Face Mask Detection
Using Image Classification

Group Members: Benjamin Alpert, Parisa Suchdev, Shaurya Swami, Zunyi Liao

1. Introduction

It is a well-known fact that Coronavirus has become a worldwide health pandemic, causing 4.55 million deaths worldwide [1], Throughout the crisis, this virus has evolved into a variety of hazardous variants, and while different vaccinations have been released, the efficiency rate of each vaccine on variant varies meaning that even if 100% population of a certain area is vaccinated, it does not guarantee the virus would stop spreading. Amid this crisis, one of the effective preventive methods is to wear filtered masks to cover mouth and nose in order to prevent virus from entering the body or stopping onward transmission of the virus as it is contagious [2].

Through this project, we aim to implement an Image Classification model that identifies if the person is wearing a mask or not. It is important to identify this classification problem as it lays the foundation of further research that can be done using this model, such as finding out the impact of masks on preventing the virus. It also can be used in organizations where wearing masks is compulsory and can be used on the entrance protocol such as, screening process, that can prevent non-mask-wearers from entering the building.

The related work (further discussed in section 4) emphasizes that a majority of research work has focused on Deep Learning and Artificial Intelligence techniques for facial recognition, whereas, the approach used for this project aims to use classification methods to identify if a person is wearing a mask or not. Depending on the timeliness of the project, we hope to be able to incorporate classification labels for input images that yield to figure out if the person is wearing masks correctly or not, such that the nose and mouth are both covered properly.

We were able to find a labelled dataset worth 500GB, where each of the images is labeled with image size, photo type, person's age, gender, and user ID (further discussed in Section 2). This project will constitute only a subset of this huge dataset. Looking at this dataset and the problem statement, we identified that it is multiclass classification model as it will be trained based on four classes:

Class 1 - “The mask is worn correctly, covering the nose and mouth.” [6]

Class 2- “The mask covers the mouth, but does not cover the nose.”[6]

Class 3 - “The mask is on, but does not cover the nose or mouth.” [6]

Class 4: - “There is no mask on the face.” [6]

For such multiclass dataset, the accuracy of the model will be measured according to the positive class which is class 1. The algorithm that will be used is CNN (Convolutional neural network) due to its efficiency towards high accuracy rate and it will be implemented using TensorFlow.

Dataset will be split into a training set and testing set, where the training set will be used to train the model and testing set will be used to test the accuracy of that model (Further discussion of accurate numbers discussed in Section 2). The results we aim to achieve through this model are:

1. High accuracy if a person is wearing a mask and belongs to class 1 and low accuracy if the person is not/incorrectly wearing a mask.
2. High accuracy on testing set to avoid overfitting

2. Problem Definition and Algorithm

2.1 Task Definition

Problem and Solution:

The problem space of this project is not only to detect the mask but to check if it properly covers the face. As mentioned, we will be using labeled images of people categorized into 4 classes, and the model will be trained according to that data in order to label other unknown images of people for mask detection purposes. This could be helpful in airport, hospital, or university camera security systems to disallow unmasked individuals from entering.

Inputs:

Using a kaggle dataset, we retrieved a large set of images worth 500GB which included 4 images of each person with 4 classes. For this project we are using only an 80GB subset that we shrunk to 2GB. We reduced the size of the subset by reducing all the images to be 50 by 50 pixels. We ended up with 40,000 images of 10,000 people. The purpose of reducing the dataset is to make it more manageable and uniform in image size. Out of this data, 80% is used to train the model and 20% is used to test the model.

Outputs:

Each photo is labeled with the person ID, the **type of photo** it is (e.g. wearing a mask, not wearing a mask, etc.) and gender. There are four photos associated with each person: wearing a mask, not wearing a mask, wearing a mask without covering their nose, and wearing a mask below their chin.

2.2. Dataset

Dataset retrieved from Kaggle contains 500GB worth of data. It has images of people and each person has 4 different images with 4 classes. It also has a dataframe in CSV format. The data subset is relatively huge (about 80GB), due to which, we will access it over http within the python code. If this proves to be too slow, we will use a smaller dataset of roughly 1GB and store it locally.

