Learning Pollution Maps from Mobile Phone Images

Ankit Bhardwaj , Shiva Iyer , Yash Jalan , Lakshminarayanan Subramanian
Courant Institute of Mathematical Sciences, New York University
{bhardwaj.ankit, shiva.iyer, lakshmi}@nyu.edu

Abstract

Air pollution monitoring and management is one of the key challenges for urban sectors, especially in developing countries. Measuring pollution levels requires significant investment in reliable and durable instrumentation and subsequent maintenance. On the other hand, there have been many attempts by researchers to develop image-based pollution measurement models, which have shown significant results and established the feasibility of the idea. But, taking image-level models to a city-level system presents new challenges, which include scarcity of high-quality annotated data and a high amount of label noise. In this paper, we present a low-cost, end-to-end system for learning pollution maps using images captured through a mobile phone. We demonstrate our system for parts of New Delhi and Ghaziabad. We use transfer learning to overcome the problem of data scarcity. We investigate the effects of label noise in detail and introduce the metric of in-interval accuracy to evaluate our models in presence of noise. We use distributed averaging to learn pollution maps and mitigate the effects of noise to some extent. We also develop haze-based interpretable models which have comparable performance to mainstream models. With only 382 images from Delhi and Ghaziabad and single-scene dataset from Beijing and Shanghai, we are able to achieve a mean absolute error of 44 $\mu g/m^3$in $PM_{2.5}$ concentration on a test set of 267 images and an in-interval accuracy of 67% on predictions. Going further, we learn pollution maps with a mean absolute error as low as 35 $\mu g/m^3$and in-interval accuracy as high as 74% significantly mitigating the image models' error. We also show that the noise in pollution labels emerging from unreliable sensing instrumentation forms a significant barrier to the realization of an ideal air pollution monitoring system. Our codebase can be found at https://github.com/ankitbha/pollution_with_images.

1 Introduction

Ambient air pollution is a major threat to global health and climate change today. Ambient air pollution is the cause of stroke, heart disease, lung cancer, acute and chronic respiratory diseases, resulting in an estimated 4.2 million deaths globally [(WHO), 2021]. Even short-term exposure to poor air containing higher than permissible levels of $PM_{2.5}$ and ozone has been shown to cause non-accidental and respiratory deaths [Lei *et al.*, 2019]. Particulate matter (PM), both $PM_{2.5}$ and PM_{10}, refers to fine airborne particles of various sizes, generated due to anthropogenic causes such as vehicles, chimneys etc. Fine PM or $PM_{2.5}$ refers to particles with diameter smaller than 2.5 microns [(EPA), 2021]. These are the most hazardous to the health of living organisms, since they can potentially get deep into the lungs and the bloodstream. In many urban areas, especially in developing nations, average $PM_{2.5}$ levels are high due to the presence of vehicles, industries and other human activities. Typically, $PM_{2.5}$ is accurately determined and measured by expensive industrial-grade monitors, which require significant maintenance and calibration. However, in many instances, outdoor images can often reveal important atmospheric information, such as the amount of haze in the ambient atmosphere, which is correlated to the level of air pollution. Many researchers have explored the idea of image-based pollution measurement models, and we present their findings in section 2. Thus, it is imperative that one would consider building an end-to-end image based air pollution monitoring system.

In this paper, we tackle the problem of learning city-wide pollution maps from a small set of mobile phone images taken across various locations in the city.

- **Scarcity of Labeled Data**: An important shortcoming of all previous studies is their reliance on high-quality annotated image data, which is notoriously difficult to obtain. Many past works use multi-scene dataset crawled from Beijing tourism website[1], where exact time and location of the image capture is available. It should be noted that most image-data-rich platforms (like Instagram) consider location and timestamp as private features and strip this information before uploading to their websites. Such data sources are unavailable for

[1] https://www.tour-beijing.com/real_time_weather_photo/

most cities around the world. Another publicly available resource are single-scene datasets from Beijing and Shanghai, but such datasets are few and not very useful for learning pollution maps because of lacking variety of images. Generalizing from single-scene to multi-scene is a very challenging problem in general.

