Face Mask Detection and Recognition in Multi-Face Images

Eric Vistnes

Abstract: The COVID-19 pandemic is the impetus for widespread research into facial and face mask detection around the world. Convolutional Neural Networks (CNN) have proven themselves to be extremely performative in this research, but many are not usable in the real world due to lack of breadth or facial detection systems that cannot adapt to face masks. There needs to be more work on facial detection with face masks, or wide-spread programs that can recognize a person regardless of if they have a face mask. This project works to expand this through a tuned facial detection system that passed in images to a performative CNN that accurately classifies the presence of a face mask. Most previous literature has problems with finding faces that are wearing masks in a group image, and so this model specifically works focused on this issue. Though the problem is still present by the end, there are possible solutions and improvements presented and proposed that could lead to usable face mask recognition models.

Introduction

The COVID-19 pandemic is the impetus for widespread research into facial and face mask detection around the world. Many countries around the globe have instituted face mask mandates for public health and safety. This has proven to be an effective protection against COVID-19 infections, and so use of face masks has proliferated to the point that most people are wearing them in public, especially during surges of the pandemic. However, this change has greatly interrupted the performance of facial recognition systems of all kinds, one of which is the relatable issue of not being able to use Face ID systems to log into cellular devices. Other surveillance systems have similarly suffered and there is no simple way to create a model to find unmasked faces for public safety.

Facial recognition has become a major topic of research and of new technology in recent years. Major companies, governments, and individuals all make use of facial recognition systems to control access to locations, phones, and more. The technology is becoming more and more ingrained in our lives with every passing year. On the other hand, the COVID-19 made much of that technology unusable as more face mask mandates were instituted. This has made it necessary to pursue new models and systems to now find accurate models.

One of the techniques that has been pursued is Convolutional Neural Networks. Convolutional Neural Networks (CNN) have proven themselves to be extremely performative in this research, but many are often not usable in the real world due to lack of breadth of training or due to facial detection systems that cannot adapt to face masks. To be widespread and useful in the real world, models need to be able to accurately classify images even when they are not the best quality, are at angles, or similarly suffer from poor facial recognition. Once we overcome these obstacles, we can not only return to the old normal of facial recognition from before the pandemic, but also implement more rigorous public safety measures through software that can correctly identify people who are not wearing a mask.

Background

Previous work and attempts have has success with creating performative models that are robust when identifying if a given face is wearing a face mask or not. A study by Talahua, Buele, Calvopiña, and Varela-Aldás showed how a Convolutional Neural Network could be trained to identify a given person, regardless of if they are wearing a mask. This model had 99.52% accuracy on the validation test set and can tell who a face is, as well as if they are wearing a face mask. However, the authors note that this model does not work on videos, struggles with faces at an angle or tilt, and must be trained extensively on each person. The model was only trained on a dataset of 10 people and needed hours of video turned into images to be trained. Image (a) showcases how the model performs.

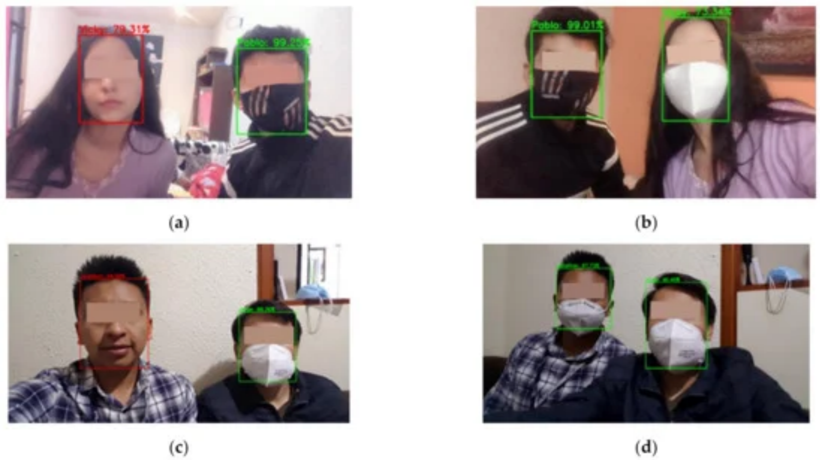

Image (a): Facial recognition system from Talahua, Buele, Calvopiña, and Varela-Aldás

Another study by Batagelj, Peer, Štruc, and Dobrišek also created a performative model that had an accuracy of 97-98.93% and could additionally classify if a face mask was worn correctly, incorrectly, or not at all. Additionally, the authors gave labelling information to each image that included race, gender, and pose to better identify bias or failures in certain populations of the model. For instance, the model is more performative on males and on front facing poses rather than side views and least performative on Asian faces.

