Monitoring compliance to face mask regulations

Increasing the effectiveness of COVID-19 preventive measures

Springboard Capstone 2 Project Presentation
September 2021

INTRODUCTION

While authorities enforce the mandatory use of face masks as a preventive measure against COVID-19 transmission, “one can observe many cases of half-compliance or sham compliance.” Many would wear masks, but “slide them down onto their chins or take them off completely while talking to someone on the street or speaking on the phone”. In other cases, the mask is not properly positioned, exposing the nose, mouth or chin.

Such widespread non-compliance is severely undermining the effectiveness of measures put in place and to the extent it provides a false sense of confidence to the public, stricter adherence to the regulation need to be implemented especially with more infectious variants spreading.

The project will explore the feasibility of using computer vision to identify whether a person is wearing a face mask and if so, whether this is being worn properly.

PROBLEM

Improper use and constant removal of face masks hamper effectiveness of measures and can even be detrimental if it provides a false sense of confidence to the public

SOLUTION

Leveraging on computer vision for constant and automated monitoring of *proper* compliance to face mask regulations

THE PROJECT

AIMS

The main objective is to develop an image classifier which can identify whether the subject is:

- NOT WEARING A MASK
- CORRECTLY WEARING A MASK
- INCORRECTLY WEARING A MASK

To prove its potential for production, it is crucial that the model achieves a high level of accuracy across all the three classes. For this project, a minimum threshold of 90% accuracy for each class is used.

SCOPE

Within the scope

- ✓ incorrectly worn mask defined as not covering the mouth or nose
- ✓ individual subjects only

Deliverables

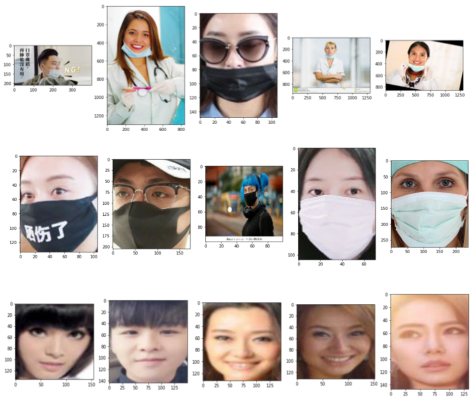
- ✓ image classifier model taking an image as input and outputs a label
- ✓ report

Out of scope

- X No bounding boxes / location of mask
- X Does not account for the type of mask worn
- X Deployment considerations

THE DATASET

Due to challenges in obtaining a sufficiently large sample size for incorrectly worn masks, various sources have been combined to create a more balanced dataset.



FaceMaskDetection[1]

a collection of real images from Bing Search API, Kaggle datasets and RMFD dataset

MaskedFace-Net[2]

images of real faces based on the dataset Flickr-Faces-HQ (FFHQ), edited to exhibit improper use of face masks (uncovered chin, nose and/or mouth)

Aliyun[3]

a collection of augmented real images showing incorrectly worn masks; only the original pictures were used for the project

[1] https://github.com/chandrikadeb7/Face-Mask-Detection/tree/master/dataset
[2] https://github.com/cabani/MaskedFace-Net
[3] https://tianchi.aliyun.com/dataset/dataDetail?dataId=93724

MODELLING STEPS

01
IMAGE PRE-PROCESSING
- RGB to BGR
- Zero-centred by mean pixel
- Rescaling

02
ESTABLISHING BASELINE
- 3-layer VGG style architecture

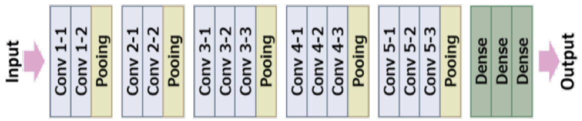
03
MODELLING
- Transfer learning using pre-trained models (VGG16, ResNet50, EfficientNetB0, Inception V3)
- 50 epochs with early stopping
- Adam optimiser

