Environmental Science & Technology

pubs.acs.org/est

Article

Urban Air-Quality Estimation Using Visual Cues and a Deep Convolutional Neural Network in Bengaluru (Bangalore), India

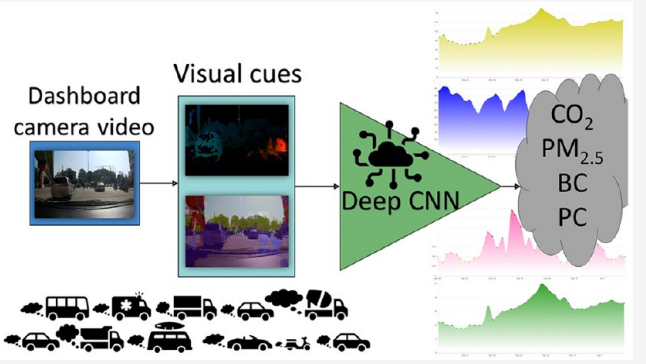
Alon Feldman,* Shai Kendler,* Julian Marshall, Meenakshi Kushwaha, V. Sreekanth, Adithi R. Upadhya, Pratyush Agrawal, and Barak Fishbain*

Cite This: *Environ. Sci. Technol.* 2024, 58, 480−487

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Mobile monitoring provides robust measurements of air pollution. However, resource constraints often limit the number of measurements so that assessments cannot be obtained in all locations of interest. In response, surrogate measurement methodologies, such as videos and images, have been suggested. Previous studies of air pollution and images have used static images (e.g., satellite images or Google Street View images). The current study was designed to develop deep learning methodologies to infer on-road pollutant concentrations from videos acquired with dashboard cameras. Fifty hours of on-road measurements of four pollutants (black carbon, particle number concentration, $PM_{2.5}$ mass concentration, carbon dioxide) in Bengaluru, India, were analyzed. The analysis of each video frame involved identifying objects and determining motion (by segmentation and optical flow). Based on these visual cues, a regression convolutional neural network (CNN) was used to deduce pollution concentrations. The findings showed that the CNN approach outperformed several other machine learning (ML) techniques and more conventional analyses (e.g., linear regression). The CO_2 prediction model achieved a normalized root-mean-square error of 10−13.7% for the different train-validation division methods. The results here thus contribute to the literature by using video and the relative motion of on-screen objects rather than static images and by implementing a rapid-analysis approach enabling analysis of the video in real time. These methods can be applied to other mobile-monitoring campaigns since the only additional equipment they require is an inexpensive dashboard camera.



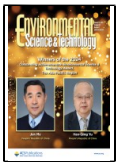
KEYWORDS: *traffic-related air pollution, deep learning, computer vision, mobile monitoring*

1. INTRODUCTION

The rise in the world's population and the constant increase in energy consumption make the degradation of air quality (AQ) a growing threat to global wellbeing. Air pollution has been associated with chronic and acute disease, higher mortality, global warming, and the deterioration of wildlife and agriculture.[1−5] Urban environments typically suffer from traffic-related air pollution (TRAP).[6] Effective AQ monitoring is the first step toward addressing this issue and considering the courses of action. Since the installation and maintenance of reliable monitoring equipment come at a high cost, sensory arrays are sparsely deployed. Thus, the resulting AQ map has a spatial resolution at the neighborhood scale at best.[7] Pollution levels between monitored points are usually assessed by land-use regression, interpolation, and dispersion models,[8] but these methods cannot replace actual measurements. Low-cost mobile sensors are often implemented to improve spatial resolution.[9] However, such sensors must be calibrated periodically, especially for outdoor applications, and are subjected to harsh conditions.[10,11] Hence, despite the impressive progress in the research and development of miniature, low-cost sensors, their operation still involves an economic and technical burden, thus underscoring the need to develop methods and devices for chemical sensing that can exploit other signals such as visual cues from imagery data to complement sensory data.

Previous studies have used satellite imagery for AQ predictions[12−14] with promising results for AQ assessment in large-scale areas with limited accessibility. Etzion et al.[15] made one of the first attempts to evaluate AQ using a ground camera. The rationale was that light waves are distorted and distant objects appear blurry in the presence of aerosols; hence, the relative blurriness of the images was used to estimate the particulate matter ($PM_{2.5}$) concentrations. In their study, a camera was placed to capture a wide landscape. Then, $PM_{2.5}$

Received: June 12, 2023
Revised: November 5, 2023
Accepted: November 6, 2023
Published: December 17, 2023



and blurriness were compared in three regions of interest in the image. Each region differed from the others in terms of the distance between the observed objects and the camera. The findings suggested that a blurriness-based approach failed at distances under 3.1 km because there was not enough medium for a clear-cut change in blurriness to estimate typical pollution states.[15]

Schroeder et al.[16] implemented an object detection algorithm to estimate traffic density and relate it to AQ measurements. Their setup consisted of a $PM_{2.5}$ sensor and a camera focused on a road. As cars drove by, the motion was estimated by comparing object locations in consecutive frames. The underlying assumption was that emissions could be predicted according to the vehicle type and movement parameters. While the basic premise is correct,[17] a detailed depiction of air pollution distribution must also account for topography and meteorological conditions. To incorporate structural (topographical) parameters into the estimation problem, Qi and Hankey[18,19] used a semantic segmentation neural network to analyze visual AQ-related data. Their year-long average pollution measurements were compared against Google Street View images at specific locations. The images were used to extract street structure features, such as the surrounding buildings, that could create a canyoning effect. However, averaging year-long data does not enable temporal analysis. In other studies, researchers have utilized deep convolutional neural networks (CNNs) to predict AQ directly from images.[20−23] These methods all used the black box approach, where the model extracts features from the raw data without physical reasoning.

