SPECIAL TOPICS IN AI – PROJECT REPORT

FACE MASK DETECTION FOR COVID-19

Team: Mounica Subramani, Arjun Prasanth, Vivek Kandregula **Advisor**: Hongyang Zhang

INTRODUCTION

The Covid-19 pandemic is causing a global health crisis due to its rapid spread across millions of people all around the world. A study backed by the World Health Organization found that wearing a mask over your nose and mouth can cut the chances of transmitting or being infected by the coronavirus by more than 80%. So, while healthcare professionals and scientists are learning more every day, it is essential that we follow the mask regulations diligently to effectively keep the infection rates from rising again, especially as cities and countries come out of lockdown.

However, it is not feasible to manually track the implementation of the mask policy in public places and crowded areas. Hence, we implemented a Deep Learning based system that can detect masked, unmasked, and improperly masked faces. This approach is called Face Mask Detection (FMD). The proposed model is a transfer learning model that is built by fine-tuning pre-trained state-of-the-art deep learning models such as RESNET101.

Why is it interesting and important?

With the ongoing global pandemic still negatively affecting the whole world, it is important to work on research problems that could help us automate and detect face masks due to their ease and effectiveness at controlling the spread and help in prevention of the virus. If deployed correctly, this system could potentially be used to help ensure the safety of millions of people around the world.

According to many healthcare professionals, people are expected to wear masks for up to the next 2 years and that masks will be necessary until something happens to change the viruses threat level, such as:

- Society developing widespread immunity.
- Widespread implementation of a successful vaccine, or the emergence of a cure of sorts, such as an antiviral drug.

We know that there has been an increased number of data science use cases for the covid-19 situation. Most of them are concentrated on predicting the number of cases, patterns, mortality rate, etc., but the approach of detecting face masks is not widely explored. Considering the longevity of this research project's application, its positive impact on society, and the scope of exploration of multiple deep learning approaches we have chosen to work on this research project.

Why hasn't it been solved before?

Since the requirement for a face mask detection model is fairly new, there is not much research and literature available regarding the approaches that would provide good results. So, our primary goal is to explore multiple transfer learning models (Simple RESNET101, Simple RESNET101 with data augmentation, Extended RESNET101, Extended RESNET101 with data augmentation), choose a metric that best encapsulates the success of our model and perform a comparative research based on the selected metric to learn about the best performing models and the reason a model is performing better than the other models.

How is our work different?

Most of the research that has been conducted on this research problem has only been performed on small datasets that consist of data less than 5000 images or on massive datasets that were generated by artificially simulated images. Our project presents the results obtained from utilizing the world's largest real-world face mask detection dataset that has been made publicly available recently.

