Facemask Detection Using Convolutional Neural Networks

D. De Luca, H. Alsurayhi,

***Abstract*. The Coronavirus Disease (COVID-19) has rapidly spread, gripping the world in a pandemic. The World Health Organization (WHO) has proposed measures to halt the airborne spread of the virus such as wearing a facemask. Wearing the facemask is important to slow the spread of the virus, however the enforcement of this rule has posed to be difficult. Thus, we are developing a facemask detection system to detect if a facemask is being worn, not worn, or worn incorrectly. We chose to use a convolutional neural network (CNN) architecture to correctly identify the three classes. Our methodology is as follows: data pre-processing, image cropping for facial detection, data augmentation for class imbalance, and facemask classification. We are training our model on a dataset containing 3275 images belonging to the three classes. We evaluated our model using the following metrics: accuracy, precision, recall, and F1 score. Our model can accurately detect the classes facemask is worn and facemask not worn.**

Keywords—facemask, detection, CNN

I. Introduction

SINCE 2020, the world has been under a pandemic due to the spread of COVID-19. According to the World Health Organization (WHO), more than 5 million deaths have been reported [1]. In attempts to try slowing down the spread of the virus, the WHO has put many measures in place such as wearing a facemask to minimize the airborne spread of the virus. However, in public spaces or large workplaces, it is impractical to manually monitor if individuals are following the rules and wearing facemasks correctly. A better alternative is to use Machine Learning and Deep Learning to automatically detect instances of improper use of facemasks. Thus, we propose a facemask detection model using images captured from different public spaces. Our dataset has three classes: *facemask is worn, not worn, or worn incorrectly.*

To build this model, we used a convolutional neural network (CNN) to extract features from the image data. The CNN is considered the state-of-the-art for computer vision and image classification, making their use suitable for our project. The model is trained to identify subjects who are wearing a mask, not wearing a mask, or wearing a mask incorrectly. The paper is outlined as follows: Section 2 states some of the related work in the field of deep learning for facemask detection. Section 3 describes the dataset used in this project. Section 4 explains our CNN model structure and parameters and the evaluation metrics. The results of the proposed model are presented in Section 5. Finally, Section 6 states the conclusion and the future work.

II. Related Works

Over the duration of the COVID-19 pandemic, similar work has been carried out for the detection of wearing facemasks in public areas. G. Jignesh Chowdary et al. [2] designed a facemask detection model using transfer learning of InceptionV3, achieving a training accuracy of 99.9% and testing accuracy of 100%. However, their model performs binary classification only. Most of the face-masked samples used are synthetic mask that is applied on the images.

In another related work, A. Chavda et al. [3] used a dual-stage CNN architecture to detect if a person is wearing a facemask or not. In the first stage, face detection method is used to detect faces from the images. In the second stage, an image classifier is used to classify the faces detected in stage 1. Binary classification is applied only, either wearing a mask or not. Another facemask detection model was proposed by B.Qin and D.Li [4]. They developed a facemask-wearing condition identification method which quantifies three classes: with mask, without mask, and mask worn incorrectly. Their facemask detection technique has four steps: Image pre-processing, facial detection and cropping, image super-resolution, and facemask-wearing condition identification. They used transfer learning with MobileNet-v2. Their model trained and evaluated using the public Medical MasksDataset [5] which contains 3835 images and achieved 98.70% accuracy result. The Medical MasksDataset has 3030 images of correct facemask-wearing, 671 images of no facemask-wearing, 134 images of incorrect facemask-wearing. So, there is class imbalance in the data used to build the model and they did not consider addressing it.

Different approaches have been proposed to address the issue of class imbalance problem [6]. There are the data level approaches, algorithm level approaches, and the hybrid approaches. In the data level method, the class imbalance is addressed by over-sampling the minority class, or under-sampling the majority class or combine both sampling techniques. In the algorithm level approach, a weight is applied to each class to increase the cost of the minority class and decrease the likelihood that the model will incorrectly classify samples from this class. The hybrid methods combine both data-level and algorithm-level methods to address class imbalance problem.

III. Methodology

A. Dataset

For this project, we used the Facemask Detection dataset that is available on Kaggle [7]. The dataset is composed of 853 images captured from different public spaces. Each image

has one or more samples (faces). One image may contain different samples that belong to the three stated classes: wearing a mask, not wearing a mask, wearing a mask incorrectly. The dataset also has xml files as annotations for the faces in each image.