This study advances previous studies, which used cameras and image features, by introducing a cue-based approach. This approach capitalizes on video features, such as motion and scenery, to infer air pollution. The method exploits the strong capabilities of CNNs to extract pollution-related features such as motion and segmentation from the original image. These features are used to construct a near-real-time AQ estimation model based on visual cues in the images acquired by commercial dashboard cameras. In this configuration, the scene and background change rapidly, which makes the problem extremely complex. This method enhances existing model frameworks, such as land use regression modeling, by implementing known relationships between land use and air quality, often from static data, with information from on-road vehicles. The advantages of this technique include the fact that it can cover a large area with high resolution and frequency and can account for the dynamic nature of traffic emissions and meteorological conditions. In addition, it relies on technology that will probably become part of autonomous vehicles in the near future so that the deployment of many data collection agents is already readily available. This paper describes the methodology and data processing steps involved in this technique and presents results based on data collected in Bengaluru, India. Bengaluru serves as an excellent environment to explore and deploy such technology due to its dense population, rapid urbanization, high traffic congestion, and significant number of vehicles on the road.

2. METHOD

2.1. Data Acquisition—Mobile Monitoring. Data were recorded using a commercial, off-the-shelf, Red-Blue-Green (RBG) dashboard camera, GPS, and air pollution instrumentation mounted on a car driving along a predetermined route in Bengaluru (Bangalore), India.[24] For the sake of completeness, a short description is provided here. Table 1 lists the measurement devices used in the survey to evaluate the fine particulate matter ($PM_{2.5}$), black carbon (BC), carbon dioxide (CO_2), and particle concentration, which subsumed particles measuring less than 0.1 μm.

Table 1. Devices Used during the Mobile Monitoring Campaign

parameter	instrument and model	units
video footage	Noise Play 2 action camera	
location (latitude, longitude)	Garmin GPSMAP64s	degrees
black carbon (BC)	microAeth-AE51	ng/m^3
particulate matter 2.5 ($PM_{2.5}$)	DustTrak-8530	mg/m^3
carbon dioxide (CO_2)	LI-850	ppmv
particle concentration (PC)	CPC-3007	$\#/cm^3$

The monitoring devices were installed in a compressed-natural-gas-powered hatchback car (Maruti Suzuki Celerio) with low tailpipe emissions compared to other cars. The temporal resolution of the air pollution measurements was set to 1 s (1 Hz). The car had a rear window that was kept open to allow air to enter the instruments mounted in it. This procedure was used to minimize self-pollution from the car's exhaust. The instruments were also cushioned and strapped with bungee cords to reduce the impact of road vibration on their performance. All instruments were portable, battery operated, and factory calibrated.

The monitoring campaign started in May 2019 and continued through the first week of March 2020. The surveys included video footage taken during 43 measurement days over 6 months. A typical measurement day started at around 9−10 am and finished around 12−13 pm. Figure 1 shows a map of Bengaluru, where the green shading indicates the regions where the air pollution levels were monitored. The study area was divided into four daily routes: the central business district (CBD), Kannuru (KAN), a peri-urban transecting route, and two routes in Malleshwaram (MAL), a residential neighborhood in Bengaluru. The entire study area was typically covered once a week.[24,25] There were high variations in pollutant concentrations along the routes, consistent with previous findings.[7,26] The highest pollution levels were found on the major roads, followed by arterials and residential areas. More information about the air quality monitoring campaign, device setup, and data correction have been described by Upadhya et al.[25] The air pollution instrumentation measurements were used here as the ground truth, whereas the video footage was used for feature set crafting that was then fed into the CNN model.

2.2. Data Preprocessing. Figure 2 presents the big picture of the process introduced here. Features were crafted from the raw videos in the following way: frames were sampled from the videos at a constant rate of 1 frame every 5 s (0.2 Hz) to avoid data redundancy and to smooth out noisy readings. Next, the pixel motion fields (for technical details, see section 2.2.1) and the pixel labels (section 2.2.2) were extracted from each frame. These features were concatenated along with the original frame to create a multichannel image. Finally, each multichannel image was associated with its air-pollution measurement: i.e., the ground truth, obtained during the same time as the frame. Figure 2 illustrates this process. Each multichannel image was composed of a 16-channel matrix consisting of three intensity