Each photo is labeled with the person's ID, the type of photo (e.g., wearing a mask, not wearing a mask, etc.), and the person's gender. Many gender fields are designated "none," but others are not. In order for the field to be relevant to us, we may elect to discard the images that do not have a gender label. Moreover, with the small subset of the dataset, there might be no need for hardware to process the data. However, if needed, we will require a GPU.

2.3 Algorithm Definition

We will be using Tensorflow's convolutional neural network library as our algorithm for solving this problem because it has very good accuracy with image classification. We hope to expand our knowledge of how this works by the end of the project. We use this tutorial to create our baseline model following it very closely. This CNN model starts from loading the data to calculating the prediction including split and plots. We plan on following some more advanced TensorFlow tutorials to improve the results of the baseline code. Additionally, keeping the timeliness of the project in mind, we expect to be able to code our own convolutional neural network without using Tensorflow. Hopefully, we'll be able to compare Tensorflow's results to our own. Neural networks are one of the course's last units, we may not be able to complete this in time.

Pseudocode Description:

1. Import tensorflow
2. Bring our training data into memory and split into validation and training subsets using 80% for training and 20% for validation.
3. Shuffle the dataset.
4. Normalize the data
5. Create the model
6. Compile the model
7. Train the model
8. Run predictions on unseen data
9. Plot training and validation epochs versus accuracy
10. Plot training and validation epochs versus loss

3. Experimental Evaluation

3.1 Methodology

What are the criteria you are using to evaluate your method? What specific hypotheses do your experiment test? Describe the experimental methodology that you used, and why do you think it is the right way to measure the performance. What is the training/test data that was used, and why is it realistic or interesting? Exactly what performance data did you collect and how are you presenting and analyzing it? Comparisons to competing methods that address the same problem are useful.

We are using tensorflow and its respective libraries to utilize the power of Convolutional Neural Networks without having to code the deep learning method from scratch. Moreover, we are implementing the "One vs all" approach to create class labels for defining our 4 classification scenarios.

Hypothesis - Given an input image of a person wearing a face mask in various ways (as described in our class labels above) or not wearing a face mask at all, our model will be able to correctly output the appropriate image classification label with at least 80% validation accuracy for each class label input.