- **Label Noise**: One important point that has been neglected in all the past literature on the problem is the difficulty in measuring air pollution, which shows huge variations across short intervals in space and time, and the label noise introduced in datasets due to this. Experimentally, we have found that even 2-kilometer distance can lead to significantly different readings in pollution sensors for the city of Delhi.
- **Image Quality**: Most of the past work uses high-quality images, which limits their performance on mobile phone images collected by untrained photographers.

Our contributions in this paper are outlined below:

- We use data augmentation and transfer learning framework to train deep learning models for predicting air pollution using extremely few and low quality images.
- We develop interpretable models with comparable performance to mainstream models. For this, we use haze as a representative feature of air pollution and pre-train a U-Net architecture for predicting haze density maps, before moving onto predicting $PM_{2.5}$ values.
- We investigate the effect of noise in detail, and provide a suitable metric of In-Interval accuracy to measure our models' performance in presence of noise.
- We learn pollution maps using distributed processing and aggregation on top of our model predictions, through which we mitigate the effects of noise to some extent.

2 Related Work

There has been some previous work which have used images to estimate or classify $PM_{2.5}$ concentration levels. Many past approaches have used deep learning to solve this problem.

Purely deep learning approaches: [Bo *et al.*, 2018], [Zhang *et al.*, 2016; Zhang *et al.*, 2017] and [Chakma *et al.*, 2017] use different deep convolutional neural networks to estimate or classify the $PM_{2.5}$ concentration levels. While [Bo *et al.*, 2018] combine the output of the CNN on the images with weather information in a regression model to produce the final estimate, [Zhang *et al.*, 2016] and [Chakma *et al.*, 2017] feed a CNN to the input images to produce a result. All three deep neural network approaches do not provide any understanding or intuition as to why the network learns the correct features and don't generalize to noisy data sources. [Ma *et al.*, 2018] uses an interesting approach, in which they also try to leverage the information from classical features from the dark channel map, which have been shown to be correlated to haze density and have been used in classical methods to dehaze images ([He *et al.*, 2009]). However, they have also formulated the problem as a classification problem while reporting only the accuracy, and not the precision, recall or F-1 scores. Also, they predict only three categories, whereas air quality index is recommended to be divided into at least 6 categories based on health impact [(WHO), 2021]. [Rijal *et al.*, 2018] try to use ensemble technique to improve their model performance and robustness, but don't explicitly tackle label noise.

Hybrid approaches: [Wang *et al.*, 2014] and [Liu *et al.*, 2016] use domain specific knowledge to model scattering of light and accordingly generate features as input to a regression model to determine haze level, air quality or particulate matter concentration levels. Both additionally require a depth map of the image. While [Wang *et al.*, 2014] also require a sequence of the same image as input to their model, the model proposed by [Liu *et al.*, 2016] is trained specifically for three single scene images. [Li *et al.*, 2015] propose a method that relies on modeling light propagation to produce an estimate of $PM_{2.5}$ concentration levels, but uses combination of a depth map obtained using a deep convolutional neural field and a transmission function obtained using a dark channel prior to find correlation between the derived features and the $PM_{2.5}$ concentration levels.

In section 4, we have evaluated several models derived from the past work on the novel setting of transfer learning with limited data and highly noisy labels, and compared their performance on our test set.

