



Research article

Real-time image-based air quality estimation by deep learning neural networks

Pu-Yun Kow^a, I-Wen Hsia^a, Li-Chiu Chang^b, Fi-John Chang^{a,*}^a Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, 10617, Taiwan^b Department of Water Resources and Environmental Engineering, Tamkang University, New Taipei City, 25137, Taiwan

ARTICLE INFO

Keywords:

PM_{2.5}
 PM₁₀
 Air quality index (AQI)
 Image (photo) classification
 Deep learning
 Convolutional neural network (CNN)

ABSTRACT

Air quality profoundly impacts public health and environmental equity. Efficient and inexpensive air quality monitoring instruments could be greatly beneficial for human health and air pollution control. This study proposes an image-based deep learning model (CNN–RC) that integrates a convolutional neural network (CNN) and a regression classifier (RC) to estimate air quality at areas of interest through feature extraction from photos and feature classification into air quality levels. The models were trained and tested on datasets with different combinations of the current image, the baseline image, and HSV (hue, saturation, value) statistics for increasing model reliability and estimation accuracy. A total of 3549 hourly air quality datasets (including photos, PM_{2.5}, PM₁₀, and the air quality index (AQI)) collected at the Linyuan air quality monitoring station of Kaohsiung City in Taiwan constituted the case study. The main breakthrough of this study is to timely produce an accurate image-based estimation of several pollutants simultaneously by using only one single deep learning model. The test results show that estimation accuracy in terms of R² for PM_{2.5}, PM₁₀, and AQI based on daytime (nighttime) images reaches 76% (83%), 84% (84%), and 76% (74%), respectively, which demonstrates the great capability of our method. The proposed model offers a promising solution for rapid and reliable multi-pollutant estimation and classification based solely on captured images. This readily scalable measurement approach could address major gaps between air quality data acquired from expensive instruments worldwide.

1. Introduction

Air quality is closely related to people's daily life, and good air quality will lead to a healthy life. Unfortunately, air quality has gotten worse in recent decades, which has attracted much public attention. PM_{2.5} is considered a dangerous threat to human health because it can penetrate into the lungs through breathing without being trapped by the cilia in the respiratory tract (Li et al., 2020; Sugiyama et al., 2020; Tseng et al., 2019; Wang et al., 2018 & 2021; Zhang et al., 2019; Zhou et al., 2018). Another important air quality indicator is the Air Quality Index (AQI) defined by the Environmental Protection Administration (TW EPA). AQI composed of six pollutants (PM_{2.5}, PM₁₀, O₃, NO₂, SO₂ and CO) is a comprehensive index to reflect air pollution levels more objectively than an index merely involving a single air pollutant (Kumar and Goyal, 2011; Ruggieri and Plaia, 2012). With rapid industrial development and urbanization in recent decades, Taiwan inevitably encounters severe air pollution problems caused by industrial emissions and vehicular exhausts, especially in Kaohsiung City of southern Taiwan

(Kow et al., 2020; Tsai et al., 2003).

The severe threat of air pollution to human health makes air quality a focus of public attention, and timely air quality monitoring is critical to pollution control and greatly beneficial for human health protection. At present, air quality data collection relies mainly on monitoring stations. However, such in-situ monitoring is less feasible to carry out at the majority of areas of concerns due to the high material and set-up cost of sophisticated sensors, which becomes a big burden for poor or large countries (Rijal et al., 2018). Cost-effective image-based methods can serve as an auxiliary to monitor air quality at an ungauged area of interest or when air quality monitoring devices at gauge stations have malfunctioned. There have been various research efforts to develop inexpensive instruments for air pollution monitoring recently (e.g., Babari et al., 2011; Chakma et al., 2017; Zhang et al., 2016; Zhao et al., 2019).

Recently, in-depth machine learning (ML) has made significant breakthroughs in many aspects, such as voice and image, with increasingly prominent advantages. The convolutional neural network (CNN) is

* Corresponding author.

E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

<https://doi.org/10.1016/j.jenvman.2022.114560>

Received 28 January 2021; Received in revised form 5 January 2022; Accepted 16 January 2022

Available online 24 January 2022

0301-4797/© 2022 Elsevier Ltd. All rights reserved.

regarded as one of the most commonly used ML methods extensively applied to image processing and computer vision research recently, with plausible performance in tackling several challenging tasks on estimation/classification (e.g., Bo et al., 2018; Giyenko et al., 2018; Kopp et al., 2019; Soh et al., 2018; Vahdatpour et al., 2018; Wang et al., 2018; Yuan

et al., 2020; Zhang et al., 2020; Zhong et al., 2019). In recent years, using ML techniques to analyze air quality has also received increasing attention (e.g., Chang et al., 2020; Zhou et al., 2019 and 2020), and several studies based on image processing have been carried out for the classification or estimation of PM_{2.5} concentrations (Li et al., 2015; Liu

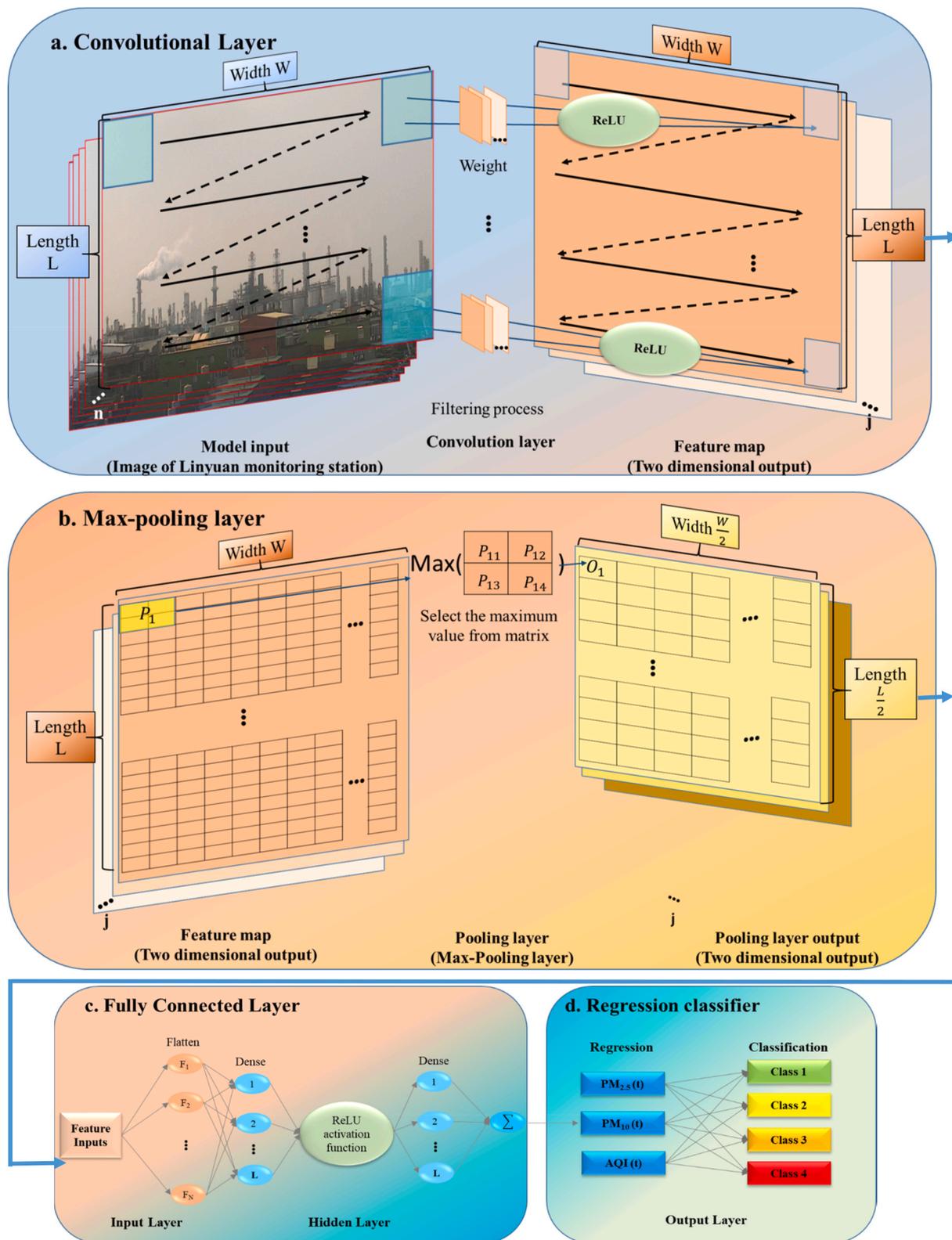


Fig. 1. Architecture of the CNN-FC deep learning approach (fusing the convolutional neural network with a regression classifier). (a) Convolutional layer. (b) Max-pooling layer. (c) Fully-connected layer. (d) Regression classifier.

et al., 2016; Chakma et al., 2017; Ma et al., 2018). Image-based air pollution estimation paves a promising direction, while only limited works have been carried out up to now. Therefore, it is worth exploring in-depth research on image-based air quality estimation for improving estimation reliability and accuracy.

This study proposes a hybrid deep learning model (CNN–RC) that seamlessly integrates CNN and a regression classifier (RC) to simultaneously estimate PM_{2.5} and PM₁₀ concentrations and the AQI as well as classify these estimates according to pollution levels based on air quality photos (Fig. 1). Model inputs contain images and HSV (Hue, Saturation, Value) statistics for improving model reliability and estimation accuracy. The HSV value denotes the correlation between HSV histograms of the current image and the baseline image. Therefore, the proposed approach is of a multi-input multi-output (MIMO) structure. Two CNN learning schemes utilized are VGG (Visual Geometry Group) and ResNet (residual network). The hybrid CNN–RC models built upon VGG and ResNet learning schemes with different input combinations are independently constructed. This study is organized to outline the study area and materials in Section 2, introduce the methods adopted in Section 3, show and discuss the image-based regression classification results in Section 4, and make concluding remarks in Section 5.

2. Study area and materials

2.1. Study area

The worst air quality usually occurs in southern Taiwan, especially in Kaohsiung City, due to the increase in industrial activities and the number of vehicles. The Linyuan air quality monitoring station in Kaohsiung City is located in an industrial area, where several petrochemical plants operate full day. The scene captured by the monitoring camera at this station for air pollution control purpose reflects pollution emissions from petrochemical operation, with main emission sources containing SO_x, NO_x, VOCs, PM_{2.5}, PM₁₀, CO₂ and others. The Linyuan station is selected as our case study for the following reasons: 1) it is located in an industrial area with pollutant emissions from petrochemical production; 2) factory chimneys can be clearly identified from image collected; and 3) smoke from factory chimneys is clearly visible, and the difference in image clarity is observed between the baseline image and the image taken when air pollution occurs.

