
Predicting Ground-Level Ozone Concentration from Urban Satellite and Street-Level Imagery using Multimodal CNN

Andrea Vallebuena (avaimar@stanford.edu)¹, Nina Prakash (nprakas1@stanford.edu)¹, and Nicolas Suarez (nsuarez@stanford.edu)²

¹Department of Statistics, Stanford University

²Department of Economics, Stanford University

Abstract

Understanding the impact of the built environment and urban populations on climate change and air quality is a growing challenge as the percentage of the world's population that resides in urban areas continues to rise. In particular, human activity levels have been shown to be directly linked with ground-level ozone, which is responsible for several health and climate effects. We explore this relationship through the use of a multimodal learning architecture that predicts ozone concentrations (parts per billion) in urban areas from satellite and street-level imagery. The model comprises two Convolutional Neural Networks (CNN), one trained on satellite images of each location to learn higher-level features such as geographical characteristics and land use, and another trained on multiple street-level images of each location to learn ground-level features such as motor vehicle activity. The feature representations learned from each sub-model are concatenated and passed through several fully connected layers to predict the ozone level of the location. This concatenated model achieves a test RMSE of 11.70 ppb. This approach can be used to inform urban planning and policy by providing an insight into the particular urban features that aggravate ozone concentrations.

1 Introduction

Air Quality Ozone (O_3) is one of the air pollutants with the strongest evidence of associated health risks [18], and is used by several countries in their computation of air quality indices. As a secondary pollutant, it is not directly emitted into the air but rather formed as a result of chemical reactions between nitrogen oxides and volatile organic compounds such as methane and carbon monoxide. In turn, the presence of these primary pollutants is highly associated to human activity levels and can be aggravated by features of the physical landscape resulting from inadequate urban planning [18].

The Urban Landscape Although less than 3% of global land surface is classified as urban, these areas house over 50% of the world's population today and this percentage is expected to increase in the coming years. Characterizing the relationship between specific features of the urban landscape and ozone concentrations is thus essential to inform policy and urban planning decisions in order to build more sustainable and resilient environments.

Proposed Model In this work we use a dual-input CNN approach to characterize the relationship between ozone concentrations and learned urban features such as land use patterns, geography, buildings, and traffic and motor vehicle patterns. Existing work describes the use of machine learning to forecast air quality measurements over time or characterize the relationship between air quality and weather patterns, but there is a lack of literature surrounding the use of imagery to quantify air quality in urban centers. In addition, related works have used CNNs to predict land use classification and socioeconomic features using satellite imagery or street-view imagery separately, but we implement a dual-input model where the features from both types of images are concatenated to predict ozone concentrations in an urban location.

2 Related work

CNNs and Satellite Imagery. The application of CNNs to satellite imagery has been well-explored in recent years, particularly to predict land use classification. Examples of this include Castelluccio et al. who trained a CNN on the UC Merced Land Use dataset [15] and Bragilevsky et al. who trained a CNN on imagery of the Amazon Rainforest [4]. Albert et al implemented transfer learning to predict land use classification in urban areas [3].

Experiments in recent years have also extended the use of CNNs on satellite imagery to extract information about social and economic features of the physical landscape. For example, Oshri et al. use a ResNet pretrained on the ImageNet dataset to predict infrastructure quality in Africa [19], Maharana et al. developed a CNN pretrained on ImageNet to predict neighborhood crime rates [14], and Perez et al. developed a dual-input model trained on nighttime and daytime satellite imagery and pretrained on ImageNet [22].

CNNs and Street-View Imagery. Literature on the application of CNNs to street-level imagery is much more limited. One similar work by Gebru et al. proposes a CNN trained on Google Street View imagery to predict demographic characteristics at the neighborhood level such as race, income, and voting patterns, pretrained on ImageNet [6].

Emissions Predictions from Sensing Data. There have been several studies on the use of machine learning directly on air pollution sensor measurements, typically to make short-term air pollution forecasts ([5], [17]) or to understand the relationship between air pollutants and weather patterns [11]. However, to the best of our knowledge, there is no work that aims to use Deep Learning to understand the relationship between air quality and urban features by predicting measurements from satellite or street-view imagery.