However, while the model is both useful and performative, the authors were able to identify that the largest point of failure is in the facial recognition system used, rather than the classifier model. The paper finds that the Dual Shor Face Detector (DSFD) and RetinaFace face detection models perform the highest on the dataset, under all performance metrics. RetinaFace and DSFD were the most successful, at 99% accuracy for non-masked faces and around 85% accurate for masked faces. This is a large disparity between masked and unmasked faces, and accounts for most of the classification failures in the results of the paper.

Methods

The Data

The dataset used in this project came from Kaggle and contains 11,042 images of masked and unmasked faces, with annotations. Of these images, they are split evenly where 5,521 contain masks and 5,521 do not contain masks. Each image contains a single face, with generally similar size images. Additionally, the images in the dataset are set in various poses, angles, lighting, obstructions, and quality. Given the breadth of images, the model can more successfully classify inputs as masked and unmasked even if the images all differ in the above qualities.

The Model

Once that data was gathered and preprocessed to the same 64x64 pixel dimensions, it ran through the model. The model is a CNN formatted after the Very Deep Convolutional Network (VGGNET). VGGNet takes an RGB image and passes it through 13 convolutional layers and 3 fully connected layers. The filters are all 3x3 and the smaller filter is mitigated by the greater depth. All VGGNet based models use 3x3 filters and a large depth of shrinking layers. Image (b) shows the architecture of the full VGGNet model.

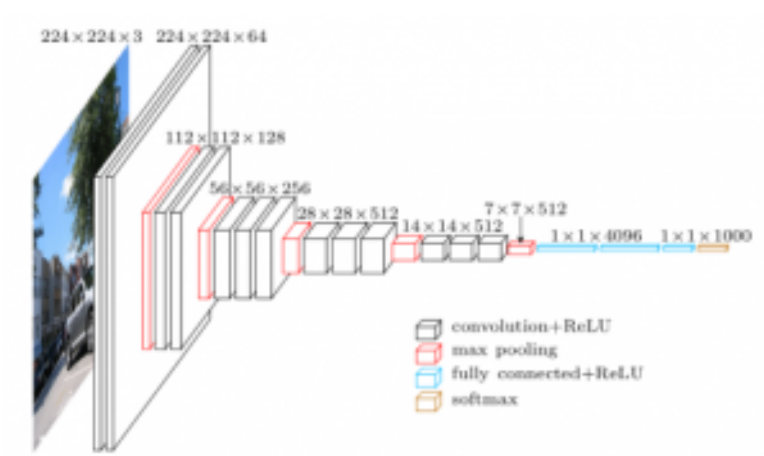

Image (b): Visualization of the VGGNet architecture

The smaller VGGNet based model in this project is built on only 5 layers- 3 convolutional layers and 2 fully connected layers. The model was trained on 80% of the data while the remaining 20% was used as a test set. After 75 epochs, the model was fully trained to classify masked and unmasked faces. It could even classify images not from the test set that contained a single face.

Creating this aspect of the model and project was more a matter of finding good data to train the model on than the struggle of creating a model architecture. In past literature, most have succeeded in training a performative model but struggled with providing input to the model. The other half of this project is creating the system that can feed input images to the model from everyday life.

Facial Recognition

This system uses the Haar Cascade, which is an algorithm that is used to identify faces in images or videos in real time. The algorithm uses edge and line detection features to identify faces. It was one of the first facial detection algorithms created and is still in use today. It is used in this project to identify faces and face locations in images with more than one person. Sadly, if it were that simple to use current facial recognition software on images of faces with face masks, then this would be a trivial problem. When run with no specifications and fine-tuning, the Haar Cascade algorithm misses some masked faces, especially if they are darker, at angles, or blurry. However, attempting to overcompensate by lowering the thresholds needed to find faces also means that the algorithm finds faces where there are none. Image (c) below is an example of both issues.

Image (b): Visualization of the VGGNet architecture

The Haar Cascade needs to be fine-tuned for each image and some images will not work properly where some faces are never detected. At the moment, it is a functional workaround for testing on group photos when the Region of Interest (ROI) bounds are expanded to find a larger face. With this change, we do not see the issues from above where faces are found where there is no face. Each face detected is surrounded by box when found and outlined in green if the person is wearing a mask and red if not. The system also writes the classification in the box.

There are multiple Haar cascade algorithms that can be used for faces in different poses, angles, expressions, and so on. In this system, we implemented a frontal face recognition to simplify the problem to only people who were facing the camera. In testing, most images could use a tuned Haar Cascade to identify each of the faces, regardless of if they wore a mask.

Results

The system performed with 99.15% accuracy on the testing set and a training set accuracy of 98.35%, which indicates that the model was not overfit, even given the high number of epochs and steps in each epoch (75 epochs of 276 steps). The high accuracy with a simple CNN indicates that the problem of identifying if a face is wearing a mask is a trivial task. However, as the previous literature on the topic has found, the mask recognition model is not the most difficult part of the system. Rather, the true obstacle is recognizing the locations of masked faces in images.