2.2. Data collection and statistical analysis

This study collected a large number of datasets, where each dataset composed a photo (image) and its corresponding air quality indexes (PM_{2.5}, PM₁₀, and AQI). The TW EPA provides a web-based open data platform with on-demand local air quality monitoring datasets accessible to the public, which greatly facilitates the collection of reliable data for research use. This study first implemented a Web crawler to extract from this open data platform hourly air quality monitoring images with corresponding data collected at the Linyuan station, spanning between 8 p.m. on July 1st and 11 p.m. on November 30th in 2019. A total of 3549 hourly datasets (=24 h × 153 days - 123 h (missing images due to problems of transmission, frontend (camera)/backend (storage facility) device malfunction, or else) were used for model construction. Models were constructed based on daytime (6 a.m.–6 p.m.) and nighttime (7 p.m.–5 a.m.) images separately. 1946 datasets out of 3549 datasets were categorized as daytime datasets, where 1556 datasets (80%) and 390 datasets (20%) were used for training and testing, respectively. The remaining 1603 datasets were categorized as nighttime datasets, where 1282 datasets (80%) and 321 datasets (20%) were used for training and testing, respectively. The collected dataset is available on the website <http://hyinfo.bse.ntu.edu.tw/apci/>.

Table 1 presents the results of the statistical analyses on air quality data used in this study. The results indicate that the classification of air quality images for the Linyuan station should be very challenging due to

Table 1

Results of statistical analyses on hourly air quality data for the Linyuan air quality monitoring station (July 1, 2019–November 30, 2019).

	Indicator	PM _{2.5} (µg/m ³)	PM ₁₀ (µg/m ³)	AQI
Daytime	Mean	20.79	45.51	67.07
	Max	64.00	305.00	185.00
	Min	0.00	0.00	13.00
	Std ^a	11.70	24.54	32.71
Nighttime	Mean	20.61	47.79	71.44
	Max	59.00	305.00	192.00
	Min	1.00	7.00	16.00
	Std	11.25	26.43	35.90
Whole day	Mean	20.71	46.54	69.05
	Max	64.00	305.00	192.00
	Min	0.00	0.00	13.00
	Std	11.50	25.43	34.25

^a Standard deviation.

high variations in concentration. Furthermore, the statistical results of daytime and nighttime datasets imply that emission mechanisms should have a significant difference between daytime and nighttime. According to the mean and standard deviation values, PM_{2.5} had slightly larger values in the daytime than in the nighttime, whereas PM₁₀ and AQI, has higher values in the nighttime. Furthermore, the collections of colors in daytime and nighttime images are very different. Daytime images are more colorful than nighttime images (close to black-and-white). Therefore, we decided to establish hybrid deep learning models based separately on daytime and nighttime datasets.

3. Methodology

3.1. Problems and motivations

Image-based regression classification can serve as an auxiliary technique to monitor air quality at ungauged areas or when monitoring devices have malfunctions. Our goal was to estimate multiple pollutants and classify the estimates simultaneously based on the images collected at the Linyuan station. We proposed a hybrid deep learning model (CNN–RC) that integrated CNN and a regression classifier (RC) based on multiple inputs (photos) and their HSV statistics) for estimating/classifying multiple air pollutants (PM_{2.5}, PM₁₀, and AQI). Two CNN learning schemes (ResNet and VGG) with weight adjustment were utilized. Fig. 2 illustrates the architectures of the RN_H (ResNet-HSV) scheme (Fig. 2(a)) and the VN_H (VGG-HSV) scheme (Fig. 2(b)). The methods used in this study are briefly introduced as follows.

3.2. Convolutional neural network (CNN)

The CNN is a type of feed-forward ANNs configured by a deep learning algorithm. It has been widely used in image processing, video recognition, and time series forecasting (Bai et al., 2019; Hamrani et al., 2020; Hatami et al., 2018; Kow et al., 2020; Milošević et al., 2020; Persello et al., 2019; Pyo et al., 2019; Qian et al., 2020; Wang et al., 2021; Yu et al., 2020; Zhang et al., 2017). The implementation of the CNN is briefly introduced below.

In this study, there are 3549 samples collected from the Linyuan station. The filtering process of each CNN–RC model was conducted on each sample, where the number of filters varied from model to model. It is noted that the CNN has a concept of “weight sharing”, that is to say, a filter does not change its weight values when screening each sample during the training stage. This leads to lesser parameters required during model construction for the CNN than for other feed-forward ANNs. As a result, the CNN is easier to train and can avoid overfitting, which makes the CNN an attractive deep learning algorithm. The max-pooling layer that filtrates the maximum value from each screen is usually connected behind a convolutional layer.

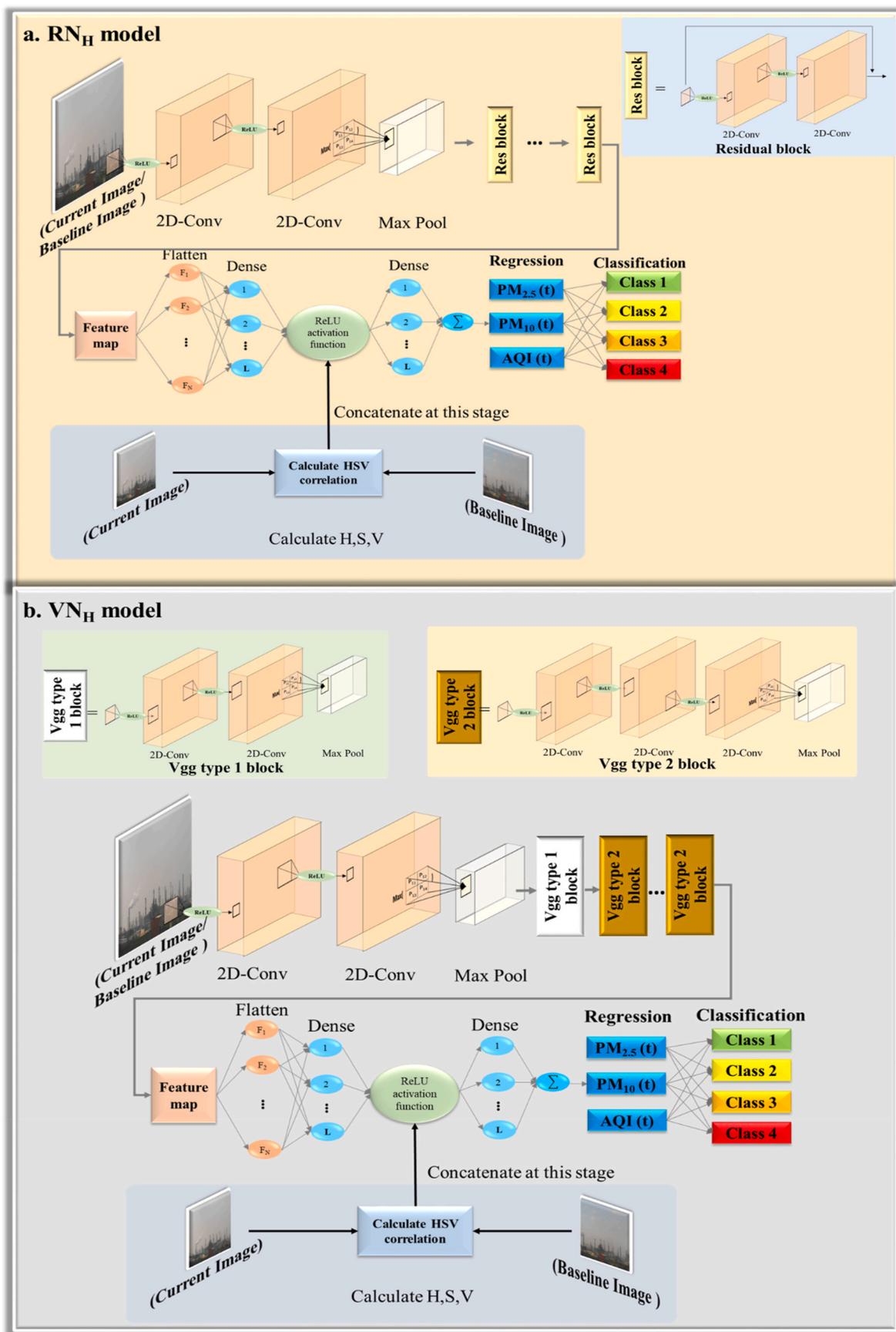


Fig. 2. CNN–RC deep learning approach in a hybrid architecture. (a) RN_H (RESNET-HSV) scheme. (b) VN_H (VGG-HSV) scheme.

A filter window determines the number of pixels in an image to screen at a time. Because this study utilized a two-dimensional CNN (Fig. 1(a) and (b)), the length of the filter window was the only parameter to determine (Table 2). The number of connections between the input layer and the convolutional layer conforms to the length of the filter window, while there is only one connection between the convolutional layer and the output layer (feature map). More details of the CNN can be found in Chen et al. (2019).

3.3. Fully connected layer and regression classifier

In this study, the final max-pooling layer of the CNN–RC approach was followed by three fully connected layers (Fig. 1(c)), in accordance with three outputs. A fully connected layer enables two main actions: propagation and weight adjustment. Propagation involves both forward propagation and backward propagation. In the forward propagation, each input signal is assigned a weight by the activation function (i.e., the ReLU function in this study) in the hidden layer, and then the weighted signal is passed to the output layer for calculating the output value. Next, a backward propagation follows if the difference between the output value and the target output value falls outside the tolerable error range. The network training stops when the error falls within the tolerable error range. Furthermore, a regression classifier is linked to the end of the CNN to generate regression outputs to be classified into different pollution levels (levels 1, 2, 3, and 4).

3.4. Hybrid of CNN and regression classifier

The proposed CNN–RC model that seamlessly fuses CNN with a regression classifier aims to classify multiple outputs, and its two-phase implementation engages feature extraction by CNN and multi-output estimation by the regression classifier (Fig. 1). The CNN is so powerful a tool for feature extraction that the similarity in patterns among samples can be considered as an auxiliary to classify air quality. Following feature extraction, the flatten layer links the feature map of the convolutional layer with the fully connected hidden layer to reshape each multi-dimensional input into a one-dimensional vector. Then, the fully connected hidden layer and the output layer constitute the CNN–RC model’s estimation phase. We set up three neurons in the output layer for producing three-dimensional outputs simultaneously (i.e., PM_{2.5} and PM₁₀ concentrations as well as AQI).

3.4.1. Hue, saturation and value (HSV)

In color image processing, HSV is one of the commonly used color models. HSV stands for hue, saturation, and value. The hue is the color type of an image and is normally expressed as a number from 0 to 360°, where each degree corresponds to one color. The saturation is the

Table 2
CNN–RC model description.

Deep Learning						
Architecture	Convolution Neural Network w/Regression Classifier (CNN–RC)					
Type	Residual Network (ResNet)			VGG Net (Visual Geometry Group Net)		
Layers	6, 8, 10, 12	4, 6, 8, 10				
Model	RN ^a	RN _H ^b	SRN _H ^c	VN ^d	VN _H ^e	SVN _H ^f
Model inputs						
Current image	✓	✓	✓	✓	✓	✓
Baseline image	✓	✓		✓	✓	
HSV value		✓	✓		✓	✓

^a ResNet.

^b ResNet-HSV.

^c Simplified ResNet-HSV.

^d VGG.

^e VGG-HSV.

