

Research Article

Vision-Based Deep Q-Learning Network Models to Predict Particulate Matter Concentration Levels Using Temporal Digital Image Data

SungHwan Kim ,¹ SeHee Jung ,² SeongMin Yang ,¹ JiSeong Han,¹ ByungYong Lee,³ JaeHwa Lee ,^{4,5} and Sung Won Han ,⁶

¹Department of Applied Statistics, Konkuk University, Seoul, Republic of Korea

²AI Analytics Team, Deep Visions, Seoul, Republic of Korea

³Marketing Team, INSUNGIT, Seoul, Republic of Korea

⁴Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

⁵NASA Goddard Space Flight Center, Greenbelt, MD, USA

⁶School of Industrial Management Engineering, Korea University, Seoul, Republic of Korea

Correspondence should be addressed to Sung Won Han; swhan@korea.ac.kr

Received 2 April 2019; Revised 21 May 2019; Accepted 27 May 2019; Published 26 June 2019

Guest Editor: Sidike Paheding

Copyright © 2019 SungHwan Kim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Particulate matter (PM) has been revealed to have detrimental effects on public health, social economy, agriculture, and so forth. Thus, it became one of the major concerns in terms of a factor that can reduce “quality of life” over East Asia, where the concentration is significantly high. In this regard, it is imperative to develop affordable and efficient prediction models to monitor real-time changes in PM concentration levels using digital images, which are readily available for many individuals (e.g., via mobile phone). Previous studies (i.e., DeepHaze) were limited in scope to priorly collected data and thereby less practical in providing real-time information (i.e., undermined interpretation). This drawback led us to hardly capture drastic changes caused by weather or regions of interests. To address this challenge, we propose a new method called Deep Q-haze, whose inference scheme is built on an online learning-based method in collaboration with reinforcement learning and deep learning (i.e., Deep Q-learning), making it possible to improve testing accuracy and model flexibility in virtue of real-time basis inference. Taking into account various experiment scenarios, the proposed method learns a binary decision rule on the basis of video sequences to predict, in real time, whether the level of PM10 (particles smaller than 10 in aerodynamic diameter) concentration is harmful ($>80\mu\text{g}/\text{m}^3$) or not. The proposed model shows superior accuracy compared to existing algorithms. Deep Q-haze effectively accounts for unexpected environmental changes in essence (e.g., weather) and facilitates monitoring of real-time PM10 concentration levels, showing implications for better understanding of characteristics of airborne particles.

1. Introduction

Particulate is a minute particle that is in liquid or solid phase in the atmosphere and often refers to a particulate material having an aerodynamic diameter of $10\mu\text{g}/\text{m}^3$ or less (PM10). This originates from anthropogenic sources, such as combustion of fossil fuels such as coal, oil, the exhaust gas of manufacturing factories, and automobile engines as well as natural sources, such as desert and ocean (mineral dust and sea salt). Particulates are also known to affect climate

and precipitation as well as human health [1, 2]. Moreover, confronting threats of PM to Asian countries becomes no longer negligible to the point that the media and research groups consistently reveal detrimental effects [3]. To our surprise, it is notable that the World Cancer Institute in October 2013 analyzed a large-scale cohort of 2,095 lung cancer patients out of 312,944 people in the nine European countries [4]. Evidences that PM was determined as primary carcinogens were due to the fact that the risk of lung cancer increased by 22% at an PM increment of $10\mu\text{g}/\text{m}^3$.

Of late days, air pollution remains intractable to be resolved in massively populated regions like Seoul in South Korea, where the presence of fine dust is easily detected in vision. It is reported that PM10 concentration in South Korea is measured twice as high as that of OECD countries on average [5]. This record is even higher than the major cities such as New York City and Paris. To circumvent air pollution, the government has put significant efforts for better forecasting and developing benchmarks. And yet, we still encounter many challenges, for instance, inaccurate reporting system particularly at a specific location because of the limited metering sites, costly gadgets, and so forth. No wonder the most complete way to resolve fine dust is to eliminate the sources. However, this strategy obviously takes demanding costs and time-consuming tasks. Under this circumstance, concerns to public health have increased at an unprecedented rate. Civilians believe that hourly reporting of PM levels might not be sufficient for real-time air quality [6]. Thus, it is necessary to suggest a method applicable to prompt measurements of PM concentrations without expensive devices and spacious place to install. This is the point where our research motivation comes in.

The predictive models of PM concentration are proposed in various ways. A majority of methods adopted an explorative way: (1) elementary statistic [7], (2) time-series visualization [8], (3) histogram on a yearly basis [9], and (4) image data [10]. Another choice is to use predictive models such as logistic regression, support vector machine (SVM), and deep neural network (DNN) [11]. To construct training data set, a majority of previous methods typically utilized regional, climatic, or daily publicly available weather data (e.g., humidity, insolation, etc.), whereas the image data-based method makes an exclusive use of RGB data (Red, Green, and Blue) calibrated on true PM levels.