Novel Contributions. This paper adds to the current body of work with two novel contributions: the first, to understand the relationship between features of the urban landscape and air quality, specifically ozone concentrations, and the second, to use a dual-input multiscale approach with both satellite and street-level imagery of urban areas.

3 Dataset and Features

The predicted labels for our dataset are ozone level measurements in parts per billion (ppb). The inputs for our dataset comprise a set of satellite images and a set of street-level images.

3.1 Ozone Measurements

The ozone dataset was constructed by scraping data from the AirNow API [2], which centralizes data from air quality agencies in the U.S., Canada and Mexico and data provided by U.S. Embassies and Consulates on monitoring sites around the world. AirNow also reports the U.S.' Air Quality Index (AQI), which classifies the level of health concern into six categories ranging from 'good' to 'hazardous' according to a location's pollutant levels. Ozone measurements were gathered in ppb for each monitoring site on an hourly basis for the year 2020, and averaged to obtain an annual ozone level for each site. We use the annual ozone level in a location in order to eliminate seasonal variations in pollution, which prompts our model to learn about urban features that are associated with a baseline level of pollution for each location.

Our primary dataset comprised 2020 average ozone levels for 1,423 unique global monitoring sites. As a form of data augmentation, we used monitoring site to U.S. zip code and monitoring site to U.S. county mappings in order to expand the size of our dataset and obtain information at a more granular level across locations in the U.S. This preprocessing increased our dataset to 12,976 semi-unique locations with ozone level information. This augmentation modified the geographic distribution of our dataset, with over 89.5% of data points being located in the U.S. A summary of statistics on the ozone readings dataset is reported in the Appendix in Table 5.



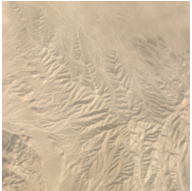

3.2 Satellite and Street-Level Imagery

The satellite imagery dataset was constructed using the Google Earth Engine API. For each location labeled with an ozone reading, we retrieve one satellite image centered at that location from the Landsat 8 Surface Reflectance Tier 1 Collection with a resolution of 224×224 pixels which represents $6.72 \text{ km} \times 6.72 \text{ km}$. We use 7 bands from this collection: RGB, ultra blue, near infrared, and two shortwave infrared bands. We preprocess each of our images by adding a cloud mask per pixel and then computing the per pixel and band mean composite of all the available images for the year 2020. As a consequence of the cloud masking process, around 5.7% of our images have missing pixels

that were clouded during all of 2020, so we impute those pixels with the per channel mean of the image. Among the affected images, on average only 0.22% of the pixels were fully clouded, and the most affected image had about 5% of its pixels clouded. Landsat 8 has a resolution of 30 meters and our patches have dimensions $6.72 \text{ km} \times 6.72 \text{ km}$, so our images contain 224×224 pixels, and 7 channels.

The street-level imagery dataset was constructed using the Google Maps Street View API. For each location labeled with an ozone level, we randomly sample 10 geospatial points within 6.72 km from the measurement point. For each point we retrieve an image with a resolution of 224 pixels and a field of view (FOV) of 120 which represents the widest zoom, filtered for outdoor images. Each image has dimension $224 \times 224 \times 3$, where the 3 channels represent RGB.

Table 1: Data points of locations with the lowest and highest ozone levels

	Rangoon, Myanmar		Mojave National Preserve, U.S.	
Ozone Level (ppb)	8.15		54.14	
Ozone AQI	good		moderate	
	Satellite	Street	Satellite	Street
Image sample				

Note: Images have size (224, 224, C) but have been resized for visualization purposes

Satellite and street-view images were mapped to an ozone reading using a unique identifier of the location. Satellite images were available for all locations, while street images were available for 99.7% of locations. We used a split of 85% Train, 7.5% Validation and 7.5% Test in order to randomly generate the following datasets. These splits were built for the satellite and street-level imagery such that they both share the same locations in each split.