^f Simplified VGG-HSV.

intensity of the color and is expressed by a number between 0 and 1 (0: no color; and 1: primary color). The value is the brightness and ranges between 0 and 1 (0: black; 1: brightest (white)) (Cantrell et al., 2010; Ngoc et al., 2019; Sural et al., 2002). In this study, the correlation (R^2) of the current image and the baseline image was calculated based on their HSV statistics, and then the correlation value became another input to the CNN–RC approach for enhancing model accuracy.

3.4.2. RN_H (ResNet-HSV) scheme

ResNet, one of the famous CNN architectures, comprises several residual blocks and fully connected layers (Fig. 2(a)). A residual block is composed of two convolutional layers, where the input of the block is passed to the next layer by skipping the convolutional layers in between and then is summed with the input of the next layer. The function of the residual block is to avoid the problem of vanishing gradient for increasing classification accuracy (He et al., 2016; Kálin et al., 2019; Wu et al., 2019). In the fully connected phase, the HSV correlation value is concatenated with the output of the flatten layer, where air quality estimation are carried out.

The ResNet-HSV scheme is constructed by stacking the convolutional layer and the dense layer. The inputs (e.g. the current and the baseline images) are fed into the convolutional layer, and image features are extracted by the filter that screen images (Fig. 1(a)). The HSV correlation is fed into the flatten layer and concatenated with the flatten output of the convolutional layer (Fig. 1(c)). Consequently, the outputs (estimations) of PM_{2.5}, PM₁₀ and AQI from the final dense layer are then converted into different classes (Fig. 1(d)). In this study, three ResNet models were built based on different input combinations: ResNet (RN); ResNet-HSV (RN_H); and simplified ResNet-HSV (SRN_H), as shown in Table 2. Besides, ResNet6 (RN6) had 6 residual blocks. This naming style was applied to ResNet8 (RN8), ResNet10 (RN10), and else.

3.4.3. VN_H (VGG-HSV) scheme

The use of building blocks was first proposed by the Visual Geometry Group (VGG) from Oxford University. The implementation of building blocks in the formation of repeated sequences in code with deep learning architecture through loops and subroutines is easy. VGG is made up of several VGG blocks and fully connected layers (Fig. 2(b)). A VGG block consists of two or three convolutional layers, and the end of the block is joined by a max-pooling layer. The advantage of VGG is that this scheme has small filters and pooling windows, which allows it to build a deeper model. A deeper model can learn more information from each input image and gain a higher possibility to improve classification accuracy (Jun et al., 2018; Li et al., 2020; Wang et al., 2015). Similar to RN_H, the HSV correlation value is concatenated with the output of the flatten layer in the fully connected phase, where air quality estimation and classification are carried out.

The VGG-HSV scheme is constructed by stacking the convolutional layer, the max pooling layer, and the dense layer. The inputs (e.g. the current and the baseline images) are fed into the convolutional layer, and image features are extracted by the filter that screen images (Fig. 1(a)). The extracted feature map then enters the max pooling layer. The HSV correlation is fed into the flatten layer and concatenated with the flatten output of the max pooling layer (Fig. 1(c)). Consequently, the outputs (estimations) of PM_{2.5}, PM₁₀ and AQI from the final dense layer are then converted into different classes (Fig. 1(d)). The VGG-HSV scheme has the same training procedure as the ResNet-HSV scheme. Three models of VGG were also built based on different input combinations: VGG (VN); VGG-HSV (VN_H); and simplified VGG-HSV (SVN_H) as shown in Table 2. This naming style of ResNet was applied to VGG, e.g. VN₄, VN₆, and else.

3.5. Evaluation indicators

We used three indicators to evaluate model performance, which were the Root Mean Square Error (RMSE), MAPE (Mean Absolute Percentage

Error), and the coefficient of determination (R^2). The formula of the three indicators are given below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^L (o_i - p_i)^2}{L}} \tag{1}$$

$$MAPE = \frac{1}{L} \sum_{i=1}^L \left| \frac{o_i - p_i}{o_i} \right| \tag{2}$$

$$R^2 = \frac{L \sum_{i=1}^L o_i p_i - \sum_{i=1}^L o_i \sum_{i=1}^L p_i}{\sqrt{L \sum_{i=1}^L o_i^2 - (\sum_{i=1}^L o_i)^2} \sqrt{L \sum_{i=1}^L p_i^2 - (\sum_{i=1}^L p_i)^2}} \tag{3}$$

where o_i denotes observed data, p_i denotes estimated values, and L denotes the data length.

4. Results and discussion

The CNN–RC model was proposed to estimate and classify air quality. Two types of CNN learning schemes (ResNet and VGG) were employed for comparison purpose. The results and findings were presented in the order of data preprocessing, model parameter setting, and model comparison, shown as follows.

4.1. Data preprocessing

A total of 3549 datasets (color images and their corresponding air quality monitoring data) were used for model training (80%) and testing (20%). The original dimension of each image was 1280×720 . After preprocessing, the (input) dimension of each image was reduced to 224×224 , where image pre-processing engaged removal of insignificant boundaries and preservation of useful parts of each image. In other words, this classification problem needs to tackle the curse of dimensionality induced by high dimensional datasets (3549 images with 224×224 dimensions). Such modelling is a computation-intensive task. Therefore, we proposed a deep learning framework to handle this classification. Our computation also benefited from Graphics Processing Units (GPUs). Besides, to well train a deep network, it is necessary to find the most suitable depth (the number of layers) of the network to avoid underfitting or overfitting. For instance this study considered 6, 8, 10, and 12 layers for the ResNet and 4, 6, 8, and 10 layers for the VGG Net. It is also crucial to explore more information of each image to assist model construction. In this study, HSV correlation between the current image and the baseline image was considered to be extra information adopted in training and testing phases.

Table 3
Parameter settings of CNN–RC models constructed in this study.

Model	Parameters						
	Epochs	Number of filters/neurons	Learning rate	Batch size	Kernel size	Patience (Early stopping)	Optimizer
VN6 series ^a	50	64 filters (VGG block), 128 filters (VGG block), 256 filters (VGG block), 16 neurons (FC layer), 3 neurons (FC layer)	0.0001	8	8	8	Adam
RN6 series ^b	50	64 filters (2 Conv ^d layer), 64 filters (ResNet block), 64 filters (ResNet block), 16 neurons (FC layer), 3 neurons (FC layer)	0.0001	8	8	8	Adam

VGG4, VGG6 and VGG8 have two, three and four VGG type 1 blocks, respectively.

VGG10 has two VGG type 1 blocks and two VGG type 2 blocks.

ResNet6, ResNet8, ResNet10 and ResNet12 have two, three, four and five ResNet blocks, respectively.

^a Consist of VN6, VN6_H, and SVN6_H.

^b Consist of RN6, RN6_H, and SRN6_H.

^c FC: Fully-connected.

^d Convolutional.

4.2. Model parameter setting

Table 3 presents the parameter settings of CNN–RC models. The VN series consisted of VN4, VN6, VN8, and VN10 only because overfitting would occur when more building blocks are employed. Similarly, the RN series contained RN6, RN8, RN10, and RN12. The learning rate of both models were set to be 0.0001 through trial and error. A small learning rate may lead to a local optimum. The batch size of both models was set to be 8. Batch size refers to the number of training samples utilized in one iteration. The larger the batch size, the faster the training speed. The kernel size of each filter was set to be 8, which meant the filter length and width were 8. Patience (early stopping) was set to be 8, which meant the training process of the model would terminate earlier if no further reduction in the error was made after 8 consecutive iterations. An early stopping mechanism allows a model to avoid overfitting problems.

4.3. Comparison of different CNN–RC architectures for image-based air quality regression classification

4.3.1. Model selection

Table 4 gives the comparison of six CNN–RC models constructed based on daytime and nighttime datasets separately, where ResNet and VGG are the two main CNN training schemes with three input combinations. The comparative results indicate that the SRN6_H and SVN6_H models (inputs: current image and HSV correlation value) performed the best in the daytime and nighttime cases, respectively, because they produced the majority of the highest R^2 and the lowest RMSE values. Moreover, the SVN6_H model was superior to the SRN6_H model. The reasons are that daytime images (more colorful) contain more useful information than nighttime images (less colorful), while nighttime images usually contain noises because the camera would recommend a long exposure due to insufficient light intensity in the nighttime. The reason for the SRN6_H model to serve as the most adaptive model for the daytime case is due to its ability to retain the properties of the previous layer. In other words, this model can more effectively learn deeply and extract useful information from higher-dimensional datasets, as compared to VGG6-related models. The reasons for the SVN6_H model to serve as the most adaptive model for the nighttime case are that nighttime images are in general nearly black-and-white and contain less information than daytime images (more colorful). Therefore, a shallower and simpler model like SVN6_H might be more suitable to apply in the nighttime. Besides, VGG models implemented max-pooling in each building block, which could effectively decrease the input dimension and reduce the noise of each building block.

Whether it was daytime or nighttime image, the best models (SRN6_H and SVN6_H) have two inputs, i.e., the current image and the HSV correlation value. The reason not to include the baseline image as a model input is that its high dimensionality would cause noises during model

Table 4

Performance (R^2 and RMSE) of 6 CNN–RC models constructed separately based on daytime datasets (images) and nighttime datasets (images) of the Linyuan air quality monitoring station.

Datasets	Model	$PM_{2.5}$ ($\mu g/m^3$)		PM_{10} ($\mu g/m^3$)		AQI	
		Training	Testing	Training	Testing	Training	Testing
Daytime (6 a.m.–6 p.m.)	RN6	0.6 ^a (7.6 ^b)	0.7 (7.4)	0.5 (19.0)	0.5 (18.8)	0.6 (20.8)	0.6 (19.5)
	RN6 _H	0.6 (8.0)	0.7 (7.5)	0.5 (20.1)	0.5 (16.6)	0.8 (15.6)	0.7 (17.5)
	SRN6_H^c	0.8 (5.9)	0.7 (6.8)	0.5 (17.7)	0.5 (16.6)	0.8 (15.6)	0.7 (18.4)
	VN6	0.7 (6.5)	0.6 (7.1)	0.6 (17.8)	0.5 (16.8)	0.8 (15.8)	0.7 (18.3)
	VN6 _H	0.7 (7.4)	0.7 (7.2)	0.5 (18.2)	0.5 (16.5)	0.8 (16.4)	0.7 (17.4)
	SVN6 _H	0.7 (6.9)	0.6 (7.7)	0.5 (18.4)	0.5 (16.5)	0.7 (17.3)	0.6 (19.6)
Nighttime (7 p.m.–5 a.m.)	RN6	0.8 (5.6)	0.7 (6.3)	0.9 (11.4)	0.7 (14.7)	0.9 (10.3)	0.8 (16.8)
	RN6 _H	0.9 (4.6)	0.7 (6.1)	0.8 (13.7)	0.7 (16.1)	0.9 (15.1)	0.8 (19.0)
	SRN6 _H	0.4 (10.1)	0.5 (9.8)	0.8 (15.2)	0.6 (17.9)	0.8 (22.7)	0.7 (26.1)
	VN6	0.9 (4.0)	0.8 (5.9)	0.9 (9.0)	0.7 (14.5)	0.9 (9.7)	0.8 (17.3)
	VN6 _H	0.8 (6.1)	0.7 (7.0)	0.9 (9.3)	0.7 (14.1)	0.9 (10.0)	0.8 (16.9)
	SVN6 _H	0.9 (3.4)	0.8 (5.4)	0.9 (7.5)	0.7 (14.6)	0.9 (9.3)	0.8 (17.1)

^a R^2 value.