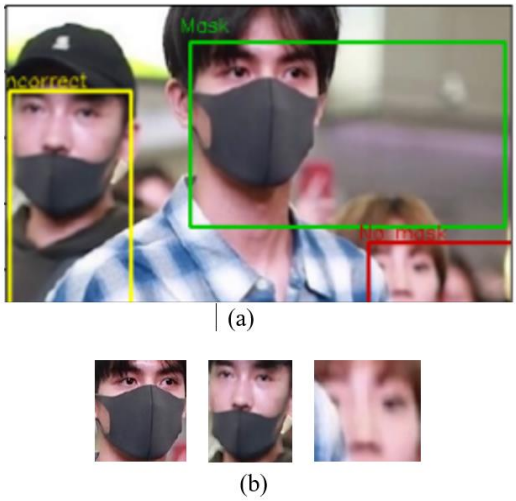

Figure 1: a) An example of image contains three samples where each sample belong to different class. b) The samples after cropping

B. Data Pre-Processing

As we mentioned in the previous section each image in the dataset contains more than one sample (face). So, different pre-processing steps has been applied to extract all faces from the images. First, the annotation provided by the xml files are applied to the images. The xml files contain the coordinates of each sample in the image as well as the corresponding label for each sample. After applying those annotations, the samples are cropped to create images of faces only (Figure 1). All cropped images are resized to 32 × 32 before feeding them to the model. Images that have width or height less than 7 are discarded from our data. The total number of extracted samples is 4072 where 3232 samples belong to *face with mask* class, 717 samples are face without mask, and only 123 samples belong to *facemask worn incorrectly* class (Figure2). So, there is class imbalance in our dataset. The class imbalance problem affects the training process. When there is class imbalance in the training data, the model is more likely to overclassify the majority class (*face with mask*). The higher probability of the majority class is the reason for the over classification. As a result, samples belonging to the minority class are more frequently misclassified than those belonging to the majority class.

C. Class Imbalance and Data Augmentation

In our project, we applied the oversampling approach to the classes *facemask not worn*, and *facemask worn incorrectly*. As we are dealing with image data, we can oversample the images by applying different transformations to the image to create new transformed images. For training and validation sets, different transformations are applied to all images from all classes to expose the model to different transforms of an image. The transformation methods that were applied were rotations with a range of 20 degrees, sheers, and shifting with a factor of 0.2. Some samples from our transformed data are shown in figure 3.

The whole dataset after the initial augmentation for the minority classes are split into a training dataset (80%) and a testing dataset (20%).

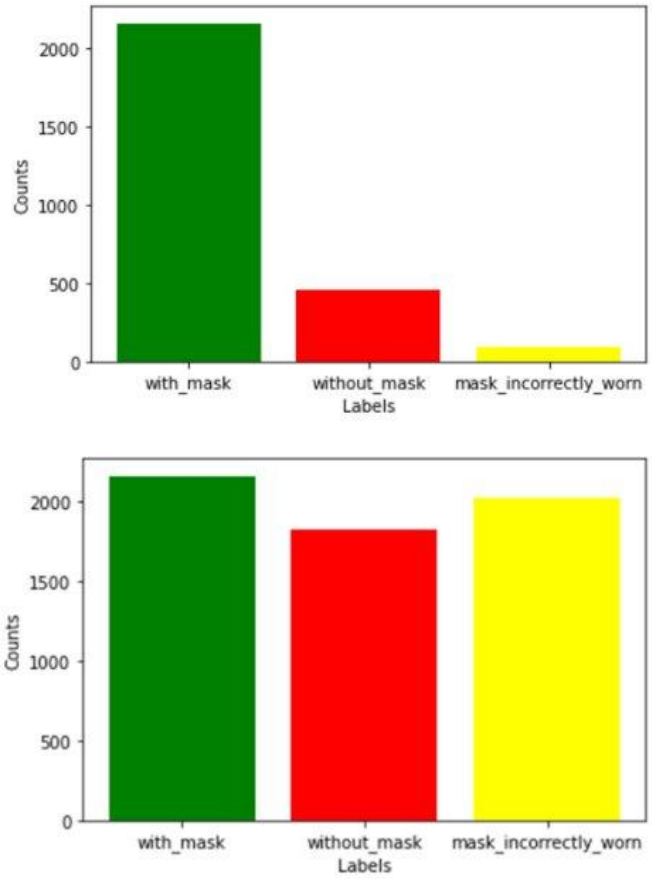

Figure 2: Distribution of samples among classes. Top: before the augmentation step, bottom: after the augmentation step

Figure 3: Augmentation, two image samples with three different image transformation.

