Face Mask Detection Classifier Using CNN

-For multiclass classification problems with imbalanced data

Raphael Du, Ray Yang, Mike Cui

Abstract

The use of face mask detection systems can help enforce mask-wearing policies and reduce the spread of infectious diseases like COVID-19. The team is working on a face mask detection system using Convolutional Neural Networks (CNNs). The goal is to use deep learning tools to recognize faces in images and determine if they are wearing masks. CNNs are well-suited to image classification tasks and can achieve high accuracy with appropriate network architecture. However, the data is extremely imbalanced and three possible solutions were proposed which turns out that the solution for data augmentation performs the best. But there are still issues with this solution which leads to the model cannot classify the label mask_weared_incorrectly correctly.

Problem Description

The occurrence of the COVID-19 epidemic has abruptly affected our lives. Wearing a facial mask is one of those habits we developed during the pandemic. Yet for the past few months, masking has gradually become optional as more areas are below the high-risk category. However, it does not necessarily mean COVID is any less contagious; as a matter of fact, the latest variant of the virus is even more transmissible according to Kathy Katella (2022), a senior clinical writer at Yale Medicine.

And in order to mitigate the transmission of the virus, wearing a mask is still one of the most efficient ways, especially in those crowded and confined spaces. At Duke University, masks are still required on Duke buses and vans and in all clinical settings. Off the campus, there are many places like airports that continue to require people to wear a mask considering the huge traffic.

Hence the face mask detection system. The goal of this project is to achieve face mask detection utilizing some fundamental machine learning tools, and the suggested technique will recognize the face in the image and then determine whether it has a mask on it. Knowing people not wearing a mask in advance can help reduce the workload and give people an opportunity to allocate resources ahead of time. This is essentially to mandate the use of masks for the sake of people's health. Thus, apart from implementing machine learning knowledge to the project, contributing at least something helpful, if not significant, to the community's health security is another little goal that our team wants to realize.

The technique that our team applies to achieve face mask detection, in general, is the Convolutional Neural Network, also known as CNN. Our input test data shares common characteristics with images that are used to train the network; therefore, we hypothesize that the classifier we develop should be able to achieve an accuracy higher than 85% with an appropriate CNN architecture.

There are several enlightening resources. The latest Face mask recognition system using CNN model published this September is one of the most related and helpful articles. As the title implies, the authors utilize CNN along with Python script and tensor flow to develop an efficient network for face mask recognition (Kaur et al., 2022). Some of the ideas in feature extraction and dimension reduction from that paper can potentially help our team to develop our own neural network.

Data Description and Preprocessing

The data gathered from the Kaggle website consists of 853 colored PNG files and corresponding XML files which include the information of each PNG file (See figure 1).

Figure 1: PNG file(left) and XML file(right)

The PNG files include human face images, which can be divided into three categories: with masks, without masks, and wrongly wearing masks. The data comprises images of individuals wearing masks taken from various perspectives (e.g., lateral, frontal, and rear side), from a single person to multiple people in one image, and different people might have different mask-wearing conditions in one image, as well as different types and colors of masks. It also covers persons of various ages, genders, ethnicities, etc.

The XML files corresponding to each of the PNG files include the information about the size of the PNG file, denoted by <size>, and the coordinates of faces of each people in the PNG file, denoted by <bndbox> under <object> as well as the corresponding label, with_mask, without_mask, and mask_weared_incorrect.

The sufficient data provided could enable us to train our model that fits various images with complex masks and wearing conditions.

According to the coordinates of faces in each image, the faces are cropped from each PNG file to form an individual face with a corresponding label and the size of each face is increased by 10% to include more features. Therefore, the dataset of 853 colored PNG files and corresponding XML files is transformed into 4072 faces and its label (see figure 2). In this case, each data point has only one label.