^b RMSE value.

^c Best models and optimal values are marked in bold.

training. In contrast, the HSV correlation value is numeric, i.e., a relatively low dimension. The calculation of the HSV correlation value is equivalent to noise removal from images and can extract the difference between the current image and the baseline image effectively.

Therefore, models that adopted the HSV correlation value could improve estimation/classification accuracy.

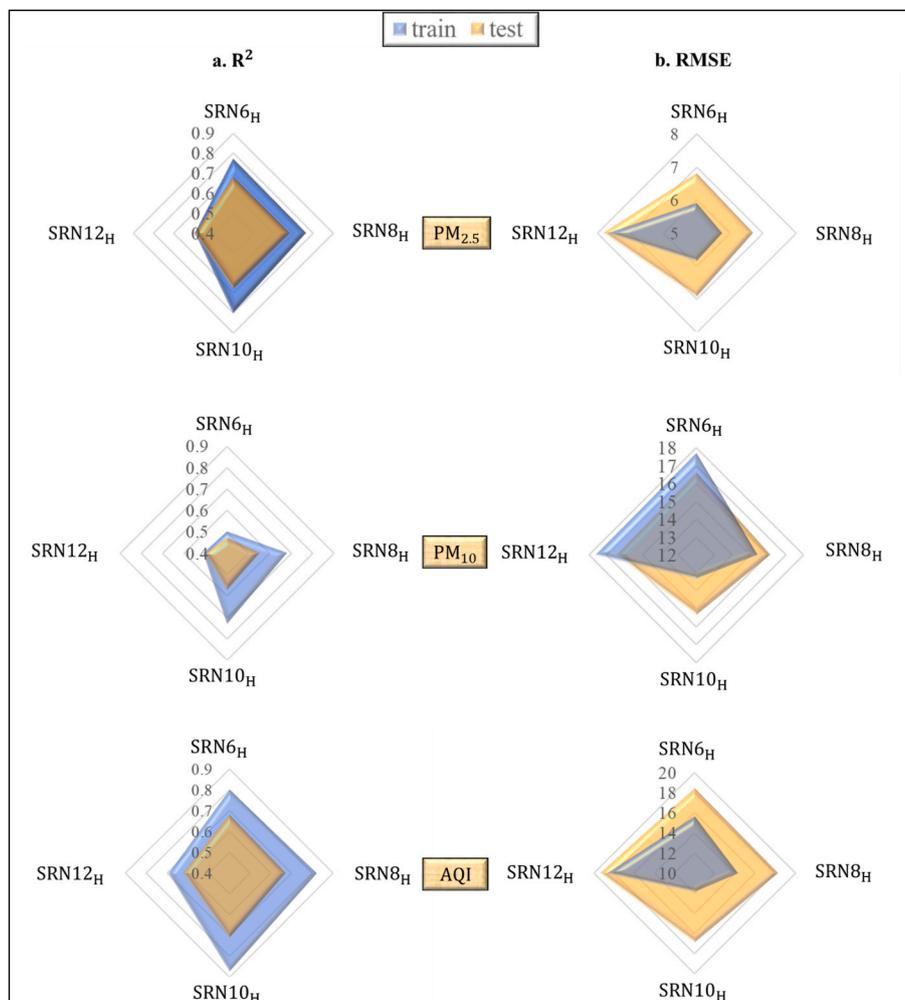


Fig. 3. Performance of SRN_H (simplified ResNet-HSV) model that are 6, 8, 10 and 12 layers deep based on the daytime input images of the Linyuan air quality monitoring station. (a) R^2 between observed and estimated air quality. (b) RMSE between observed and estimated air quality.

4.3.2. Determination of the optimal number of building blocks

Following the results of Table 4, this study further explored the optimal number of building blocks for the SRN_H and SVN_H models in the interest of improving classification accuracy. Fig. 3 shows that the SRN10_H model, in general, outperformed (highest R² and lowest RMSE values, except for PM_{2.5}) the other three models for the daytime case. Fig. 4 reveals that the SVN6_H model performed better (highest R² and lowest RMSE values, except for AQI) than the other three models for the nighttime case. In brief, the SRN10_H and SVN6_H models are considered the best models for the daytime and nighttime cases, respectively.

4.3.3. Classification performance

Table 5 shows the four pollution levels of each output (PM_{2.5}, PM₁₀, and AQI) for classification purpose in this study. Air quality standards for PM_{2.5}, PM₁₀, and AQI established by the TW EPA based on moving average values of pollutant concentrations can refer to EPA (2020). This study alternatively classified images into pollution levels based on the estimates of PM_{2.5} and PM₁₀ concentrations and AQI, rather than their moving average values.

Table 6 shows the classification results of the SRN10_H model based on the daytime images (1556 training datasets and 390 test datasets). The classification accuracies in training (testing) stages were 86% (76%) for AQI, 87% (84%) for PM₁₀, but a bit low (79% (76%)) for PM_{2.5}. Table 6 also shows the classification results of the SVN6_H model based on the nighttime images. The classification accuracy of PM_{2.5} achieved 87% and 83% in the training and testing stages, respectively. PM₁₀ had

Table 5

Classification level of AQI, PM_{2.5} and PM₁₀ (according to TW EPA standard: <https://taqm.epa.gov.tw/taqm/en/b0201.aspx>).

Pollution Level	AQI	PM _{2.5} (µg/m ³)	PM ₁₀ (µg/m ³)
Level 1 (Good)	0–50	0.0–15.4	0–54
Level 2 (Moderate)	51–100	15.5–35.4	55–125
Level 3 (Unhealthy)	101–150	35.5–54.4	126–254
Level 4 (Very Unhealthy)	151+	54.5+	255+

classification accuracies of 93% and 84% in the training and testing stages, respectively. The classification accuracy of AQI reached 89% and 74% in the training and testing stages, respectively. In addition, classification accuracy was higher in the nighttime than in the daytime. It seemed that colorfulness in images did not provide useful information for this image classification problem. Classification of air quality image is considered more dependent on the smoke size and density as well as the clearness of the image. For example, the model can well estimate pollutant concentrations when the image is less clear and shows much smoke, CNN considers there is air pollution.

Regardless of daytime and nighttime, there was an obvious underestimation of air pollutants under the severe, harmful condition (i.e., level 4), especially true for PM_{2.5}. This was because deep learning required a lot of data to learn, whereas there were only very few training data (PM_{2.5}: 8 and 7 images for daytime and nighttime, respectively; and AQI: 30 and 49 images for daytime and nighttime, respectively) in level 4. Besides, the estimation of PM₁₀ concentration was the most accurate,

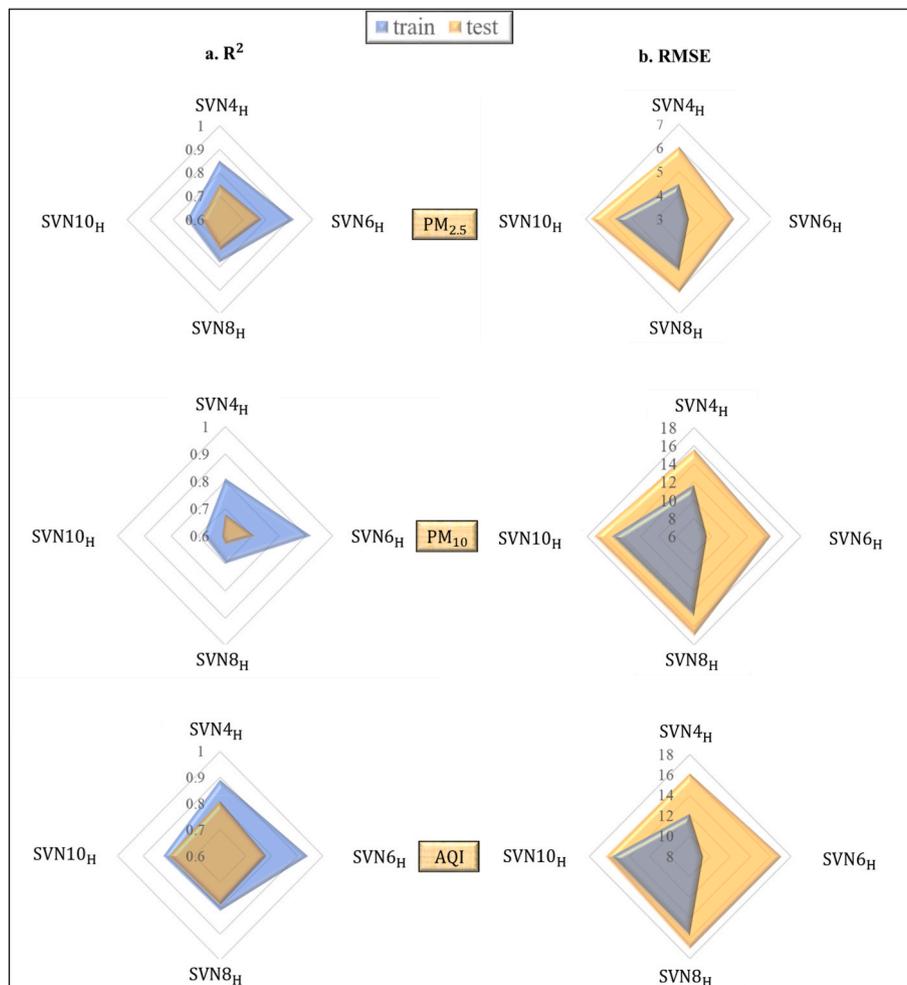


Fig. 4. Performance of SVN_H (simplified VGG-HSV) model that are 6, 8, 10 and 12 layers deep based on the nighttime input images of the Linyuan air quality monitoring station. (a) R² between observed and estimated air quality. (b) RMSE between observed and estimated air quality.

Table 6
Classification results of the SRN10_H and SVN6_H models based on daytime images and nighttime images.

