Behind the Mask: What Photos on Twitter Reveal

By Melissa Anthony,
Tim Rauen
and
David Sutton

Latests masks in the headlines:

- NPR - [Joe Biden: For The Next 3 Months, All Americans Should Wear A Mask When Outside](#)
- VOX - [Do you really have to wash your mask after every use? Short answer: Yes.](#)
- CNN - [Think your mask makes you invisible to facial recognition? Not so fast, AI companies say](#)
- Politico - [Biden, Harris call for all states to mandate masks after first joint Covid-19 briefing](#)
- USA Today - [Florida sheriff forbids employees, visitors to wear masks: If they do, 'they will be asked to leave'](#)

Data science problem:

Can mask wearing in photos on Twitter help predict COVID infection rates, giving hospitals and medical workers insight to prepare for spikes.

The data:

- Created a Twitter crawler using Tweepy
- Filtered by certain hashtags:
  - #nomaskselfie, #maskup
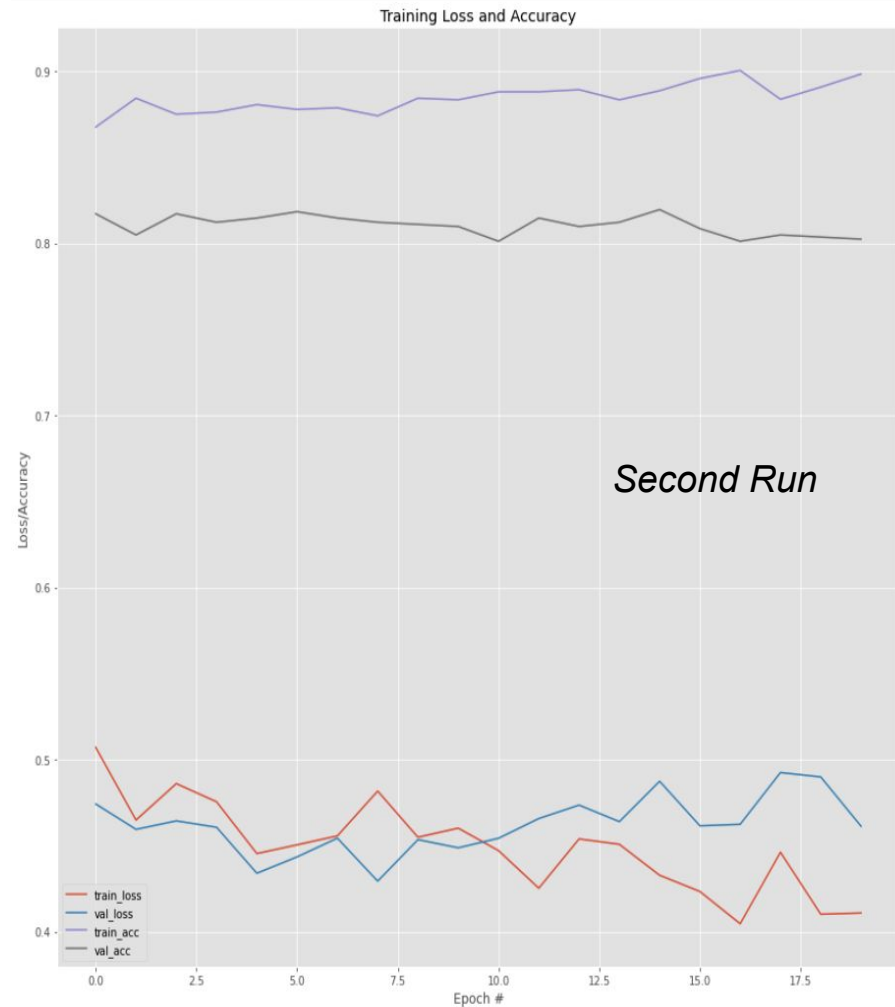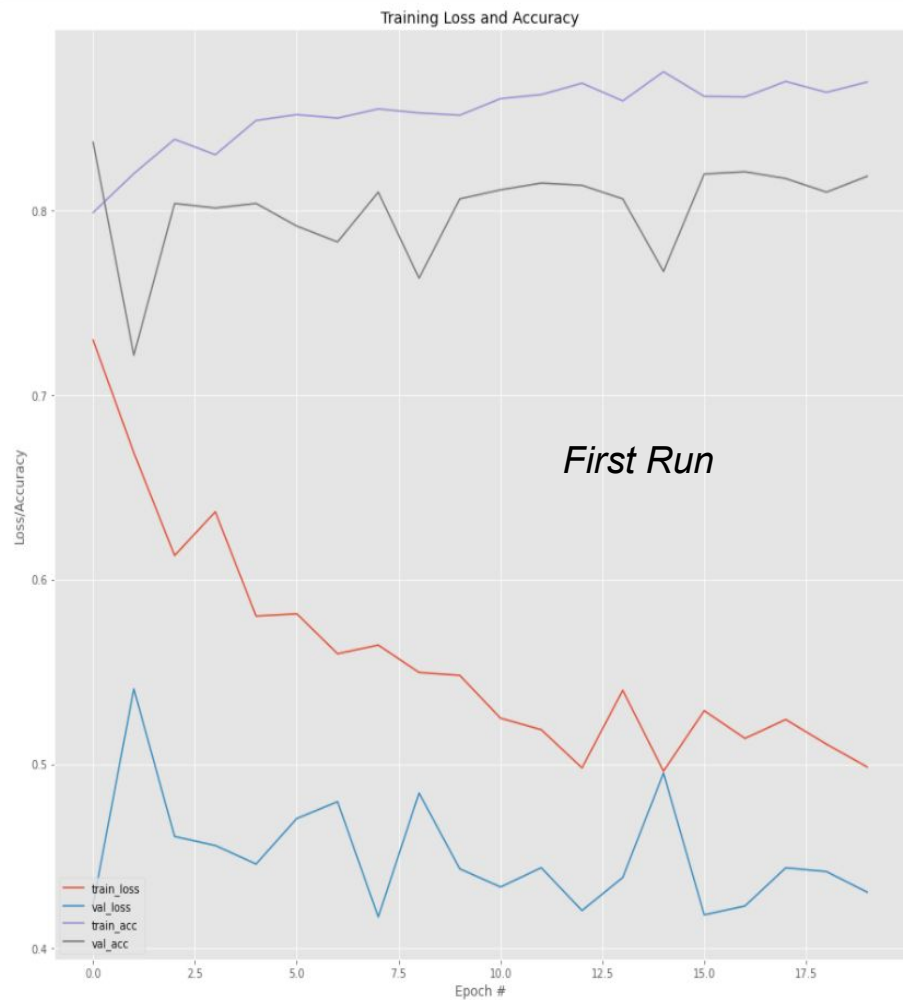- Manually filtered out photos that did not contain human face