Table 2: Summary statistics of the image datasets

Dataset	Resolution	Train Examples	Validation Examples	Test Examples	Total Examples
Satellite	$224 \times 224 \times 7$	11,029	973	974	12,976
Street-View	$224 \times 224 \times 3$	109,609	9,658	9,627	128,894

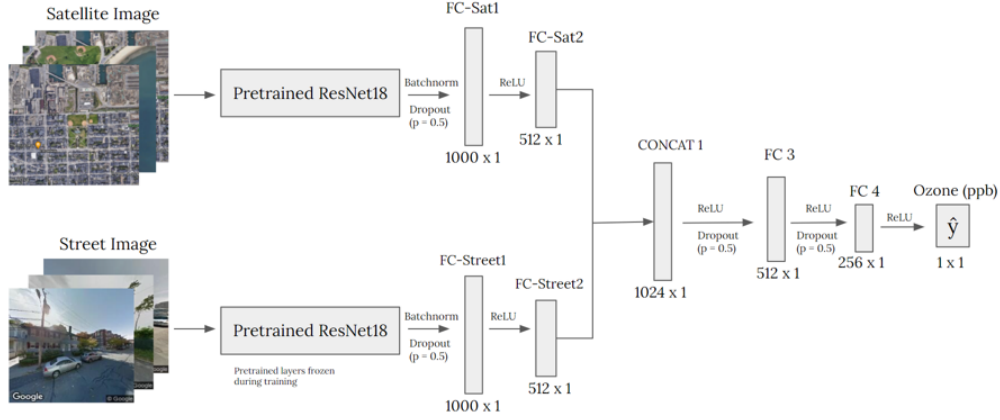
4 Model Architecture and Methods

We train the two CNNs separately on the satellite and street-level imagery, both using a ResNet-18 architecture [10] implemented in PyTorch [20] and pretrained on the ImageNet dataset. The models are trained separately as the nature of the features they need to learn to associate with ozone concentration is quite different for each dataset. Transfer learning is used for both CNNs to leverage lower-level features learned on the ImageNet dataset. The ResNet-18 architecture was slightly adapted for our particular task; in the case of the satellite imagery, the CNN’s input layer was modified to accommodate for the image’s seven channels and was initialized using Kaiming initialization [10].

Due to overfitting concerns, we experiment with adding Dropout and additional Fully Connected layers at the end of the model, prior to a final layer consisting of a single neuron outputting the location’s ozone concentration (ppb). For the prior layers, we keep the BatchNorm layer that the ResNet-18 architecture uses after each convolution and prior to activation, which has an implicit regularization effect. We also use data augmentation to combat overfitting. Satellite images are loaded applying random horizontal and vertical flips, and a random rotation of up to 20 degrees. These transformations leverage the fact that these different spacial visualizations remain true to the location’s physical representation. Due to the nature of the street imagery only random horizontal flips were used for data augmentation.

After training both CNNs separately to predict the ozone reading for each location, we extract 512 features for each satellite and each street image. These are concatenated to create a feature vector of size 1,024 representing the satellite image and a particular street view of a given location. We then train a Concatenated Feedforward Neural Network (NN) using these multiple representations of each location to predict the location’s average ozone level in 2020.

Figure 1: Model Architecture



5 Experiments

When training the CNN on the satellite imagery, we tuned several hyperparameters including the optimizer, the batch size and the learning rate, holding other hyperparameters constant. We also experimented training satellite images with and without cloud masking. Although the models using satellite images without cloud masking generally performed better on the validation set, we chose to use the images with cloud masking as it prompted the model to learn about the urban features of the location rather than potentially learn from weather patterns or visible pollution to predict ozone levels. We then chose the best hyperparameters based on their RMSE on the validation set. The hyperparameters selected as per this criteria were the Adam optimizer, a batch size of 64, and a learning rate of $1e-3$. To reduce overfitting, we added dropout and tuned both the dropout rate and the number of additional layers with dropout. The model including one additional dropout layer with a rate of 0.5 was selected as it reduced overfitting the most. The final hyperparameters for the satellite model were selected as described in Experiment 8 in Table 6.