		Training					Testing						
		obs'est*	1	2	3	4	Total	obs'est	1	2	3	4	Total
	PM _{2.5}	1	439	162	0	0	601	1	99	49	0	0	148
		2	21	656	101	0	778	2	9	168	22	0	199
		3	2	29	137	1	169	3	0	14	28	0	42
		4	1	1	6	0	8	4	1	0	0	0	1
		Accuracy count	1232 (79%)					1556	Accuracy count (rate)	295(76%)			
MAPE	0.21						MAPE	0.25					
Daytime	PM ₁₀	obs'est	1	2	3	4	Total	obs'est	1	2	3	4	Total
		1	916	154	0	0	1070	1	226	41	0	0	267
		2	44	432	0	0	476	2	21	100	0	0	121
		3	1	7	0	0	8	3	2	0	0	0	2
		4	1	0	1	0	2	4	0	0	0	0	0
Accuracy count	1348 (87%)					1556	Accuracy count (rate)	326 (84%)					390
MAPE	0.14						MAPE	0.17					
	AQI	obs'est	1	2	3	4	Total	obs'est	1	2	3	4	Total
		1	506	47	0	0	553	1	115	26	0	0	141
		2	44	665	52	0	761	2	20	150	22	0	192
		3	1	44	167	0	212	3	0	21	32	0	53
		4	0	2	21	7	30	4	0	1	3	0	4
Accuracy count	1345 (86%)					1556	Accuracy count (rate)	297 (76%)					390
MAPE	0.14						MAPE	0.24					
	PM _{2.5}	obs'est	1	2	3	4	Total	obs'est	1	2	3	4	Total
		1	449	42	0	0	491	1	101	19	0	0	120
		2	47	614	13	0	674	2	13	146	5	0	164
		3	0	55	55	0	110	3	0	17	19	0	36
		4	0	0	7	0	7	4	0	1	0	0	1
Accuracy count	1118 (87%)					1282	Accuracy count (rate)	266 (83%)					321
MAPE	0.13						MAPE	0.17					
Nighttime	PM ₁₀	obs'est	1	2	3	4	Total	obs'est	1	2	3	4	Total
		1	809	52	0	0	861	1	190	31	0	0	221
		2	27	382	0	0	409	2	17	79	0	0	96
		3	0	8	3	0	11	3	1	3	0	0	4
		4	0	0	0	1	1	4	0	0	0	0	0
Accuracy count	1195 (93%)					1282	Accuracy count (rate)	269 (84%)					321
MAPE	0.07						MAPE	0.17					
	AQI	obs'est	1	2	3	4	Total	obs'est	1	2	3	4	Total
		1	399	34	0	0	433	1	84	16	0	0	100
		2	21	573	13	0	607	2	21	119	17	0	157
		3	0	33	160	0	193	3	0	17	34	0	51
		4	0	0	34	15	49	4	0	0	13	0	13
Accuracy count	1147 (89%)					1282	Accuracy count	237 (74%)					321
MAPE	0.11						MAPE	0.26					

* Estimate

Min  Max

as compared to those of $PM_{2.5}$ concentration and AQI, which proves the black smoke in images has relatively high correlation with PM_{10} . Table 6 shows that the regression classification of images produced better results for AQI than for $PM_{2.5}$ and PM_{10} . This could be because AQI is a comprehensive indicator and can objectively reflect the degree of environmental pollution.

Fig. 5 shows the results of air quality ($PM_{2.5}$, PM_{10} and AQI) estimated continuously by SVN_{6H} for nighttime and SRN_{10H} for daytime during October 15th, 2019 and October 30th, 2019 to observed data. In addition, three enlarged plots presented on the right-hand side of each subfigure were given to clearly show the models' performance at three selected times that were adopted in Figs. 7 and 8 as well, where significant variations in value (concentration or AQI index) occurred. In general, both models could well keep the tracks and produce accurate estimation even though the value changed drastically. Nevertheless, there were occasions difficult to estimate. For instance, the $PM_{2.5}$ trend was captured but $PM_{2.5}$ was underestimated on October 27th. Similarly, PM_{10} was underestimated on October 21st.

4.3.4. Discussion on air quality images

Table 1 indicates that the pairs of (mean, standard deviation) for $PM_{2.5}$, PM_{10} , and AQI at the investigative station are (20.7, 11.5) $\mu g/m^3$,

(46.5, 25.4) $\mu g/m^3$, and (69.1, 34.3), respectively. These mean values are far away from their individual lower limits of level 4, which are 54.5 $\mu g/m^3$, 255 $\mu g/m^3$, and 151 for $PM_{2.5}$, PM_{10} , and AQI, respectively. Recognizing the difficult, if not impossible, problem related to the precise classification of the extreme (harmful) conditions with high variations in concentrations of air pollutants and only very limited datasets in level 4, the analytical results produced by the proposed model could be beneficial and useful in practice.

Table 4 shows the smallest RMSE values for $PM_{2.5}$, PM_{10} , and AQI based on daytime (nighttime) images in the test phase were 6.8 (5.4), 16.6 (14.1), and 17.5 (16.9), respectively (Table 4). It appears that the estimation errors of the proposed approach were much smaller than the standard deviation (about half) of their corresponding pollutants. These results suggest that the proposed CNN-RC approach has a great potential to accurately real-time estimate and/or classify multiple air pollutants based solely on in-situ photos. Nevertheless, there were still several misjudgements, and they deserved to have a close check.

Figs. 6 and 7 show a total of twelve photos representative of well-classified and poorly-classified images for $PM_{2.5}$, PM_{10} , and AQI in the daytime case (SRN_{10H}) and in the nighttime case (SVN_{6H}). It is observed that CNN models could well recognize images that are less clear and have much smoke. This could be because the weights of the model could

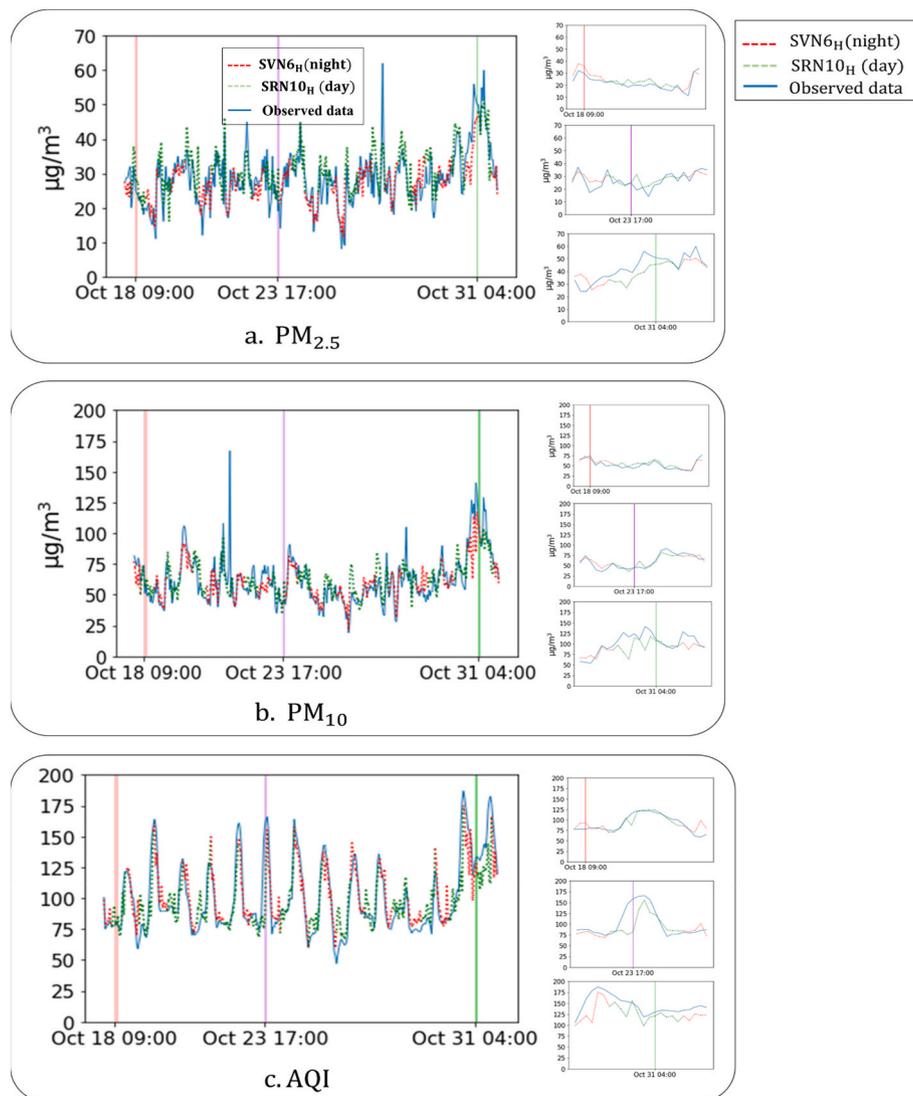


Fig. 5. Air quality estimation performance of SVN_{6H} (nighttime) and SRN_{10H} (daytime) during October 15th and October 30th of 2019, in comparison with observed data. (a) $PM_{2.5}$. (b) PM_{10} . (c) AQI.

		Well estimated (classified)			Poorly estimated (classified)				
PM _{2.5}	a1. 2019/9/27 11:00		b1. 2019/7/30 17:00						
		Obs ¹ (class)	Est ² (class)	Error	Obs(class)	Est(class)	Error		
		PM _{2.5} *	43(3)	40.1(3)	0	PM _{2.5} *	64(4)	11(1)	-3
		PM ₁₀	76(2)	74.7(2)	0	PM ₁₀	73(2)	33.3(1)	-1
	AQI	109(3)	103.9(3)	0	AQI	69.4(2)	49.9(1)	-1	
Daytime (SRN10 _H)	a2. 2019/10/18 09:00		b2. 2019/7/4 09:00						
		Obs(class)	Est(class)	Error	Obs(class)	Est(class)	Error		
		PM ₁₀ *	74(2)	66.5(2)	0	PM ₁₀ *	146(3)	36.8(1)	-2
		PM _{2.5}	30(2)	36.2(3)	1	PM _{2.5}	28(2)	6.7(1)	-1
	AQI	78(2)	92.8(2)	0	AQI	48(1)	32.3(1)	0	
AQI	a3. 2019/11/4 17:00		b3. 2019/10/23 17:00						
		Obs(class)	Est(class)	Error	Obs(class)	Est(class)	Error		
		AQI*	135(3)	117.3(3)	0	AQI*	159(4)	82.3(2)	-2
		PM _{2.5}	54(3)	46.8(3)	0	PM _{2.5}	25(2)	24.6(2)	0
	PM ₁₀	86(2)	89.5(2)	0	PM ₁₀	45(1)	44.2(1)	0	

Fig. 6. Air quality classification results for PM_{2.5}, PM₁₀ and AQI based on the daytime images of the Linyuan air quality monitoring station. The image in each sub-figure was selected according to the * item. ¹ Observation value. ² Estimation value.

be trained by the smoke and the clearness of images. Taking Fig. 6(b1) as an example, the estimate of PM_{2.5} concentration was classified into level 1, which resulted in a huge classification error (observed concentration falls into level 4). The model apparently misjudged this image because it is very clear and does not have much smoke. Therefore, it is not surprising why the classification result of Fig. 6(b1) had a huge error (-3), i.e., underestimated. PM₁₀ concentration in Fig. 6(b2) was also underestimated (error = -2). This image is slightly blurry, but the model could not distinguish clouds from smoke in the sky. From Fig. 6(b3), AQI

was also underestimated (error = -2). This image shows blurs but has little smoke.