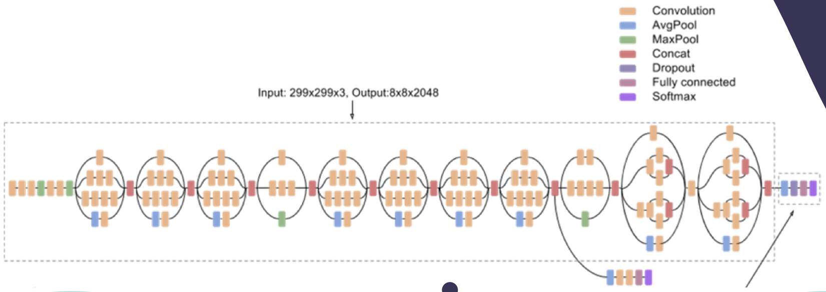
04
FINE-TUNING OF SELECTED MODEL
- Unfreezing top n layers
- Hyperparameter tuning

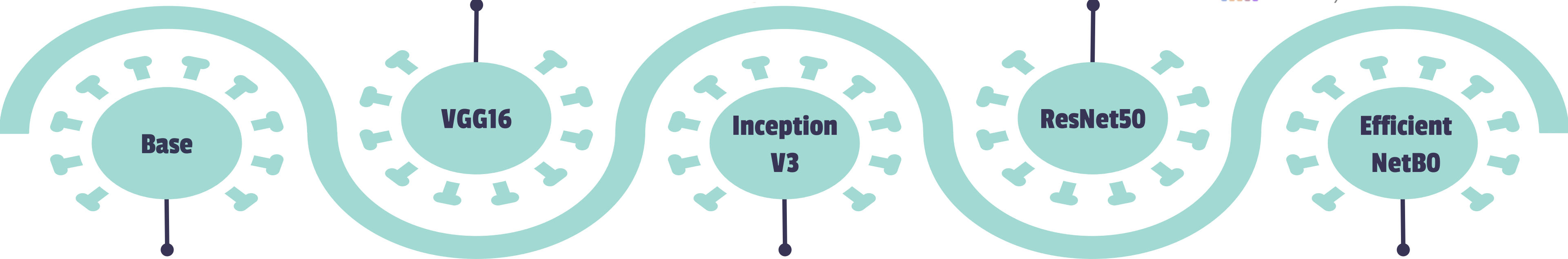
05
EVALUATION ON TEST
- Confusion matrix, accuracy, precision, recall

