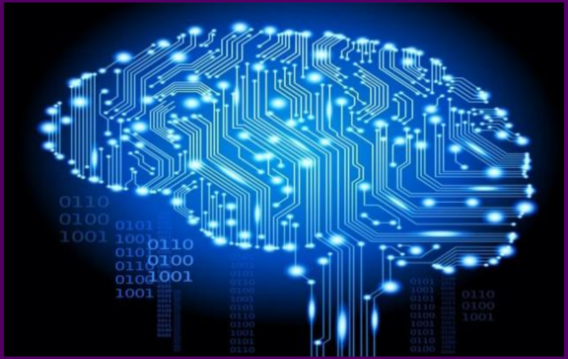
Deep Learning
MSiA 432



Mask On - Face Mask Detection

NORTHWESTERN
UNIVERSITY

NORTHWESTERN UNIVERSITY

MSIA 432: Deep Learning. Spring 2020.
Instructor: Dr. Ellick Chan. TA: Adam Sandler

Mask On - Face Mask Detection

Summary:

Using image data, our project aims to detect whether a person in the image is wearing a face mask or not, and if the image subject is wearing a mask, whether it is a normal protection mask or a N95 mask.

We define face masks as cloth masks, surgical masks and N95 respirator masks.

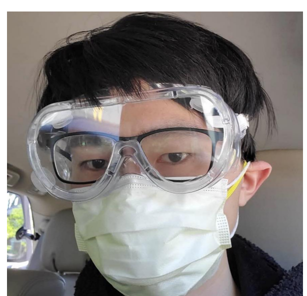
Group members:

Xuefei Liu

Ellie Tan

Aria Wang

Jue Wang



Joe Zhang

Zach Zhu

Problem Statement

- The COVID-19 pandemic is spreading virally and affecting people's daily life greatly. Wearing a mask in public will help prevent people from spreading the virus
- Our project detects whether a person in the image is **wearing a face mask or not**, and recognizes the **mask type** (N95 or non-N95) if the image subject is wearing a mask.

- By implementing our project, local governments can easily detect the mask rate from CCTV footage and be prepared for potential outbreaks. Automated public services, such as parking lots, can better control any risks.

Dataset

Data sources:

- *MAFA (for masked human faces)*
- *WIDER FACE (for not-masked human faces*
- *Kaggle (for natural scenes and animal images)*
- *Online collecting (for N95 images)*







The dataset comprises 4-class images: *no human face, human face without mask, human face with mask, human face with N95 mask*.

- All images were transformed into the same size (64*64) and normalized based on RGB scales. The class sizes are as follows:

	No human faces	*Human faces w/o masks*	*Human faces w/ non-N95 masks*	*Human faces w/ N95 masks*	**Total**
Training	1400	2321	2293	168	6182
Test	400	780	805	56	2041

NORTHWESTERN UNIVERSITY

Dataset

Data cleaning procedure:

- Checked image quality and ensured classification labels were correct.

Additional Data Labelling:

- Used *LabelImg* to label the locations of faces and faces with masks in all the images
- Stored the location and label information in separate xml files

Challenge:

- The big image size of the training set slowed down the data loading process
- Limited number of N95 mask images available online
- Limited time and computational power (each epoch sometimes took an hour to train, and there were times when our team failed to train the network on either the deepdish server or Google Colab)

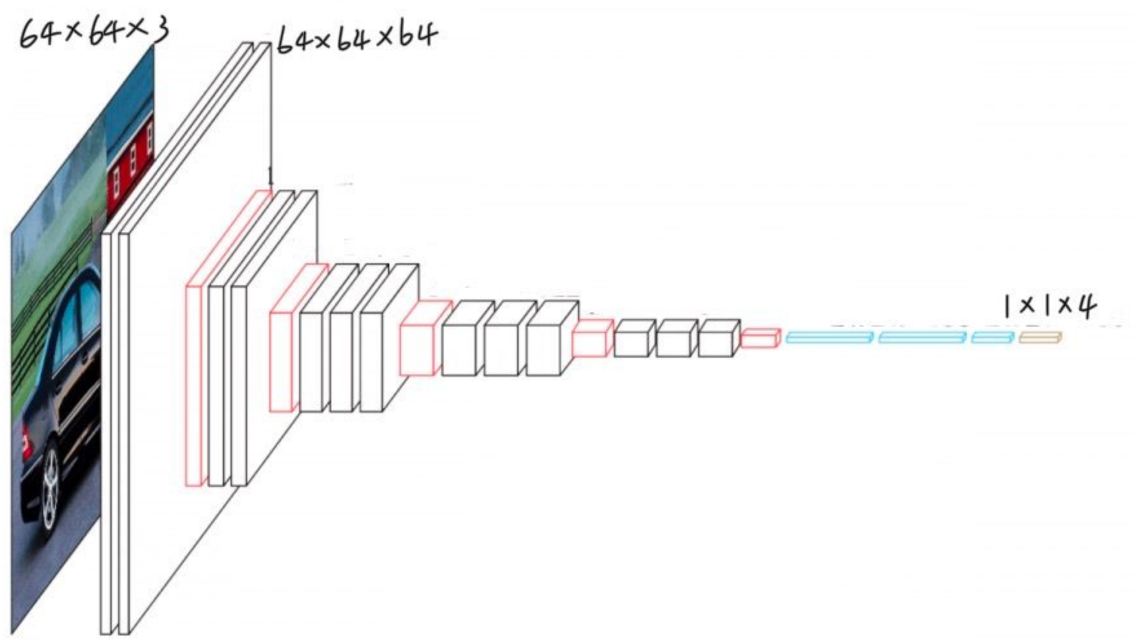
Technical Approach (VGG16)