In contrast, Fig. 6(a1), 6 (a2), and 6 (a3) present well-classified images in the daytime case, where blurriness can be observed, as compared to Fig. 6(b1). Therefore, the CNN model could classify these images accurately. As for the nighttime case, PM_{2.5} concentration in Fig. 7(b4) was underestimated and had a classification error of -2. The CNN model made a misjudgment on this image because it is very clear and has a bit of smoke only. Fig. 6(b2) reveals that PM₁₀ concentration was also

		Well estimated (classified)			Poorly estimated (classified)				
PM _{2.5}	a4. 2019/11/24 02:00		b4. 2019/9/26 23:00						
		Obs(class)	Est(class)	Error	Obs(class)	Est(class)	Error		
		PM _{2.5} *	36(3)	37.1(3)	0	PM _{2.5} *	56(4)	32.8(2)	-2
		PM ₁₀	66(2)	78.8(2)	0	PM ₁₀	87(2)	55.9(2)	0
		AQI	103(3)	97(2)	0	AQI	84(2)	99.9(2)	0
Nighttime (SVN6 _H)	a5. 2019/9/30 05:00		b5. 2019/9/30 19:00						
		Obs(class)	Est(class)	Error	Obs(class)	Est(class)	Error		
		PM ₁₀ *	60.5(2)	55.1(2)	0	PM ₁₀ *	130(3)	53.5(1)	-2
		PM _{2.5}	33(2)	34.6(2)	0	PM _{2.5}	29(2)	29.9(2)	0
		AQI	88.9(2)	81.5(2)	0	AQI	79.2(2)	90.4(2)	0
AQI	a6. 2019/10/31 04:00		b6. 2019/10/7 23:00						
		Obs(class)	Est(class)	Error	Obs(class)	Est(class)	Error		
		AQI*	129.8(3)	120.7(3)	0	AQI*	85(2)	107.8(3)	1
		PM _{2.5}	51(3)	45.5(3)	0	PM _{2.5}	32(2)	32.9(2)	0
		PM ₁₀	111(2)	107.1(2)	0	PM ₁₀	45(1)	54.2(2)	1

Fig. 7. Air quality classification results for PM_{2.5}, PM₁₀ and AQI based on the nighttime images of the Linyuan air quality monitoring station. The image in each sub-figure was selected according to the * item. ¹ Observation value. ² Estimation value.

underestimated and misclassified into level 1, which caused a huge error of -2 (observed concentration fell into level 3). It seems there were no obvious classification errors for AQI. It is worth noting that a classification error occurred in Fig. 7(b6) even though the estimated AQI (54.2) was quite close to the upper limit (50) of level 1.

Fig. 7(a4) 7 (a5), and 7 (a6) also show nice classification results (error = 0), where the flow of smoke is clearly visible. Besides, the image in Fig. 7(a6) is somewhat blurry. It is apparent that the proposed model can well classify non-defective images accurately even though it is in the

nighttime.

We next take a close look at the relationship between PM_{2.5}, PM₁₀, and AQI. According to the observed PM_{2.5} and PM₁₀ concentrations in Fig. 6(a1), 6(b1), 7(a4), and 7 (b4), we notice that PM₁₀ concentration was high when PM_{2.5} concentration was high. However, high PM₁₀ concentration did not necessarily correspond to high PM_{2.5} concentration (Fig. 6(b2) and 7 (b5)). Besides, the observed AQI values in Fig. 6 (a3), 6(b3), 7(a6), and 7 (b6) had no obvious relationship with PM_{2.5} and PM₁₀ concentrations. Fig. 6(a1), 6(a2), 7(a4), and 7 (a5) show that it

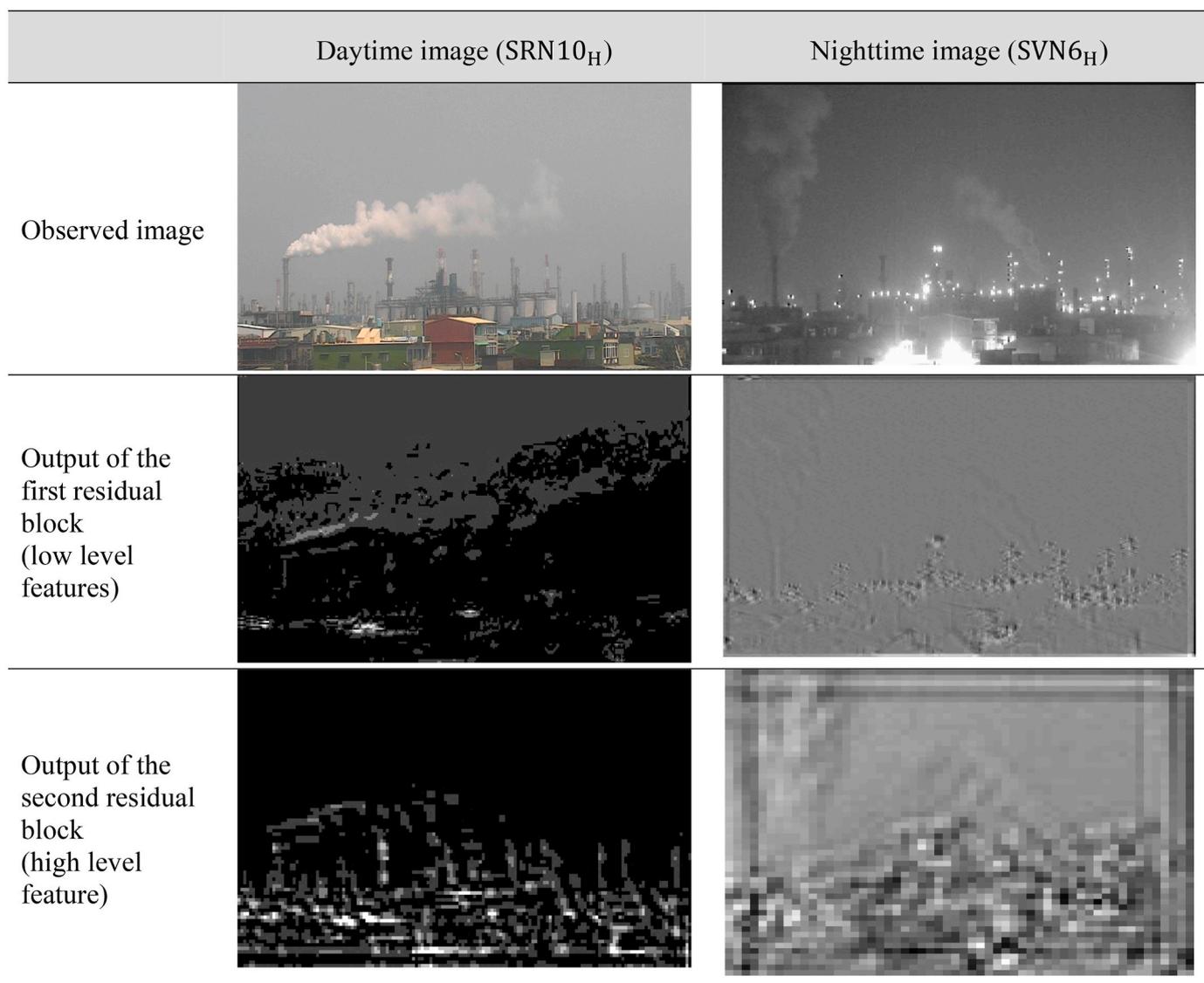


Fig. 8. Comparison between outputs of different residual blocks based on the daytime (SRN10_H) and nighttime (SVN6_H) images taken at 2019/03/03 10am and 2019/10/31 4am, respectively, at the Linyuan air quality monitoring station.

is difficult to have an intuitive look at the relationship between AQI with PM_{2.5} and PM₁₀. This is because AQI is a composite index converted from six pollutants, including PM_{2.5}, PM₁₀, SO₂, NO₂, CO, and O₃. It is observed that high AQI values occurred when the concentrations of PM_{2.5} and PM₁₀ were high. Furthermore, the formation of PM_{2.5} and PM₁₀ in the air may not happen until particulate matters react with climatic factors and other chemical substances in the air. These reasons may explain the estimation accuracy with regard to PM_{2.5}, PM₁₀, and AQI for the images in Figs. 6 and 7.

Furthermore, the reason for poor estimation shown in Figs. 6 and 7 is also related to the composition of the emissions, which are obviously associated with petrochemical production. The patterns of poor estimates are similar in Fig. 6(b1) and 6 (b2) as well as Fig. 7(b4) and 7 (b5), where both models underestimated the concentrations of PM_{2.5} and PM₁₀. However, the AQI value was well-estimated. We speculate that the main substances of the emissions from the chimneys at that time were not PM_{2.5} or PM₁₀ but were more likely to be SO_x, NO_x, or other pollutants from petrochemical production.

4.3.5. Discussion on extracted features from images

We next explore how the proposed network extracted the air quality features contained in the images. Fig. 8 shows the outputs of different

residual blocks based on the daytime (SRN10_H) and nighttime (SVN6_H) images taken at 2019/03/03 10am and 2019/10/31 4am, respectively, at the Linyuan air quality monitoring station. For the daytime image, the output of the first residual block contains low level feature, where the edge of smoke is quite obvious but the edges of clouds are scarcely distinguishable. It is rather difficult to identify chimneys, too. In contrast, the output of the second residual block contains high level features, where the edges of smoke, chimneys and buildings are obvious. For the nighttime image, the edges of smoke and buildings in the output of the first residual block are not distinguishable due to insufficient light. The trajectory of smoke in the output of the second residual block is more obvious than that of the first residual block. According to these results, we consider that the trajectory of smoke would be important information for model training. Besides, the output of the second residual block contains more significant features than that of the first one because it combines the output of the previous (first) block.

4.4. Discussion and limitation

The VGG model (SVN6_H) achieved better results on nighttime images than the ResNet models, whereas the ResNet model (SRN10_H) achieved better results on daytime images than the VGG models (Figs. 3 and 4).

Because daytime images (more colorful) contain more useful information than nighttime images, the ResNet model was further compared with another image classification model established by ShuffleNet (Zhang et al., 2018) based on daytime images. ShuffleNet has been used to tackle image classification problems (Ghosh et al., 2020; Guan, 2019; Liu et al., 2020; Miao et al., 2020). Besides, real-time tasks usually aim to reach the best accuracy with limited computational budgets. Therefore, ShuffleNet that has a lightweight architecture can be implemented to cope with real-time tasks for achieving favorable accuracy at a faster speed.

The comparative estimation results between the ResNet (SRN_{10H}) and ShuffleNet (ShuffleN_{9H}) models based on daytime images with respect to PM_{2.5}, PM₁₀ and AQI indicate that ShuffleN_{9H} for PM_{2.5} performed slightly worse than SRN_{10H} while ShuffleN_{9H} for PM₁₀ and AQI performed significantly worse than SRN_{10H}. The reason why ShuffleN_{9H} performed inferior to SRN_{10H} was that the function of channel shuffle in ShuffleN_{9H} failed to efficiently improve either the recognition on the edge of the object (smoke) or the ability to learn the difference between the current image and the baseline image.