3 Methodology

Our proposed system uses images taken by a mobile phone camera to estimate ambient $PM_{2.5}$ levels. Note that we are only interested in the "average" amount of pollution in the region where the photo is taken, rather than the exact level of $PM_{2.5}$ at a particular location coordinate. The exact level of pollution could vary depending on how finely we divide up the spatial region. An image could be captured right next to an open fire or a food cart, and this would result in a spike at that precise location. For the purposes of pollution monitoring, we are not interested in such minute fluctuations, but are more interested in the average $PM_{2.5}$ in a particular region. It is important to be able to obtain trustworthy ground truth information on the pollution at the approximate location where the image is taken. We say "approximate" since the image would cover a sizeable region in space, about 100-200 meters, which is more than the error margin of most GPS devices and mobile phones. Hence, we impose certain constraints on our data. First, the images have to be of outdoor scenes only, with a bit of background and the sky, in order to capture the ambient environment on a large-scale and limit the effect of local variations. Second, we only take images from the vicinity of established government sensors ($<$ 2 km from the nearest sensor), because even average pollution levels can vary wildly at larger distances. We assume that at sufficient altitude and with sufficient dispersion, a region of radius 2 km should have the same average ambient pollution. Note that, as we will see in section 4.2, the choice of 2 km is somewhat arbitrary, but even if we choose a smaller distance, the noise estimate remains unchanged.

3.1 Dataset

Our dataset consists of outdoor images taken in Beijing, Shanghai, Delhi and Ghaziabad. For Beijing and Shanghai, we have single-scene images, i.e. several images taken of the exact same scene each over several days in a year and at various times (figure 1). There are in total 323 images for Beijing. For Shanghai, we have 1906 images of the Oriental Pearl Tower in Shanghai City captured at different times of the day and days of the year [Liu *et al.*, 2016]. The date and times of the captured images are exact, and the $PM_{2.5}$ concentrations obtained from US Mission China [2] are matched with the date and time of the image. This constitutes the clean data that we send into our model.

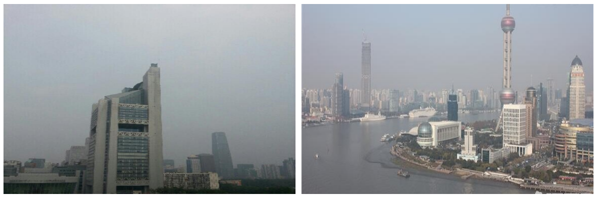

Figure 1: Single-scene image dataset from Beijing (323 images) and Shanghai (1906) used for transfer learning. The entire dataset consists of the same scene captured using a stationary camera on different dates throughout a year.

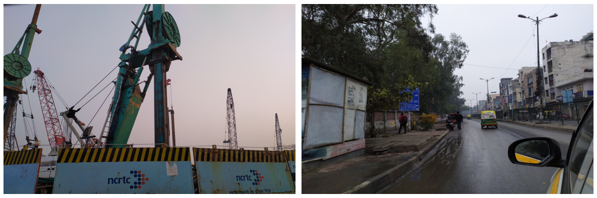

Figure 2: Multi-scene image dataset from Delhi and Ghaziabad. These images (382 total) have been collected by us using mobile phone cameras in the vicinity of government deployed air pollution monitoring sensors.

For Delhi and Ghaziabad, we have collected a total of 382 multi-scene images, which satisfy the above stated constraints. Figure 2 shows some sample examples from the dataset. These images are then matched to their corresponding $PM_{2.5}$ values obtained from the CPCB data portal[3] using the capture timestamp and GPS coordinates measured by the mobile phone. We divide our total data into two sets. The first set contains 80% of the data from China and 30% of the data from India. We also add random crops covering 50% area of images for India images to the first set to form our training and validation set. The remaining data constitutes our test set.

3.2 Models

We developed multiple deep learning models to process images and compute the $PM_{2.5}$ concentration levels. Most of these models have been adapted from previous works, and applied to our specific problem case of transfer learning and noisy labels. Most of our models use a popular feature extractor backbone tied to a fully connected network (FCN) acting as the estimator. The FCN estimator we use largely remains constant in structure across models. The first layer of the estimator reduces the total number of features to 120 dimensions, which is further reduced to 84 before being reduced to 1 dimensional output. We employ the following models on our problem statement:

- A Resnet-50 [He *et al.*, 2016] backbone feature extractor, motivated by [Bo *et al.*, 2018].
- A VGG-16 [Simonyan and Zisserman, 2015] backbone feature extractor, motivated by [Chakma *et al.*, 2017].
- An Inception-V3 [Szegedy *et al.*, 2016] backbone feature extractor.
- An ensemble of Resnet-50, VGG-16 and Inception-V3 feature extractors, motivated by [Rijal *et al.*, 2018].
- An EPAPLN feature extractor based on [Zhang *et al.*, 2016] and [Zhang *et al.*, 2017], but using the ReLU activation function.