https://doi.org/10.1021/acs.est.3c04495
Environ. Sci. Technol. 2024, 58, 480−487

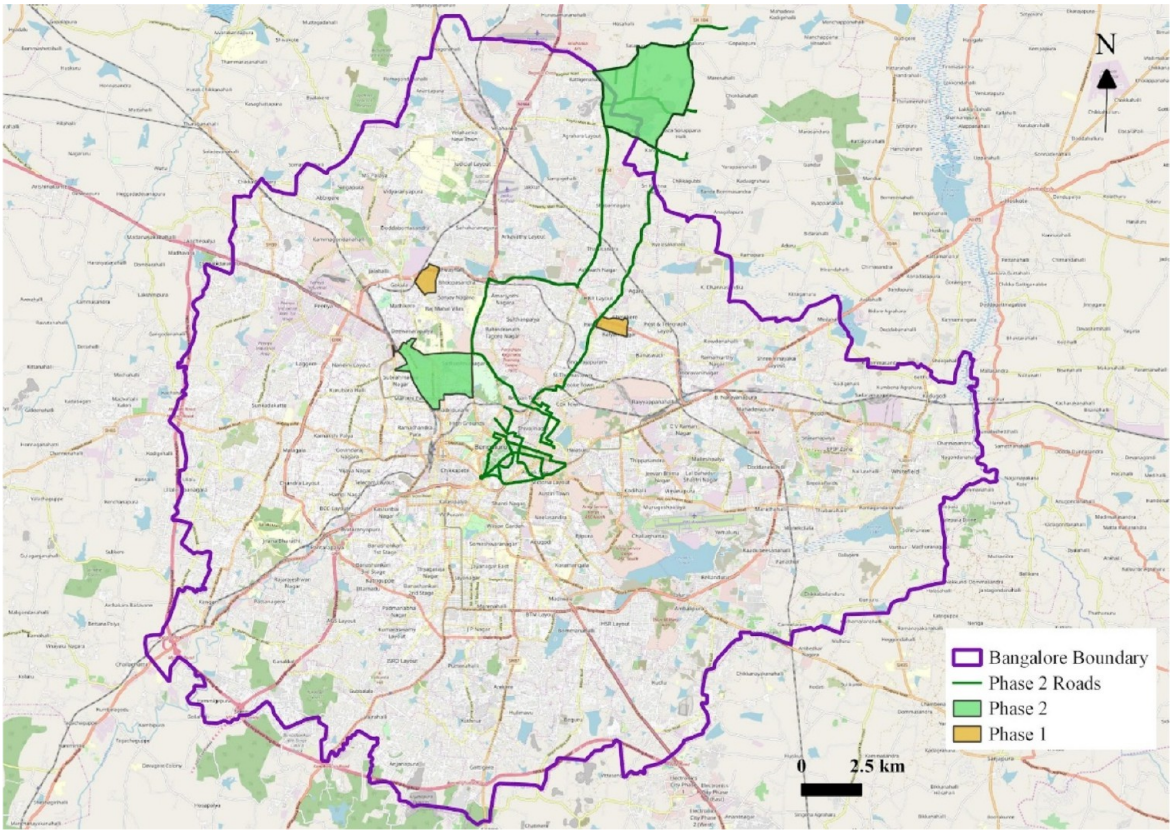


Figure 1. Map of the study area: Bengaluru (Bangalore), India.[25,27]

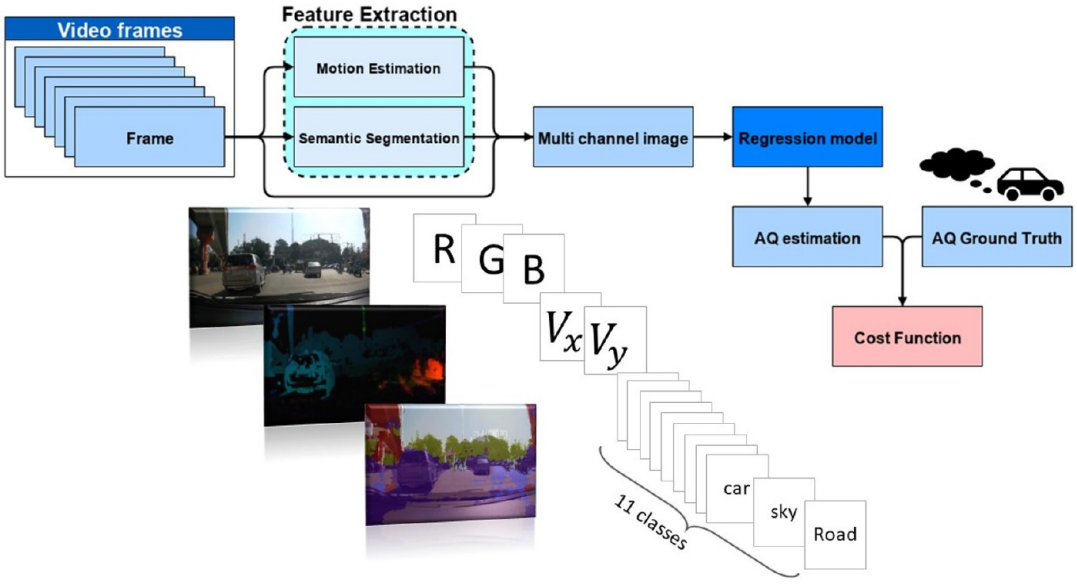


Figure 2. Model flowchart. The video was broken down into frames. Computer vision algorithms were applied to craft the features, which were fed into a regression model. The risk function compares the prediction to the ground truth. Minimizing the risk function improves the model performance over the training set.

RGB channels, two pixel-velocity channels (V_x, V_y), and 11 channels constituting the one-hot-vector for the pixel labels. One-hot encoding is a standard method of representing categorical variables in numerical form by creating a separate binary column for each category and assigning a value of "1" to the column corresponding to the category and "0" to all other columns. The AQ ground truth measurements here corresponded to the concurrent 5 s average pollution concentration.