Model Settings

- Number of VGG blocks = 5
- Enable transfer learning up to layer 2
- Size of fully connected (FC) layers = 128
- 1 FC layers

Optimizer settings

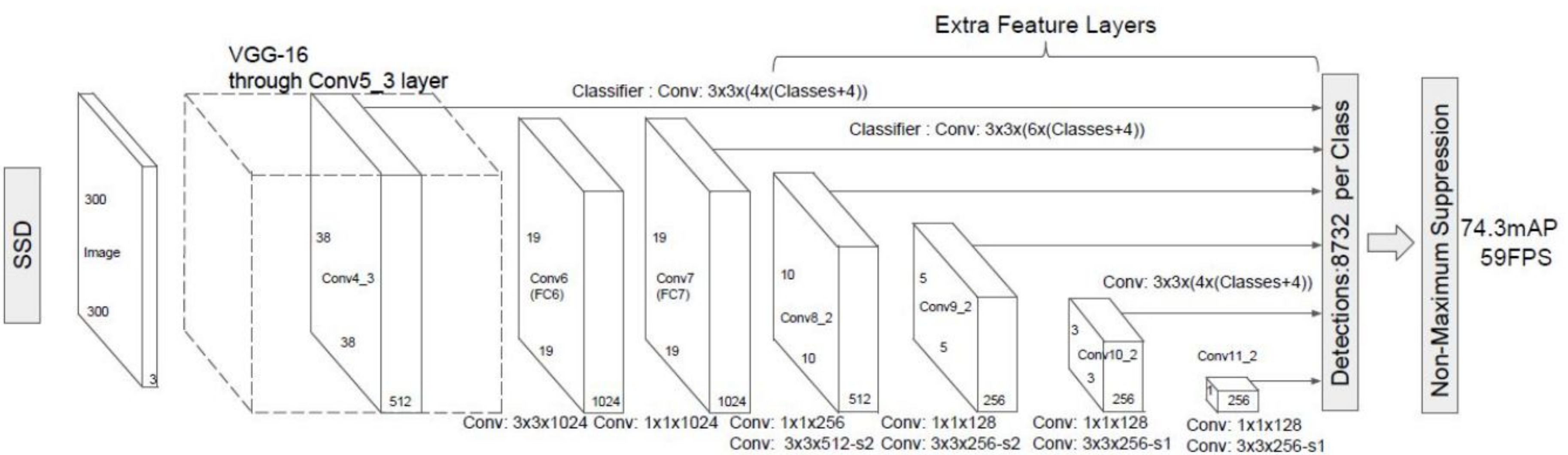
- optimizer = Adam()
- Batch size = 32
- Number of epoch = 196
- Batch normalization



NORTHWESTERN UNIVERSITY

Technical Approach (SSD)

- SSD (Single Shot MultiBox Detector)
- Using pre-trained VGG16 (structure modified, excluding FC layers) as its backbone to extract features
- SSD used VGG16 without FC layers allows for input images of different sizes
  - **3 Object Classes** (Face, Face_Mask, N95) and **1 Background Class**
  - Outputting **(Xmin, Xmax, Ymin, Ymax)** as bounding boxes based on feature map dimensions

