Using Social Media Imagery to Nowcast Air Quality

Anthony Mockler, Muhammad Rizal Khaefi

United Nations Global Pulse Lab, Jakarta

The Challenge

- **Air pollution** is one of the leading environmental risks impacting sustainable health, development and economic growth.
- Existing real time air quality monitoring data is expensive, sparse and unreliable in the developing world

Background

We lack air quality data in the developing world

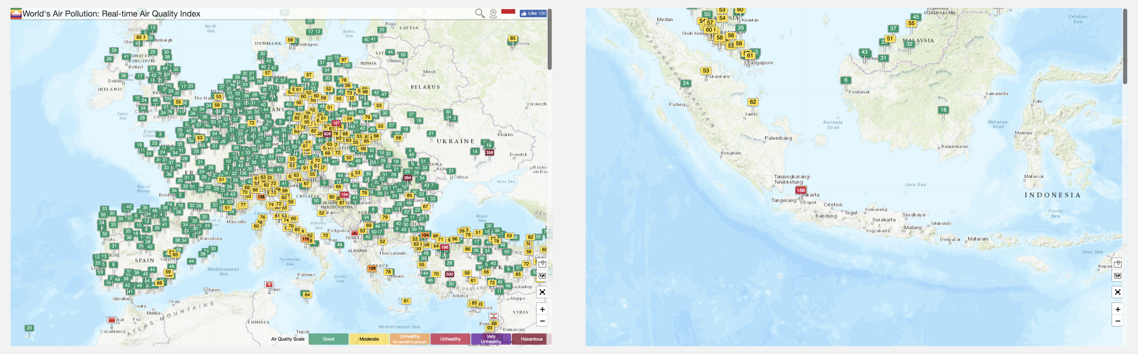


Figure: Official Air Quality Data, Europe vs Indonesia (http://www.waqi.info)

Information from official air quality monitoring stations is patchy and frequently only available from 1 or 2 locations in the entire country. **However, there's an abundance of social media imagery.**

Figure: Public social media images from a major social media platform

Method

In this experiment, we propose to use a deep learning model to:

1. Classify indoor and outdoor images

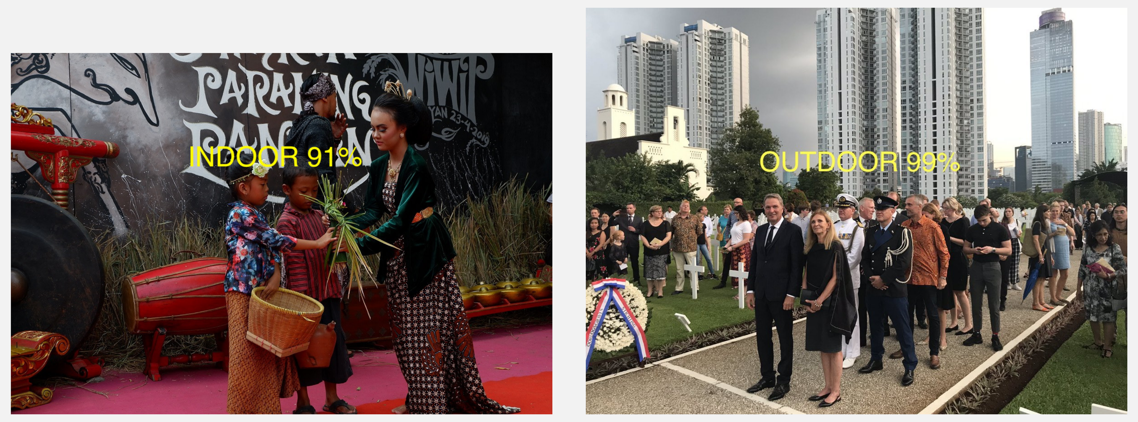


Figure: Indoor / Outdoor scene detection, trained on Imagenet + VGG-16 + Places365