ABOUT THE DATASET

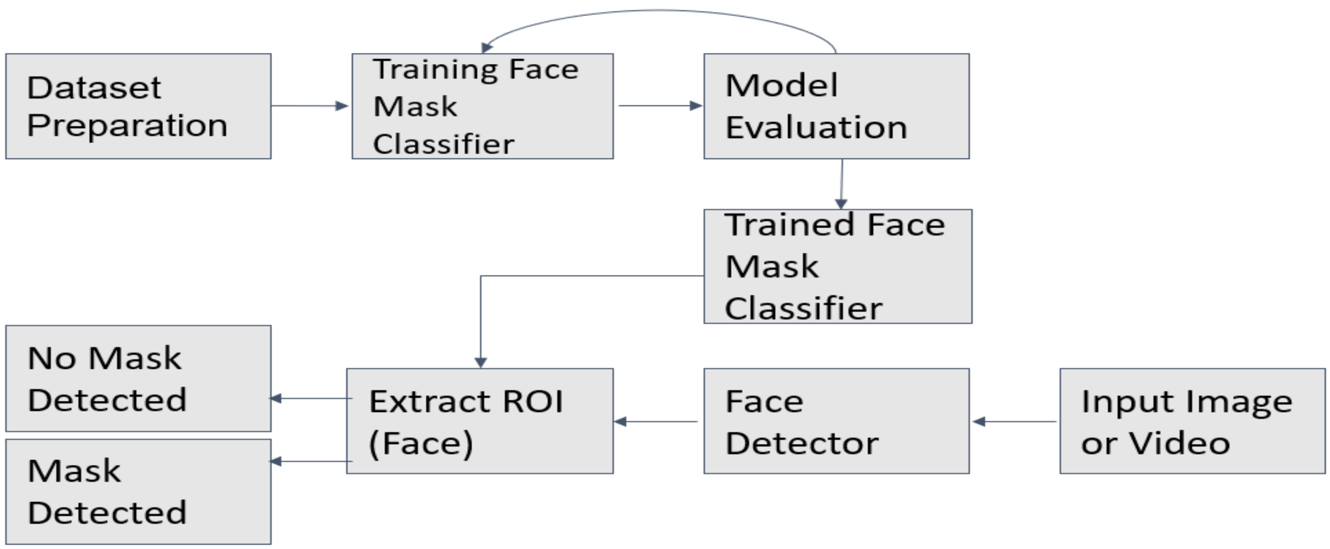
Before diving into the details of the architecture and approaches taken let's discuss the dataset. We worked with a public dataset known as the Real-world masked face recognition dataset, which was generated by the researchers at Wuhan University and it is currently the world's largest real mask occlusion dataset. The real-world masked face recognition dataset (RMFRD) contains 5,000 masked faces of 525 people and 90,000 normal faces. Our data set is highly unbalanced (5,000 masked faces VS 90,000 non-masked faces)

Link: https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset

Few samples from the RMFD dataset

The authors of this literature claim that this dataset is currently the world's largest real-world face mask image dataset. This was the reason why we selected this dataset for our project.

ARCHITECTURE



DATASET PREPARATION

Before we get into the modeling aspect of our architecture, we must preprocess the data to obtain clean and reformatted data that is easy to train and test. The dataset is present in two folders that consist of images of people with masks and images of people without a mask, we iterate over all of the images in both of the folders label the data as 1 for masked image and 0 for unmasked images. Once we complete the labeling process, we perform a stratified split in the data and obtain the training and testing set where 70 % of the data is split into the training set and 30 % of the data is split into the testing set. By performing a stratified split, we ensure that the train and test sets have approximately the same percentage of samples of each target class as the complete set.

TRAINING FACE MASK CLASSIFIER

We have trained multiple models to compare their performance and these were the best performing models:

- Simple RESNET101: Our first approach was to use Transfer Learning to train the model. We started with RESNET101. We did a simple fine-tuning where we removed the last fully connected layer and replaced it with a layer matching the number of classes in the dataset.
- Extended RESNET101: We also tried adding a few more layers for the fine-tuning to see how it performed. Therefore, we removed the last fully connected layer and replaced it with a Linear Layer, ReLU activation, Dropout and a Linear layer matching the number of classes in the dataset.

To further understand if the performance would improve, we re-trained the above models with data augmentation, as it would increase the diversity of data available for training models and will enhance the size and the quality of a given dataset. To perform data augmentation, we used transformations such as:

- Horizontal Flip Augmentation: where we reverse the entire rows and columns of an image pixels randomly and horizontally.

- Random rotation augmentation will randomly rotate the images from 0 to 360 degrees in clockwise direction.
- Adjusted the brightness, saturation, contrast and hue of the images.
- Shearing: is the process the image is distorted along an axis , mostly to create or rectify the perception angles. It was used to augment images so our model can see the image as humans see things from angles.
- Normalization: to convert an image into a range of pixel values based on their intensities which helps increase the training speed of our model.

We have also performed hyper parameter tuning to find the optimal combination of hyperparameters that minimize the loss function to obtain better results. We obtained the best results when we used stochastic gradient descent as the optimization function and set the learning rate at 0.001 where cross entropy was a loss function.