2.2.1. Motion Estimation—Horn–Schunk Optical Flow. To determine whether the concentrations were related to motion in the image,[17,28] the frames' motion fields were estimated by the Horn–Schunk optical flow algorithm[29,30] to obtain the entire image velocity field. Figure 3 illustrates the

https://doi.org/10.1021/acs.est.3c04495
Environ. Sci. Technol. 2024, 58, 480−487

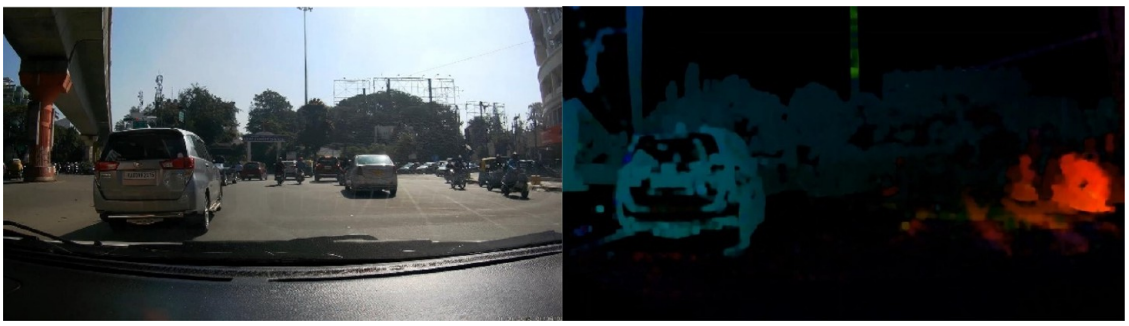


Figure 3. Dense optical flow: the video frame appears on the left, and the motion field is illustrated on the right. Hue represents the direction of movement, and intensity represents the magnitude.

Figure 4. Semantic segmentation: the original frame appears on the left, and the pixel labeling by category on the right.

frame velocity vector field, where the pixel hue represents the vector direction (red to blue for right to left) and its magnitude is determined by the color intensity. The car on the left is moving to the right (in blue), and the motorbikes on the right are moving to the left (in red).

2.2.2. Pixel Labeling—Image Semantic Segmentation. Image semantic segmentation was used to label each pixel in the video frame. Semantic segmentation provides information about sources of pollution, such as cars and trucks, and enables street structure analysis. For example, if tall buildings are on either side of a road, a canyoning effect may occur where particles get trapped and pollution levels rise. Alternatively, the unimpeded airflow may allow particles to disperse faster.

To label each pixel, a ″DeepLabv3+″ pretrained model,[3132] was used. The model contained a dictionary of 11 labels in total. The model was trained over the ″Cam-Vid data set″ from Cambridge University,[33] which contains street-level images obtained while driving that were suitable for the needs of this study. Figure 4 depicts the 11 pixel labels by color detected by this pretrained model; namely, ″Sky″, ″Building″, ″Pole″, ″Road″, ″Pavement″, ″Tree″, ″Sign-Symbol″, ″Fence″, ″Car″, ″Pedestrian″, and ″Bicyclist″.

2.3. Data Train-Validation-Split Strategies. Several strategies were implemented to divide the data into training and validation sets.

(1) The measurement data set was divided such that every "*n*" consecutive samples were assigned to the training set and the following "*m*" samples to the validation set. In this study, the *n* and *m* values were 100 and 20, respectively, for a 1000/200 temporal split.

(2) The data set was also split geographically, such that two routes were considered as the training set and the other two were the validation set. These divisions neutralized the temporal and spatial dependences between the train and validation sets.

Since the database was obtained in the same region and for consecutive times, many samples could have almost identical features. Therefore, these different splitting strategies were implemented to estimate the model's capability to generalize beyond the training set domain.

2.4. Deep-Convolutional Neural Network (Deep-CNN) Model Architecture and Training Process. A Deep-CNN is a neural network typically used for image input, which exploits its geometric structure. This model creates feature maps that represent complex scenarios capturing traffic behavior and street structure. The model here was based on the "Resnet-50" CNN[34] where the input size was modified to fit a 224-by-224 16-channel image, and the output layer was changed from a classification layer to a regression layer. The remainder of the network was initialized with the pretrained weights from a pretrained Resnet-50,[32] such that the network was trained using transfer learning (fine-tune). This architecture was chosen for its ability to parse simple and complex feature structures while maintaining a reasonable computational load.

One of the hurdles that needs to be addressed when applying ML methodologies is the computational complexity and execution time of the method (training and application). The training process was applied to several training sets, as described above. Each model learned to predict one air pollution target. The loss function for training was the mean square error function (MSE). The training was conducted in a MATLAB environment with a Nvidia Quadro-8000 GPU. The "Adam" optimizer[35] was used as an optimization algorithm, with a batch size set to 64 and an epoch number set to 10. With these computational resources, a typical training runtime

https://doi.org/10.1021/acs.est.3c04495
Environ. Sci. Technol. 2024, 58, 480−487

Table 2. Model Performance over the Different Targets and Train-Validation Division Strategies

pollutant-division method	RMSE	MAE	Validation SD	NRMSE (%)	Pearson coefficient
CO_2-Temporal [100 20]	51	37	59	10	0.49
CO_2-Temporal [1000 200]	66	44	72	13	0.40
CO_2-Spatial	48	39	51	10	0.37
PC-Temporal [100 20]	66960	52089	70930	81	0.34
PC-Temporal [1000 200]	71905	53584	72333	83	0.26
$PM_{2.5}$-Temporal [100 20]	189	93	198	100	0.32
$PM_{2.5}$-Temporal [1000 200]	171	88	167	101	0.17
BC - Temporal [100 20]	115	54	117	190	0.20
BC - Temporal [1000 200]	121	51	122	197	0.11

was under ten hours. The MATLAB code is publicly available at https://fishbain.net.technion.ac.il/home-page/projects-software.