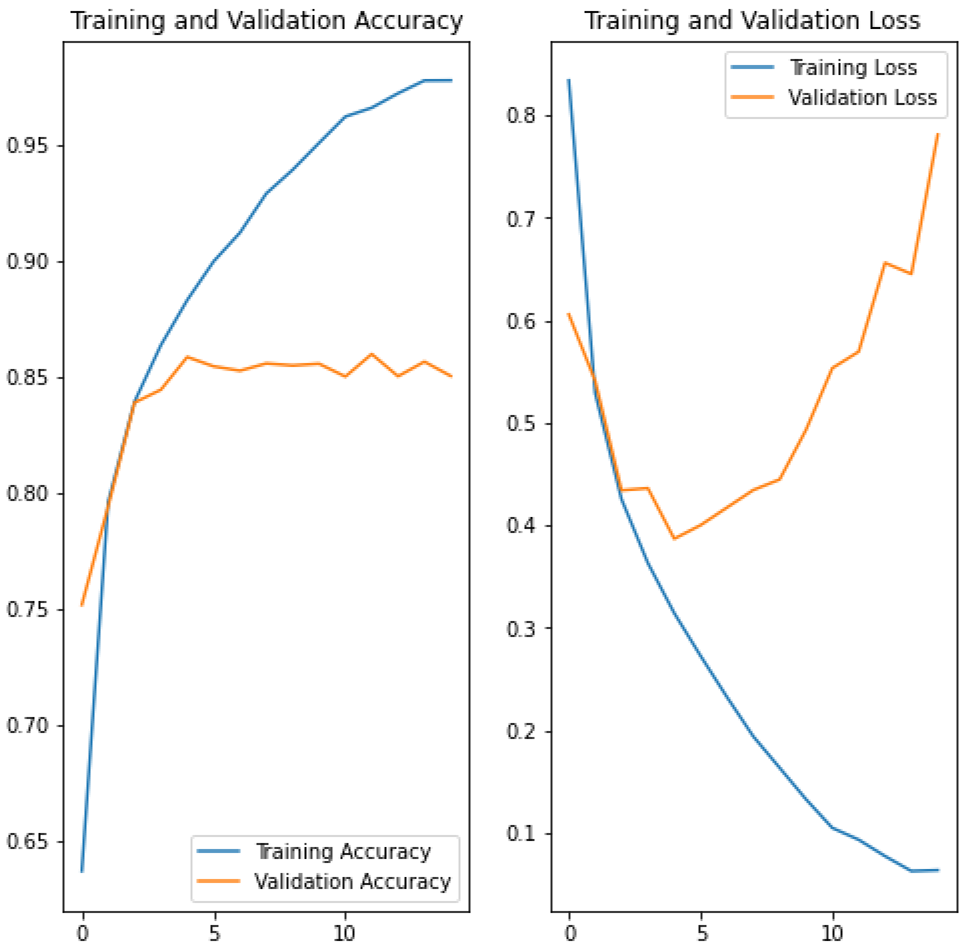
We are using an accuracy metric to measure our performance because we have a well-balanced dataset. More specifically, our dataset is organized such that every 4 consecutive images of one person cover all 4 classes that we have defined above. We choose to use 40,000 total samples from our dataset with a 80% and 20% distribution for training and testing respectively. Since we have a large dataset, this training and testing distribution allows us to train our model well while keeping a good proportion of images in the testing set to evaluate its performance. Our group is analyzing the overall accuracy of the model as well as individual accuracy for each class. Our model achieves an individual face mask wearing class accuracy of 95.91%, while similar work [5] has yielded a face mask wearing accuracy of 98.3%. While our accuracy is slightly lesser than the referenced work [5], we achieved almost equivalent accuracy using tensorflow and without having to code deep learning algorithms like YOLO-V4 and CSPDarkNet53 by ourselves. Additionally, this is the first iteration of our results, so we can only make more improvements and get higher accuracy from this point on.

3.2 Results

Present the quantitative results of your baseline system experiments. Graphical data presentation such as graphs and histograms are frequently better than tables.

Individual accuracy scores of our results are as follows (based on the group members' faces in the 4 face mask wearing/not wearing scenarios as described in classes):

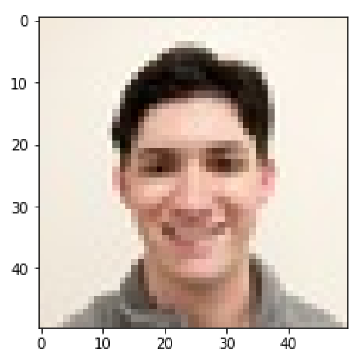
- Input image: Person wearing a mask
  → Output: Correct (Accuracy = 95.91%)
- Input image: Person wearing a mask under nose
  → Output: Correct (Accuracy = 80.88%)
- Input image: Person wearing a mask under chin
  → Output: Correct (Accuracy = 68.46%)
- Input image: Person not wearing a mask
  → Output: Correct (Accuracy = 98.05%)



epochs: 15 - loss: 0.0635 - accuracy: 0.9775 - val_loss: 0.7811 - val_accuracy: 0.8501

```
In [27]: predict("/home/benjamin/Downloads/IMG_20200113_085742_reduced.jpg")
```

```
Access Denied - Please wear a mask (99.99% confidence)
```



3.3 Discussion

Is your hypothesis supported? What conclusions do the results support the strengths and weaknesses of your method compared to other methods? How can the results be explained in terms of the underlying properties of the algorithm and/or the data? What is your target/goals to achieve by the end of the semester?