While these models have been motivated by past work, they are hard to understand intuitively, as particulate pollution in images is not visible to the naked eye. Here, we hypothesize that haze can be considered as a human interpretable representation of pollution in images. To establish this, we developed haze-based models that are pre-trained heavily on haze density maps. The first step is to estimate the amount ”haze” in an image and create a haze density map. We employ a U-Net [Ronneberger *et al.*, 2015] for this. Following this, the haze density map along with the image is passed to a LeNet-5 [Lecun *et al.*, 1998] to extract any relevant features. These features are passed to our FCN estimator for final predictions. We call this pipeline LeUNet (figure 3). For training this architecture, a pixel-wise haze density approximation for numerous web-crawled images was obtained from a transmission function using computer vision algorithm of dehazing an image via a dark channel prior [He *et al.*, 2009]. Then, a U-Net was pre-trained using this data to predict haze density maps from input images. This U-Net along with its learned weights is loaded into the LeUNet model instance and trained end-to-end to predict $PM_{2.5}$ values. We also ensemble the LeUNet model with Resnet-50, VGG-16 and Inception-V3 models. Based on the results in 4.1, we claim that our hypothesis holds merit.

3.3 Transfer Learning and Data Augmentation

The total number of images in our India dataset is only 382, which is much smaller than the requirements of various deep learning models in today’s data-driven landscape. As we have mentioned earlier, data scarcity is a huge challenge in the field and there are almost no benchmark datasets available in the public domain. Thus, we have formulated the problem as a transfer learning problem, where we train our models using a combination of single-scene data from China and a small subset of data ($\sim$ 30%) from India. The number of single-scene images in the training set is 1783 while the number of multi-scene images is only 115. We added randomly cropped

[2] http://www.stateair.net/web/historical/1/1.html

[3] https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing/caaqm-comparison-data

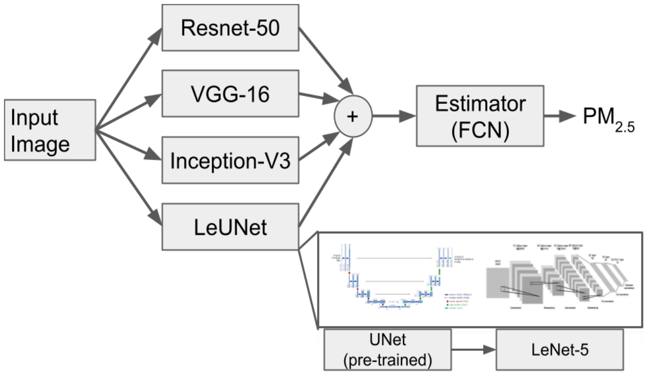

Figure 3: Architecture diagram for ensemble model on LeUNet pipeline. The LeUNet pipeline is pre-trained using haze density maps, which are correlated to $PM_{2.5}$ values. This model combines elements of interpretability with robustness.

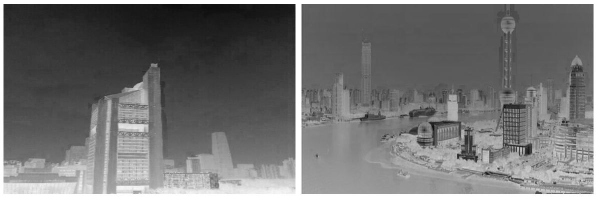

Figure 4: Haze map densities of sample images from single-scene datasets. We claim that haze can be considered as human-understandable representation of particulate pollution.

segments of images covering 50% area of the original image into the training set.