To train the CNN on the street-level imagery, we began with the best hyperparameter values from the model trained on satellite imagery and added an additional hyperparameter to control the number of pretrained ResNet18 layers to freeze in order to reduce model runtime. Increasing the number of frozen layers had a regularizing effect and reduced overfitting. We also tuned the number of examples per grid cell location as a form of data augmentation to improve both train and validation accuracy. The best hyperparameters were selected as the values from Experiment 13 in Table 6.

To tune the final neural network on the concatenated feature encoding from both the satellite and street-level CNNs, we began with the best hyperparameter values from previous experiments and tuned the model architecture, dropout, and the activation function. To tune the model architecture we experimented adding 1, 2, and 3 fully connected layers. To tune dropout we tested a rate of 0.5 and 0.75. Finally, we tuned the activation function by testing both ReLU and Tanh. The best model was selected as Model 16 in Table 6.

6 Results/Discussion

Table 3 presents the performance of each of the CNN submodels and the concatenated NN on the test set as measured by RMSE. The Satellite model registers better performance on a standalone basis with a test RMSE of 12.48 ppb, compared to the Street-level model's RMSE of 20.64. This may be driven by the increased complexity of capturing the general urban characteristics of a region from a single ground-level view. Moreover, we found that in the case of several locations, the monitoring site could perhaps be far from an urban center or the 6.72 km radius used to sample images of the environment surrounding the measurement site could have been too large for these locations. This resulted in various images which mostly reflected more rural views of the outskirts of urban locations from which our model could not have been able to associate urban features and ozone measurements. The concatenated model appears to have mainly leveraged the information from the satellite image of each location, coupled with some information from a sample street-level image of the location, and presented a test RMSE of 11.70 ppb.

To contextualize this error, we recall that ozone levels in the dataset fall in the range $[8.15, 54.14]$. Locations with levels falling in the $[0, 50]$ range are classified as having 'good' ozone AQI, while those in the $(50, 100]$ range are classified as 'moderate', indicating that a population sensitive to ozone may begin to present adverse health effects. No data points

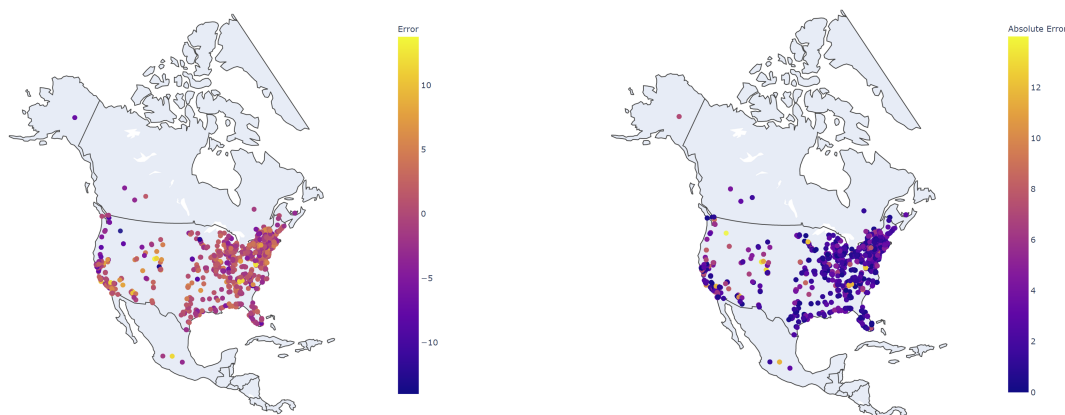
are available with an annual average superior to 54.14 ppb, as this would have required the location to sustain such levels throughout a large portion of 2020. This limits our ability to evaluate the model on a wider range of levels of health concern and to discern between locations that can be susceptible to having more concerning ozone levels.

