrees with the face turned upwards the detection of the mask and without the mask cannot detect it.

4. Conclusion

Based on the results of testing and analysis of mask detection using the Convolutional Neural Network (CNN) algorithm, this research has succeeded in getting the conclusion that this research succeeded in building a face identification model that uses a mask and does not use a show with the Convolutional Neural Network algorithm. This study succeeded in testing facial identification using a mask and not using a mask with the Convolutional Neural Network algorithm with several scenarios. The model made has an accuracy of 99%. Mask and no mask detection testing with several scenarios resulted in an accuracy of 89%. The value of confidence in the bounding box gets 100% accuracy. As for suggestions in this study, namely being able to develop models into compact forms such as CCTV, alarms and others and being able to make mask and non-mask detection detectable from any angle.
Authors' Declaration

Authors' contributions and responsibilities – The authors made substantial contributions to the conception and design of the study. The authors took responsibility for data analysis, interpretation, and discussion of results. The authors read and approved the final manuscript.

Funding – No funding information from the authors.

Availability of data and materials – All data are available from the authors.

Competing interests – The authors declare no competing interest.

Additional information – No additional information from the authors.

References