2.5. Machine Learning (ML) Models for AQ Inference.

The method above was based on reconstruction of the image by adding several features to the RGB light intensity. The underlying assumption was that the distribution of these features in the scene would indicate the air pollution level. This assumption was verified by rearranging the features so that spatial information was omitted. The input dimension was reduced to a semantic segmentation pixel-label histogram, and the mean and standard deviations over the entire frame were obtained for the velocity. Various ML models were trained to predict the AQ measurements. Linear regression models, regression trees, boosted trees, bagged trees, the Support Vector Machine (SVM) and Gaussian Process Regression (GPR), both with different kernel functions, and fully connected neural networks were all applied. All of these models were trained and evaluated using the first temporal train-validation split of 100 training points to 20 validation points.

3. RESULTS AND DISCUSSION

The initial attempts to predict air quality measurements using simpler models, such as linear regression models, regression trees, boosted trees, bagged trees, Support Vector Machine (SVM) with different kernel functions, Gaussian Process Regression (GPR) with different kernel functions, and fully connected neural networks, did not yield satisfactory results. The validation R^2 values obtained for all of the pollutants and models were consistently below 0.1, indicating a poor fit between the predicted values and the actual measurements. Detailed information on these subpar results can be found in the Supporting Information.

Given the limitations of the simpler models and their inability to effectively capture the complexity of the problem, a more advanced approach was deemed necessary, and CNN was selected as the next step. This decision was motivated by CNN's capability to leverage spatial information by taking the specific locations and patterns within the scene into account for AQ estimation.

Table 2 shows the performance of the trained Deep-CNN models using different training-validation divisions and pollution targets. In the first row, the n and m values were 100 and 20, respectively, and 1000/200 in the second row. The first and second rows depict the results obtained using the train/test split strategies presented in paragraph 1 of section 2.3. The third row shows the results for the spatial split strategy presented in paragraph 2 of section 2.3. The other rows present other target labels: namely, PC, BC, and $PM_{2.5}$. For example, $PM_{2.5}$-Temporal [100 20] refers to a model predicting $PM_{2.5}$ concentrations. The columns show the RMSE and MAE loss values calculated over the validation set. The standard deviation (SD) of the validation set is also provided for comparison as well as the normalized root-mean-square error (NRMSE), which is the RMSE divided by the mean concentration. The last column lists the Pearson coefficients comparing the actual pollution to the model predictions in the validation set.

As shown in Table 2, the Deep-CNN model was the most effective in finding CO_2-related features. This was unexpected because the literature on TRAP typically concentrates on BC and $PM_{2.5}$. These variations in results are presumably due to different sensor response functions and pollutant properties, including the extent to which the variability in concentrations for a given pollutant was related to attributes discernible from the videos. As can be seen in Table 2, the CO_2 levels were predicted with a network trained on two different configurations. In the CO_2-Temporal model, the training/test split strategy was such that samples were drawn for each set from the same geographic area but at different times. In the CO_2-Spatial model, the samples were divided by their geographic regions. In this configuration, the network inferred the CO_2 levels from the routes where training was conducted on samples from other survey routes. In addition, the accuracy decreased as the distance between the train and validation data sets became larger. This decrease was expected because air pollution levels are also associated with geographic and climatic conditions that were not considered in the current CNN model. A more elaborate model analysis including a confusion matrix suitable for a regression task is provided in the Supporting Information.

To evaluate the significance of the motion feature, a separate model was trained without incorporating any motion information, since most previous studies have ignored this feature. The results showed that the model without motion had a 20% lower correlation with the ground truth than did the model with motion, which confirms our hypothesis that motion is an important feature for this task.

One frequently used method for comparing the accuracy of different sensors is the Bland–Altman analysis.[36] The Bland–Altman plot is a graphical tool used to assess the level of agreement between two measurement methods. In this case, it compares the image-based model against a reference sensor. The plot displays the difference between the two measurements on the y-axis and the mean of the two measurements on the x-axis. The black lines on the graph represent the limits of agreement, which are calculated as the mean difference $\pm$ 1.96 times the standard deviation of the differences. These limits indicate the range within which most of the differences

