Predicting Air Quality Index Using Multimodal Neural Networks

Zoaib Sihorwala
North Carolina School of Science and Mathematics
Charlotte, NC
Under the direction of Xiaoming Liu, Feng Liu, and ZiAng Gu
Computer Vision Lab
Michigan State University

Abstract:

Air pollution is a pressing ecological issue with significant impacts on both public health and the environment. Poor air quality can cause respiratory complications and disease, which result in millions of deaths every year. To address this challenge, we present a multimodal model to predict air quality levels in densely populated areas. Our research leverages both satellite imagery and meteorological data to create accurate air quality predictions. We used urban satellite imagery from the National Agriculture Imaging Program, meteorological data from Open-Meteo, and air quality data from OpenWeatherMap. The goal was for the model to implicitly learn spatial features in each image, such as roads, greenery, and buildings, and then combine this info with weather data to make a prediction of AQI. Using multiple computer vision techniques, the model was able to predict AQI with a mean average error of 16 AQI and a classification accuracy of 77\% based on the EPA's AQI standards categories. Our results establish a baseline for AQI prediction from satellite imagery, and are a vast improvement over state-of-the-art pretrained general computer vision models.

Background/Introduction:

Oxygen is an absolute necessity for living organisms, as it allows them to carry out respiration in order to produce energy. As such, access to clean air is very critical in order for life to thrive. Exposure to air pollution can cause various health problems such as respiratory complications, heart and lung disease, and even death [1]. Despite global efforts in the past decades to reduce air pollution, 99% of the world population were living in areas that failed to meet the World Health Organization's air quality standards in 2019 [2]. This led to 4.2 million deaths worldwide, which were especially prevalent in low and middle income countries who made up 89% of those ambient air pollution related casualties.

Air Quality Index (AQI), is a metric of air pollution developed by the United States Environmental Protection Agency (EPA), in order to quantify air pollution [3]. AQI is calculated based on measurements of several pollutants, with the major ones being particulate matter, nitrogen dioxide, ozone, carbon monoxide, and sulfur dioxide. These pollutants can originate from multiple sources, such as man-made fossil fuels and emissions as well as natural smoke from volcanoes or wildfires. Air quality monitors on the ground are typically used to measure the concentration of each pollutant over a certain time period. These monitors can range from miniature household monitors, to medium and large static monitors placed around cities which can provide continuous data from certain locations within cities. These kinds of sensors use active measurement, a means of using physical or chemical methods to automatically analyze the air in a given area. These sensors are fairly expensive to manufacture and operate, and can range from \$15,000 to \$40,000 per sensor [4]. This high cost is particularly devastating to lesser developed regions of the world where governments simply cannot afford to buy several monitors to cover urban and suburban areas. In fact, only about 6% and 24% of children live within 50km of an air quality monitor in Africa and South America respectively [5].

It is evident that there is a need for thorough and inexpensive air quality monitoring around the world. Satellite imagery and meteorological data are resources that are widely available. By using this data, we can utilize neural networks and computer vision techniques in order to predict air quality in regions of the world with sparse air quality monitoring. With this approach, we can circumvent the need for expensive equipment and labor, and have inexpensive access to widespread air quality data around the world. In this project, we will develop a multimodal neural network capable of accurately predicting the air quality of a given area, using only satellite imagery and weather data.

Previous Work:

Other scientists have conducted experiments in using satellite data to predict air pollution levels in certain areas. Scheibenreif et al. used imagery and remote sensing measurements from the ESA's Sentinel satellites and a two stream deep neural network to predict levels of NO_2, a key pollutant in finding the AQI [14]. constructed for Sentinel 5P's tropospheric measurements. Rowley et. al built off of this approach and used the Sentinel satellites along with NO_2 ground monitoring stations to predict levels of NO_2, O_3 and PM_{10} [13]. They added in a fully connected network utilizing additional data such as population and altitude. This project aims to provide a more general prediction of air quality by directly predicting the AQI rather than its components.