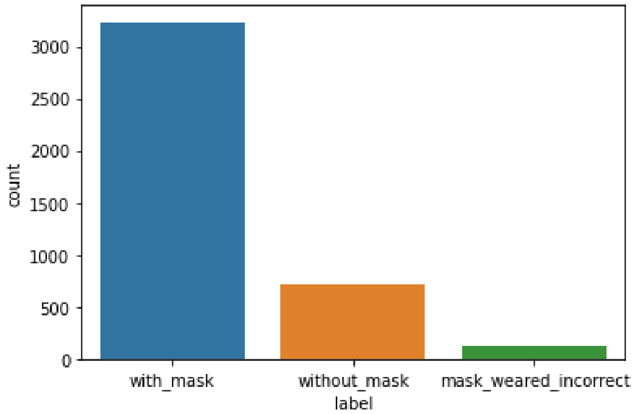

Figure 2: Imbalanced data

Over 70% of the data are labeled with_mask, therefore, it is an extremely imbalanced dataset that will be handled; otherwise, the model we choose cannot learn anything.

Methods

Since this is a classification problem of image recognition, we are planning to use the Convolutional Neural Network (CNN) model to deal with this problem. In deep learning, a Convolutional Neural Network is a class of artificial neural networks, most applied to analyze visual imagery (Value *et al.,* 2020). Its built-in convolutional layer reduces the high dimensionality of images without losing their information. That is why CNNs are especially suited for this use case. Another reason is that the number of parameters in a neural network grows rapidly with the increase in the number of layers. This can make training for a model computationally heavy (and sometimes not even feasible). The time taken for tuning these parameters is diminished by CNNs (Mishra, 2019). In this project, we chose AlexNet to be our CNN model. AlexNet was one of the first large, deep Convolutional Neural Networks (CNNs) to be trained on a very large dataset, such as ImageNet. It was able to outperform previous state-of-the-art image classification models by a significant margin.

In order to address the extremely imbalanced data issue, we proposed three solutions.

- Use existing mask_weared_incorrect and without_mask data to recreate more data to the same amount of with_mask data by rotating, stretching, or lowering the resolution.

To artificially increase the number of data samples for the initial answer, we employ data augmentation. We use the ImageDataGenerator from Keras to do that. Using this method, we may modify current training photos to produce new ones (scaling, rotating, flipping, etc.). We will only employ a tiny fraction of the potential alterations because not all of them make sense in the context of a face image, for example, flipping a face vertically is not a viable option for a person strolling down the street.

- Change the problem into a binary problem, which only has two classes, with_mask and without_mask.

The second solution involves converting the issue to binary classification, which entails combining all instances of mask_weared_incorrect with with_mask into one class. because they resemble one another more than without_mask class.

➢ Freeze a few layers of Alexnet and only train the last few layer settings.

The third solution, freezing layers in transfer learning. We accomplish this by freezing the weights in the network and only training the fully connected layers.

We will be comparing the three different solutions' performance.

Final Results

After applying the proposed solution, the results for each solution are shown below.

➢ Data augmentation (solution 1)