https://doi.org/10.1021/acs.est.3c04495
Environ. Sci. Technol. 2024, 58, 480−487

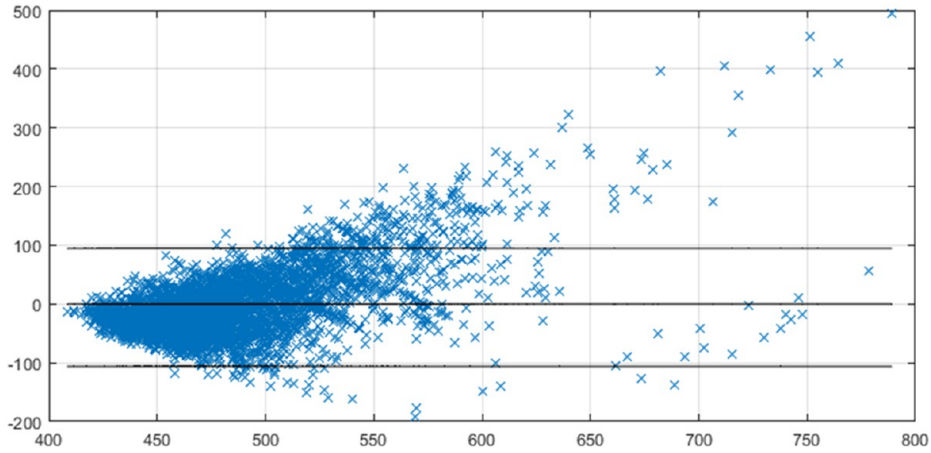


Figure 5. Bland–Altman plot comparing the difference between our image-based air pollution estimator and a reference sensor against their means. The black lines indicate that the agreement was within acceptable limits.

Figure 6. Most polluted (left) and least polluted scenes (right) identified by the model.

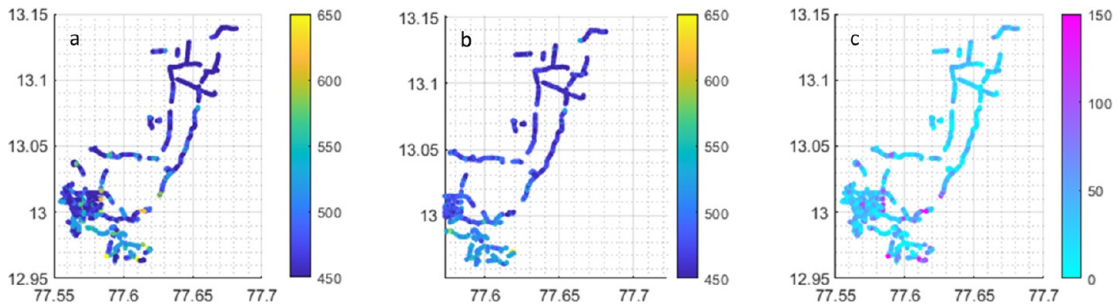


Figure 7. Ground truth CO_2 map (a), computed CO_2 map (b), and absolute error map (c).

between the two measurements would fall if the methods agree. Figure 5 shows that the majority of the data (95.12%) fell within these limits, indicating good agreement between the model and the ground-truth sensor.

To further validate the reliability of our methodology, we compared the distribution of the data in our study to that of Diez et al.[37] Their paper compared air pollution measurement errors in low- and high-end sensors. The similarity between the distribution patterns observed in Figure 5 and those presented in the study further confirms the robustness and accuracy of our approach.

https://doi.org/10.1021/acs.est.3c04495
Environ. Sci. Technol. 2024, 58, 480−487

Aside from formal mathematical evaluations of the method's performance, some images were subjected to human visual inspection. While such inspections cannot be used for a systematic study, they do provide insights into the nature of the method.

Figure 6 presents images of scenes classified by the CNN method as high and low CO_2 levels. The images taken in highly polluted scenes (left side of Figure 6) are characterized by dense traffic, wide roads with more than one lane, and surrounding buildings. On the other hand, the images taken in scenes with low pollution levels (right side of Figure 6) are characterized by fewer buildings, vegetation, and an open environment. Hence, the model can use indirect, albeit explainable, cues to predict pollution levels.

Figure 7 presents a map indicating CO_2 levels where (a) presents the ground truth measurement, (b) is the model estimations (using a Temporal 100/20 split), and (c) is the absolute value of their difference. Even though maps (a) and (b) are not identical, they evidence the method's potential to provide an estimation of CO_2 levels in high spatial/temporal resolution in large areas without dedicated air pollution instrumentation. The fine spatial resolution of the model estimations can also be useful in improving Land Use Regression methods, which rely on spatial predictors to estimate air pollution levels in areas where monitoring data are unavailable.

Thus overall, this paper explored the prediction of air quality measurements using a range of models based on visual features. The performance evaluation demonstrated the effectiveness of the Deep-CNN model in predicting CO_2-related features, thus showcasing its potential for indirect air pollution estimation. While limitations and challenges remain, this study lays the foundation for future research directions and advances in image-based air pollution estimation. The model presented here can distinguish between polluted and non-polluted scenes based on these visual features. Future work should examine why some scenes may have high pollution levels but do not show obvious signs of pollution in the images or why some scenes have low pollution levels but show misleading signs of pollution. These cases may require additional information or modalities to improve the accuracy of the method. This paper focused on TRAP, whose sources are typically found on the road. Future research could expand the analysis to off-road pollution sources. Further, a model trained in one environment will not achieve the same accuracy in an entirely different environment, because the relationship across variables may change.