To the best of our knowledge, the attention to artificial intelligence revives through the diverse fields due to the rethinking of reinforcement learning. AlphaGo broke down at the 9th stage against Lee Sedol. The level of artificial intelligence is much better than expected. AlphaGo is based on Google's deep Q algorithm [12]. It is an artificial intelligence algorithm system exploiting reinforcement learning. Originally reinforcement learning is inspired by behavioral psychology, in which an agent defined in an environment recognizes the current state and selects a behavior or sequence of actions that maximizes compensation among the selectable behaviors. These problems are so comprehensive that they are also studied in areas of game theory, control theory, operational science, information theory, simulation-based optimization, multiagent systems, flock intelligence, statistics, and genetic algorithms [13, 14].

The deep Q-network algorithm (a.k.a DQN) learns the optimal policy by learning the Q function predicting the expected value of the utility that would result from performing a given action in a given state. After learning the Q function, we can derive the optimal policy by performing the action that gives the highest Q in each state. The goal of the agent (decision maker) is to maximize the sum of the rewards. The choice is the action of getting the greatest reward in that state in the long run. The DQN predicts the Q-value

using the action-value function CNN (convolutional neural networks), one of the neural network-type decision rules. It is well known that the convolutional neural network (CNN) is an efficient image processing algorithm adapted for vision analysis and image recognition.

In this paper, we proposed a predictive model that builds on the deep Q-network algorithm in spirit of reinforcement learning in order to predict particulate levels. We call this algorithm Deep Q-haze. Inspired by conventional reinforcement learning, this predictive model assigns the state an image to evoke multiclass actions on the basis of the prespecified calibration of particulates (e.g., $80\mu\text{g}/\text{m}^3$ less or more). Subsequent to this, the reward and action to get the best reward are determined. Taken together, the proposed Deep Q-haze serves as an effective tool to predict particulate levels solely subject to image data. We hypothesize that superior predictive performance of Deep Q-haze leads to less chance of false detection compared to previous classification model (e.g., SVM, RF, and DeepHaze) and consequently improves practical utility.

2. Datasets and Related Work

2.1. Datasets. Below we describe particulate data that a predictive model learns on. For the most part, we collect the video sequence data in the major cities of South Korea (e.g., Seoul and Daegu), where the cities are featured with a large-scale industrial complex, automobiles, highly populated counties. In such mega cities, gas emission has been a years-long environmental challenge, and high-concentration dusts occupy the peninsula throughout the year. More importantly, it is asserted that air pollution is primarily attributed to contaminants of eastern and southern China [15] and thus this problem, at present, remains out of control. When it comes to data collection, we gauge particulate levels via a high-performance device (Aerosol Mass Monitor (AEROCET-831) manufactured by Met One Instruments; <http://metone.com/>), whose perceptible dust size ranges from PM2.5 to PM10. In this paper, we purposely focus on the level of PM10. Regarding nonfixed image sequences (i.e., manually taken via mobile phone), we retrieved image data in our recent research, where we take into account residential areas, a group of trees, and building complexes featured with only nonatmospheric information (i.e., absence of sky). The interested regions largely include diverse categories: (1) outdoor parking spots, (2) building complex on campus, (3) indoor office environment, (4) street regions by exhausts emission, (5) vicinity of construction sites, and (6) residential areas. On average, video sequences are recorded with 5~25 frames per second for a total of 268 sequences. To take a glance, thumbnails of each video sequence are presented in Table 3. The video sequences are taken by Samsung phone cameras (S7) and its built-in IP webcam. Data and programming codes are available online (<https://sites.google.com/site/sunghwanshome/>).

2.2. Deep Q-Network Algorithm. Briefly, the deep reinforcement learning (Deep RL) system combines reinforcement

TABLE 1: The proposed architecture of simplified CNN models.

type	patch size/stride	input size
conv	4×4/1	200×200×9
conv	2×2/1	200×200×10
flattening	-	1×1×400000
linear	-	1×1×100
softmax	classifier	1×1×2

learning and neural networks. As aforementioned, reinforcement learning relates to an area of machine learning, in which an agent defined in an environment recognizes the current state and selects a behavior or sequence of actions that maximizes the expectation of the sum of the rewards among the selectable behaviors as below:

$$R_t = r_t + \gamma (r_{t-1}) + \gamma^2 (r_{t-2}) + \cdots + \gamma^T (r_{t-T}) \quad (1)$$

The objective of the agent is to find a strategy (a.k.a. policy) so as to maximize the expected sum of discounted rewards. In theory, the optimal policy is defined as the expectation of rewards that potentially earn in the future when continuing the actions along the policy π at the current state $s_t = s$.