Table 3: Test performance for final CNN submodels and concatenated NN

	Satellite Model	Street-level Model	Concatenated Model
Test RMSE (ppb)	12.48	20.64	11.70

An analysis of the concatenated model’s prediction errors indicates that the errors are approximately Normally distributed, with errors centered around zero and primarily falling in the $[-3, 3]$ ppb range. The model’s predictions seem to largely follow the distribution of ozone levels in the test dataset, as evidenced in Table 7. As observed in Table 8, our model is best at predicting moderate ozone levels in the range $[25, 30]$, and performs more poorly when predicting extreme levels in the ranges $[15, 20]$ and $[35, 55]$. With 50% of the data pertaining to locations with ozone levels in the range $[25.55, 30.72]$, the model’s performance could be attributed to the challenging task of discriminating between the urban features belonging to locations with very similar levels as summarized by an average annual ozone reading. As observed in Table 4, the model does not seem to systematically perform predictions with high error rates for particular regions. However, it may be observed that locations with high absolute error rates tend to be clustered together, reflecting the fact that if a region in general has extreme levels, our model will generally perform poorly on the locations within the region.

Table 4: Geographic distribution of ozone reading errors (ppb) and absolute errors (ppb)



Note: The randomly-generated test set did not include any of the few locations outside of North America.

7 Conclusion/Future Work

In this work, we explored modeling the relationship between urban features and ground-level ozone concentrations through the use of satellite and street-level imagery. This is a challenging task, as it requires the model to automatically identify complex urban characteristics such as transport infrastructure, the presence of industrial facilities and the level of motor vehicle activity, and to associate them to an average ozone reading in an effort to capture a location’s nonseasonal pollution level. The features extracted from the satellite images were more successful in reflecting this type of information when compared to the street images; however, our high test RMSE of 11.70 ppb on the concatenated features underlines the challenging nature of this problem.

Several improvements and explorations could be performed in future work to obtain a more accurate portrayal of the association between urban features and ozone levels. Modeling this relationship more explicitly through the use of object detection methods and pre-trained models used to classify land use could aid in better understanding what type of urban characteristics drive higher pollution. It could also be beneficial to extract multiple pollution measures from each location in order to model their joint distributions. Lastly, the use of ozone and satellite image data disaggregated in time, as opposed to averaged over a full year as in this work, could also be helpful. This could widen the range of ozone readings observed by the model and help it understand the role urban features may play in locations with concerning ozone levels, even if these take place only for a few periods throughout the year.

8 Contributions

All three members of the team contributed to each aspect of the project, but each focused more on different areas. Nicolas focused on gathering and preprocessing the satellite imagery dataset including performing cloud masking as well as image visualization, Nina focused on gathering the street-view imagery dataset and running experiments for hyperparameter tuning, and Andrea focused on gathering and preprocessing ozone emissions data as well as taking the lead on the code for the CNN and NN model definitions and final error analysis. All team members contributed to the literature review, proposal, milestone, and final reports.

