Using Satellite Images to determine AQI Values in California

Simon Bumm
Management Science and Engineering
Stanford University
simon.bumm@stanford.edu

Abstract

In this project, Deep Neural Networks are used in an attempt to predict Air Quality Index (AQI) differences for pairs of satellite images where images are taken from the same location at different times.

https://github.com/bummy93/cs229_SPR19

1 Introduction

The goal of this project was to develop a Deep Neural Network to infer AQI (air quality index)[1] based on a satellite image by providing a baseline image and AQI data in a focal location. Since today, satellite images are more broadly available than AQI readings, having a model that can determine AQI based on just a satellite image can significantly help people to monitor air quality in their region without the need for a physical measuring station. This can be especially helpful in less developed regions with severe air quality problems, (e.g. in Africa).

2 Related work

There is plenty of literature on using satellite images to predict a plethora of different outcome variables. Other than this particular project, the majority of said research falls under the category of image classification rather than regression. Below is a brief discussion of academic work that most closely relates to this project's goal and compares:

Neal Jean et al. [2], utilized a multi-step transfer learning approach to identify and extract nighttime features from satellite images to predict poverty in certain regions. Their setup is similar to the one used in this project since extracting night light imagery features depends only on very small variations within the picture, which is the goal of this project, too. Additionally, transfer learning is used by Neal Jean et al. to help with feature extraction and to compensate for incomplete survey data. The rationale for using transfer learning in this project stems from a pursuit to speed up the learning process, rather than making up for underrepresented/missing features in the training data.

Work other than Neal Jean et al. has used satellite images to e.g. perform oil spill detection [3], wheat yield prediction [4], habitat mapping [5], biomass estimation [6] and vegetation height estimation [7]. For these regression-type problems [6][7], various regression schemes are used for the machine learning model with (root) mean squared error ((R)SME) emerging as the paramount evaluation metric.

Other research at the interface of Deep Learning and satellite imagery falls under the category of classification [4][5] or feature detection [3]. In general, while a substantial literature body on satellite imagery exists, there seems to be no existing work attempting to predict AQI values using satellite imagery.

,

Figure 1: Left: Examples of locations which planet images depict. Ceneter: Locations of AQI measurements. Right: Example image area matched to AQI data

3 Dataset and Features

A lot of time throughout the initial phase of this project had to be dedicated to pre-processing both satellite imagery and AQI data. The satellite images were acquired from planet.com [8]. AQI data was acquired from EPA [9]. A three-step pre-processing pipeline was set up to retrieve and merge image and AQI data effectively.

Before querying any API for either image- or AQI data, the scope of search was narrowed down by defining a restricted geo-location and a restricted time period. Given the limitation and accessibility of planet's image data set, California was chosen as the main geo-location, while limiting the time period to years after 2016 (i.e. 2017 - present).

Based on said geo- and date filters, the AQI- and the image API were queried. However, only metadata was retrieved from the image API in order not to waste bandwidth and local storage capacity on imagery that would not match any available AQI readings. The data-matching procedure is illustrated in Figure 1.

The image metadata contained all data points needed to match the available AQI data (see 1 for reference) with the underlying geo-location and date of the images. Since a pre-trained model on 200x200 images is used in this project, geo-locations and the pixel ranges of the satellite imagery had to be split accordingly before matching both data sources.

Date	PM2.5 Concentration	UNITS	DAILY_AQI_VALUE	Site Name	STATE	COUNTY	SITE_LATITUDE	SITE_LONGITUDE
01/01/2019	5.7	ug/m^3 LC	24	Livermore	California	Alameda	37.687526	-121.784217
01/02/2019	11.9	ug/m^3 LC	50	Livermore	California	Alameda	37.687526	-121.784217

Table 1: Example AQI data acquired from EPA

Matching sub-images and AQI data base on date and location yielded valid sub-image/AQI matches. Based on this exhaustive set of image-metadata, a representative test set of images could be easily sampled before actually downloading the underlying images in bulk.