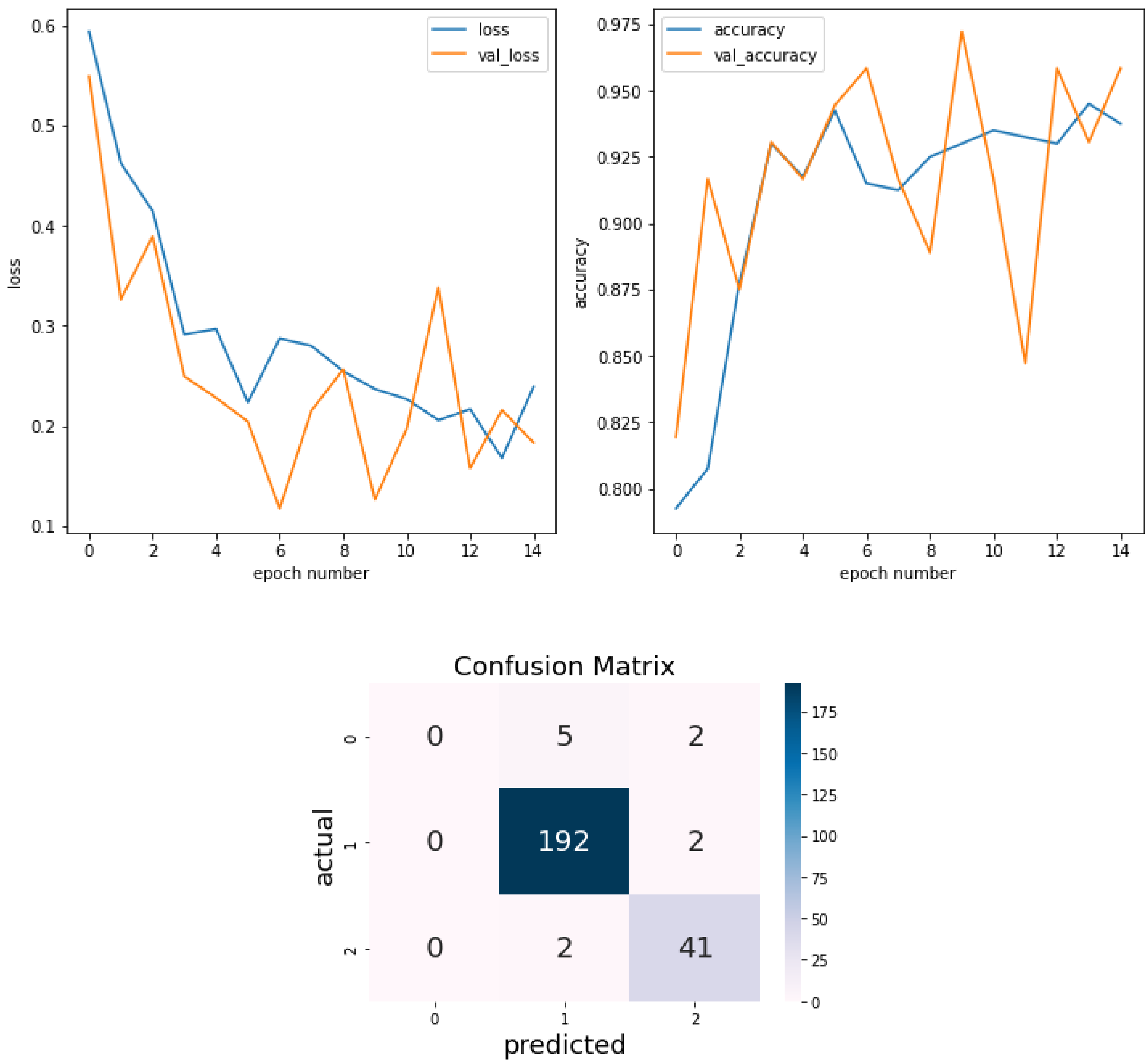

Figure 4: Final results for solution 1

The results for loss and accuracy versus the number of epoch show that the model does not overfit which reaches over 90% accuracy. The result of this model looks promising. But for the confusion matrix, there is still no chance of predicting label 0 which is the mask_weared_incorrectly correct.

➢ Convert to Binary class problem (solution 2)