$$\pi^*(s) = \arg \max_a E(R_t | s_t = s, a_t = a) \quad (2)$$

The action a (i.e., $\pi^*(s)$) is selected such that the expectation of the sum of the rewards is maximized. Instead of above, we learn $Q(s, a)$ and thereby find the optimal action $a_t (= a)$ in state $s_t (= s)$.

$$\begin{aligned} Q^{\text{new}}(s_t, a_t) &\leftarrow \\ Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right] \end{aligned} \quad (3)$$

where $\gamma \in [0, 1]$ is the discount factor. This is the Q-learning method proposed by Watkins [15]. Stepping up beyond Q-learning, [12] proposed the Deep Q-network algorithm (a.k.a. DQN) that learns the optimal policy by the Q function on the basis of the deep convolution neural network (CNN) and approximates the action-state function. In this paper, we use the customized CNN (see Table 1) to detect the characteristics of the image and to determine the behavior of the agent.

$$\min_{\theta} \sum_{t=0}^T \left[\widehat{Q}(s_t, a_t | \theta) - \left(r_t + \gamma \cdot \max_a \widehat{Q}(s_{t+1}, a | \theta) \right) \right]^2, \quad (4)$$

where θ is the set of model parameters and \widehat{Q} is an estimated Q-function.

The learning process optimizes the cost function updating the weight to minimize the above equation. Importantly, two techniques designed to enhance predictive power get involved in the learning process. The first stage is called the capture and replay method. To put this plainly, this performs repetitive tasks between storing and taking data at random.

Due to the fact that sequential samples are likely to be strongly correlated, randomness of replay memory attenuates correlation and reduces the variance of updates. In the second stage, the networks learn on a target network and main network one after the other (i.e., constructing two networks). Meanwhile, the target network is fixed and only the main network is updated. The target network updates the values of the main network once every predetermined step. This trick tackles the problem of moving targets and continuously updates the Q-function to maximize the expectation of rewards in the future. All things taken together, the optimal behavior is determined by the updated main Q-function.

3. Methods

3.1. Augmented Temporal Image Features. In context of big data analytics, it is interesting to boost power of our predictive model. To this end, the proposed model combines multiple feature channels, each containing RGB, HSV, and its haze-related features (i.e., dark channel, color attenuation, and hue disparity; [16–18], (Fattal et al., 2008, and Koschmieder et al., 1925)), for a total of 9 channels. Needless to say, it is generally true that the larger data set we apply, the more potential signals the model may decipher. To take a glance, Figure 1 illustrates how we form augmented image data, which serve as a building block to measure the amount of dusts. Saturation index in HSV ranging from 0 to 255 represents the degree of saturation, which are closely linked to noises attributed to particulates. Combining all channels above, the state s_t in the Q-function at time t takes a $200 \times 200 \times 9$ multidimensional array. To account for particulate levels, we create difference values of two consecutive arrays followed by standardization and filtering outliers exceeding 90th quantile. These arrays of difference in image sequences play a role as building blocks of our predictive model (see Figure 2).

3.2. Resampling-Based Reinforcement Learning. Here, we propose the resampling-based reinforcement learning algorithm. Typically, environmental data are prone to being sequential, time-dependent, and seasonal. These characters naturally invite reinforcement learning-type models to come into play. In one sense, an atmospheric model is suited to reinforcement learning as consecutive variability relates to atmosphere. To the contrary, it is also found that reinforcement learning is hardly exploited to natural environment data, in the sense that repetitive tasks to mimic natural environment are challenging to be implemented, as compared to training for robot arms or video games to which reinforcement learning widely applies. However unlikely it may seem, we can create an artificial environment with regard to particulates such that we arbitrarily maneuver weather conditions in purpose (e.g., dust quantity) via bootstrap sampling. In doing so, we initially build an integrated data pool consisting of real image sequences in a proportion to balanced class labels (e.g., safe and harmful) to stably perform bootstrap sampling (e.g., with replacement). Importantly, such a sampling process allows consecutive learning tasks to construct a vast number of predictive models whose training