Our hypothesis was supported and our model exceeded our expectations by yielding an above estimated overall accuracy of 85.01%. The trained model did a great job predicting precise class outputs for most of the group members' faces as well. Our accuracy results were comparable in comparison to other deep learning methods, although they are not as high as some of the works referenced below. Considering we are using tensorflow and have formulated a well-performing model without having to spend numerous hours/days on the implementation of actual deep learning techniques from scratch, we believe our initial results are promising. Furthermore, we have ideas on how to increase our model accuracy and precision in correctly identifying the 4 classes of images which we will implement to boost our model performance. Since we have not hard coded a CNN in our approach, we are taking more subtle steps to increasing model accuracy some of which (like increasing number of epochs) has already elevated overall and individual class accuracies. Our aim is to incorporate more tensorflow capabilities to improve input image quality and possibly crop images to focus on the face to improve our overall model accuracy to over 90% by the end of the semester.

4. Related Work

Answer the following questions for each piece of related work that addresses the same or a similar problem. What are their problems and methods? How is your problem approach different or better?

I. **[3]** The problem of the Research article deals with facial recognition and face mask detection using ML techniques. They mention the use of any supervised learning technique where the training dataset input would be images (preferably labeled). Next, any unlabeled images will be classified based on what the model learned from the training dataset. The generic supervised learning algorithm equation used here is Y = f(x) where Y is the predicted output and x is the input. The testing phase incorporated the use of sample test images being passed through a convolutional neural network and classification results being compared for accuracy. Our methodology would focus on facial mask image classification instead of narrowing our focus on facial recognition and object detection. Additionally, we would incorporate the use of tensorflow and its libraries instead of coding deep learning techniques from scratch.

II. **[4]** This article hopes to resolve the issue of people wearing face masks inappropriately (incorrectly positioned such as under their nose, etc.). Furthermore, the authors suggest approaches to classify images with labels of how the subject's face mask is worn. The described approach employs the use of an image recognition model in PerceptiLabs which would aid in the classification of the different ways in which people wear masks. The methodology incorporated the use of a dataset that included images of numerous different ways of wearing a face mask as input for training the model. The dataset was distributed into subdirectories based on how people were wearing their masks and numeric classification labels were drafted (as a .csv file) to output the appropriate type of mask wearing style as output (for eg, 0 - mask above chin, 1 - mask worn correctly,etc.). Similar to this paper, we aim to classify face mask images to label them as appropriately worn or inappropriately worn and specify how the mask was incorrectly positioned. If we have enough time in the semester then we would like to include more classification labels as suggested here. Our approach differs, however, since we would incorporate the use of tensorflow and its libraries instead of coding deep learning techniques from scratch or using PerceptiLabs.

III. **[5]** This research paper aims to remedy the issue of low accuracy of face mask recognition in real-time performance using the YOLO-V4 object detection algorithm. At first, the work discussed using a CSPDarkNet53 convolutional neural network into the feature extraction network. Next, an adaptive image scaling algorithm was used to reduce computation and redundancy. The object detection algorithm was then used to compare and evaluate the accuracy of the model. While the approach described in this paper did yield a high face mask recognition percentage of 98.3%, the complexity of the algorithm used may be highly complex for us (since it is a deep learning algorithm) which would take a lot of time to effectively code, train, and test on our model. This approach uses object detection and deep learning techniques to get higher face mask recognition accuracy, whereas, our project will utilize tensorflow and focus on the classification of how users are wearing face masks (using output labels) given an input image.

5. Next Steps

- Precisely describe your next steps toward the final project objectives and how you will approach them. Elaborate enough to recognize your next steps including methods, timeline, etc…

  Our focus will be to augment our model accuracy via various methods that we have thought of thus far. Some of these methods have already proven effective, such as increasing the number of epochs which in turn increased our model's training time. Other accuracy boosting methods that we are considering include improving the input image quality, reducing background content of images to focus on the face, exploring more tensorflow tools to aid in the process, possibly introducing a multiple class cross-validation K-folds method, etc. Considering we have about a month and a half left in the semester, we would keep our timeline to testing out these suggested approaches as well as others that we come up with to a month from now. Since we would continue meeting at least twice a week (like we usually have for the entire semester now), we could confirm and cross off accuracy improving techniques that work/do not work. Additionally, we will be able to approach Dr. Wshah with our doubts for better guidance during this time period.