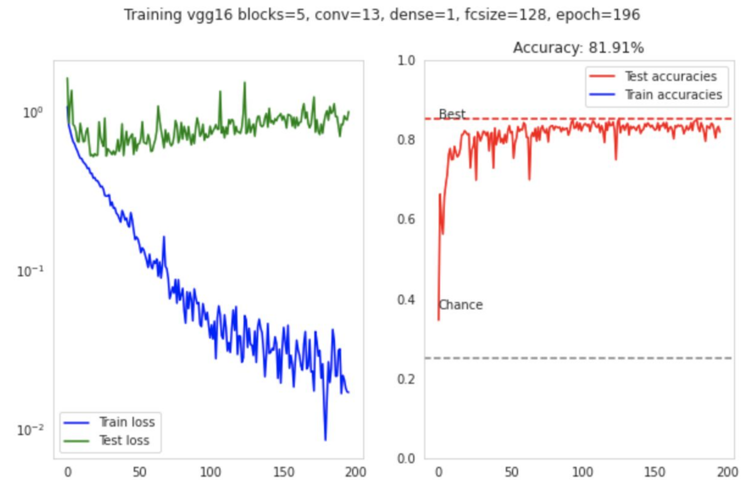


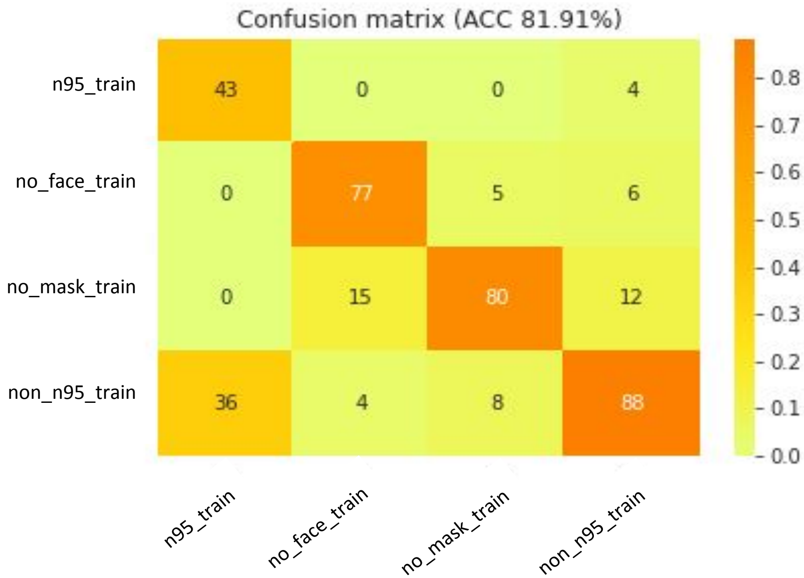
NORTHWESTERN UNIVERSITY

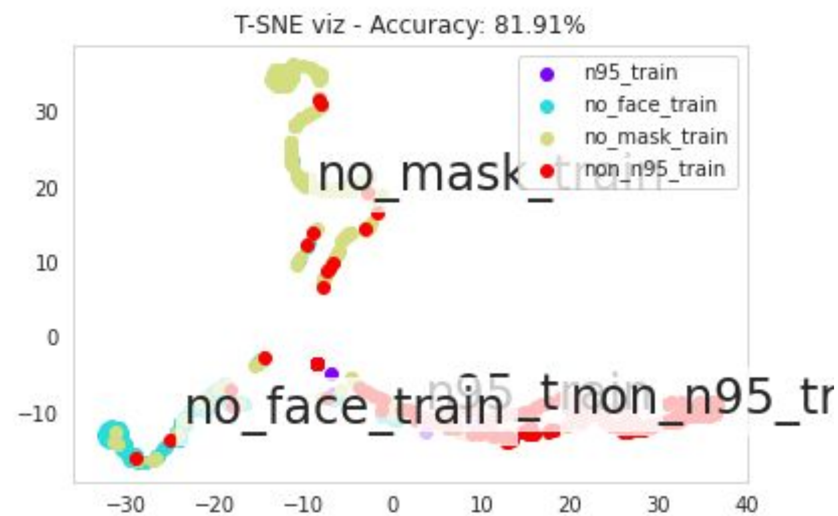
Results (VGG16)

The network accuracy almost stabilizes at 196 epochs

- Overall accuracy: **81.91%**
- High accuracy on classes of *no face*, *no mask* and *non-N95 mask*, but low accuracy on *N95 mask* images