Dataset Creation:

Due to the relative novelty of this research, there are no currently publicly available datasets that contain satellite imagery, AQI, and weather data. Thus, we will have to retrieve this data from multiple sources and construct our own dataset. We start by collecting the most inflexible of the 3, satellite imagery. Since 2002, the US Department of Agriculture has operated the National Agriculture Imagery Program, which collects high resolution imagery across the United States, including urban and suburban regions [6]. This data is accessible through Google Earth Engine, a geospatial analysis platform which allows users to write JavaScript code to export imagery [7]. We gathered longitude and latitude data for the 100 largest cities in America, and then wrote a script to collect 100 1km by 1km images from each city, each with 1 m per pixel resolution. We utilized 25 parallel processes to export all of the imagery. Since the NAIP dataset does not have total coverage of all the areas we targeted, only 7,621 images were gathered out of an attempted 10,000, but this amount is still suitable for our dataset. Along with the imagery, we also recorded the time of capture, so that we could use the time and location of each image to pair up AQI and weather information with each image.

Air quality data was retrieved via API calls to OpenWeatherMap, a weather data service [8]. However, OpenWeatherMap provides pollutant concentrations rather than directly providing AQI values, so we must do the conversions in the data preprocessing. The first step is to convert the units of each pollutant from $\mu g/m^3$ to ppb [9]. The equation is as follows:

$$ppb = \frac{\mu g/m^3}{12.187 * M}(273.15 + C)$$

Where M is the molecular weight of the pollutant and C is the surface temperature in ℃. Then, we can convert these concentrations to an AQI value using an equation and a table of breakpoints provided by the EPA [10]. The equation is as follows [11]:

$$AQI = \frac{I_{hi} - I_{lo}}{C_{hi} - C_{lo}}(C - C_{lo}) + I_{lo}$$

Where C is the pollutant concentration, C_{hi} is the concentration breakpoint greater than or equal to C, C_{lo} is the breakpoint less than or equal to C, and I_{hi} and I_{lo} are the corresponding AQI breakpoints of C_{hi} and C_{lo} respectively. Finally, the overall AQI is simply the maximum AQI calculated for each pollutant. These AQI values were then paired with each image based on time and location.

For our meteorological data, we collected four different factors: wind speed, humidity, temperature, and rain. Rain was further split up into 3 categories, a sum of precipitation over 24 hours, 8 hours, and 1 hour. A table of their respective effects on AQI is below.

Weather Factor	Effect on AQI
Rain	Can wash away particulates, cleansing the air.
Temperature	Greater temperatures accelerate photochemical reactions, resulting in more ground level O_3.
Humidity	Water vapor forms as a nucleus for small particles, which can lead to the formation of larger particulates.
Wind Speed	Low wind speed leads to stagnant air, whereas high wind speed can blow particulates away.

Table 1. Weather factors and their effects on AQI.

This data was retrieved via API calls to Open-Meteo [12], a weather data service, and then paired up with each image in a similar manner to the AQI data.

Methodology:

The basis of our model is ResNet-50, a pre-trained model very suited for image classification tasks [15]. Although our problem is fundamentally that of regression, the main goal of the image network is to be able to implicitly learn certain features of an image in order to produce an accurate prediction. These features could include the amount of greenery, roads, cars, buildings, and bodies of water present in the image. All of these features can have effects on the AQI, so it is important to dedicate part of our model to discerning them. To handle the non-image data, we have a multilayer perceptron which takes two inputs: humidity and wind speed. These two were selected out of four aforementioned meteorological factors due to their strong correlation with AQI. The outputs from the image network and the weather data network are then combined, passed through a few more layers, and eventually to a single output neuron.

