



# A Shallow ResNet with Layer Enhancement for Image-Based Particle Pollution Estimation

Wenwen Yang, Jun Feng<sup>(✉)</sup>, Qirong Bo<sup>(✉)</sup>, Yixuan Yang,  
and Bo Jiang

Northwest University, Xian 710127, China  
{fengjun, boqirong}@nwu.edu.cn

**Abstract.** Airborne particle pollution especially matter with a diameter less than  $2.5\ \mu\text{m}$  (PM<sub>2.5</sub>) has become an increasingly serious problem and caused grave public health concerns. An easily and reliably accessible method to monitor the particles can greatly help raise public awareness and reduce harmful exposures. In this paper, we proposed a shallow ResNet with layer enhancement for PM<sub>2.5</sub> Index Estimation, called PMIE. An inter-layer weights discrimination of convolutional neural networks method is proposed, providing a meaningful reference for CNN's design. In addition, a new method for enhancing the effect of the convolution layer was first introduced and was applied under the guidance of the CNN inter-layer weights discrimination method we proposed. This shallow ResNet consists of seven residual blocks with last two layer enhancements. We assessed our method on two datasets collected from Shanghai City and Beijing City in China, and compared with the state-of-the-art. For Shanghai dataset, PMIE reduced RMSE by 11.8% and increased R-squared by 4.8%. For Beijing dataset, RMSE is reduced by 14.4% and R-squared is increased by 23.6%. The results demonstrated that the proposed method PMIE outperforming the state-of-the-art for PM<sub>2.5</sub> estimation.

**Keywords:** Particulate matter · Image enhancement · Shallow ReNet  
Layer enhancement

## 1 Introduction

Air pollution has become a serious issue globally and threaten public health. Many studies [1, 2] have shown that these pollutants especially fine particles with diameters less than  $2.5\ \mu\text{m}$  (PM<sub>2.5</sub>) has very complicated and harmful effect on human body more susceptible to respiratory diseases (such as asthma, emphysema, pneumonia, etc.), and also likely to increase cardiovascular and cerebrovascular diseases (such as ischemic heart disease, coronary heart disease, myocardial infarction, high blood pressure and cerebral infarction, etc.). Thus, how to measure and reduce the air pollution effectively becomes an important and practical problem.

Nowadays, smart phones and camera surveillances are widely available to obtain images, which together with the ever-increasing computational power for sophisticated image processing, provide a great opportunity to quality and analyze airborne particle

based images. The studies that have been reported in the literature can be divided into two categories: image-feature based approaches and deep learning approaches.

In image feature-base methods, Li *et al.* [3] proposed a method to estimate haze levels from images in image feature-base methods. They get two features, depth map and transmission matrix from haze images. And they use two features to estimate haze levels by statistical methods. Mao *et al.* [4] proposed a method by detecting numerical haze image by using the statics of various images and the atmospheric scattering model. And this method can estimate the haze factor from a single image. Liu *et al.* [5] first extracts 6 image features for each image, transmittance, overall and local image contrast, sky color and smoothness, and entropy, and two non-image features, solar zenith angle and humidity, and then applies principal components analysis (PCA) and Sequential Backward Feature Selection (SBFS) to optimize the feature set. Finally, creating a SVR model to predict PM<sub>2.5</sub> indices.

Recently, deep learning has become the state-of-the-art solution for solving typical computer vision problems. In CNN based methods, Zhang *et al.* [6] built a CNN and classify images. The CNN has 9 convolution layers, 2 pooling layers, and 2 dropout layers. And they solve vanishing gradient problem by using a modified rectified liner unit as the activation function. In order to adapt to air pollution problem, they also have to use a negative log-log ordinal classifier to replace softmax classifier. Chakma *et al.* [7] proposed method applies a VGG-16 CNN model for image-based PM<sub>2.5</sub> level analysis. The images are classified into three classes according to their PM<sub>2.5</sub> concentration levels based on two major transfer learning strategies, CNN fine-tuning and CNN features-based random forest. Bo *et al.* [8] first uses a Residual convolutional neural network (ResNet50) to predict the PM<sub>2.5</sub> index based on image information, and achieved the-state-of-the-art performance.

Compared with traditional image feature-based PM<sub>2.5</sub> analysis, deep learning-based approaches tends to achieve better results due to the simple preprocess and complete feature extraction. The existing networks such as VGG [9], Inception [10], ResNet50 [11] achieved great performance on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). However, the existing networks were designed for object recognition, the complexity of these networks make it more difficult to optimize and easy to get over-fitting for PM<sub>2.5</sub> estimation task.