D. CNN Model

Convolutional Neural Network (CNN) is commonly used for image classification due to their capabilities in handling

imagery data. Features are extracted from images through the convolution process.

The three classes in our data have high similarity. Especially for the classes *facemask worn* and *facemask worn incorrectly*. A deeper CNN architecture will extract more discriminative features. On the other hand, as the neural network becomes deeper, the optimization becomes harder, and it might end up with the problem of vanishing gradient. Deep residual learning is one of the most powerful deep learning approaches [8]. It solves the problem of vanishing gradient by introducing residual connections between the layers. Thus, instead of having layers one after another to obtain the complex desired mapping, a residual mapping is learned from a few stacked layers with skip connections as shown in Figure 4. The skip connections add the output from previous layers to the output of the stacked layers.

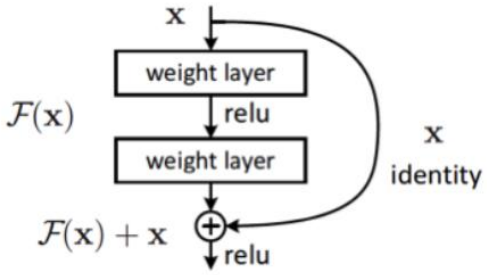

Figure 4: Residual Learning [6].

a. Model Structure and Parameters

The first layer in our proposed CNN model is a convolutional layer CONV2D with kernel size 7 × 7 and a stride of one followed by a ReLU (Rectified Linear Unit) activation function (figure 5). The number of feature maps (channel) is 16. Then, a max polling layer MaxPooling2D is applied to down-sample the image. The purpose of these two layers is to extract the low-level features such as edges, semi-circles, etc.

The resulted down-sampled image is passed to 4 residual blocks. Each residual block consists of two convolutional layers with kernel size 3×3 and two ReLU activation functions. The identity map size is adjusted and added after the second convolutional layer. This set of layers intended to extract the high-level features (e.g., eye, nose).

The resulted feature maps from the residual blocks are flattened and passed to three FCL (Fully Connected Layers). The number of nodes in these layers are 16, 7, and 3 respectively. The *SoftMax* function, which estimates the probability of all classes using the outputs of its immediate ancestors, is the activation function of output layer. As we are dealing with multi-classification problem, the *categorical cross-entropy* is used as the loss function along with Adam optimizer which works with adaptive learning rate.