We used the Haar Cascade facial recognition and passed faces detected to the CNN to classify them. From manual testing of group images, we found that the model performed successfully. Each image needed to have its own parameters for the Haar Cascade algorithm, which means this system is not automated, but it is a proof of concept of the model working on real time data.

[INFO] evaluating network...

	precision	recall	f1-score	support
0	0.99	0.99	0.99	1120
1	0.99	0.99	0.99	1089
accuracy			0.99	2209
macro avg	0.99	0.99	0.99	2209
weighted avg	0.99	0.99	0.99	2209

Table (a): Classification Report for small VGGNet Model

Graph (a): Accuracy and Loss Plot for small VGGNet Model

The results in Image (d) and Image (e) show that each face is correctly recognized, and only actual faces are recognized. Additionally, each classification is correct.

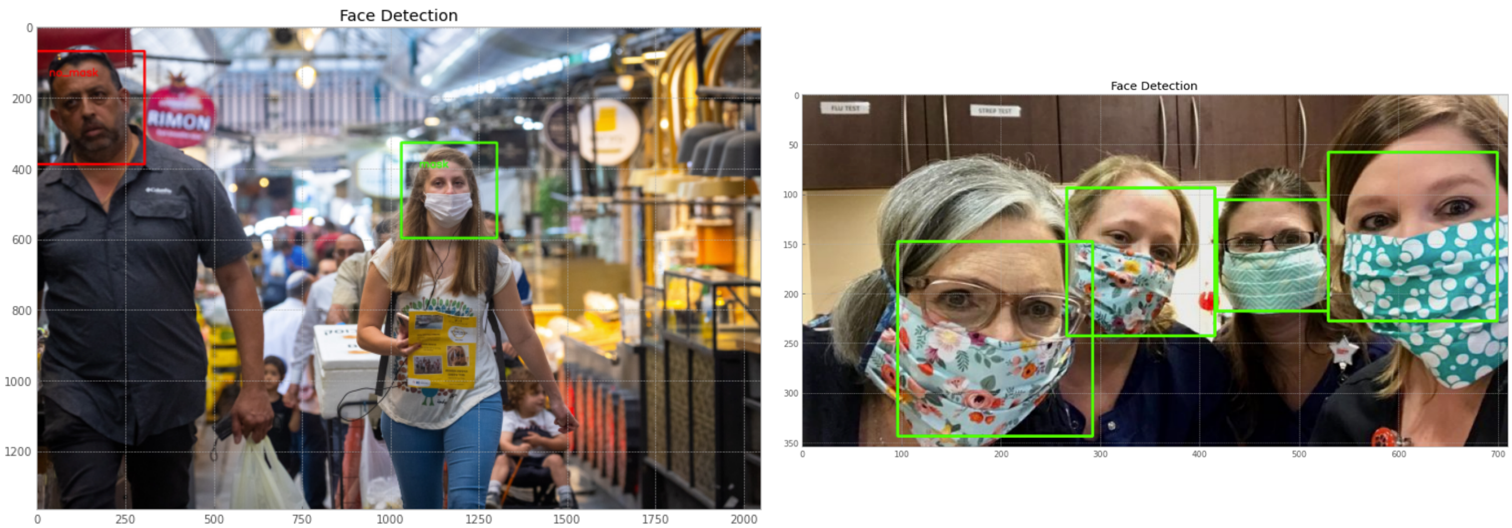

Image (d) and (e): Full usage of system to detect faces and face masks

Given that the Haar Cascade algorithm can also work in real time on a video stream, and the model itself classifies each image in as little as 200 milliseconds, this shows that the model itself could be run on real-time, real-world data once a facial detector is made autonomous.

Conclusion

Previous literature all came to the same conclusion, the current facial detection algorithms are not robust enough when competing against face masks. This project attempted to solve the additional issue that the models struggle to identify faces at an angle or with poor lighting by utilizing a dataset with more breadth of images. However, this robustness in the model does not solve the dilemma of not being able to present input to the model in the first place.

Future iterations of this work should include building an additional model to feature vectors associated with faces wearing masks. This addition to the work will allow CNNs to be fed more accurate input images of faces to classify. However, the work as it is can still be used in current technology if it is used in limited settings. These could include individual faces looking in a camera to determine if they are masked, such as FaceID.

References

Talahua JS, Buele J, Calvopiña P, Varela-Aldás J. Facial Recognition System for People with and without Face Mask in Times of the COVID-19 Pandemic. *Sustainability*. 2021; 13(12):6900. https://doi.org/10.3390/su13126900

Batagelj B, Peer P, Štruc V, Dobrišek S. How to Correctly Detect Face-Masks for COVID-19 from Visual Information? *Applied Sciences*. 2021; 11(5):2070. https://doi.org/10.3390/app11052070

Dr. Vandana S. Bhat , Arpita Durga Shambavi , Komal Mainalli , K M Manushree, Shraddha V Lakamapur, 2021, Review on Literature Survey of Human Recognition with Face Mask, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 01 (January 2021)