2. Infer visibility levels from outdoor photos

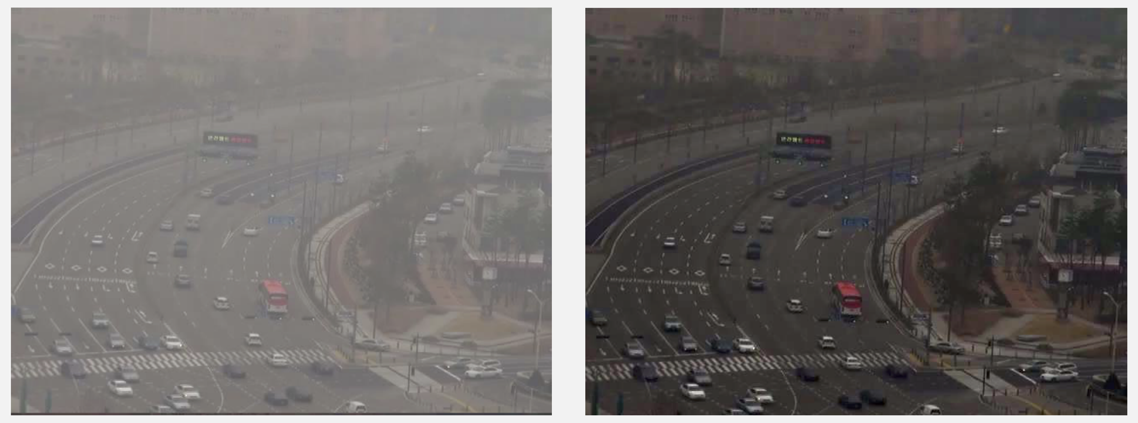


Figure: Hazed and Dehazed images from Dehazenet

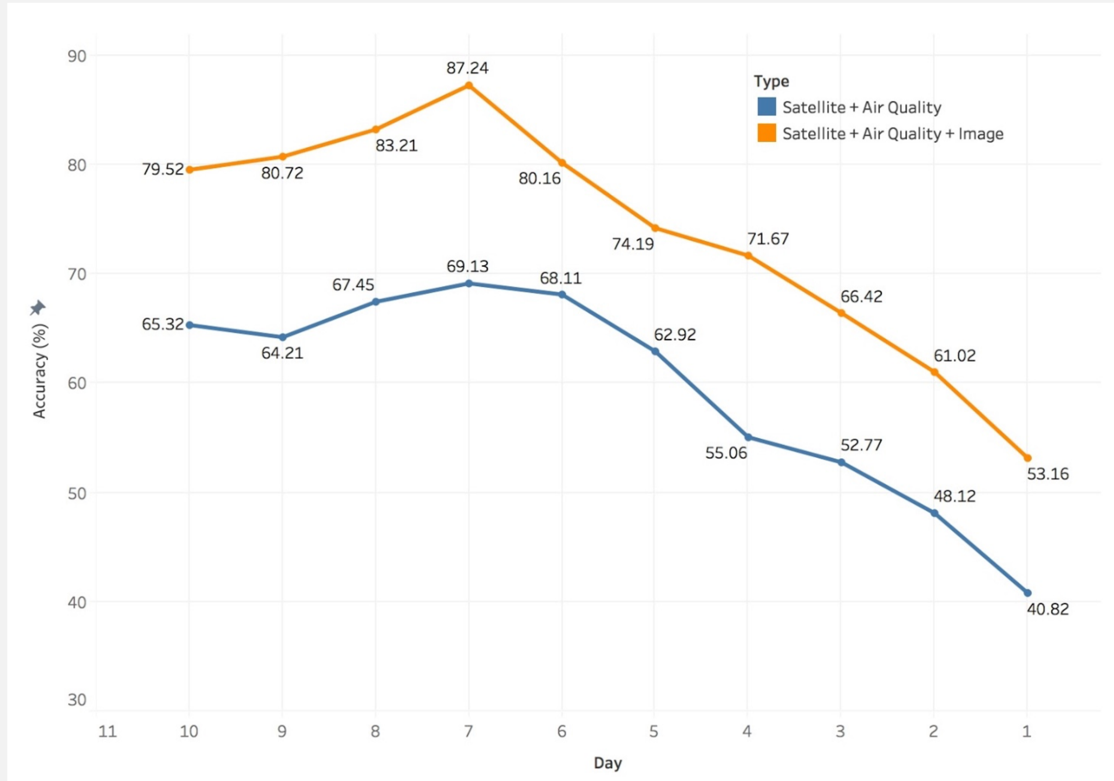
For outdoor images, we trained *DeHazeNet* and *AOD-Net* models with original images, as well as simulated (*NYU-Depth V2*) hazed images. The haze feature is defined as the distance (MSE or SSIM) between the original and dehazed images.

3. Combine haze and meteorological features

We then combined the haze feature with existing realtime satellite and meteorological data, and trained a LSTM model, using official PM2.5 air quality as the target.

Results

Comparison to baseline Adding image haze as a feature to our LSTM model, we substantially outperform a baseline of meteorological + satellite data only



Visualised in the chart above, the model which integrates social media images consistently outperforms the baseline model. The model with visibility information produces a best result 0.8724 forecast accuracy (using data aggregated from the previous 7 days), an improvement of 0.181 compared to the baseline model.

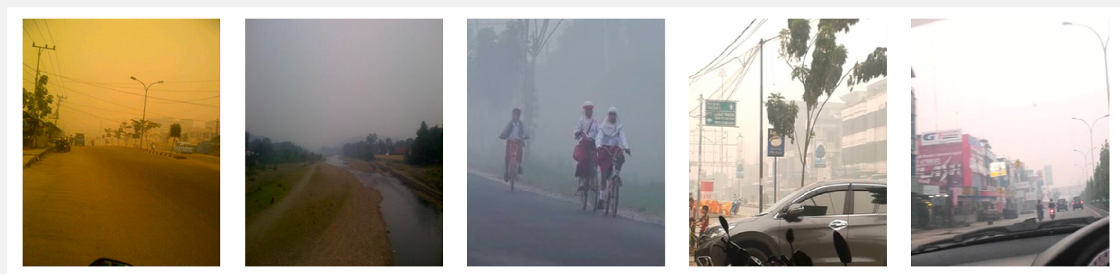


Fig. 3. Outdoor Classification Results

Conclusion

By combining conventional data with social media imagery, we can nowcast air quality across the developing world, without deploying tens of millions of dollars of expensive hardware. We can use this data to improve people's health and wellbeing, and assist in disaster management.

In future work, we hope to test the generalisability of this model to other cities, explore the inclusion of other data sources (such as road networks and structures from the Open Street Map project), and work to integrate our nowcast in to our *Haze Gazer* social listening platform.

We're excited to talk to Computer Vision experts, mentors, data owners and funders about our next steps.

Contact Information

- Web: http://www.pulselabjakarta.org/
- Email (Anthony): anthony.mockler@un.or.id
- Email (Rizal): muhammad.khaefi@un.or.id
- Email (PLJ): plj@un.or.id

PULSE LAB JAKARTA