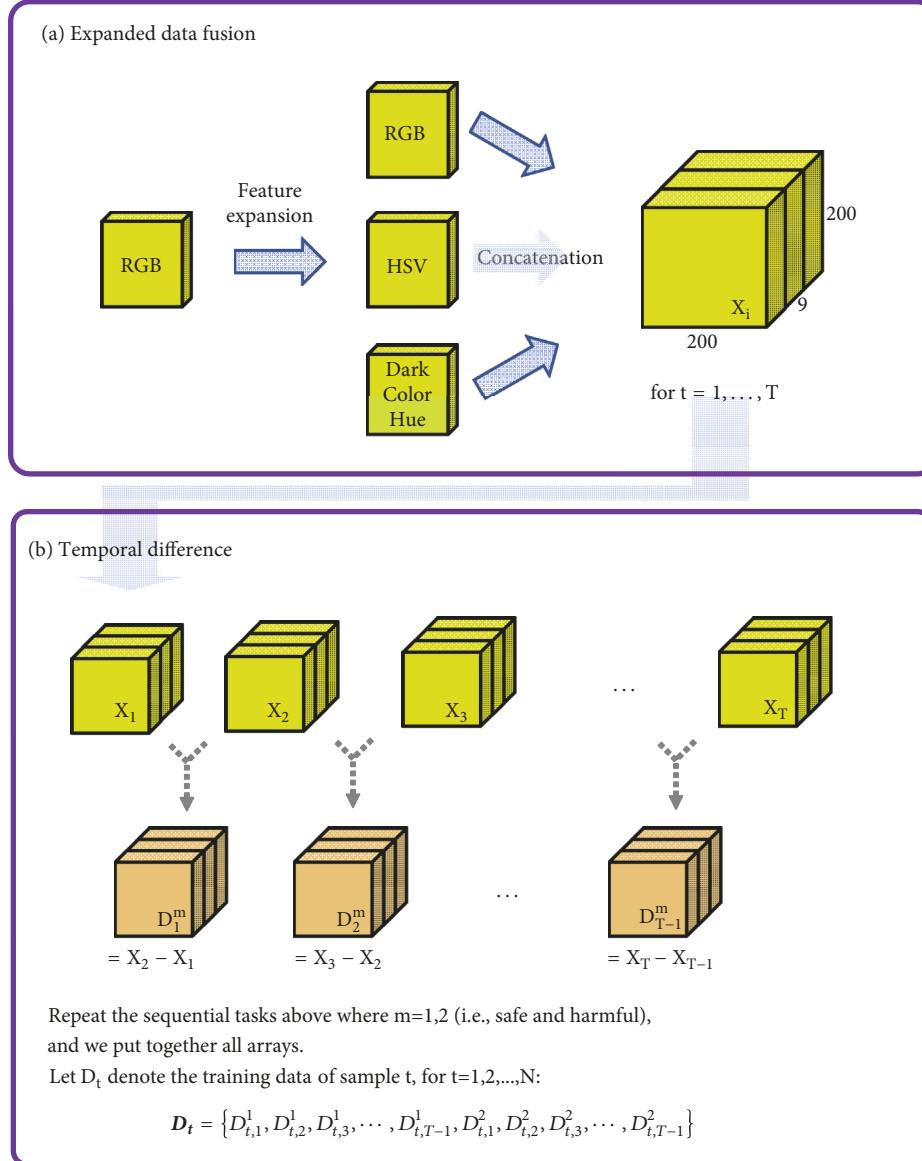


FIGURE 1: A workflow to implement Deep Q-haze and augmented image-related arrays constituting temporal differences between two consecutive arrays.

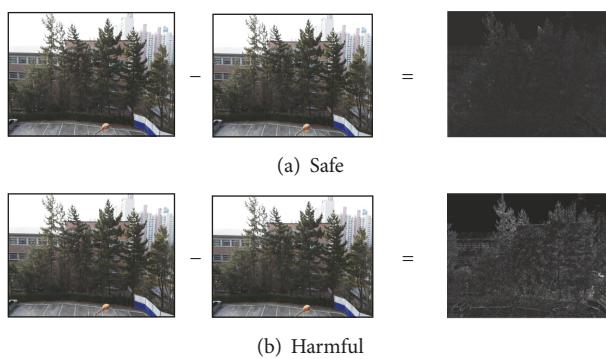


FIGURE 2: Shown are temporal differences between two consecutive arrays with regard to two scenarios (e.g., safe and harmful), respectively. The visual distinction of two difference arrays looks obvious in the sense that the harmful case generates explicit discordance as compared to the safe case.

TABLE 2: Deep Q-haze algorithm.

-
- (1) Initialize model configuration
- Initialize action-value function Q with random weights
 - Construct N sequence arrays (i.e., s_t at time t) of nine channels and randomly sample bootstrap batch out of the integrated data pool (i.e., $N = 20$ as default)
 - Initialize sequence s_t and preprocessed sequenced (i.e., standardization and filtering outliers exceeding 90th quantile) via ϕ , namely $\phi(s_t)$.
-

- (2) Create difference values of two consecutive arrays

$$\phi_t = \phi(s_t) - \phi(s_{t+1})$$

- (3) Repeat the following for $t = 1, \dots, T$:

- To derive the optimized action, select a random action a_t , where $a_t \in \{0, 1\}$ (i.e., safe or harmful) with probability ϵ
- Otherwise select $a_t = \max_a Q(\phi_t)$
- Execute action a_t in the predictive rule and observe reward r_t and new incoming sequence s_{t+1}
- Set a_t, s_{t+1}, s_{t+2} and process $\phi_{t+1} = \phi(s_{t+1}) - \phi(s_{t+2})$ and calculate rewards determining actions and impose the weight according to testing outcome (i.e., true or false) and update Q every 10 times.
- For $j = 1, \dots, N - 1$, set as follows

$$y_j = \begin{cases} r_j & \text{if terminal } \phi_{j+1} \\ r_j + \gamma \cdot \max_a Q(\phi_{j+1}, a; \theta) & \text{otherwise} \end{cases}$$

where

$$r_j = \begin{cases} v, & \text{if } z_j = \arg \max_a Q(\phi_j, a_j) \\ -v, & \text{otherwise} \end{cases}$$

and $z_j \in \{0, 1\}$, a true class label monitored via a device and set $v = 1$ in this paper.