3.4 Pollution Map Derivation

Image-based models are not supposed to work individually, but as a distributed sensing system that generated a pollution map as its end result. The part where we generate the pollution map is extremely important, not only because it gives us a larger picture of the entire city, but it also allows us to mitigate the error in our image-model to some extent. For generating a prototype pollution map of Delhi and Ghaziabad using the test images, we collected the predictions of our models on the test set and averaged the predictions at different sensor locations. The averaged predictions reflect the average score of the models on all images corresponding to the neighborhood of a particular sensor location in a unit time interval. Since we collected images in the vicinity of each sensor location for only a few minutes, which is less than the time granularity of these sensors (15 minutes), all such images have the same ground-truth value. Then, for each model, we used the ground-truth and prediction values to generate two different pollution maps of Delhi and Ghaziabad regions, one as our gold standard based on sensor readings and the other derived from the predictions of image-models. The pollution maps were generated by fitting a Gaussian mixture model onto the sensor data and using the same model to generate the image-based pollution map.

One important aspect of an image-based estimation models is our ability to deploy them on mobile devices. For demonstration, we implemented a Java-based mobile application where we ported our model into Android using PyTorch Mobile. Our app has the ability to capture images, note down the date, time and location information, predict the ambient air pollution level from the image, and bundle and send this information to the server where we compute the pollution map. In light of internet connectivity issues, we also store the relevant information in storage before we have received the same on the server. This app can not only be used for the purposes of generating pollution maps and data collection, but can also be extended to do personal pollution exposure monitoring.

4 Evaluation

All the models are trained with mean square error loss function. The final test set consists of 20% of the China data equalling 446 images and 70% of the India data equalling 267 images. With optimizer as Adam (algorithm is based on AdaGrad, which uses adaptive gradients for each iteration and is known to work well for large datasets) and learning rate as 0.0001, we train using 80% of the training set and validate using 20% of the training set. The $PM_{2.5}$ concentration levels range between 0 and 1000. The related papers either use different datasets (mainly multiple-scene Beijing dataset) for evaluation or treat the task as a classification task, so the results are very hard to compare against previously published work. Also, the lack of open-sourced code and full implementation details make it harder to do an accurate replication. In this section, we evaluate the trained models on our test set, both in absolute terms and with respect to noise.

4.1 Prediction Performance

Table 1 shows the mean absolute test error (MAE) of our models on different partitions of the test data. The results show that ensemble and inception feature extractors perform the best on our test set, with their MAE around 45 and 44 $\mu g/m^3$respectively. For the China test set, we get a standard deviation of $\sim$ 12 $\mu g/m^3$. The relatively low standard deviation of the errors show that the models are able to, with good accuracy and certainty, approximate the $PM_{2.5}$ concentration levels for the single-scene image dataset. The relatively higher mean and standard deviation ($\sim$ 41 $\mu g/m^3$) on the multiple-scene test dataset of Delhi and Ghaziabad indicate that even though the models are able to approximate relatively well for many images in the set, there are many images where the performance is far from optimal. Note that haze-based models perform comparable to other deep learning approaches, giving strong indication that there is strong correlation between haze and particulate pollution.

Given the challenging task of training image-based estimation models using such a small dataset, this result can be considered satisfactory. To improve the performance further, we will need a larger source of data. It should also be noted that we have not yet analyzed the label noise, which will further shed light on our models' performance.

4.2 Noise Evaluation

Looking at the error numbers for India dataset in table 1 shows us that there is significant room for improvement. But,

Model	China test set (446)		India test set (267)	
	MAE	Std. Dev.	MAE	Std. Dev.
LeUNet	12.44	16.06	51.90	42.02
Resnet-50	9.94	14.97	46.76	38.53
EPAPLN	14.57	18.50	54.25	45.29
VGG16	10.52	13.67	48.08	36.94
Inception V3	9.77	12.15	**44.08**	40.74
Ensemble(VGG16, ResNet-50, Inceptionv3)	**9.73**	12.75	44.82	41.12
Ensemble(LeUNet, VGG16, ResNet-50, Inceptionv3)	9.92	10.94	47.94	40.32