The planet image API can simply be queried with a pre-composed list of image IDs obtain from the previous step. These images are importet as as 5000 x 5000 x 4 .tif files and then broken down into 25 distinct 200 x 200 x 3 .png images. For every particular geo-location, one anchor image is defined and paired up with other images from the same location but different dates. Finally, these pairs get stacked together, labeled with their AQI deltas to compose one holistic data set that got divided into a 20% test and 80% training set.

4 Method

The basis of this project is a pre-trained network developed by Neal Jean et al. [2]. Said network contains 8 convolutional layers and a pooling layer at the end. Since the goal of this project is to predict AQI values using baseline data that contains satellite images and AQI values corresponding on both date and location, the most effective way to train the underlying model was to use a Siamese

Network architecture that discriminates the difference between two input images. A Siamese network is set up as two parallel convolutional neural networks that both use the same weights when two different images are passed through. This makes sure that the network extracts features the same way for both input images. This transfer learning approach was chosen in hope that the feature extraction characteristics of Neal Jean et al.'s legacy network can help augment the difference between two input images with fewer training epochs. An illustration of the neural network can be seen in Figure 2.

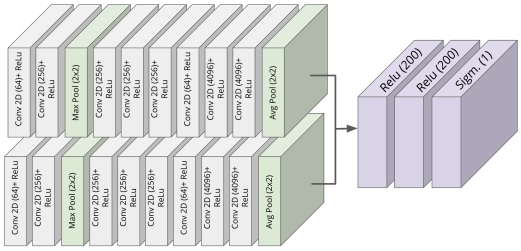

Figure 2: An illustration of the neural network setup

For the sake of this project, the loss function is defined as such: After two images have gone through the Siamese Network, their flattened vector representations are vertically concatenated and fed into an additional, one-layer neural network (i.e. dense layer with linear activation and single output unit) to map the image representation against a real-valued outcome label, which will be the difference between the AQI values of two images. AQI values are defined on a range between 0 and 500, whereas higher values indicate higher degrees of pollution. Hence, the predictions expected from the network are defined on a range between -500 and 500. The only meaningful metric that lends itself to evaluate a model against such a real-valued outcome variable is mean squared error (MSE). Other hyper parameters are derived from established best practices and set as such: batch size of 60 images (30 image pairs and their opposite), learning rate of 0.01 (default ADAM optimizer).

By training the network with this setup, the goal is to accurately predict the difference in AQI values between given images. A reference image can then be used to infer absolute AQI values for any image in the same location without relying on a measurement station.

5 Experiments and Results

5.1 Training Acceleration

The code base for this project was implemented using tensorflow [10] and keras [11]. In the beginning, the entire network was trained without any parallelization (neither in the data pipeline, nor in the training process itself). With a total of approx. 300,000 pairs of labeled training images, this approach was hardly feasible in the finite amount of time. Especially the batch-wise pre-processing of every 60 image batch took way too much time (about 5 minutes per batch). By leveraging keras' sequential interface [12], to parallelize both pre-processing and training the network, iterations became much faster ($\approx$ 30 min/epoch).

5.2 Improvement of Training Predictions

The initial predictions after having trained the full model using entire data set for around 20 epochs were hardly meaningful in the sense that they were not even close to the actual AQI deltas (MSE > 500). In conversations with members of the teaching team, four measures to help guide the network closer towards a meaningful range of predicted values were determined.

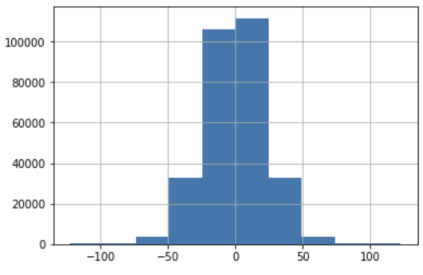

Figure 3: Empirical distribution of AQI delta of the training set

First, given the empirical distribution of AQI deltas (Figure 3), the network should have the tendency of predict small AQI deltas for any given image pair. To leverage this intuition in favor of making faster learning progress, a highly skewed training set was sampled to more forcefully manipulate the legacy knowledge of the pre-trained network. Highly skewed in the sense that we only provide the