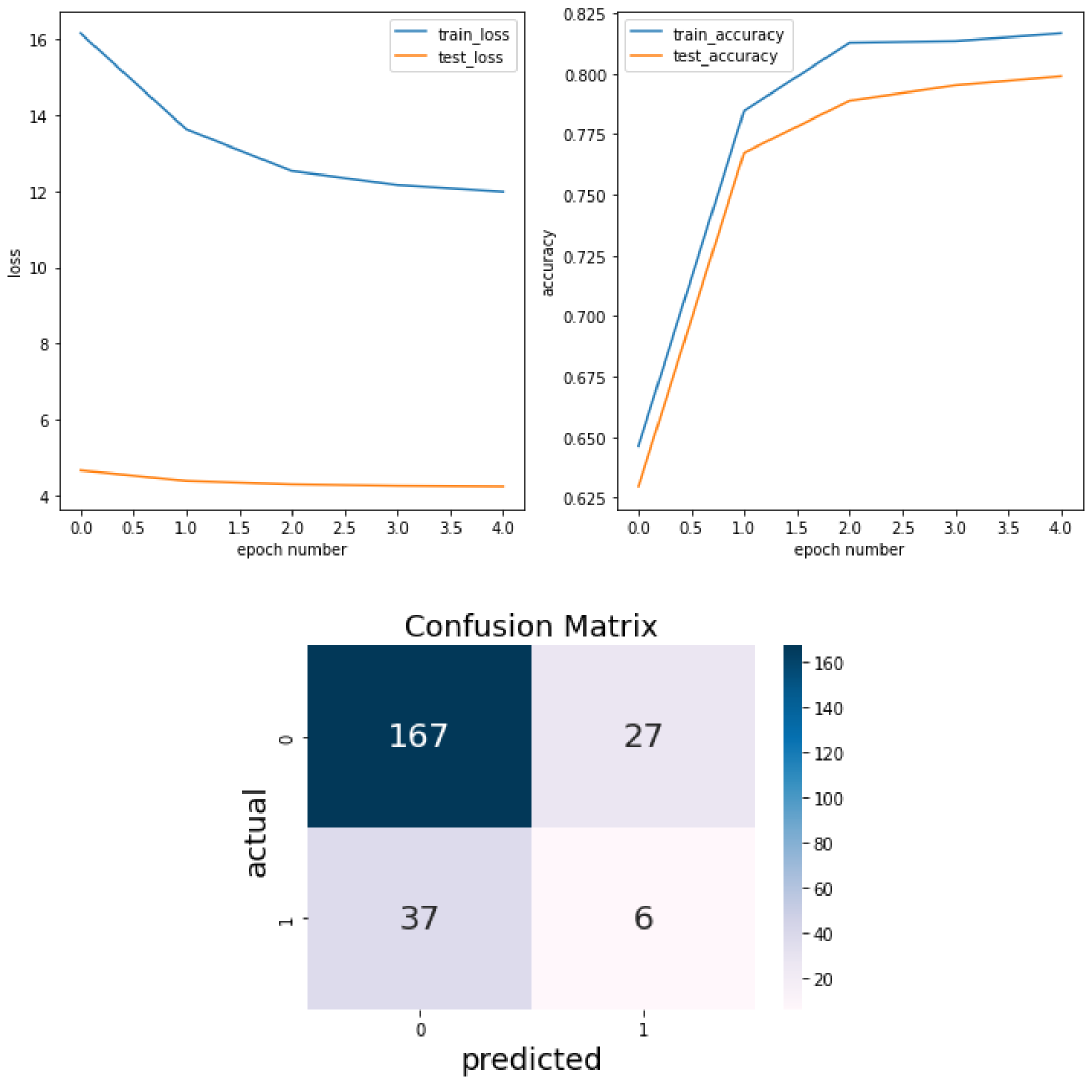

Figure 4: Final results for solution 2

After combing the mask_weared_incorrect data with with_mask and fit into the model, the results show that the accuracy for training data and test data has reached 80% and the model is generalized. But for the confusion matrix, there are very few chances to predict label 1 correctly. This model confuses label 1 with label 0.

➢ Freezing layers in transfer learning (solution 3)