MODEL ARCHITECTURES

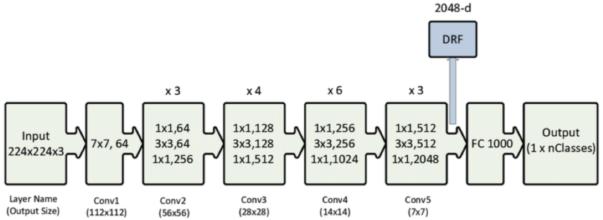
Leveraging on transfer learning





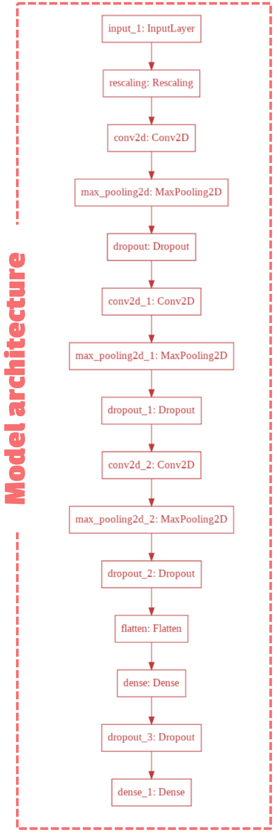


3-layer VGG style



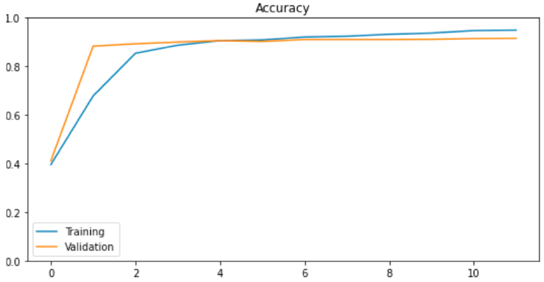
THE BASELINE

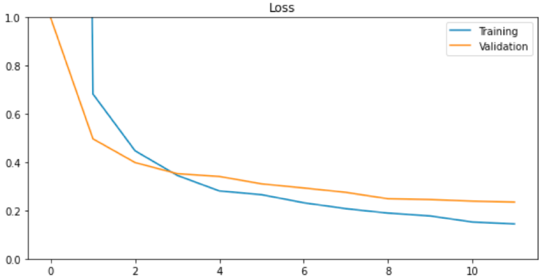
Model architecture



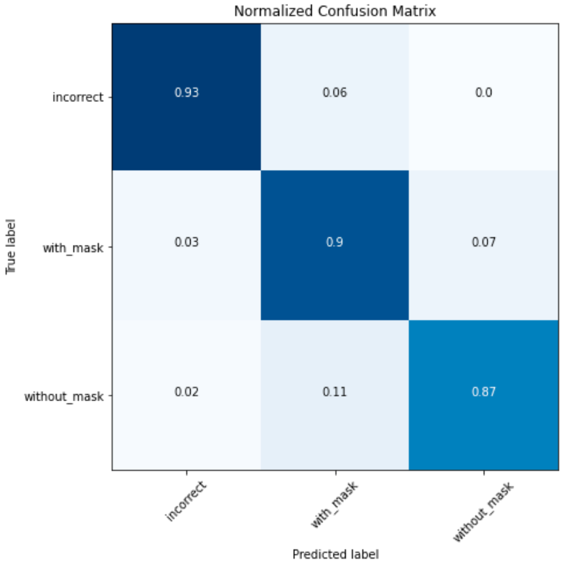
Model Training

- With early stopping, training ended on the 12th epoch as the model reached a 91.57% accuracy on the validation set.
- Model improvement started to plateau as shown in both charts below.





Model Performance

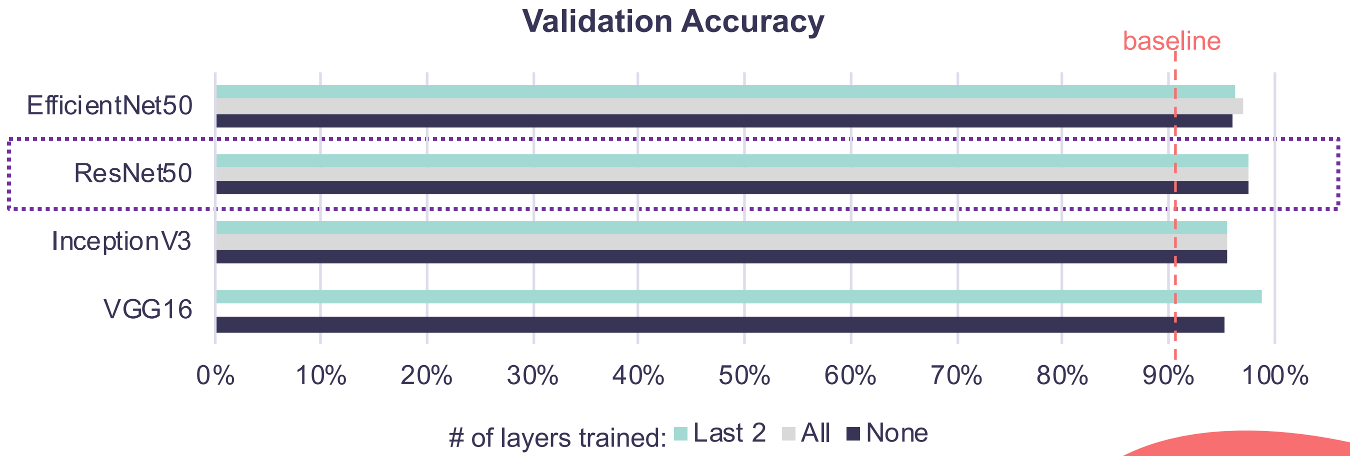


- While achieving a decent performance of 90% accuracy, classification of subjects without masks requires further improvement as 13% are erroneously predicted as having masks.