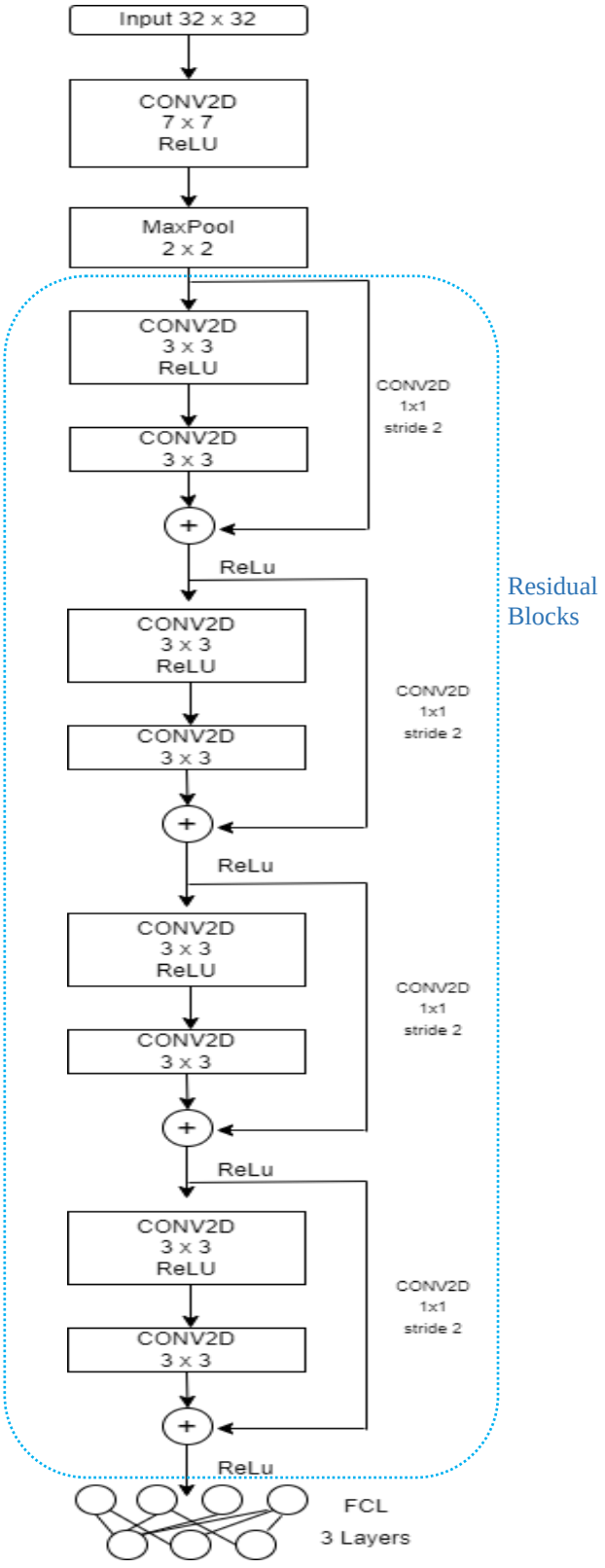

Figure 5: CNN structure

b. Model Training and Validation

The model is trained using 5566 images of size 32 × 32 and evaluated on 983 images. Initially, we trained our model with 50 epochs and a batch size of 100. The training/validation loss and accuracy in each epoch are reported to assess the model performance during training (figure 6).

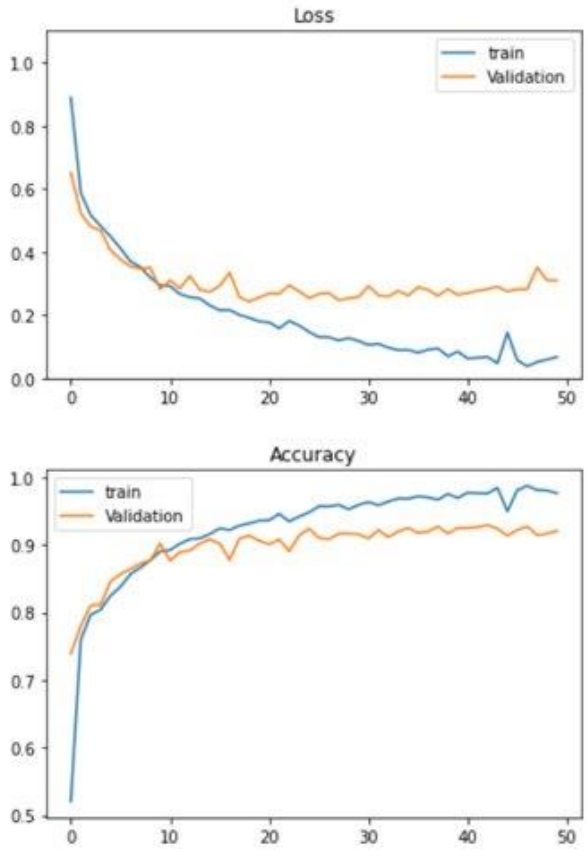

Figure 6: Training and Validation Loss (Top), Accuracy (Bottom)

The loss plot shows that the training loss are decreasing, and the validation starts to decrease and then remain steady in the last few epochs.

Therefore, an Early Stopping technique is used to assign the best epoch number and prevent the slight overfitting of the initial model. After applying the early stopping, the model training phase stopped at epoch number 41. Table 1 shows the accuracy results for training set, validation set, and test sets for both the initial model and the model after applying early stopping. The optimized model has better accuracy result for validation, and testing sets. The accuracy results improved by 3% in the validation and test sets.