Table 1: Mean Absolute Error (MAE) in $\mu g/m^3$on different partitions of our test set. The prediction range goes from 0 to 1000. The India partition (multi-scene) of test set is the partition of interest and the China partition (single-scene) only serves validation purposes. As can be seen, the Inception-V3 based model is the best performing model, with the ensemble model based on Rijal *et al.* coming a close second. We use the ensemble model for generating the pollution map.

as pointed before, without accounting for label noise in images, it would be wrong to jump to conclusions. To estimate the label noise in images, we used the data from government air pollution sensors deployed in the city of New Delhi and Ghaziabad from the CPCB data portal[4]. In this dataset, we took all pairs of sensors less than 2 kilometers from each other, and measured the mean absolute substitution error between them. Substitution error refers to the error in the readings if one sensor was to be used to predict the value of the other sensor. Table 2 shows the substitution error for different sensors as well as the corresponding distance between them.

Sensor Location 1	Sensor Location 2	Distance (km)	MASSE ($\mu g/m^3$)
Ashok Vihar - DPCC	Wazipur - DPCC	1.66	60.16
Dilshad Garden - CPCB	Vivek Vihar - DPCC	1.58	29.24
Jawaharlal Nehru Stadium - DPCC	Lodhi Road - IMD	1.43	47.53
Pusa - DPCC	Pusa - IMD	0.0001	54.59
Pusa - DPCC	Shadipur - CPCB	1.32	49.53
Pusa - IMD	Shadipur - DPCC	1.32	53.71

Table 2: Mean Absolute Sensor Substitution Error (MASSE). The table shows all pairs of government monitors in Delhi and Ghaziabad that are less than 2 km away from each other. To calculate the substitution error for every pair, we use one sensor to predict the reading of another sensor, and get the mean of absolute error. This gives us an idea of the difference in sensor readings even at close distances.

There are many important things to note about table 2. First, Delhi Pollution Control Committee (DPCC), Central Pollution Control Board (CPCB) and Indian Meteorological Department (IMD) are three different government bodies that monitor air pollution in the area through their own deployed sensors. As we can see in the table, despite a relatively short distance between the sensors, their readings vary from 30 $\mu g/m^3$to 60 $\mu g/m^3$. Next, note that the location of Pusa has two pollution sensors deployed by DPCC and IMD respectively, which are extremely close. Even in this case, there is a significant difference between the two sensors' readings. These differences in sensor readings might be due to a variety of reasons, including difference in altitude, hardware, calibration error and so on. This data shows that pollution measurement is not an easy task even using industrial grade sensors, thus it is almost certain that there is significant label noise in our dataset. Also, even with better and more elaborate data collection strategies, we are likely to struggle if we want to obtain noise-free annotations. Clearly, these results are not free from errors, but this is the best possible estimate we can have for the label noise in our dataset. Averaging across these sensors, we estimate that the label noise could be as high as 50 $\mu g/m^3$. It must be emphasized that this is not an artifact of our data collection process, but results from the lack of reliable technology for measuring air pollution. Thus, we are unable to ascertain the true performance of our models beyond a certain limit. To elaborate, say that a hypothetical model were to have zero error on the test set. We can't claim that this model is better than another model that makes an error of almost 50 $\mu g/m^3$on every image in the test set. We call this result as the uncertainty principle of pollution measurement.

4.3 Normalized Performance

Given the uncertainty principle, we know that we can not distinguish between two models which are able to predict within 50 $\mu g/m^3$of the ground truth value. Thus, how many times our models predict values under 50 $\mu g/m^3$of ground truth can be considered as a metric indicative of model performance. Let's define the interval of size 100 centered around the ground truth value as the target interval. We are interested in the in-interval accuracy of our models, which is equal to the number of images when our model's prediction lies in the target interval divided by the total number of images. The ideal model would achieve a 100% in-interval accuracy. Table 3 lists the in-interval accuracy of different models. Note that we are taking a hard maximum on the noise value to be 50 $\mu g/m^3$, which is stricter than what we have observed in the sensor readings.