ARCHITECTURE PERFORMANCE

For each model architecture, a separate model has been trained on the ff. basis:

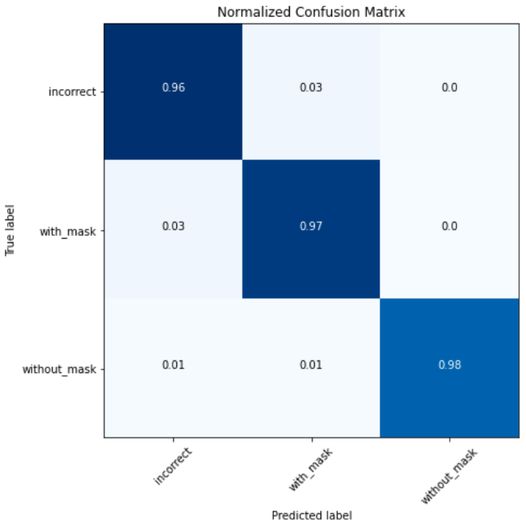
- Using the pre-trained model as a feature extractor using the weights learned on the ImageNet dataset
- Reconfiguring the pre-trained model to the dataset by unfreezing and retraining only the top 2 layers
- Training and learning the weights from scratch (unfreezing all layers of the model)



Transfer learning has enabled us to achieve superior predictive performance, surpassing the baseline accuracy of 91%. Whilst all model architectures showed high accuracy, ResNet50 is ahead by a slight margin so this model architecture has been chosen for fine-tuning.

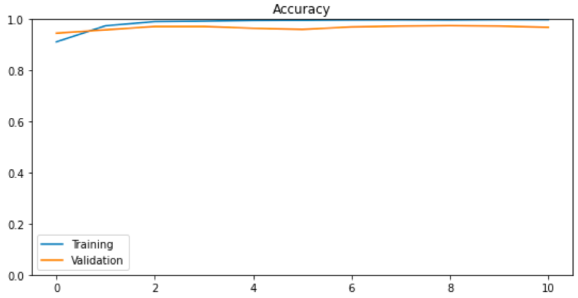
THE FINAL MODEL

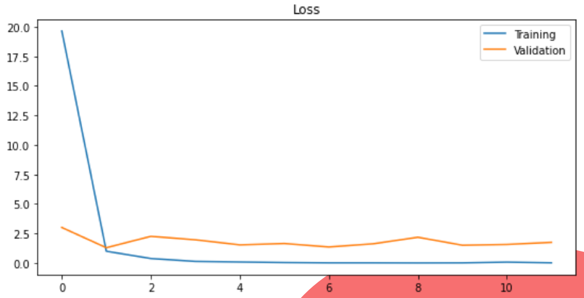
Model Performance



- 97% accuracy
- Top performance on identifying subjects without masks
- The error mainly comes from misclassifying between the correctly worn masks and incorrectly worn masks

Model Training





- All layers unfrozen
- Learning rate = 0.001

LIMITATIONS

- While I was able to obtain enough images for the improper use of face masks, majority of these were synthetic, edited surgical face masks overlaid on the subjects' faces. This raises possible challenges in the generalizability of the model, specifically in detecting *non-surgical* incorrectly worn masks. Additionally, the images used were not as diverse in terms of age and race. Hence, it is advisable to test the performance of the model against these possible weaknesses.

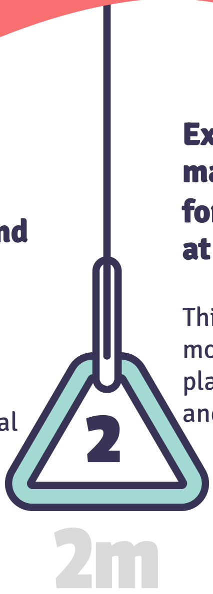
- The model was also trained on images of individuals only so the application of the model does not extend to multiple subjects per image.

NEXT STEPS

Transfer learning has been key to unlocking superior model performance. With an average accuracy of 97% on the test set, the model demonstrates the potential of using computer vision to enforce stricter compliance to face mask regulations.

Test model generalisability and and diversify the training set as required

(following the potential limitations mentioned earlier)



Extend the model to make classifications for multiple subjects at one time

This will be valuable in monitoring crowded places like train stations and airports.

Deployment considerations

With future deployment in mind, a follow-up study can focus on making predictions in real-time as opposed to using images. Speed of the model also need to be considered.

Thank you.

Feel free to connect with me!

linkedin.com/in/nikki-amurao

github.com/nikamurao

amurao.frances@gmail.com

Any questions?