- (vi) Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$.
-

data determine rewards, policy, and actions. Particulate levels and image data are monitored over the years, and models all in one, aiming at exclusion of possible seasonal and climate effects. In what follows, Table 2 encapsulates the major implementation schemes one at a step, in short, each including the kernels of deep Q-network [12] and vision-based DeepHaze [11] learning on differences vision of neighboring sequences. In our simulation, we, for simplicity, make in rewards to be small, equivalently adjusting future rewards to be quite negligible. With regard to the Q-function, we adopt the CNN architectures of the predictive model as presented in Table 1, and the CNN architectures are implemented by TensorFlow 1.10 in Python.

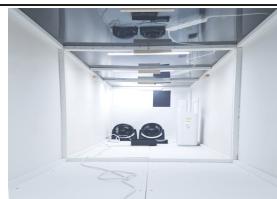
4. Results and Discussion

In this experiment, we evaluate the variants of the Deep Q-haze models learning on a range of frame numbers and compare them to other popularly used classifiers (e.g., DeepHaze, random forest, and SVM). With varying parameters, diverse experiment scenarios are considered to mimic real environments and to fortify universal applicability of the model. Tables 4 and 5 encapsulate the predictive performance of the Deep Q-haze and its competitor classifiers. It is evident to say that the proposed algorithm, when using all datasets,

outstandingly distinguishes a harmful atmospheric condition with high accuracy and low false detection (i.e., Youden index = sensitivity + specificity - 1; e.g., 0.9817 - 0.9894 for the indoor office).

4.1. Indoor Environment (an Office and an Experimental Chamber). It is certain that clean air quality in an indoor office is a critical part to maintain health. It is sensible, with that in mind, to purposely focus on image sequences in office at Konkuk University over the several months. We collect 2,200 video sequences (i.e., 1,100 clips of each class label), each containing at least 20 image frames per 1 min. Generally robustness of the algorithm is essential for practical utility. In what follows, we performed large-scale experiments under controlled conditions to verify if the Deep Q-haze is robust against various environmental factors. The experiments were carried out largely under four conditions: presence of windiness, high temperature, high humidity, and high light intensity. To this end, we construct the experiment chamber (i.e., large container) specially designed to create artificial circumstances (see Table 3 at the bottom). Beside factors of interest, other conditions remained the same at ordinary level. Table 6 shows that Deep Q-haze consistently maintains high predictive power regardless of environmental conditions (e.g., windiness, high temperature, etc.). It is found

TABLE 3: Thumbnails of target regions where particulate matters are analyzed to construct the predictive model.

Category	thumbnails	# of video sequences
Indoor office		2,200
Outdoor parking lot of Konkuk Univ.		3,200
Outdoor parking lot of Keimyung Univ.		3,000
Mobile video clips		1,000
Experimental chamber		500

that Deep Q-haze is less likely to be deteriorated, even though varying environmental factors can promote the randomness of particulates. We collect 500 sequences of only harmful labels, consisting of at least 20 images per 1 min.

4.2. Outdoor Regions. Unsurprisingly, outdoor regions have a tendency to higher levels of particulate than indoor and, due to open space, facilitate visually gauging dust particles present in the air through a long distance. Considering that campus regions are filled up with automobiles, where population flows are relatively intense, we chose two regions: (1) a parking lot (Keimyung University) and (2) building complex (Konkuk University), where we install high-resolution cameras and dust measurement device (AEROCET-831). For several months (2017~2018), we monitored outdoor parking lots all day long and recorded image sequences. We collect 3,000 (an outdoor parking lot of Keimyung University) and 3,200 (an outdoor parking lot of Konkuk University) video

sequences of both safe and harmful labels, consisting of at least 20 image frames per 1 min. We focus on image captured from fixed camera and mobile phone camera in a different way due to perturbation that occurs when a mobile phone is manually controlled. To take a glimpse, refer to the thumbnail image in Table 3.