[4] https://cpcb.nic.in/real-time-air-qulity-data/

Model	In-Interval Accuracy
LeUNet	0.558
Resnet-50	0.603
EPAPLN	0.599
VGG16	0.581
Inception V3	**0.674**
Ensemble (VGG16, ResNet-50, Inceptionv3)	0.614
Ensemble (LeUNet, VGG16, ResNet-50, Inceptionv3)	0.614

Table 3: In-Interval Accuracy Comparison. In-interval refers to the event when the model's prediction lies at a distance of less than 50 $\mu g/m^3$from the ground truth value. Due to label noise, we cannot effectively compare two predictions inside this interval. Here, we list the accuracy of our models if in-interval predictions are deemed correct.

4.4 Pollution Map Evaluation

The oversized estimate of label noise at 50 $\mu g/m^3$is a sobering realization, but it must be kept in mind that this result applies equally to the current pollution monitoring systems. If the sensor readings were gold standard, we would not see such a large divide in the readings between Pusa - DPCC and Pusa - IMD sensors. The pollution maps obtained after aggregating predictions in a neighborhood give us the most broad picture of pollution in the city. For evaluating pollution maps, we compare the gold standard maps obtained using sensor readings to the maps obtained on averaged predictions. The mean error between the sensor readings and averaged predictions for different models is summarized in table 4. One of the pollution maps, generated using the ensemble model, are contrasted in figure 5.

Model	Averaged Prediction MAE	In-Interval Locations
LeUNet	38.09	11
Resnet-50	40.31	11
EPAPLN	35.43	**14**
VGG16	37.99	11
Inception V3	35.99	13
Ensemble(VGG16, ResNet-50, Inceptionv3)	**35.13**	13
Ensemble(LeUNet, VGG16, ResNet-50, Inceptionv3)	44.76	10

Table 4: Pollution Map Evaluation. We compute the pollution map over 19 locations across Delhi and Ghaziabad using multiple models' averaged predictions. The Averaged Prediction Mean Absolute Error is the mean of absolute error made by averaged predictions across all 19 locations. In-Interval Locations count the number of locations where our learned map is indistinguishable from the gold standard map due to uncertainty in measurements.

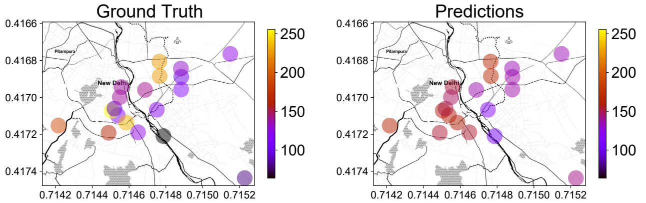

Figure 5: Ground truth and predicted pollution maps for Delhi and Ghaziabad derived using Ensemble model. The x and y axes are the latitude and longitude coordinates, while the color scheme denotes the pollution levels.

5 Conclusions and Future Work

In this paper, we present our end-to-end system of image-based air pollution monitoring and evaluate it using a pilot study in Delhi and Ghaziabad. We identify the two main challenges in prototyping and eventual deployment of such systems, namely the scarcity of labeled data and the noise in said labels. We show that transfer learning can overcome the data scarcity to a large extent, while distributed processing can somewhat mitigate the label noise. We also estimate the amount of label noise in our data, while also showing the limit of current methodology. We also evaluate our models on absolute and normalized scales, giving detailed insight into the system's performance.

In our pilot, we are able to achieve an error of $\sim$ 35 $\mu g/m^3$in the generated pollution map using only 382 images. This can be considered an impressive result by all means. Based on this, we think that image-based air pollution monitoring is a feasible concept. With sufficient resources spent on data collection and new innovations on the AI front of things, image-based monitoring systems can be a viable alternative to sensor-based monitoring systems.

There are many avenues for future work for this paper. One could do a large-scale experiment by deploying the image-model and empirically measure the error in pollution maps generated. One could come up with better data collection strategies and methods to circumvent the problems of data scarcity and noise. One could also look into weak-supervision methods that predict the pollution using noisy labels and other domain-specific features like weather, visibility, scattering of light etc.