One direction would be to optimize the model architecture and performance and perhaps consider applying this method to a setting where external conditions are controlled (such as airflow, climate, and self-motion). To enhance performance, models could be built to incorporate additional information such as weather conditions, nearby direct measurements, and Lidar signals (which are likely to be standard in autonomous vehicles) or to use short videos as input (where more information about the scene can be extracted as opposed to a single frame). Another enhancement would be to include a wider range of pixel labeling options: for example, categories such as types of vehicles, vegetation, and roads. Incorporating these specific label categories could provide more detailed insights into the air pollution levels associated with each element in the scene. Future model architectures could consider cutting-edge advances in computer vision such as 3D reconstruction, visual transformers, and other image processing methodologies. These should be considered carefully before being applied to this task to ensure that enough data and computational resources are available before heavier models are trained.

■ ASSOCIATED CONTENT

Data Availability Statement

Code samples related to the study are available at https://fishbain.net.technion.ac.il/home-page/projects-software for creating a database, model training, and model evaluation.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.3c04495.

Results of simpler ML models and a detailed confusion matrix demonstrating the performance of the CNN model (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Alon Feldman − *Department of Mathematics, Technion−Israel Institute of Technology, Haifa 3200003, Israel;* orcid.org/0009-0007-8692-6697; Phone: 972-4-829-3177; Email: alonfeldman@campus.technion.ac.il

Barak Fishbain − *Department of Environmental, Water and Agricultural Engineering, Faculty of Civil & Environmental Engineering, Technion−Israel Institute of Technology, Haifa 3200003, Israel*; Phone: 972-4-829-3177; Email: fishbain@technion.ac.il

Shai Kendler − *Department of Environmental, Water and Agricultural Engineering, Faculty of Civil & Environmental Engineering, Technion−Israel Institute of Technology, Haifa 3200003, Israel; Environmental Physics Department, Israel Institute for Biological Research, Ness Ziona 7410001, Israel*; Phone: 972-4-829-3177; Email: skendler@technion.ac.il

Authors

Julian Marshall − *Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0003-4087-1209

Meenakshi Kushwaha − *ILK Laboratories, Bengaluru 560046, India*

V. Sreekanth − *Center for Study of Science, Technology & Policy, Bengaluru 560094, India;* orcid.org/0000-0003-0400-6584

Adithi R. Upadhya − *ILK Laboratories, Bengaluru 560046, India; Department of Public Health, Policy & Systems, University of Liverpool, Liverpool L69 3GF, England*

Pratyush Agrawal − *Center for Study of Science, Technology & Policy, Bengaluru 560094, India;* orcid.org/0000-0003-4298-7854

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.3c04495

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was partially funded by the Israeli Ministry of Science and Technology, the Israeli Water Authority, the Israeli Data Science Initiative (IDSI), and the Council for Higher Education of Israel.

REFERENCES

(1) Kampa, M.; Castanas, E. Human health effects of air pollution. *Environ. Pollut.* **2008**, *151* (2), 362−367.

(2) Zhang, K.; Batterman, S. Air pollution and health risks due to vehicle traffic. *Science of The Total Environment* **2013**, *450*−*451*, 307−316.

(3) Smith, W. H. Air pollution—effects on the structure and function of the temperate forest ecosystem. *Environmental Pollution (1970)* **1974**, *6* (2), 111−129.

(4) Newman, J. R. Effects of industrial air pollution on wildlife. *Biol Conserv* **1979**, *15* (3), 181−190.

(5) Ramanathan, V.; Feng, Y. Air pollution, greenhouse gases and climate change: Global and regional perspectives. *Atmos. Environ.* **2009**, *43* (1), 37−50.

(6) Schwela, D.; Zali, O. *Urban Traffic Pollution*; CRC Press: 1998. DOI: 10.1201/9781482272093

(7) Moltchanov, S.; Levy, I.; Etzion, Y.; Lerner, U.; Broday, D. M.; Fishbain, B. On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Sci. Total Environ.* **2015**, *502*, 537−547.

(8) Jerrett, M.; Arain, A.; Kanaroglou, P.; Beckerman, B.; Potoglou, D.; Sahsuvaroglu, T.; Morrison, J.; Giovis, C. A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol* **2005**, *15*, 185−204.

(9) Kizel, F.; Etzion, Y.; Shafran-Nathan, R.; Levy, I.; Fishbain, B.; Bartonova, A.; Broday, D. M. Node-to-node field calibration of wireless distributed air pollution sensor network. *Environ. Pollut.* **2018**, *233*, 900−909.

(10) Castell, N.; Dauge, F. R.; Schneider, P.; Vogt, M.; Lerner, U.; Fishbain, B.; Broday, D.; Bartonova, A. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environ Int* **2017**, *99*, 293−302.

(11) Kendler, S.; Zuck, A. The Challenges of Prolonged Gas Sensing in the Modern Urban Environment. *Sensors* **2020**, *20* (18), 5189.

(12) Hong, K. Y.; Pinheiro, P. O.; Minet, L.; Hatzopoulou, M.; Weichenthal, S. Extending the spatial scale of land use regression models for ambient ultrafine particles using satellite images and deep convolutional neural networks. *Environ Res* **2019**, *176*, 108513.

(13) Hong, K. Y.; Pinheiro, P. O.; Weichenthal, S. Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data. *Environ Int* **2020**, *144*, 106044.

(14) Hong, K.; Weichenthal, S. A New Method of Estimating Global PM2.5 Concentrations using Satellite Images. *Environmental Epidemiology* **2019**, *3*, 432.

(15) Hong, K.; Weichenthal, S. A New Method of Estimating Global PM2.5 Concentrations using Satellite Images. *Environmental Epidemiology* **2019**, *3*, 432.