4.2.1. Image Sequences of Fixed Camera. Tables 4 and 5 show that Deep Q-haze outperforms DeepHaze, SVM, and RFs. Note that Deep Q-haze performs with high accuracy (0.9839 ~ 0.9914 of Keimyung Univ., 0.9040 ~ 0.9220 of Konkuk Univ.; hereafter this order is kept the same) as opposed to DeepHaze (0.6300 ~ 0.6336, 0.4560 ~ 0.4580), random forest (0.4100 ~ 0.4581, 0.4380 ~ 0.4690), and SVM (0.3800 ~ 0.4236, 0.6240 ~ 0.6560). It is interesting to see that predictive power tends to be increasing as the frames augmented from 5 to 20. Besides, Deep Q-haze suffers less from the false detection (i.e., high Youden index; Deep Q-haze: 0.9658

TABLE 4: Sensitivity (Sen), specificity (Spe) and Youden index (= sensitivity + specificity - 1) of predictive models with an application to various video sequences.

# of frames	5 frames				10 frames				15 frames				20 frames			
	Sen	Spe	Youden	Sen	Sen	Spe	Youden	Sen	Sen	Spe	Youden	Sen	Spe	Youden	Sen	Spe
Deep Q Haze	0.9873	0.9963	0.9836	0.9927	0.9963	0.9890	0.9927	0.9967	0.9894	0.9927	0.9927	0.9890	0.9927	0.9817	0.9817	0.9817
Deep Haze	0.9575	0.4650	0.4225	0.9850	0.4300	0.4150	0.9750	0.5150	0.4900	0.9800	0.5175	0.4900	0.5175	0.4975	0.4975	0.4975
RF	0.9850	0.2675	0.2525	0.9825	0.2774	0.2599	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SVM	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
<hr/>																
Deep Q Haze	0.8550	0.9366	0.7916	0.8500	0.9500	0.8000	0.8500	0.9600	0.8100	0.8600	0.8600	0.8100	0.8633	0.8233	0.8233	0.8233
Deep Haze	0.3760	0.5360	-0.0880	0.4120	0.4880	-0.1000	0.3640	0.5400	-0.0960	0.3740	0.5420	0.3740	0.5420	-0.0840	-0.0840	-0.0840
RF	0.5240	0.4140	-0.0620	0.5300	0.4820	0.0120	0.4080	0.4679	-0.1241	0.4540	0.4679	0.4540	0.4679	-0.0781	-0.0781	-0.0781
SVM	0.7320	0.5800	0.3120	0.7180	0.5360	0.2540	0.7020	0.5460	0.2480	0.7380	0.5500	0.7380	0.5500	0.2880	0.2880	0.2880
<hr/>																
Deep Q Haze	0.9871	0.9814	0.9685	0.9885	0.9842	0.9727	0.9885	0.9871	0.9756	0.9914	0.9914	0.9828	0.9828	0.9828	0.9828	0.9828
Deep Haze	0.8760	0.4250	0.3010	0.8980	0.3950	0.2930	0.8820	0.4216	0.3036	0.8900	0.4200	0.3036	0.8900	0.3100	0.3100	0.3100
RF	0.8200	0.0683	-0.1117	0.8580	0.0900	-0.0520	0.8480	0.1030	-0.0490	0.8740	0.1116	0.1030	0.8740	-0.0144	-0.0144	-0.0144
SVM	0.8375	0.1185	-0.0440	0.8375	0.1585	-0.0040	0.8375	0.2000	0.0375	0.8375	0.1871	0.2000	0.8375	0.0246	0.0246	0.0246
<hr/>																
Deep Q Haze	0.9733	0.7874	0.7607	0.9866	0.7632	0.7498	0.9822	0.7487	0.7309	0.9777	0.7439	0.7309	0.9777	0.7216	0.7216	0.7216
Deep Haze	0.9130	0.1288	0.0418	0.9178	0.1288	0.0466	0.9130	0.1244	0.0374	0.9082	0.1244	0.9082	0.1244	0.0326	0.0326	0.0326
RF	0.7004	0.2711	-0.0285	0.7681	0.2622	0.0303	0.6908	0.2311	-0.0781	0.7198	0.2577	0.2311	0.7198	-0.0225	-0.0225	-0.0225
SVM	0.6280	0.0533	-0.3187	0.5990	0.0666	-0.3344	0.5990	0.0533	-0.3477	0.5893	0.0622	0.5893	0.0622	-0.3485	-0.3485	-0.3485

TABLE 5: Prediction accuracy of predictive models with an application to various video sequences.