Table 1: Accuracy results for training, validation, and testing sets for the initial model and the optimized model

	Initial Model	**Optimized Model**
Training	97%	98%
Validation	92%	95%
Test	91%	94%

IV. RESULTS

To evaluate the model classification performance, we used precision, recall, and F1 score. For each class, the precision measures the proportion of correctly classified samples in a certain class to the total number of samples classified to the class. The recall for each class measures the number of correctly classified samples to the total number of samples belong to a certain class in the data. The F1 score serves as the harmonic average between the precision and the recall, it identifies the overall performance of the classification for each class. To calculate the precision, recall, and F1 score for all classes, we used *classification_report* function from *Sklearn*. Table 2 shows the resulted precision, recall, and F1-score for each class. *Face with mask* class has high value of precision, recall, and F1-score. *Facemask not worn* class has relatively high recall but lower precision, only 84% of samples classified *as face without mask* are belong to this class. *Facemask worn incorrectly* has very low values of both precision and recall. Only 53% of *facemask worn incorrectly* samples are identified by the model, and 28% of the samples classified as they belong to this class are correctly classified. The low results observed in this class is due to the high similarity between this class and the other two classes. It shares the existence of the mask in the image with the *Face with mask* class, and the existence of the nose or the mouth and nose with *facemask not worn class*. Moreover, this class has the lowest number of samples.

Table 2: Precision, Recall, and F1 score for each class.

Class	**Precision**	**Recall**	**F1score**	**Number of samples**
Face with mask	0.99	0.92	0.95	719
Facemask not worn	0.84	0.95	0.90	152
Facemask worn incorrectly	0.28	0.53	0.36	30

We also compare our proposed method with the other facemask detection methods in term of accuracy results. The two binary classification methods InceptionV3[2] and the Dual stage [3] models achieved an accuracy result of 99%. MobileNet-v2 [5] model which perform multi-classification achieved an accuracy score of 98% while our proposed method achieved an accuracy score of 94%. Our model achieved a relatively good accuracy result when compared to the other previous models, although those methods used pre-trained models.

	InceptionV3	**Dual-stage CNN**	**MobileNet-v2**	**Our Proposed method**
Classification type	Binary	Binary	Three classes	Three classes
Accuracy score	99%	99%	98%	94%

V. Conclusion and Future Works

Coronavirus Disease 2019 (COVID-19) has spread rapidly over the world, resulting in a global pandemic. Facemask covering is one of the essential ways to prevent the spread of COVID-19. In this project, we proposed a facemask detection model that identifies three classes: *facemask is worn*, *facemask is not worn*, and *facemask is worn incorrectly*.

This model could be further improved by integrating a face detection technique that could be performed before applying the classification. Moreover, image enhancement like image super resolution could be applied to improve the quality of the training data and build more robust model.

In general, our model can accurately detect the classes *facemask is worn*, *not worn*. We expect our model will have potential applications in epidemic prevention including COVID-19.

VI. Contribution Statement and Project Code

The authors contribution to the project as follows: Data preparation: H. Alsurayhi; data augmentation: D. De Luca, H. Alsurayhi; model design: H. Alsurayhi; initial model training: H. Alsurayhi; Model Optimization: D. De Luca; analysis and interpretation of results: H. Alsurayhi; draft manuscript preparation: D. De Luca, H. Alsurayhi. All authors reviewed the results and approved the final version of the manuscript.

The project code is uploaded on GitHub: https://github.com/ddeluca98/face-mask-detection.git

VII. References

1. *WHO Coronavirus (COVID-19) Dashboard*. (n.d.). Retrieved from https://covid19.who.int/

2. Jignesh Chowdary, G., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020). Face mask detection using transfer learning of inceptionv3. *Big Data Analytics*, 81–90.

3. Chavda, A., Dsouza, J., Badgujar, S., & Damani, A. (2021). Multi-stage CNN architecture for face mask detection. *2021 6th International Conference for Convergence in Technology (I2CT)*.

4. Qin, B., & Li, D. (2020). Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. *Sensors*, *20*(18), 5236. https://doi.org/10.3390/s20185236

5. Medical Masks Dataset (https://www.kaggle.com/vtech6/medical-masks-dataset)

6. Johnson, J.M., Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J Big Data* **6,** 27 (2019).

7. Larxel. (2020, May 22). *Face mask detection*. Kaggle. Retrieved March 18, 2022, from https://www.kaggle.com/andrewmvd/face-mask-detection

8. K. He, X. Zhang, S. Ren and J. Sun. “Deep Residual Learning for Image Recognition”, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, DOI: 10.1109/CVPR.2016.90.