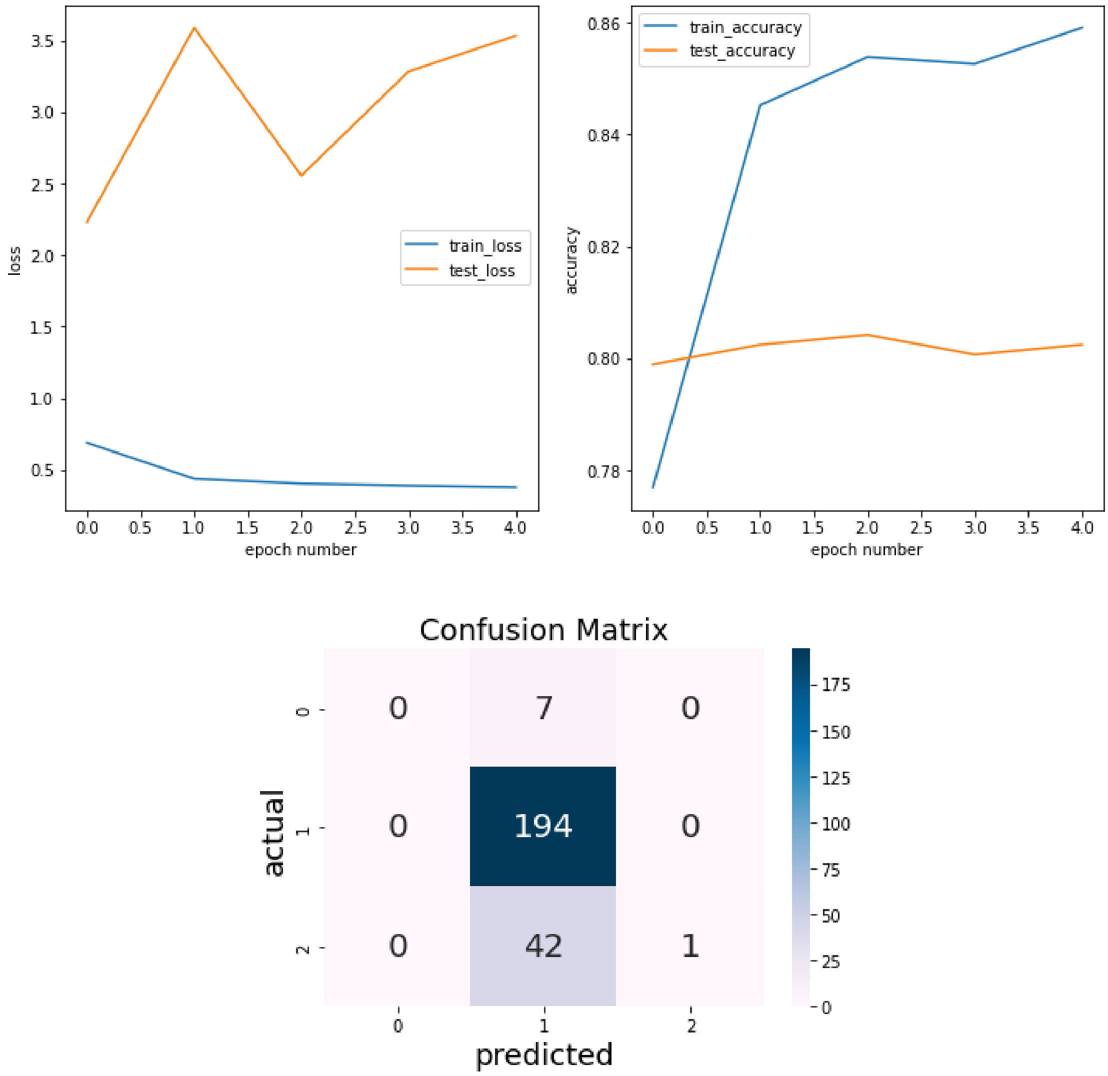

Figure 5: Final results for solution 3

The results for solution 3 do not look promising, because the model does not generalize. It has a high accuracy of over 85% on the training dataset but it has a low accuracy on the test dataset and the confusion matrix indicates that the model confuses label 2 as label 1. There is still no chance that the model can predict label 1 correctly.

Current Conclusions and Future Work

The performance for the data augmentation solution is the best, while the performance for the freezing layers in the transfer learning method is the poorest. There are still problems with the first solution, the model has a limited ability to predict the mask_weared_incorrectly class. This is because the split test data includes very few images labeled mask_weared_incorrectly before the data are augmented. To improve the performance of the model, we should collect more data and then use various oversampling/undersampling algorithms in machine learning. Another solution is to combined data augmentation and converting to binary classification. The result of it looks pretty good (Fig6)but this deviates from our original muti-classification goal.