(16) Schroeder, A. K.; Haugen, M. J.; Stettler, M. E. J.; Boies, A. M. Using Computer Vision with Instantaneous Vehicle Emissions Modelling. *2020 Forum on Integrated and Sustainable Transportation Systems, FISTS 2020*, 2020; pp 89−94. DOI: 10.1109/FISTS46898.2020.9264872.

(17) Kamarianakis, Y.; Gao, H. O.; Holmén, B. A.; Sonntag, D. B. Robust modeling and forecasting of diesel particle number emissions rates. *Transp Res D Transp Environ* **2011**, *16* (6), 435−443.

(18) Qi, M.; Hankey, S. Using Street View Imagery to Predict Street-Level Particulate Air Pollution. *Cite This: Environ. Sci. Technol* **2021**, *55*, 2695.

(19) Qi, M.; Dixit, K.; Marshall, J. D.; Zhang, W.; Hankey, S. National Land Use Regression Model for NO 2 Using Street View Imagery and Satellite Observations. *Environ. Sci. Technol* **2022**, *56*, 13499−13509.

(20) Xu, J.; Zhang, M.; Ganji, A.; Mallinen, K.; Wang, A.; Lloyd, M.; Venuta, A.; Simon, L.; Kang, J.; Gong, J.; Zamel, Y.; Weichenthal, S.; Hatzopoulou, M. Prediction of Short-Term Ultrafine Particle Exposures Using Real-Time Street-Level Images Paired with Air Quality Measurements. *Environ. Sci. Technol.* **2022**, *56* (18), 12886−12897.

(21) Weichenthal, S.; Hatzopoulou, M.; Brauer, M. A picture tells a thousand···exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. *Environ Int* **2019**, *122*, 3−10.

(22) Hong, K. Y.; Pinheiro, P. O.; Weichenthal, S. Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data. *Environ Int* **2020**, *144*, 106044.

(23) Suel, E.; Sorek-Hamer, M.; Moise, I.; von Pohle, M.; Sahasrabhojanee, A.; Asanjan, A. A.; Arku, R. E.; Alli, A. S.; Barratt, B.; Clark, S. N.; Middel, A.; Deardorff, E.; Lingenfelter, V.; Oza, N. C.; Yadav, N.; Ezzati, M.; Brauer, M. What You See Is What You Breathe? Estimating Air Pollution Spatial Variation Using Street-Level Imagery. *Remote Sens (Basel)* **2022**, *14* (14), 3429.

(24) Upadhya, A. R.; Agrawal, P.; Vakacherla, S.; Kushwaha, M. mmaqshiny v1.0: R-Shiny package to explore Air-Quality Mobile-Monitoring data. *J Open Source Softw* **2020**, *5* (50), 2250.

(25) Upadhya, A. R.; Kushwaha, M.; Agrawal, P.; Gingrich, J.; Asundi, J.; Sreekanth, V.; Marshall, J. D., Apte, J. S., MULTI-SEASON MOBILE MONITORING CAMPAIGN OF ON-ROAD AIR POLLUTION IN BENGALURU, INDIA: HIGH-RESOLUTION MAPPING AND ESTIMATION OF QUASI-EMISSION FACTORS, 2023. DOI: 10.26434/CHEMRXIV-2023-QQ7PH.

(26) Lerner, U.; Yacobi, T.; Levy, I.; Moltchanov, S. A.; Cole-Hunter, T.; Fishbain, B. The effect of ego-motion on environmental monitoring. *Sci. Total Environ.* **2015**, *533*, 8−16.

(27) ILK Labs - Publications. https://www.ilklabs.com/what-we-do/publications (accessed May 03, 2022).

(28) Sonntag, D. B.; Gao, H. O. Developing link-based particle number emission models for diesel transit buses using engine and vehicle parameters. *Transp Res D Transp Environ* **2009**, *14* (4), 240−248.

(29) Horn, B. K.P.; Schunck, B. G. Determining optical flow. *Artif Intell* **1981**, *17*, 185−203.

(30) Barron, J. L.; Fleet, D. J.; Beauchemin, S. S. Performance of optical flow techniques. *Int J Comput Vis* **1994**, *12* (1), 43−77.

(31) Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [Online]. Available: https://github.com/tensorflow/models/tree/master/.

(32) Semantic Segmentation Using Deep Learning - MATLAB & Simulink. https://www.mathworks.com/help/vision/ug/semantic-segmentation-using-deep-learning.html (accessed May 04, 2022).

(33) Brostow, G. J.; Fauqueur, J.; Cipolla, R. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Lett.* **2009**, *30*, 88.

(34) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. Accessed: May 07, 2022 [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

(35) Kingma, D. P.; Ba, J. ADAM: A method for stochastic optimization. In *ICLR 2015*, 2015. DOI: 10.48550/arXiv.1412.698.

(36) Bland, J. M.; Altman, D. G. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet* **1986**, *327*, 307.

(37) Diez, S.; Lacy, S. E.; Bannan, T. J.; Flynn, M.; Gardiner, T.; Harrison, D.; Marsden, N.; Martin, N. A.; Read, K.; Edwards, P. M. Air pollution measurement errors: is your data fit for purpose? *Atmos Meas Tech* **2022**, *15* (13), 4091−4105.