In this study, ‘real time’ reflects the time that model parameters are updated during actual operation. Taking SRN_{6H} as an example, the total

computation time of the backward propagation phase is 1455 s (=29.1s/ epochs*50 epochs), which is significantly larger than that of the feed-forward phase (<5 s). Considering the feed-forward operation consumes significantly shorter computation time than the backward propagation phase, the computation time of the feed-forward phase can be ignored. Fig. 9 presents the average computation time per image in the backward propagation phase of different image classification models constructed with 6 hidden layers (Fig. 9(a)) and with different numbers of hidden layers (Fig. 9(b)) in this study. Fig. 9(a) shows that the VGG models consumed relatively short computation time, as compared to the ResNet models with similar input combinations. It is noted that for VGG series, the model with 3 inputs (i.e., current image, baseline image, and HSV value) consumed the highest amount of computation time whereas the model with 2 inputs (i.e. current image, and HSV value) consumed the least amount of computation time. Similar results can be found in the ResNet series. Fig. 9(b) indicates that the more the number of layers in the SRNH (from 6 to 12) and SVNH (from 4 to 10) models, the longer the computation time. Besides, we also notice that the computation time of ShuffleN_{9H} was obviously longer than that of SRN_{10H}. It is because the structure of a shuffle block is composed of a group convolution unit, a channel shuffle unit, and a depth-wise convolution unit, which is more

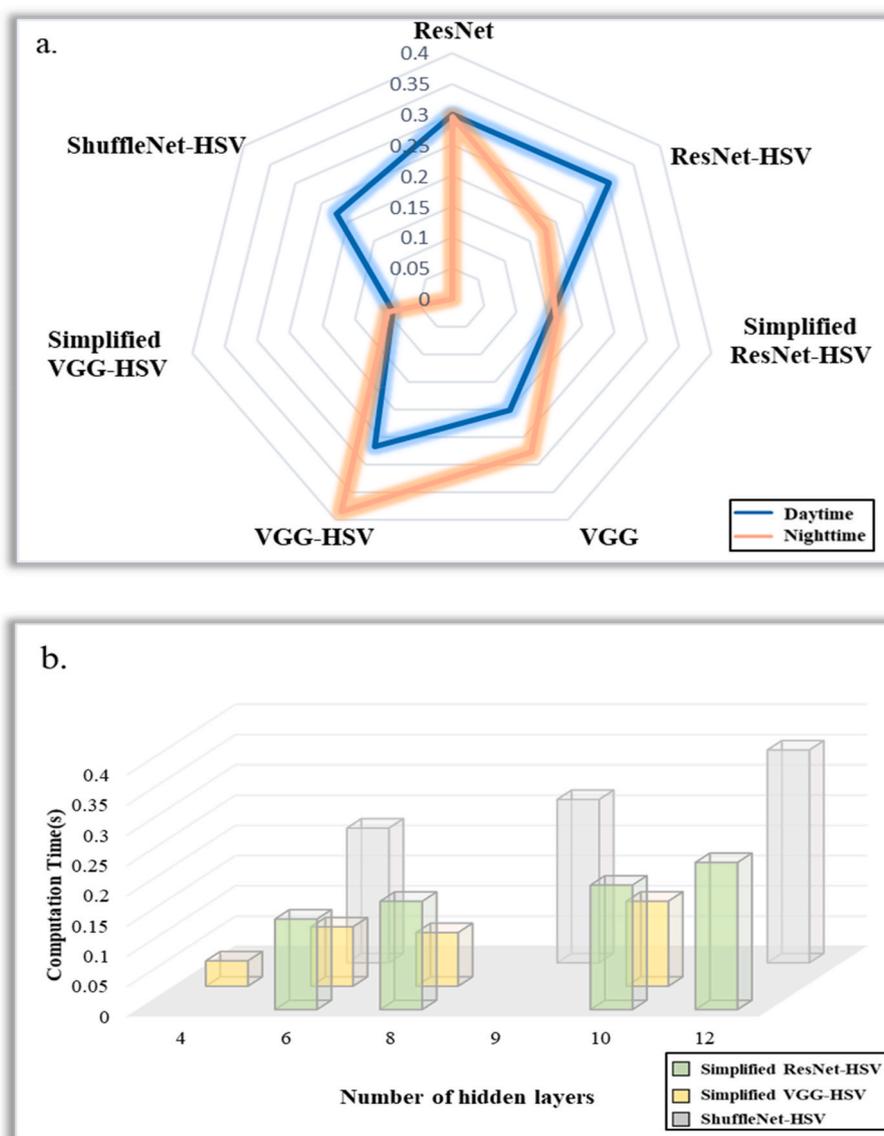


Fig. 9. Average computation time (in second) per image in the backward propagation phase of different image classification models. (a) 6 hidden layers. (b) Different numbers of hidden layers.

complicated than that of ResNet. Therefore the computation time is considered another important reason to choose the SRNH model, rather than Shuffle9NH, as the best image classification model.

Experimental results demonstrated that the proposed CNN–RC approach could well estimate air quality according to the smoke (size and density) and clearness of images and provide reliable and suitable measurements (PM_{2.5}, PM₁₀ and AQI). The CNN–RC model exhibited short computation time and easy operation. Even though this image-based model cannot perfectly reach the estimation made by the expensive instruments equipped at monitoring stations, it can play a complementary role with the advantage of obtaining timely comprehensive air quality estimation. Because the investigative datasets only consisted of air quality images at a fixed spot, the proposed model has relatively high applicability, reliability, and accuracy for the relevant local area. As known, different weather conditions would affect the lightness or darkness of air quality images. Thus, our model may not achieve the expected accuracy in other regions due to the difference in regional climatic and atmospheric conditions. It needs to train and adjust our model again using local image data when applying our model in other regions. Future research can incorporate these factors to improve model accuracy.

5. Conclusion

Air pollution imposes serious threats to human health while raising intensive public attention in recent decades, yet ambient pollution measurements are expensive, and therefore the spatial coverage of air quality monitoring stations is limited. An efficient low-cost air quality sensing device is of great benefit for human health and air pollution control. In this study, we employed the state-of-the-art computer vision techniques for analyzing photos to estimate the haze levels (air pollution), which could serve as an information source complementary to official data. We proposed a deep learning approach (CNN–RC) hybridizing CNN (under ResNet and VGG schemes with several layers deep) for image feature extraction and a regression classifier based on air quality photos collected at the Linyuan air quality monitoring station. The CNN–RC models with different input combinations were investigated, where the input sources were current images, the baseline image, and HSV statistics. The main contributions of the proposed CNN–RC approach are three-fold.

Firstly, the CNN–RC approach can well estimate and classify multiple outputs (PM_{2.5}, PM₁₀, and AQI) at the same time based on multiple inputs, i.e., it has a multi-input multi-output (MIMO) framework. The test results show that classification accuracy for PM_{2.5}, PM₁₀ and AQI based on day-time (night-time) images reached 76% (83%), 84% (84%) and 76% (74%), respectively. The results demonstrate that the proposed CNN–RC models not only could adequately handle daytime and nighttime images to tackle the curse of dimensionality but could also allow models to learn and extract useful knowledge from high-dimensional datasets more deeply than shallow neural networks. Besides, HSV statistics evidently played an important role in improving the accuracy of the proposed CNN–RC approach.

Secondly, the proposed CNN–RC approach can produce accurate estimates on pollutant concentrations for images that are less clear and have much smoke. The reason is that CNN considers such photos to be of high air pollution because the model's weights are adjusted by the smoke and the clearness of the image.

Thirdly, an interesting finding is that the CNN–RC approach performed better in the nighttime case than in the daytime case. The reason could be the colors between an image and the significant feature such as smoke. The smoke present in the images appears white, gray, or black in general. Daytime photos are colorful and/or have clouds (i.e., noises relative to smoke), whereas nighttime images appear to be black-and-white.

We conclude that the proposed CNN–RC approach can rapidly and suitably provide air pollution estimation, especially when some air

quality detectors are malfunctioning or at ungauged areas. In light of methodological transferability, future research can extend the CNN–RC approach from one single station to multiple stations for producing multiple outputs at multiple sites at the same time. Our analytical results demonstrate that this image-based approach provides reliable and accurate measurements as the ones acquired from expensive instruments, while it exhibits short computation time and easy operation, which have important implications for how air pollution is measured and managed.

Credit author statement

Pu-Yun Kow: Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft. **I-Wen Hsia:** Formal analysis, Image processing, Visualization. **Li-Chiu Chang:** Methodology, Project administration, Resources, Supervision. **Fi-John Chang:** Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study is supported by the Ministry of Science and Technology, Taiwan (MOST: 109-2119-M-002-15-A). The datasets provided by the Environmental Protection Administration of Taiwan are acknowledged. The authors would like to thank the Editors and anonymous Reviewers for their constructive comments that greatly contribute to enriching the manuscript.

References

- Babari, R., Hautiere, N., Dumont, E., Brémond, R., Paparoditis, N., 2011. A model-driven approach to estimate atmospheric visibility with ordinary cameras. *Atmos. Environ.* 45 (30), 5316–5324. <https://doi.org/10.1016/j.atmosenv.2011.06.053>.
- Bai, Y., Li, Y., Zeng, B., Li, C., Zhang, J., 2019. Hourly PM_{2.5} concentration forecast using stacked autoencoder model with emphasis on seasonality. *J. Clean. Prod.* 224, 739–750. <https://doi.org/10.1016/j.jclepro.2019.03.253>.
- Bo, Q., Yang, W., Rijal, N., Xie, Y., Feng, J., Zhang, J., 2018. Particle pollution estimation from images using convolutional neural network and weather features. In: 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3433–3437. <https://doi.org/10.1109/ICIP.2018.8451306>.
- Cantrell, K., Erenas, M.M., de Orbe-Payá, I., Capitán-Vallvey, L.F., 2010. Use of the hue parameter of the hue, saturation, value color space as a quantitative analytical parameter for bitonal optical sensors. *Anal. Chem.* 82 (2), 531–542. <https://doi.org/10.1021/ac901753c>.
- Chakma, A., Vizena, B., Cao, T., Lin, J., Zhang, J., 2017. September). Image-based air quality analysis using deep convolutional neural network. In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3949–3952. <https://doi.org/10.1109/ICIP.2017.8297023>.
- Chang, F.J., Chang, L.C., Kang, C.C., Wang, Y.S., Huang, A., 2020. Explore spatio-temporal PM_{2.5} features in northern Taiwan using machine learning techniques. *Sci. Total Environ.* 736, 139656.
- Chen, X., Kopsaftopoulos, F., Wu, Q., Ren, H., Chang, F.K., 2019. A self-adaptive 1D convolutional neural network for flight-state identification. *Sensors* 19 (2), 275. <https://doi.org/10.3390/s19020275>.
- EPA, 2020. <https://taqm.epa.gov.tw/taqm/en/b0201.aspx>.
- Ghosh, S., Mondal, M.J., Sen, S., Chatterjee, S., Roy, N.K., Patnaik, S., 2020, October. A novel approach to detect and classify fruits using ShuffleNet V2. In: 2020 IEEE Applied Signal Processing Conference (ASPCON). IEEE, pp. 163–167. <https://doi.org/10.1109/ASPCON49795.2020.9276669>.
- Guan, C.Z., 2019, August. Realtime multi-person 2d pose estimation using shufflenet. In: 2019 14th International Conference on Computer Science & Education (ICCSE). IEEE, pp. 17–21. <https://doi.org/10.1109/ICCSE.2019.8845343>.
- Giyenko, A., Palvanov, A., Cho, Y., 2018. January). Application of convolutional neural networks for visibility estimation of CCTV images. In: 2018 International Conference on Information Networking (ICOIN). IEEE, pp. 875–879. <https://doi.org/10.1109/ICOIN.2018.8343247>.
- Hamrani, A., Akbarzadeh, A., Madramootoo, C.A., 2020. Machine learning for predicting greenhouse gas emissions from agricultural soils. *Sci. Total Environ.* 741, 140338. <https://doi.org/10.1016/j.scitotenv.2020.140338>.