# of frames	5 frames	10 frames	15 frames	20 frames
Konkuk Univ. Indoor				
Deep Q-Haze	0.9918	0.9945	0.9927	0.9817
Deep Haze	0.7112	0.7487	0.7450	0.7487
RF	0.6262	0.6300	0.5000	0.5000
SVM	0.5000	0.5000	0.5000	0.5000
Konkuk Univ. Outdoor				
Deep Q-Haze	0.9040	0.9100	0.9160	0.9220
Deep Haze	0.4560	0.4580	0.4520	0.4580
RF	0.4690	0.5060	0.4380	0.4610
SVM	0.6560	0.6270	0.6240	0.6440
Keimyung Univ.				
Deep Q-Haze	0.9839	0.9861	0.9877	0.9914
Deep Haze	0.6300	0.6336	0.6309	0.6336
RF	0.4100	0.4390	0.4418	0.4581
SVM	0.3800	0.4054	0.4318	0.4236
Mobile Phone				
Deep Q-Haze	0.8842	0.8796	0.8703	0.8657
Deep Haze	0.5046	0.5000	0.5023	0.5000
RF	0.4768	0.5046	0.4513	0.4791
SVM	0.3287	0.3217	0.3148	0.3148

TABLE 6: Simulation study: sensitivity (Sen), specificity (Spe), Youden index, and standard errors in parentheses with an application to various video sequences and environmental conditions. The simulation was repeated 50 times.

# of frames	5 frames	10 frames	15 frames	20 frames
Windiness (Use of fan)				
Sen	0.8411 (0.0408)	0.8495 (0.0452)	0.9152 (0.0335)	0.9116 (0.0304)
Spe	0.8796 (0.0299)	0.9141 (0.0266)	0.9411 (0.0221)	0.9419 (0.0200)
Youden	0.7207 (0.0377)	0.7636 (0.0526)	0.8563 (0.0738)	0.8535 (0.0496)
High Temperature (40°C)				
Sen	0.8664 (0.0291)	0.8866 (0.0315)	0.8844 (0.0357)	0.9090 (0.0291)
Spe	0.9160 (0.0236)	0.9524 (0.0201)	0.9601 (0.0157)	0.9587 (0.0170)
Youden	0.7824 (0.0336)	0.8390 (0.0398)	0.8445 (0.0427)	0.8677 (0.0371)
High Humidity (50%)				
Sen	0.8758 (0.0274)	0.9317 (0.0221)	0.9478 (0.0193)	0.9392 (0.0234)
Spe	0.9562 (0.0185)	0.9647 (0.0199)	0.9961 (0.0029)	0.9821 (0.0146)
Youden	0.8320 (0.0312)	0.8764 (0.0298)	0.9439 (0.0209)	0.9213 (0.0308)
High luminous Intensity (250lx)				
Sen	0.8990 (0.0276)	0.9290 (0.0247)	0.9259 (0.0251)	0.9211 (0.0278)
Spe	0.8669 (0.0308)	0.8947 (0.0307)	0.9071 (0.0293)	0.9265 (0.0278)
Youden	0.7659 (0.0342)	0.8237 (0.0396)	0.8330 (0.0404)	0.8476 (0.0430)

~ 0.9828 , $0.7916 \sim 0.8233$). Putting another way, the low Youden index values imply that random forest and SVM are not as efficient as Deep Q-haze with regard to image-based prediction.

4.2.2. Image Sequences of Mobile Phone Camera. We hypothesize whether our predictive model effectively applies to image sequences manually taken. Admittedly, chances are that our proposed method does not work due to unexpected

minute vibration; it is sensible to assess its performance in this scenario. Coherent to experiments above, Tables 4 and 5 show that the proposed models are superior in accuracy to DeepHaze, SVM, and RFs (i.e., Deep Q-haze: $0.8657 \sim 0.8842$, random forest: $0.6500 \sim 0.6429$, SVM: $0.3148 \sim 0.3287$) and in low false detection (i.e., Deep Q-haze: $0.9733 \sim 0.9866$, random forest: $-0.0781 \sim 0.0303$, and SVM: $-0.3485 \sim -0.3187$). Additionally, it is notable to see that indoor experiment designs generally show better results

compared to outdoor ones. This gap mainly results from the difference in experimental setups. Since extra variables (e.g., light and atmosphere) are adequately adjusted indoors, predictive power of indoor models tends to be superior to models of outdoor environments, where unexpected hardly controlled variables are present.

5. Conclusion

Recently, we dove into the season of burgeoning AI. Many are fascinated with its widespread applicability and practical benefits (e.g., self-driving car, robots, healthcare, etc.). Here we tried to take advantage of the flexible, highly efficient AI technique in air quality monitoring and bring spatial scale of the monitoring down to a “room scale”. Derivation of real-time PM concentrations (even in a semiquantitative way) at a “room scale” is essential, as it can provide information on quality of air that people actually inhale in their everyday life. There is no doubt that it would be even better if the task can be done relatively easily using data readily available to the public. We presented a novel deep learning approach to determine real-time PM₁₀ level whether it is harmful or not from digital images acquired by nonindustrial level recording devices, including mobile phones. Our previous method (DeepHaze, Kim et al. [11]) triggered developing a vision-based predictive model and is found to be applicable in a range of experimental scenarios. Compared to the existing decision rule, Deep Q-haze in the model stretches to additional colorific features (e.g., RGB, HSV, and particulate related features), implicating that the predictive power noticeably improved due to the blessing of big data. Yet, there are urgent needs to synchronize pixels across image sequences (e.g., homogenous configuration), as taking images from flying drones or manual controls is possibly subject to external factors. This homogenous nature serves essential roles to make it to the exquisite differences between consecutive frames. Besides, it is recommended to ensure universal applicability regardless of the type of regions, weather, and the amount of light. Avoiding false detection is an intractable hurdle due to the fact that particulates in image are captured with weak signals for the most part. To enhance utility to the maximum extent by the public, Deep Q-haze is planned to be implemented in portable electronic gadgets in the form of mobile application software. The model needs to have advance extension; the model should be advanced toward multiclass prediction on the basis of moderate calibrations, together with aerosol-related features (e.g., image contrast or visibility [19–21]). To this end, another recurrent neural network-type architecture can potentially be a choice to improve accuracy. We leave these topics for future study.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request. Refer to author's website (<https://sites.google.com/site/sunghwan-shome/>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Acknowledgments