Explicitly define the role of each student on the project, make sure that each student will develop or participate in developing a machine learning algorithm.

Benjamin - Exploring tensorflow methods that could optimize our code and amplify model accuracy.
Parisa - Modify images to focus on peoples' faces to avoid background interference while training the model
Shaurya - Read more research papers to aid with methodological techniques that could be used to increase accuracy.

Zunyi - Review dataset and examine image improvement techniques to yield better classification accuracy, and develop a real time interactive window for recognizing and giving feedback if a person is wearing a mask or not.
Everyone - Potentially, develop our own implementation of the convolutional neural network later on.

5. Code and Dataset

Share the code and the dataset (don't share private datasets), describe in precise and clear steps how to reproduce the results you achieved. Sharing the code and the steps using GitHub or gitlab would be great and enough!

https://github.com/BenjaminAlpert/cs254a-final-project

6. Conclusion

Briefly summarize the important results and conclusions presented in the report. summarize the most important points illustrated by your work? summarize how your results will improve by enhancing the current approach or using different approaches.

Using a "one vs all" classification technique and CNN algorithm in tensorflow, we were able to successfully label and classify 4 different types of face mask wearing trends commonly noticed today. The 4 face mask labels created were to classify an input image as a person wearing a mask, wearing a mask under the nose, wearing a mask under the chin, or not wearing a mask. When we gave a compressed dataset randomly split off into 80% training and 20% testing to our model, it yielded a label classification output of how their face mask was worn/not worn. Our initial results suggest an overall accuracy of around 80% in successfully classifying the 4 face mask wearing labels.

Our input images are currently reduced to 7500 (50 x 50 x 3) dimensions/features, so one method to get higher accuracy would be to increase our features to get more lucid training images to better train our model. Additionally, we could employ other tensorflow libraries and capabilities to filter out the background distraction/noise from our images to achieve better accuracy (cropping). Finally, we could train our model to better classify the various methods of wearing/not wearing a face mask by increasing our number of epochs. Since we only trained the model for 10 epochs on 32,000 training images to get fast computation times, it is possible that our model may be showing symptoms of underfitting which can be fixed by increasing the complexity of the model and training time. Overall, we were surprised at our model giving us above 80% validation accuracy (85.01%) and happy with our model's performance for each image input class type.

Bibliography

Be sure to include a standard, well-formatted, comprehensive bibliography with citations from the text referring to previously published papers in the scientific literature that you utilized or are related to your work. Include references to the public code that you used for your project.

[1] Worldometers.info. 2021. COVID Live Update: 234,780,398 Cases and 4,801,297 Deaths from the Coronavirus - Worldometer. [online] Available at: <https://www.worldometers.info/coronavirus/> [Accessed 1 October 2021].

[2] Who.int. 2021. Coronavirus disease (COVID-19): How is it transmitted?. [online] Available at: <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted> [Accessed 1 October 2021].

[3] M. M. Boulos, "Facial recognition and face mask detection using machine learning techniques," *Montclair State University Digital Commons*. [Online]. Available: https://digitalcommons.montclair.edu/etd/728/. [Accessed: 01-Oct-2021].

[4] M. Isaksson, "Machine learning use case: Classifying ways to wear a face mask," *Medium*, 22-Jul-2021. [Online]. Available: https://towardsdatascience.com/machine-learning-use-case-classifying-ways-to-wear-a-face-mask-f90af8562530. [Accessed: 01-Oct-2021].

[5] Yu and W. Zhang, "Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4," Sensors, vol. 21, (9), pp. 3263, 2021. Available: https://login.ezproxy.uvm.edu/login?url=https://www-proquest-com.ezproxy.

uvm.edu/scholarly-journals/face-mask-wearing-detection-algorithm-based-on/docview/2530180785/se-2?accountid=14679. DOI: http://dx.doi.org.ezproxy.uvm.edu/10.3390/s21093263.
[6] Roman, K. (2021, June 14). *500 GB of images with people wearing masks. part 1*. Kaggle. Retrieved October 31, 2021, from https://www.kaggle.com/tapakah68/medical-masks-part1.