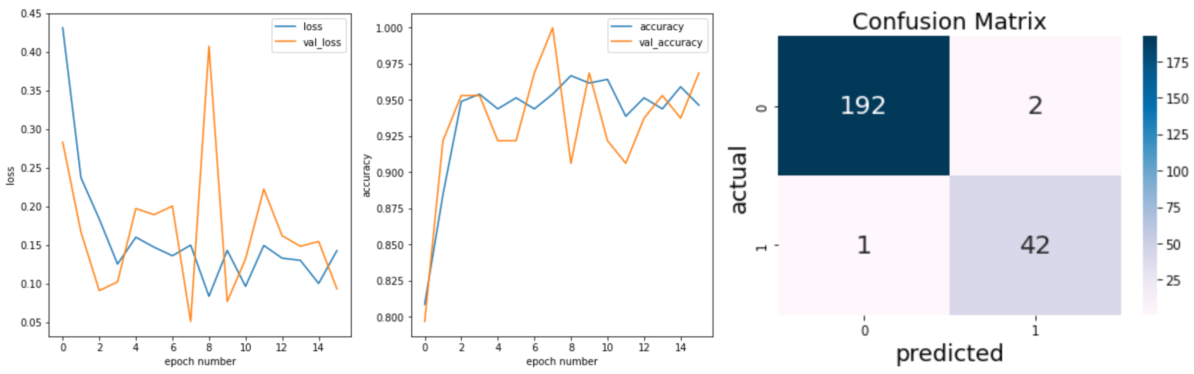

Figure 6: Final results for combined solution1 and solution2

For future studies, this model could be applied to CCTV integration with a real-time prediction as well as a percentage of prediction confidence that will help detect people without masks. Additionally, a research project on the effectiveness of mask-wearing policies in reducing the spread of infectious diseases could use face mask detection to automatically collect data on mask usage in different settings. This data could be used to evaluate the effectiveness of different mask-wearing policies and help to inform public health decision-making. Moreover, a research project on the psychological effects of mask-wearing could use face mask detection to automatically record the number of people who are wearing masks in different settings, and then use this information to evaluate the psychological impact of mask-wearing on individuals. There are many potential ways that face mask detection could be used as part of an academic research project, depending on the specific research question being investigated.

Broader Impacts

This project demonstrates a face mask detection system based on deep learning. The presented approach can be utilized with surveillance cameras to detect persons who do not wear face masks and those who do not wear them properly, hence the restriction on COVID transmission. Apart from the fact that the system can bring us convenience when implementing mask regulations at places where people are exposed to the virus, it may also raise potential risks. One of the most concerning ones is the malicious use of facial pictures. The face mask detection system, in most cases, captures people's facial pictures without their awareness or approval. These pictures, for instance, may be collected by some private agencies to analyze their customers for commercial purposes like precision advertising. There is also a possibility that criminals could use the leaked facial data to forge personal identities with the help of deepfake technology. Our privacy may be violated, and our personal property safety may be endangered. Therefore, we encourage people to improve their awareness of personal information protection, as well as industries to formulate self-discipline toward personal privacy. We do hope that the relevant policy could be more well-established so that this face mask detection system can be properly used to benefit the entire society.

Reference

Katella, K. (2022) *Omicron, delta, alpha, and more: What to know about the Coronavirus variants*, *Yale Medicine*. Available at: https://www.yalemedicine.org/news/covid-19-variants-of-concern-omicron (Accessed: October 25, 2022).

Kaur, G. *et al.* (2022) “Face mask recognition system using CNN model,” *Neuroscience Informatics*, 2(3), p. 100035. doi: 10.1016/j.neuri.2021.100035.

Mishra, P. (2019) *Why are Convolutional Neural Networks good for image classification?*, *DataDrivenInvestor*. Available at: https://medium.datadriveninvestor.com/why-are-convolutional-neural-networks-good-for-image-classification-146ec6e865e8 (Accessed: October 25, 2022).

Valueva, M. V. *et al.* (2020) “Application of the residue number system to reduce hardware costs of the convolutional neural network implementation,” *Mathematics and computers in simulation*, 177, pp. 232–243. doi: 10.1016/j.matcom.2020.04.031.

We adhered to the honor code.