neural network top 10% subset of our training data that has the highest AQI delta. Second, instead of using a linear activation function in the output layer, a scaled sigmoid function was used (i.e. $f(x) = \sigma(x) * 300 - 150$). A scaled sigmoid function is much more sensible to predictions that are close to zero, which corresponds very naturally to the empirical distribution of our data. The sigmoid function was only scaled by 150 since there were no higher AQI values present in the training data (see Figure 3). Third, the first six pre-trained layers of the neural network were freezed so that fewer high-level parameters had to be trained and the network was guided to better learn lower-level features that are more meaningful to the particular AQI prediction task. Fourth, two additional, sigmoid-activated dense layers were added to the end of the network for it to learn more complex relationships between the images' vector representations and the AQI deltas. This last part of the architecture was inspired by a blog post written by Kevin Mader [13], who used this type of prediction head to gauge image similarity on the MNIST data set. All the above features are added after training results show that the loss are not decreasing, or decreasing relatively slowly.

5.3 Improvement of Validation Predictions

With the adjustments made to the neural network mentioned in subsection 5.2, loss on the training set decreased significantly faster. However, the predictions from the half-freezed network, trained on the skewed training data using a scaled sigmoid activation and a different prediction head, revealed substantial overfitting issues. While the training MSE had come down to about 100, the validation MSE was still stuck at 800. To counter overfitting, three different forms for regularization were applied. First, 0.2 dropout layers were added to the non-pretrained part of the network. Second, random samples were added back to the training data so that there were 50% images with particularly high AQI deltas and 50% images that came from the default training data distribution. As a third measure of regularization, keras' powerful image pre-processing API was used to apply random augmentation to 50% of the training data. In particular, images were randomly flipped or rotated so that the pixel values were changed while the underlying geo-location and associated AQI labels could stay the same. This last regularization technique is what extends this project beyond a similar project that has been done for CS230 this quarter. A distinct comparison of results with and without image augmentation will be presented below.

5.4 Model Evaluation

To evaluate the trained neural networks against each other (i.e. with image augmentation vs. without image augmentation) and against a benchmark predictor, an independent sample of 18,900 labeled image pairs was fetched from the planet [8] database. Predictions on these images are used below to benchmark the resulting MSE against both, a meaningful baseline predictor (human performance) and the distribution of the outcome variable.

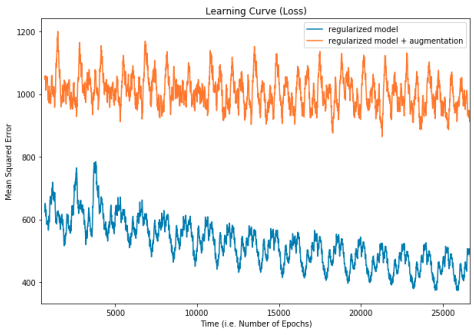

Figure 4: Training- and Validation loss for the Final Model Architecture (both with augmented and non-augmented image data)

In general, the task to predict AQI values from satellite images turns out to be much harder than initially anticipated. From the image pair in Figure 5, it is easy to see that a human could hardly tell which image has better air quality, while it is even harder to determine the actual numeric difference in AQI values (the AQI values between these images actually differ by 41, i.e. air quality on in the left image is much worse than on the right).

A benchmark predictor for this setting is chosen to replicate a human who would guess the AQI difference randomly, based on his/her knowledge about the underlying distribution. Hence, to come up with a corresponding baseline predictor for the test set (i.e. 18,900 images), 18,900 AQI deltas were sampled from the empirical distribution of all ≈300,000 training samples (Figure 3). This relatively profane approach yielded a baseline mean squared test error of 1231.53.