References

- [1] AirNow. *Air Quality Index (AQI) Basics*. accessed January 25, 2021, <https://www.airnow.gov/aqi/aqi-basics/>. 2020.
- [2] AirNow. *AirNow API*. data accessed on January 24, 2021, <https://docs.airnowapi.org/files>. 2021.
- [3] Adrian Albert, Jasleen Kaur, and Marta C. González. “Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale”. In: *CoRR* abs/1704.02965 (2017). arXiv: 1704.02965. URL: <http://arxiv.org/abs/1704.02965>.
- [4] L. Bragilevsky and I. V. Bajić. “Deep learning for Amazon satellite image analysis”. In: *2017 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*. 2017, pp. 1–5. DOI: 10.1109/PACRIM.2017.8121895.
- [5] T. Yang D. Zhu C. Cai and X. Zhou. “A machine learning approach for air quality prediction: Model regularization and optimization”. In: *Big Data and Cognitive Computing 2* (2018), p. 5.
- [6] Timnit Gebru et al. “Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States”. In: *Proceedings of the National Academy of Sciences* 114.50 (2017), pp. 13108–13113. ISSN: 0027-8424. DOI: 10.1073/pnas.1700035114. eprint: <https://www.pnas.org/content/114/50/13108.full.pdf>. URL: <https://www.pnas.org/content/114/50/13108>.
- [7] *Google Street View Static API*. Google. URL: <https://developers.google.com/maps/documentation/streetview/overview>.
- [8] Noel Gorelick et al. “Google Earth Engine: Planetary-scale geospatial analysis for everyone”. In: *Remote Sensing of Environment* (2017). DOI: 10.1016/j.rse.2017.06.031. URL: <https://doi.org/10.1016/j.rse.2017.06.031>.
- [9] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- [10] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV].
- [11] K. Demertzis I. Bougoudis and L. Iliadis. “HISYCOL a hybrid computational intelligence system for combined machine learning: The case of air pollution modeling in Athens”. In: *Neural Computing and Applications* (2016), pp. 1191–1206.
- [12] Plotly Technologies Inc. *Collaborative data science*. 2015. URL: <https://plot.ly>.
- [13] Qingshan Liu et al. “Learning Multi-Scale Deep Features for High-Resolution Satellite Image Classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* PP (Nov. 2016). DOI: 10.1109/TGRS.2017.2743243.
- [14] Adyasha Maharana, Quynh C. Nguyen, and Elaine O. Nsoesie. “Using Deep Learning and Satellite Imagery to Quantify the Impact of the Built Environment on Neighborhood Crime Rates”. In: *CoRR* abs/1710.05483 (2017). arXiv: 1710.05483. URL: <http://arxiv.org/abs/1710.05483>.
- [15] Carlo Sansone Marco Castelluccio Giovanni Poggi and Luisa Verdoliva. “Land use classification in remote sensing images by convolutional neural networks”. In: (2015). arXiv: 1508.00092.
- [16] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.
- [17] S. P. N. Kumar J. Gu A. Hauryliuk E. S. Robinson A. L. Robinson N. Zimmerman A. A. Presto and R. Subramanian. “Closing the gap on lower cost air quality monitoring: Machine learning calibration models to improve low-cost sensor performance”. In: *Atmospheric Measurement Techniques* (2017), pp. 1–36. DOI: 10.5194/amt-2017-260.
- [18] World Health Organization. *Ambient Air Pollution*. Accessed on February 24, 2021, <https://www.who.int/airpollution/ambient/pollutants/en/>. 2021.

- [19] Barak Oshri et al. “Infrastructure Quality Assessment in Africa Using Satellite Imagery and Deep Learning”. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '18. London, United Kingdom: Association for Computing Machinery, 2018, 616–625. ISBN: 9781450355520. DOI: 10.1145/3219819.3219924. URL: <https://doi.org/10.1145/3219819.3219924>.
- [20] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [21] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: <https://doi.org/10.5281/zenodo.3509134>.
- [22] Perez Anthony Driscoll Anne. et al. Yeh Christopher. “Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa”. In: *Nature Communications* (2020). DOI: 10.1038/s41467-020-16185-w.

A Appendix

A.1 Dataset information

Table 5: Summary statistics of the ozone readings dataset

Location type	Examples	Mean (ppb)	Std. Dev. (ppb)	Minimum (ppb)	Maximum (ppb)
Monitoring site	1,423	28.75	5.63	8.15	54.14
Zipcode	11,095	28.25	3.96	8.26	46.77
County	548	29.57	6.22	8.15	49.43
Total	12,976	28.35	4.28	8.15	54.14

A.2 Experiments for Hyperparameter Tuning

Table 6: Hyperparameter search for the Satellite and Street CNNs, and the Concat NN