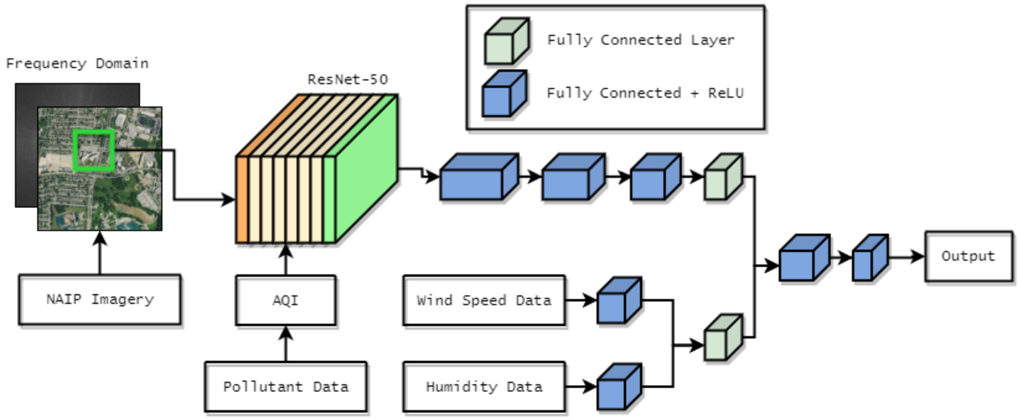

Fig.1. Full model architecture of AQINet

Before this model can be evaluated, however, we must first address a major issue in our dataset: AQI imbalance. AQI values are not evenly distributed in the real world, and this is reflected in our data.

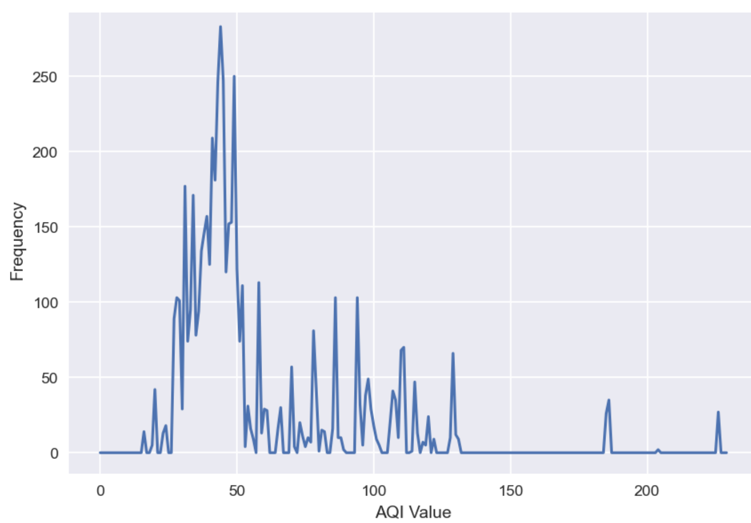

Fig. 2. AQI values and their frequencies in the dataset.

This imbalance will lead our model to simply predict from a small range of AQI values that are the most frequent. Consequently, the model will be rewarded for a majority of its predictions, and will not learn. There are two ways in which we approach this.

The first method is to use upsampling to even the AQI distribution by taking more samples of images with infrequent AQIs. Instead of feeding our model the entire 1000x1000 image, we take N 224x224 patches from the image. N is proportional to the inverse of our frequency graph, so that images with rare AQI are sampled much more than those with common AQI.

The second method we can use is to punish the model more for incorrect guesses on infrequent AQI, by adjusting our loss function. We start with a standard Mean Squared Error loss. Then, we compute some weight, W, to scale our MSE loss by. W is defined as follows:

$$W_i = [N(\mu, \sigma^2) * F(A_i)]^{-1}$$

Where N is a gaussian kernel, F is our frequency distribution, and A_i is the ground truth AQI for patch i. Note that the asterisk is a convolution, rather than multiplication. This convolution serves to "smooth" our frequency distribution such that our weights are calculated based on large changes in frequency rather than small inconsistencies. This is illustrated in the graph below.