This paper was supported by Konkuk University in 2018.

References

- [1] S. G. Narasimhan and S. K. Nayar, “Vision and the atmosphere,” *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.
- [2] D. Muir and D. P. H. Laxen, “Black smoke as a surrogate for PM₁₀ in health studies?” *Atmospheric Environment*, vol. 29, no. 8, pp. 959–962, 1995.
- [3] H. C. Kim, S. Kim, B. Kim et al., “Recent increase of surface particulate matter concentrations in the Seoul Metropolitan Area, Korea,” *Scientific Reports*, vol. 7, no. 1, p. 4710, 2017.
- [4] O. Raaschou-Nielsen, Z. J. Andersen, R. Beelen, E. Samoli, and M. Stafoggia, “Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE),” *The Lancet Oncology*, vol. 14, no. 9, pp. 813–822, 2013.
- [5] *OECD Economic Surveys: KOREA*, OECD, 2018.
- [6] F. J. Kelly and J. C. Fussell, “Air pollution and public health: emerging hazards and improved understanding of risk,” *Environmental Geochemistry and Health*, vol. 37, no. 4, pp. 631–649, 2015.
- [7] W. Yuchi, E. Gombojav, B. Boldbaatar et al., “Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city,” *Environmental Pollution*, vol. 245, pp. 746–753, 2019.
- [8] M. L. Bell, J. M. Samet, and F. Dominici, “Time-series studies of particulate matter,” *Annual Review of Public Health*, vol. 2, no. 5, pp. 247–280, 2004.
- [9] C. Liu, F. Tsow, Y. Zou, N. Tao, and H. Liu, “Particle pollution estimation based on image analysis,” *PLoS ONE*, vol. 11, no. 2, p. e0145955, 2016.
- [10] Z. He, X. Ye, K. Gu, and J. Qiao, “Learn to predict PM_{2.5} concentration with image contrast-sensitive features,” in *Proceedings of the 2018 37th Chinese Control Conference (CCC)*, pp. 4103–4106, IEEE, 2018.
- [11] S. H. Kim and S. Kim, “Vision-based predictive model on particulates via deep learning,” *Journal of Electrical Engineering and Technology*, vol. 13, no. 5, pp. 2107–2115, 2018.
- [12] D. Silver, J. Schrittwieser, K. Simonyan et al., “Mastering the game of Go without human knowledge,” *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.
- [13] L. Buşoniu, R. Babuška, and B. De Schutter, “Multi-agent Reinforcement learning: an overview,” in *Innovations in Multi-Agent Systems and Applications-I*, vol. 3, pp. 183–221, Springer, 2010.
- [14] L. Deng and D. Yu, “Deep learning: methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [15] R. A. Rohde and R. A. Muller, “Air pollution in china: mapping of concentrations and sources,” *PLoS ONE*, vol. 10, no. 8, p. e0135749, 2015.

- [16] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
- [17] P. Carr and R. Hartley, "Improved single image dehazing using geometry," in *Proceedings of the Digital Image Computing: Techniques and Applications*, vol. 1, pp. 3–10, IEEE, 2009.
- [18] N. Jacobs, N. Roman, and R. Pless, "Consistent temporal variations in many outdoor scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–6, 2007.
- [19] W. C. Malm, K. K. Leiker, and J. V. Molnar, "Human perception of visual air quality," *Journal of the Air Pollution Control Association*, vol. 30, no. 2, pp. 122–131, 1980.
- [20] W. Huang, J. Tan, H. Kan et al., "Visibility, air quality and daily mortality in Shanghai, China," *Science of the Total Environment*, vol. 407, no. 10, pp. 3295–3300, 2009.
- [21] C. A. Olman, K. Ugurbil, P. Schrater, and D. Kersten, "BOLD fMRI and psychophysical measurements of contrast response to broadband images," *Vision Research*, vol. 44, no. 7, pp. 669–683, 2004.