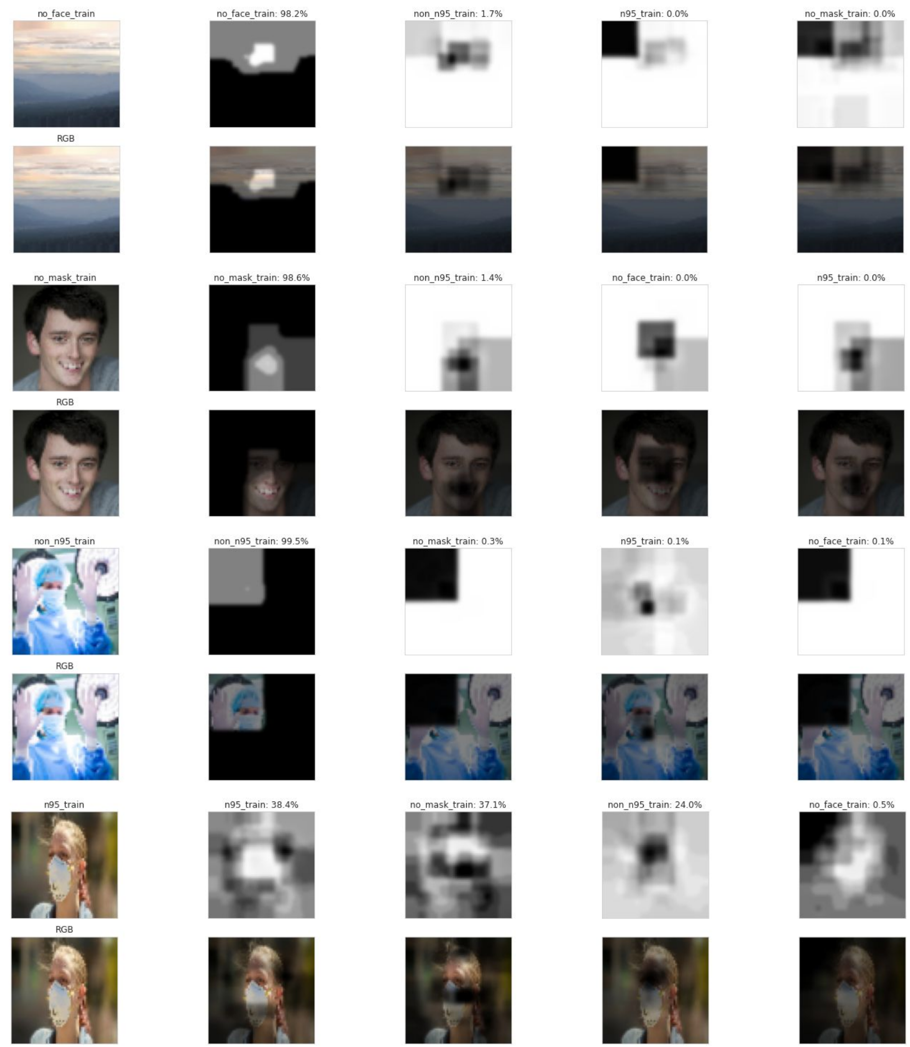


NORTHWESTERN UNIVERSITY

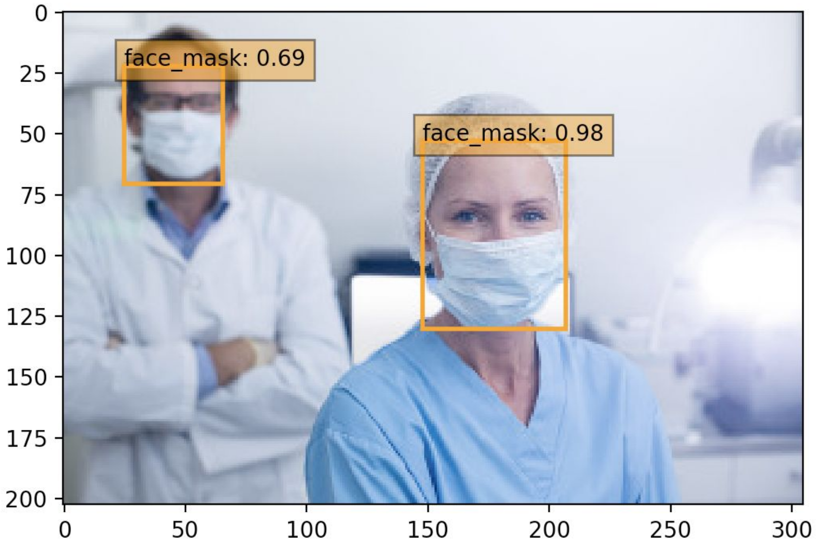
Results (VGG16)