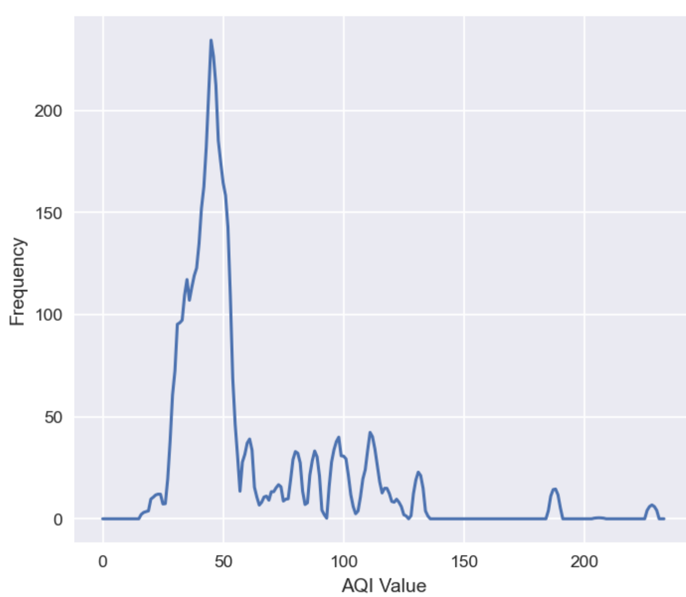

Fig. 4. Smoothed AQI frequency distribution.

After computing our weight, our total loss is simply this weight multiplied by MSE loss:

$$L = \frac{k}{n}\sum_{i=0}^{n} W_i \cdot (x_i - y_i)^2$$

Note that k is a constant scalar applied to prevent our loss from becoming very small, since W will always be significantly less than 1. A very small loss will adversely affect the training of our model, since it will diminish the gradients and slow down the learning. Together, both the upsampling and weighted loss will solve the data imbalance issue.

A smaller issue with our dataset emerges from the low variance of AQI at each location images were taken from. For each city, images taken were only a few kilometers away from each other, which led to many images in the same area having the same exact AQI. The data was ultimately retrieved from ground sensors, and if the distribution of these sensors were sparse in any given area, the AQI would be exactly the same for a majority of the images taken in that area. This is detrimental to our model because it could lead to overfitting and predicting specific AQI values that are common. We approach this issue by adding noise to each patch taken from an image. Each patch's noise is between -5 and +5 AQI. We ensure that the mean of the patch AQI's is still equivalent to the AQI of the whole image. This noise will result in a more robust model, which will generalize better to unseen data.

One aspect of poor air quality in images that we can directly extract is the haziness or slight blurring caused by dust or other particulates in the air. By using a 2D Fourier transform, we can convert our RGB input to the frequency domain, and provide a 4th channel of input to our image network. The frequency domain can reflect the haziness in an image in its low frequency components. To do this, we first convert our image to grayscale by averaging the values from the red, green and blue channels. Then, we normalize every pixel value to be between 0 and 1 for faster computations. Next, we can treat the values of the pixels as a stream of values, or a signal. Then we can decompose this signal into sine and cosine components using the Fast Fourier Transform, or FFT. This FFT will provide us with an image that will be generally darker for low frequencies (represented as smooth changes over time in the image), and lighter for high frequencies (fine details in the image). We then pass this new image to the model. Example transformations are below.

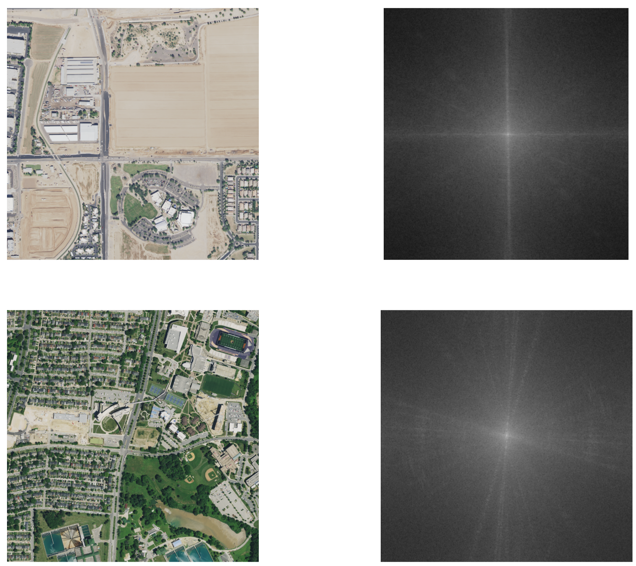

Fig. 5. Imagery and their Fourier transforms. The top image contains haze, which results in a darker frequency domain. AQI for the top image is 185, while the bottom image AQI is 52.