MODEL EVALUATION

To evaluate the models after training we have used multiple metrics such as:

- Sensitivity (True positive rate): the proportion of actual people wearing masks that got predicted as wearing a mask

  Sensitivity = (True Positive)/ (True Positive + False Negative)

- Specificity (True negative rate): the proportion of people classified as not wearing a mask out of all of the people that were actually without a mask.

  Specificity = (True Negative)/ (True Negative + False Positive)

- Precision: the proportion of people who were classified as wearing a mask and were actually wearing a mask.

  Precision = (True positive)/ (True Positive + False Positive)

Where,

- True positive is the count of people that were classified as wearing a mask and were actually wearing a mask
- False negative is the count of people that were classified as not wearing a mask but were actually wearing a mask
- True negative is the count of people that were classified as not wearing a mask and were actually not wearing a mask
- False positive is the count of people that were predicted as wearing a mask but were actually not wearing a mask

Since, our dataset is highly imbalanced with the proportion of masked vs unmasked people being 1:18. Instead of calculating the F1 score we used the Matthews correlation coefficient to understand the performance of our model as it is a more reliable metric which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories, proportionally both to the size of positive elements and the size of negative elements in the dataset.

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Using these metrics, we evaluate the performance of our model on the test set. If we felt that the performance could've been better, we went back to work on our model for hyperparameter tuning.

RESULTS ON TEST SET

We tested the metrics mentioned in the model evaluation section to understand and compare all of the models we have trained in the previous section.

	Specificity	Sensitivity	Precision	Matthews Correlation Coefficient
Simple RESNET101	99.78	**69.94**	66.85	51.93
Simple RESNET 101 with Data Aug.	99.62	25.24	43.09	32.41
Extended RESNET101	99.87	59.34	80.66	**68.96**
Extended RESNET101 with Data Aug.	**99.9**	5.82	**86.74**	21.04

From the table, we can see that all the models have high specificity. This tells us that our model is able to identify faces without masks really well. The comparatively low sensitivity and precision tells us that our model is not able to identify faces with masks as well. But since the main aim of this problem is to identify people without masks correctly, we feel that our model hits the brim with its performance. The reason we feel for the low sensitivity and precision is due to the fact that the masked face images are a minority class and hence, our model would immediately show better performance with more data. However, ideally, we would also like to have good sensitivity and recall.

FACE MASK DETECTION PHASE

We pass real-world images/videos as input to our MTCNN which is a pretrained face detection model. Multi-task Cascaded Convolutional Networks (MTCNN) is a framework developed as a solution for face detection. The process consists of three stages of convolutional networks that are able to recognize faces and landmark locations such as eyes, nose, and mouth to identify the region of interest.

Problems such as detecting multiple faces, correctly distinguishing between a face and a table are taken care of by using state-of-the-art object detection models like the MTCNN. We use this model to then extract the Region of Interest(ROI), which is the face in our case, from these real-world images/videos and feed it to our trained face mask detection model. The prediction given by our face mask detection model will be the final output for the system.

CONCLUSION

In this project, we proposed an approach that uses computer vision and RESNET models to help maintain a secure environment and ensure individuals protection by monitoring public places to avoid the spread of COVID-19. The Face Mask detection (FMD) system will operate in an efficient manner in the current situation to track the individuals, if there are following CDC guidelines by wearing face mask or not. This approach can be employed in various sectors for effective health and safety monitoring. This approach can be integrated with social distancing detector as part of future scope.

FUTURE WORK

Extending the existing classifier to three class classification as “masked”, “unmasked”, and “improperly masked” face. Another interesting approach is to use some single shot object detection technique to identify any kind of head, in any possible angle, with labels like “front face, with mask”, “front face, without mask”, and “face not visible”. We need to track heads instead of just faces, to have a continuous movement even if the person looks in the opposite direction than the camera. This overall approach might not be as robust than the current one, but would probably be simpler and faster, and it might be a good idea for devices with limited resources.

Further, The Face mask detection can be integrated with social distancing detectors. Social distancing detectors can be implemented by applying object detection to all people in the image/video input, computing pairwise distance between all detected people, based on these distances, we can check if any two people are N pixels apart.