The final models (i.e. 3-layer prediction head, dropout, scaled sigmoid, with- and without image augmentation) were trained on the entire training set for a total of 60 epochs. Over those epochs, a training loss (MSE) decreasing from 1376.2 (1423.6) after the first epoch to 441.5 (983.8) after the last (i.e. 60^{th}) epoch was observed for the non-augmented (augmented) data set. At the same time, the gap between training and validation loss remained far lower compared to when no regularization had been applied at all. What was striking is to see the extend to which image augmentation makes the model much harder to train. While the loss on the augmented image data set is still decreasing, the gradient of said loss is a lot smaller than the one one the non-augmented data set. However, this additional training effort pays off when looking at the corresponding test errors: prediction from the model trained on the non-augmented data set generalized a lot worse than the 1234.1 test error vs. 753.9 test error). Quite disappointingly, however, none of both models was able to get both training and validation loss down to meaningful levels, the non-augmented one wasn't even better than the profane baseline predictor.

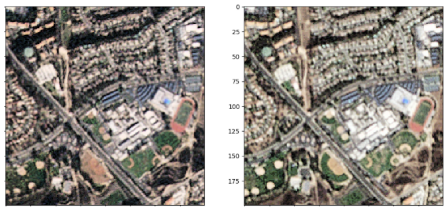

Figure 5: An example of training image pairs. The right image's AQI is 41 higher than the left image's

6 Conclusion/Future Work

The three image pairs below exemplify why both models perform so poorly. First, satellite images vary widely and can be drastically different even with the same AQI values due to clouds, landscape changes and time of the day. A lot of additional training data and -time might get the network to learn a broader range of these sparse conditions. On the other hand, one could also narrow the purpose of the model down to very specific conditions (i.e. no cloud, mid-day) and adjust training AND test data accordingly. Then, however, the training results won't be as meaningful and limited in their applications by default.

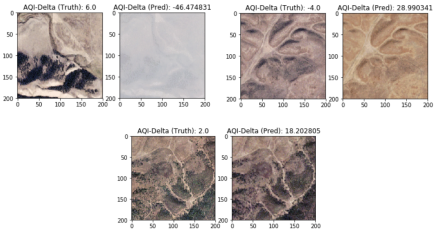

Figure 6: Image pairs illustrating problematic conditions for the NN to grasp.

There are also various hyper parameters that could possibly improve the model. For example, adding more layers with a different activation functions after the Siamese Network might help learn more complex features.

References

[1] United States Environmental Protect Agency, "Air Quality Index (AQI) Basics." https://www.airnow.gov/index.cfm?action=aqibasics.aqi. [Online; accessed 30-April-2019].

[2] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.

[3] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Machine learning*, vol. 30, no. 2-3, pp. 195–215, 1998.

[4] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Computers and Electronics in Agriculture*, vol. 121, pp. 57–65, 2016.

[5] A. Kobler, S. Džeroski, and I. Keramitsoglou, "Habitat mapping using machine learning-extended kernel-based reclassification of an ikonos satellite image," *Ecological modelling*, vol. 191, no. 1, pp. 83–95, 2006.

[6] N. R. Jachowski, M. S. Quak, D. A. Friess, D. Duangnamon, E. L. Webb, and A. D. Ziegler, "Mangrove biomass estimation in southwest thailand using machine learning," *Applied Geography*, vol. 45, pp. 311–321, 2013.

[7] D. Stojanova, P. Panov, V. Gjorgjioski, A. Kobler, and S. Džeroski, "Estimating vegetation height and canopy cover from remotely sensed data with machine learning," *Ecological Informatics*, vol. 5, no. 4, pp. 256–266, 2010.

[8] Planet Team (2017), "Planet Application Program Interface: In Space for Life on Earth." https://api.planet.com. [Online; accessed 8-Jun-2019].

[9] United States Environmental Protect Agency, *Air Quality Index Daily Values Report*. [Online; accessed 8-Jun-2019].

[10] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265–283, 2016.

[11] F. Chollet, "keras." https://github.com/fchollet/keras, 2015.

[12] F. Chollet, "The sequential model api." https://github.com/keras-team/keras/blob/master/keras/engine/sequential.py, 2015.

[13] K. Mader, "Image similarity with siamese networks." https://www.kaggle.com/kmader/image-similarity-with-siamese-networks, 2018.