Results:

While similar projects have predicted air quality from imagery in some form, none have directly predicted AQI, so this project's results serve as a baseline for future research. We compared our model's results to that of a plain ResNet-50 model, with no additional transformations or techniques.

In the following table, the Mean Absolute Error (MAE) represents the average AQI error for each prediction from the model. The accuracy metric is a classification percent based on the AQI categories provided by the EPA. If the prediction and ground truth are both in the same category, the guess was marked correct, and incorrect otherwise.

Model	**MAE**	**Accuracy**
ResNet-50	31AQI	67%
AQINet	16AQI	77%

Table 2. AQINet results.

Future Work:

Computational Optimization

One of the main disadvantages of AQINet is that it is very slow to train, and can take up to an hour to train a single epoch. I feel this is a direct result of both the short length of this program and my relatively new introduction to machine learning. For example, the weighted loss. The weights for each AQI could simply be computed once and stored in the dataset, but we compute them every single epoch while training. This is a very simple fix, but I simply did not have the time to implement this. Small changes like this can allow us to save a lot of time while training.

Model Variation

I want to test other base models other than ResNet, such as VGG-16 and EfficientNet, and then compare their performances. I would also like to experiment with vision transformers, as they seem very interesting and could produce great results.

Dataset Expansion

Our current dataset only contains imagery from locations in the USA, since that is where the most widespread high resolution imagery was available. I want to expand this to other regions of the world, particularly areas which suffer the most from air pollution such as India, Bangladesh, and China. Lower quality images would be another hurdle for the model, but it would also make it more applicable worldwide.

Publishing

I want to release the dataset I have created at some point, because I think it will help others with similar projects and also raise awareness about fighting air pollution.

Acknowledgements:

I would like to acknowledge my mentors, Dr. Xiaoming Liu, Dr. Feng Liu, and ZiAng Gu for welcoming me to the CV Lab and teaching me so much this summer. I would also like to thank Dr. Richmond and MSU for providing me with this valuable experience, as well as my peers for making these past 7 weeks one of the best summers ever.

References:

[1] U.S. EPA, "Trends Report 2003," 2017. [Online]. Available: https://www.epa.gov/sites/default/files/2017-11/documents/trends_report_2003.pdf

[2] World Health Organization, "Ambient (Outdoor) Air Quality and Health," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

[3] AirNow, "Air Quality Index (AQI) Basics," [Online]. Available: https://www.airnow.gov/aqi/aqi-basics/

[4] U.S. EPA, "FRM/FEM and Air Sensors," 2019. [Online]. Available: https://www.epa.gov/sites/default/files/2019-12/documents/frm-fem_and_air_sensors_dec_2019_webinar_slides_508_compliant.pdf

[5] UNICEF, "Silent Suffocation in Africa: Air Pollution," 2019. [Online]. Available: https://www.unicef.org/media/55081/file/Silent%20suffocation%20in%20africa%20air%20pollution%202019%20.pdf

[6] USDA NAIP, "National Agriculture Imagery Program," [Online]. Available: https://naip-usdaonline.hub.arcgis.com/

[7] Google Earth Engine, [Online]. Available: https://earthengine.google.com/

[8] OpenWeatherMap, [Online]. Available: https://openweathermap.org/

[9] Danish National Environmental Research Institute, "Conversion between ppm and µg/m3,"

[Online]. Available:

https://www2.dmu.dk/atmosphericenvironment/expost/database/docs/ppm_conversion.pdf

[10] Wikipedia, "Air quality index," [Online].

Available: https://en.wikipedia.org/wiki/Air_quality_index#United_States

[11] AirNow, "AQI Technical Assistance Document," [Online]. Available:

https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf

[12] Open-Meteo, [Online].

Available: https://open-meteo.com/

[13] Andrew Rowley, Oktay Karakus, Predicting Air Quality via Multimodal AI and Satellite Imagery, Remote Sensing of Environment 2023

[14] Linus Scheibenreif, Michael Mommert, Damian Borth, Toward Global Estimation of Ground-Level NO2 Pollution With Deep Learning and Remote Sensing, IEEE Transactions On Geoscience and Remote Sensing, Vol. 60, 2022

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)