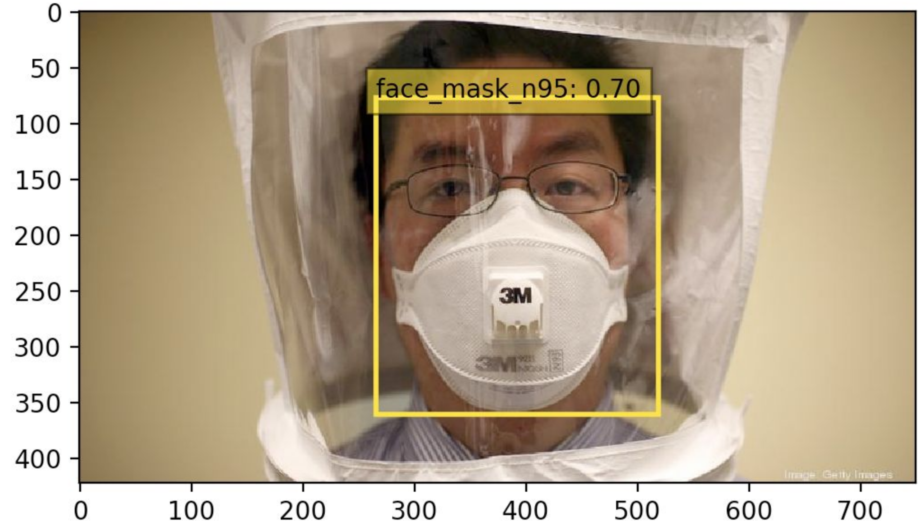
Heatmaps for correct classifications

- *No human faces*
- *Human face without mask*
- *Human face with non-N95 mask*
- *Human face with N95 mask*



Results (SSD)





	AP for *Face*	AP for *Face Mask*	AP for *N95 Face Mask*	MAP
SSD (*Mask / No Mask*)	0.249	0.821	N/A	0.535
SSD (*Mask / N95 / No Mask*)	0.285	0.811	0.7385	0.612

NORTHWESTERN UNIVERSITY

Summary

Observations

- Within limited time and hardware conditions, we have tried multiple combinations to find so far the best model which yielded the best performance. Our network has a decent performance of detecting human faces and whether the image subjects are wearing masks (especially when background is not too noisy)
- The model falls short for detecting N95 masks due to insufficient training data
- When identifying facial masks, the network learns to detect the absence of facial features on the lower part of human faces. The network distinguishes surgical and N95 masks by looking at their colors and their shapes around the subject's nose.

Limitations of the approach

- Low accuracy on detecting human faces/masks for faces not directly facing the camera
- Low accuracy on differentiating between regular surgical and N95 mask

Potential improvements

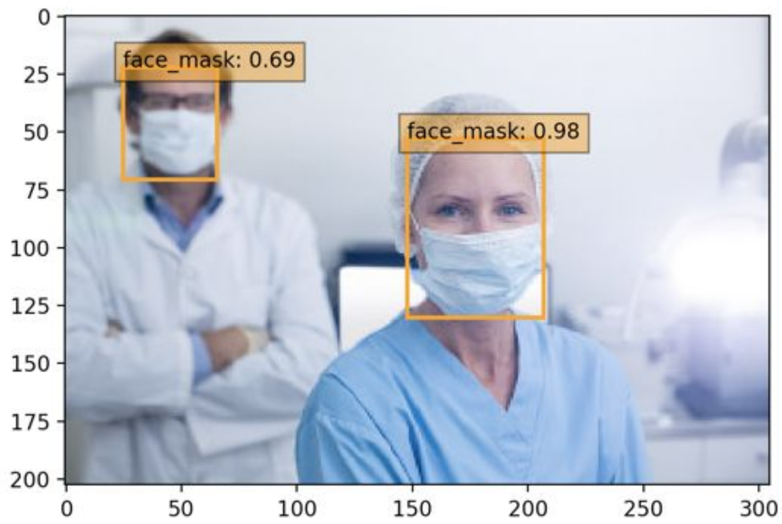
- Train the network on more images of human faces/masks facing different directions
- Increase the number of images of the N95 mask in training data
- Increase model complexity
- Add features on taking streaming input data in addition to static images and making real time detection

NORTHWESTERN UNIVERSITY

References

1. Detecting Masked Faces in the Wild with LLE-CNNs
   http://www.escience.cn/people/geshiming/mafa.html
2. WIDER FACE: A Face Detection Benchmark
   http://shuoyang1213.me/WIDERFACE/
3. FaceMaskDetection by AIZOOTech
   https://github.com/AIZOOTech/FaceMaskDetection
4. PyTorch Implementation of SSD
   https://github.com/amdegroot/ssd.pytorch

Live demo


face_mask: 0.69
face_mask: 0.98
0
25
50
75
100
125
150
175
200
0
50
100
150
200
250
300