Experiment Number	Model	Examples per location	Batch size	Learning rate	Epochs	Optimizer	Dropout	Batch Norm	Activation	Num Frozen Pretrained Layers	Cloud Masking	Train RMSE	Dev RMSE
1	Satellite	1	64	1e-3	100	Adam	None	Yes	ReLU	0	No	0.3	15.1
2	Satellite	1	64	1e-3	100	RMSProp	None	Yes	ReLU	0	No	1.6	18.3
3	Satellite	1	64	1e-3	100	SGD	None	Yes	ReLU	0	No	1.1	20.4
4	Satellite	1	32	1e-3	100	Adam	None	Yes	ReLU	0	No	0.2	15.0
5	Satellite	1	128	1e-3	100	Adam	None	Yes	ReLU	0	Yes	0.4	13.4
6	Satellite	1	64	1e-4	100	Adam	None	Yes	ReLU	0	No	0.3	12.3
7	Satellite	1	64	1e-3	100	Adam	1 layer (p=0.5)	Yes	ReLU	0	No	10.5	10.1
8	Satellite	1	64	1e-3	100	Adam	1 layer (p=0.5)	Yes	ReLU	0	Yes	5.6	12.7
9	Satellite	1	64	1e-3	100	Adam	1 layer (p=0.75)	Yes	ReLU	0	No	13.8	11.1
10	Satellite	1	64	1e-3	100	Adam	2 layers (p=0.5)	Yes	ReLU	0	No	11.9	12.9
11	Street	1	64	1e-3	100	Adam	1 layer (p=0.5)	Yes	ReLU	62	N/A	18.6	18.9
12	Street	1	64	1e-3	100	Adam	1 layer (p=0.5)	Yes	ReLU	31	N/A	2.9	22.2
13	Street	5	64	1e-3	100	Adam	1 layer (p=0.5)	Yes	ReLU	62	N/A	17.8	16.4
14	Concat (1 FC)	5	64	1e-3	100	Adam	None	No	ReLU	N/A	N/A	10.9	7.7
15	Concat (2 FC)	5	64	1e-3	100	Adam	None	No	ReLU	N/A	N/A	5.6	12.5
16	Concat (2 FC)	5	64	1e-3	100	Adam	2 layers (p=0.75)	No	ReLU	N/A	N/A	9.9	13.0
17	Concat (3 FC)	5	64	1e-3	100	Adam	2 layers (p=0.75)	Yes	ReLU	N/A	N/A	8.4	10.4
18	Concat (2 FC)	5	64	1e-3	100	Adam	2 layers (p=0.5)	Yes	ReLU	N/A	N/A	8.9	12.7
19	Concat (2 FC)	5	64	1e-3	100	Adam	2 layers (p=0.5)	Yes	Tanh	N/A	N/A	9.9	10.9

A.3 Additional Results Visualizations

Table 7: Test set distributions of ozone labels and ozone predictions from the concatenated feature representations

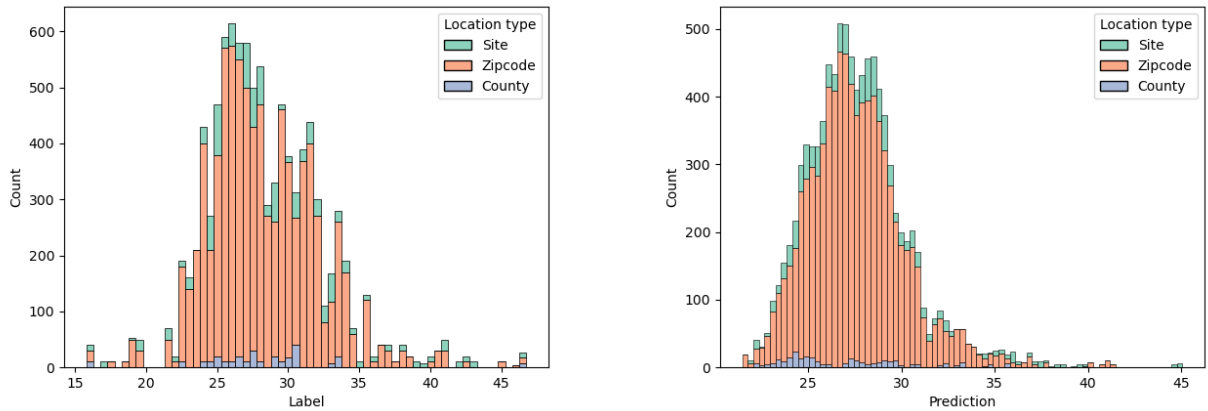


Table 8: Test set distribution of ozone reading errors and relationship to ozone labels