- Hatami, N., Gavet, Y., Debayle, J., 2018, April. Classification of time-series images using deep convolutional neural networks. In: Tenth International Conference on Machine Vision (ICMV 2017, vol. 10696. International Society for Optics and Photonics, p. 106960Y. <https://doi.org/10.1117/12.2309486>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Jun, H., Shuai, L., Jiming, S., Yue, L., Jingwei, W., Peng, J., 2018. Facial expression recognition based on VGGNet convolutional neural network. In: 2018 Chinese Automation Congress (CAC). IEEE, pp. 4146–4151. <https://doi.org/10.1109/CAC.2018.8623238>.
- Kälän, U., Lang, N., Hug, C., Gessler, A., Wegner, J.D., 2019. Defoliation estimation of forest trees from ground-level images. *Rem. Sens. Environ.* 223, 143–153. <https://doi.org/10.1016/j.rse.2018.12.021>.
- Kopp, M., Tuo, Y., Disse, M., 2019. Fully automated snow depth measurements from time-lapse images applying a convolutional neural network. *Sci. Total Environ.* 697, 134213. <https://doi.org/10.1016/j.scitotenv.2019.134213>.
- Kow, P.Y., Wang, Y.S., Zhou, Y., Kao, I.F., Issermann, M., Chang, L.C., Chang, F.J., 2020. Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead PM_{2.5} forecasting. *J. Clean. Prod.* 261, 121285. <https://doi.org/10.1016/j.jclepro.2020.121285>.
- Kumar, A., Goyal, P., 2011. Forecasting of daily air quality index in Delhi. *Sci. Total Environ.* 409 (24), 5517–5523. <https://doi.org/10.1016/j.scitotenv.2011.08.069>.
- Li, J., Hu, Y., Liu, L., Wang, Q., Zeng, J., Chen, C., 2020. PM_{2.5} exposure perturbs lung microbiome and its metabolic profile in mice. *Sci. Total Environ.* 721, 137432. <https://doi.org/10.1016/j.scitotenv.2020.137432>.
- Li, W., Dong, R., Fu, H., Wang, J., Yu, L., Gong, P., 2020. Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping. *Remote Sens. Environ.* 237, 111563. <https://doi.org/10.1016/j.rse.2019.111563>.
- Li, Y., Huang, J., Luo, J., 2015, August. 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, pp. 1–5. <https://doi.org/10.1145/2808492.2808564>.
- Liu, C., Tsow, F., Zou, Y., Tao, N., 2016. Particle pollution estimation based on image analysis. *PLoS One* 11 (2), e0145955. <https://doi.org/10.1371/journal.pone.0145955>.
- Liu, G., Zhang, C., Xu, Q., Cheng, R., Song, Y., Yuan, X., Sun, J., 2020. I3D-Shufflenet based human action recognition. *Algorithms* 13 (11), 301. <https://doi.org/10.3390/a13110301>.
- Ma, J., Li, K., Han, Y., Yang, J., 2018, August. Image-based air pollution estimation using hybrid convolutional neural network. In: 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, pp. 471–476. <https://doi.org/10.1109/ICPR.2018.8546004>.
- Miao, W., Zeng, Z., Wang, C., Chen, Y., Song, C., 2020, May. Efficient and accurate classification enabled by a lightweight CNN. In: 2020 5th International Conference on Computer and Communication Systems (ICCCS). IEEE, pp. 989–992. <https://doi.org/10.1109/ICCCS49078.2020.9118411>.
- Milošević, D., Milošavljević, A., Predić, B., Medeiros, A.S., Savić-Zdravković, D., Piperac, M.S., et al., 2020. Application of deep learning in aquatic bioassessment: towards automated identification of non-biting midges. *Sci. Total Environ.* 711, 135160. <https://doi.org/10.1016/j.scitotenv.2019.135160>.
- Ngoc, D.D., Loisel, H., Jamet, C., Vantrepotte, V., Duforêt-Gaurier, L., Minh, C.D., Mangin, A., 2019. Coastal and inland water pixels extraction algorithm (WiPE) from spectral shape analysis and HSV transformation applied to Landsat 8 OLI and Sentinel-2 MSI. *Rem. Sens. Environ.* 223, 208–228. <https://doi.org/10.1016/j.rse.2019.01.024>.
- Persello, C., Tolpekin, V.A., Bergado, J.R., de By, R.A., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Rem. Sens. Environ.* 231, 111253. <https://doi.org/10.1016/j.rse.2019.111253>.
- Pyo, J., Duan, H., Baek, S., Kim, M.S., Jeon, T., Kwon, Y.S., et al., 2019. A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. *Remote Sens. Environ.* 233, 111350. <https://doi.org/10.1016/j.rse.2019.111350>.
- Qian, Y., Xing, W., Guan, X., Yang, T., Wu, H., 2020. Coupling Cellular Automata with Area Partitioning and Spatiotemporal Convolution for Dynamic Land Use Change Simulation. *Science of The Total Environment*, p. 137738. <https://doi.org/10.1016/j.scitotenv.2020.137738>.
- Rijal, N., Gutta, R.T., Cao, T., Lin, J., Bo, Q., Zhang, J., 2018, June. Ensemble of deep neural networks for estimating particulate matter from images. In: 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE, pp. 733–738. <https://doi.org/10.1109/ICIVC.2018.8492790>.
- Ruggieri, M., Plaia, A., 2012. An aggregate AQI: comparing different standardizations and introducing a variability index. *Sci. Total Environ.* 420, 263–272. <https://doi.org/10.1016/j.scitotenv.2011.09.019>.
- Soh, P.W., Chang, J.W., Huang, J.W., 2018. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *Ieee Access* 6, 38186–38199. <https://doi.org/10.1109/ACCESS.2018.2849820>.
- Sugiyama, T., Ueda, K., Seposo, X.T., Nakashima, A., Kinoshita, M., Matsumoto, H., et al., 2020. Health effects of PM_{2.5} sources on children's allergic and respiratory symptoms in Fukuoka, Japan. *Sci. Total Environ.* 709, 136023. <https://doi.org/10.1016/j.scitotenv.2019.136023>.
- Sural, S., Qian, G., Pramanik, S., 2002, September. Segmentation and histogram generation using the HSV color space for image retrieval. *Proc. Int. Conference Image Process. 2 (II-II)* <https://doi.org/10.1109/ICIP.2002.1040019>. IEEE.
- Tsai, S.S., Goggins, W.B., Chiu, H.F., Yang, C.Y., 2003. Evidence for an association between air pollution and daily stroke admissions in Kaohsiung, Taiwan. *Stroke* 34 (11), 2612–2616.
- Tseng, C.H., Tsuang, B.J., Chiang, C.J., Ku, K.C., Tseng, J.S., Yang, T.Y., et al., 2019. The relationship between air pollution and lung cancer in nonsmokers in Taiwan. *J. Thorac. Oncol.* 14 (5), 784–792.
- Vahdatpour, M.S., Sajedi, H., Ramezani, F., 2018. Air pollution forecasting from sky images with shallow and deep classifiers. *Earth Sci. India* 11 (3), 413–422. <https://doi.org/10.1007/s12145-018-0334-x>.
- Wang, B., Yan, Z., Lu, J., Zhang, G., Li, T., 2018, December. Deep multi-task learning for air quality prediction. In: International Conference on Neural Information Processing. Springer, Cham, pp. 93–103. https://doi.org/10.1007/978-3-030-04221-9_9.
- Wang, L., Guo, S., Huang, W., Qiao, Y., 2015 (08). Places205-vggnet models for scene recognition, pp. 1–2. arXiv preprint arXiv:1508.01667v1.
- Wang, N., Zhu, H., Guo, Y., Peng, C., 2018. The heterogeneous effect of democracy, political globalization, and urbanization on PM_{2.5} concentrations in G20 countries: evidence from panel quantile regression. *J. Clean. Prod.* 194, 54–68. <https://doi.org/10.1016/j.jclepro.2018.05.092>.
- Wang, Y.S., Chang, L.C., Chang, F.J., 2021. Explore regional PM_{2.5} features and compositions causing health effects in Taiwan. *Environ. Manag.* 67 (1), 176–191.
- Wu, Z., Shen, C., Van Den Hengel, A., 2019. Wider or deeper: revisiting the resnet model for visual recognition. *Pattern Recogn.* 90, 119–133. <https://doi.org/10.1016/j.patrec.2019.01.006>.
- Yu, S., Li, H., Li, X., Fu, Y., Liu, F., 2020. Classification of Pathogens by Raman Spectroscopy Combined with Generative Adversarial Networks. *Science of The Total Environment*, p. 138477. <https://doi.org/10.1016/j.scitotenv.2020.138477>.
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al., 2020. Deep learning in environmental remote sensing: achievements and challenges. *Remote Sens. Environ.* 241, 111716. <https://doi.org/10.1016/j.rse.2020.111716>.
- Zhang, C., Yan, J., Li, C., Rui, X., Liu, L., Bie, R., 2016, October. On estimating air pollution from photos using convolutional neural network. In: Proceedings of the 24th ACM International Conference on Multimedia, pp. 297–301.
- Zhang, C., Liu, B., Yan, J., Yan, J., Li, L., Zhang, D., et al., 2017, June. Hybrid measurement of air quality as a mobile service: an image based approach. In: 2017 IEEE International Conference on Web Services (ICWS). IEEE, pp. 853–856. <https://doi.org/10.1109/ICWS.2017.105>.
- Zhang, Q., Fu, F., Tian, R., 2020. A deep learning and image-based model for air quality estimation. *Sci. Total Environ.* 724, 138178. <https://doi.org/10.1016/j.scitotenv.2020.138178>.
- Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856.
- Zhang, Y., Shuai, C., Bian, J., Chen, X., Wu, Y., Shen, L., 2019. Socioeconomic factors of PM_{2.5} concentrations in 152 Chinese cities: decomposition analysis using LMDI. *J. Clean. Prod.* 218, 96–107. <https://doi.org/10.1016/j.jclepro.2019.01.322>.
- Zhao, K., He, T., Wu, S., Wang, S., Dai, B., Yang, Q., Lei, Y., 2019. Research on video classification method of key pollution sources based on deep learning. *J. Vis. Commun. Image Represent.* 59, 283–291.
- Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Rem. Sens. Environ.* 221, 430–443. <https://doi.org/10.1016/j.rse.2018.11.032>.
- Zhou, Y., Chang, F.J., Chang, L.C., Kao, I.F., Wang, Y.S., 2019. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. Clean. Prod.* 209, 134–145.
- Zhou, Y., Chang, L.C., Chang, F.J., 2020. Explore a Multivariate Bayesian Uncertainty Processor driven by artificial neural networks for probabilistic PM_{2.5} forecasting. *Sci. Total Environ.* 711, 134792.
- Zhou, L., Chen, X., Tian, X., 2018. The impact of fine particulate matter (PM_{2.5}) on China's agricultural production from 2001 to 2010. *J. Clean. Prod.* 178, 133–141. <https://doi.org/10.1016/j.jclepro.2017.12.204>.