Acknowledgements

This work was supported in part by funding from industrial affiliates in the NYUWIRELESS research group[5]. Shiva was also funded in part by an NSF Grant OAC-2004572. We thank Aurojit Panda, Rohini Pande and Anant Sudarshan for their constructive suggestions on the work and the paper draft.

[5] https://www.nyuwireless.com

References

[Bo *et al.*, 2018] Qirong Bo, Wenwen Yang, Nabin Rijal, Yilin Xie, Jun Feng, and Jing Zhang. Particle Pollution Estimation from Images Using Convolutional Neural Network and Weather Features. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3433–3437, October 2018. ISSN: 2381-8549.

[Chakma *et al.*, 2017] Avijoy Chakma, Ben Vizena, Tingting Cao, Jerry Lin, and Jing Zhang. Image-based air quality analysis using deep convolutional neural network. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3949–3952, September 2017. ISSN: 2381-8549.

[(EPA), 2021] United States Environmental Protection Agency (EPA). Particulate matter (pm) basics. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics, 2021. Accessed: 2022-06-01.

[He *et al.*, 2009] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, June 2009. ISSN: 1063-6919.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[Lecun *et al.*, 1998] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Lei *et al.*, 2019] Ruoqian Lei, Furong Zhu, Han Cheng, Jie Liu, Chaowei Shen, Chao Zhang, Yachun Xu, Changchun Xiao, Xiaoru Li, Junqing Zhang, Rui Ding, and Jiyu Cao. Short-term effect of pm2.5/o3 on non-accidental and respiratory deaths in highly polluted area of china. *Atmospheric Pollution Research*, 10(5):1412–1419, 2019.

[Li *et al.*, 2015] Yuncheng Li, Jifei Huang, and Jiebo Luo. Using user generated online photos to estimate and monitor air pollution in major cities. In *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*, ICIMCS '15, pages 1–5, New York, NY, USA, August 2015. Association for Computing Machinery.

[Liu *et al.*, 2016] Chenbin Liu, Francis Tsow, Yi Zou, and Nongjian Tao. Particle Pollution Estimation Based on Image Analysis. *PLoS ONE*, 11(2):e0145955, February 2016.

[Ma *et al.*, 2018] Jian Ma, Kun Li, Yahong Han, and Jingyu Yang. Image-based Air Pollution Estimation Using Hybrid Convolutional Neural Network. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 471–476, August 2018. ISSN: 1051-4651.

[Rijal *et al.*, 2018] Nabin Rijal, Ravi Teja Gutta, Tingting Cao, Jerry Lin, Qirong Bo, and Jing Zhang. Ensemble of Deep Neural Networks for Estimating Particulate Matter from Images. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages 733–738, June 2018.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[Simonyan and Zisserman, 2015] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[Szegedy *et al.*, 2016] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[Wang *et al.*, 2014] Haoqian Wang, Xin Yuan, Xingzheng Wang, Yongbing Zhang, and Qionghai Dai. Real-time air quality estimation based on color image processing. In *2014 IEEE Visual Communications and Image Processing Conference*, pages 326–329, December 2014.

[(WHO), 2021] World Health Organization (WHO). Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health, 2021. Accessed: 2022-06-01.

[Zhang *et al.*, 2016] Chao Zhang, Junchi Yan, Changsheng Li, Xiaoguang Rui, Liang Liu, and Rongfang Bie. On Estimating Air Pollution from Photos Using Convolutional Neural Network. In *Proceedings of the 24th ACM international conference on Multimedia*, MM '16, pages 297–301, New York, NY, USA, October 2016. Association for Computing Machinery.

[Zhang *et al.*, 2017] Chao Zhang, Baoxian Liu, Junchi Yan, Jinghai Yan, Lingjun Li, Dawei Zhang, Xiaoguang Rui, and Rongfang Bie. Hybrid Measurement of Air Quality as a Mobile Service: An Image Based Approach. In *2017 IEEE International Conference on Web Services (ICWS)*, pages 853–856, June 2017.