The model:

- Obtained data from Kaggle competition to classify photos into two groups:
  - Wearing a mask
  - Not wearing a mask
- Borrowed code from Mirza Mujtaba's notebook on Kaggle
- Used Keras to train neural network
 - Able to achieve 89% accuracy on 20 epochs
- Made some slight alterations

TensorFlow

K Keras

Training Loss and Accuracy

First Run

Training Loss and Accuracy

Second Run

Model metrics:

```
Epoch 19/20
101/101 [==============================] - 360s 4s/step - loss: 0.4102 - accuracy: 0.8908 - val_lo
ss: 0.4900 - val_accuracy: 0.8037
Epoch 20/20
101/101 [==============================] - 323s 3s/step - loss: 0.4118 - accuracy: 0.8984 - val_lo
ss: 0.4614 - val_accuracy: 0.8025
```

	precision	recall	f1-score
0	0.58	0.45	0.50
1	0.86	0.91	0.89
accuracy			0.82
macro avg	0.72	0.68	0.70
weighted avg	0.81	0.82	0.81

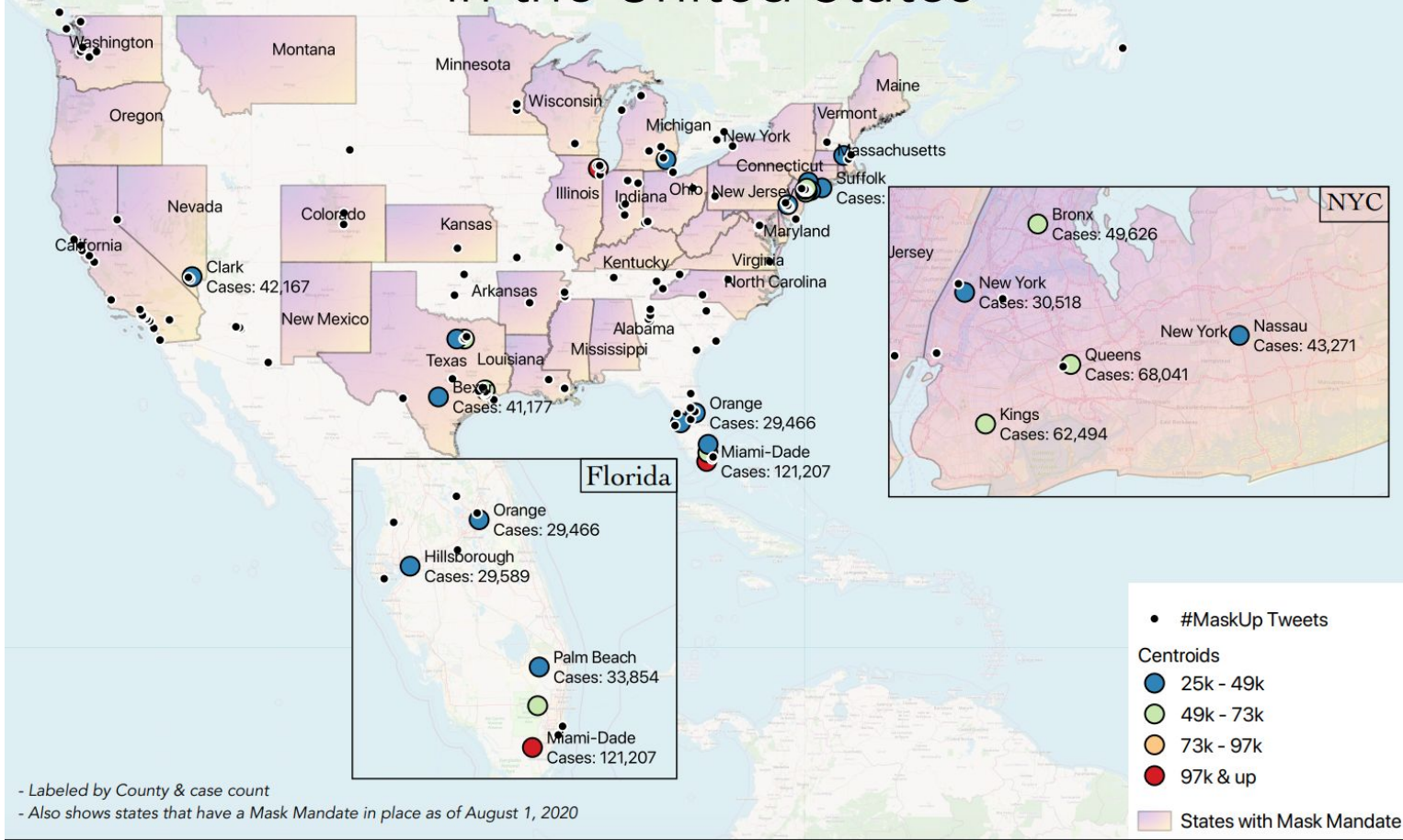Predictions on Twitter:
Mask vs NoMask
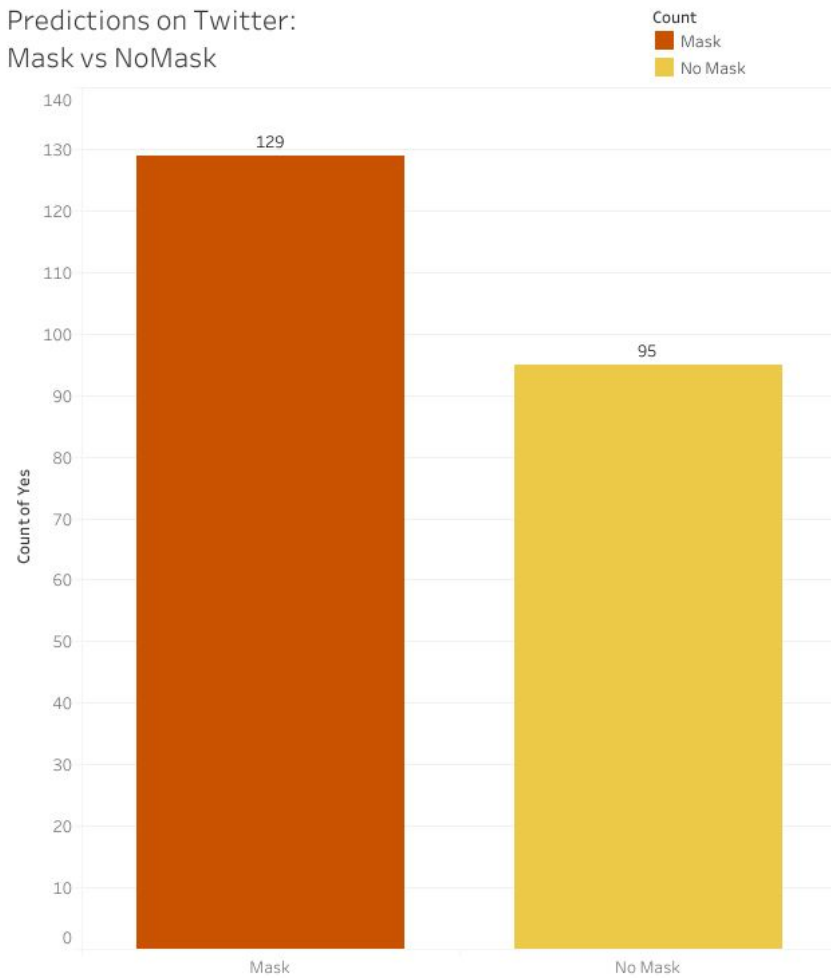
Conclusions:

Some challenges to the project:

- Conversion of the images proved challenging at times during preprocessing and conversion to/from an array.
- Running a neural network, though faster, can still be labor intensive. Especially since it needs to run multiple times to learn.

Next steps:

1. Find additional data/photo sources with geotags included which will allow us to add depth to the analysis, pin down specific locations and make inferences about the rate of virus transmission.
2. Predict dates US locales will reach specific infection thresholds.
3. Fine-tune the model to increase its usefulness and accuracy.
4. Review entire process at a later date with the benefit of knowing how the pandemic plays out, and compare that knowledge with the predictions made at the time of model construction.
5. Pitch to government agencies (especially hospitals) as a prediction tool for estimating the number of cases to expect and prepare for based on mask-wearing on social media.
6. Perform analysis on hashtags and discover which are the best predictors of mask-wearing (perhaps using NLP).

Severe Covid Cases as of August 1, 2020 in the United States

Questions?...