In this paper, we explore the way for how to design a CNN model for air particle pollution estimation. A deep network with fewer layers is presented in this paper for PM<sub>2.5</sub> Index Estimation (PMIE). Our contributions are: (1) An inter-layer weights discrimination of convolutional neural networks method is proposed, providing a meaningful reference for CNN's design. (2) A shallow ResNet with layer enhancement is proposed, which not only improve the convergence speed in the training, but also improve the over-fitting performance. Meanwhile, the training time is greatly reduced due to the shallow network in the case of the same training epochs. (3) The proposed method PMIE achieves a good performance on the dataset [5, 8]. For Shanghai dataset, PMIE reduced RMSE by 11.8% and increased R-squared by 4.8%. For Beijing dataset, our method PMIE outperformed [8], which is reported based on ResNet50. RMSE is reduced by 14.4% and R-squared is increased by 23.6%.

## 2 Methodology

The complexity of these existing deep networks designed for object recognition made it more difficult to optimize and easy to get over-fitting in our task. Therefore, we proposed a shallow convolutional neural network with layer enhancement for PM Index Estimation, called PMIE. First, we proposed a network consist seven residual block, which is shallow compared to the Residual networks such as ResNet50, ResNet101. In addition, a new method for enhancing the effect of the convolution layer was first introduced and was applied behind the convolution layer that has obvious effect on the output. In our task, we add the layer enhancement following block six and seven. The flowchart was illustrated in the Fig. 1.

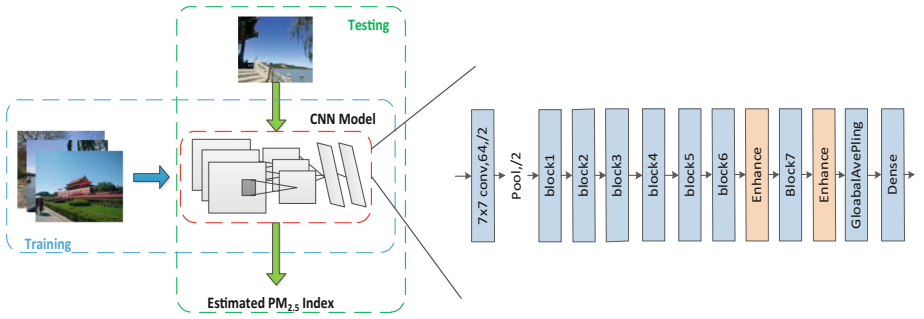
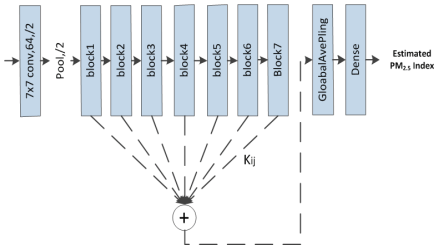


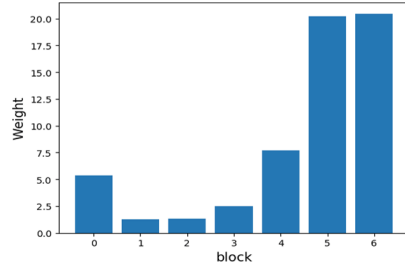
Fig. 1. Flowchart of the proposed method

### 2.1 Layer Weight Distribution Discrimination Method

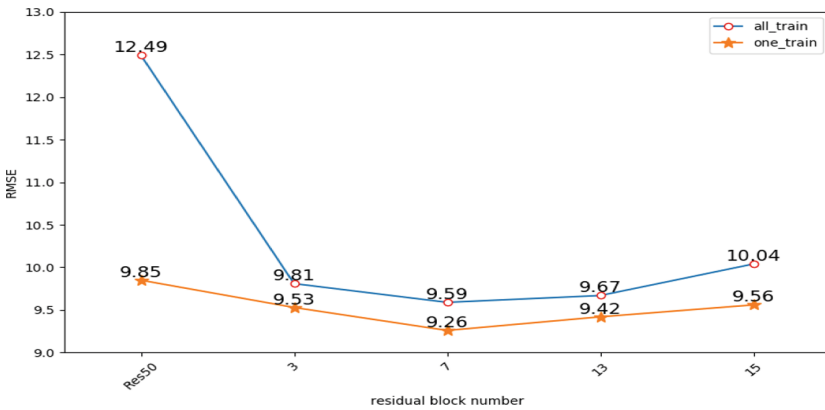
In the machine translation mechanism [12] of text deep learning, the weight of each word in the process of translation is not same, and more attention is paid to the core words. Similarly, the weight distribution of each layer is also different in a convolutional neural network. According to this consideration, we propose a CNN inter-layer weight distribution discriminant method. For a convolutional network, we assign a random weight  $K_{ij}$  to the output of each residual block and train  $K_{ij}$  by the back-propagation algorithm which is showed at Fig. 2. The specific approach is as follows: (1) Training basic convolutional neural networks; (2) Fixing the weight of the basic network, and assign a random weight  $K_{ij}$  to the output layer of each residual block,  $i$  represents the  $i^{\text{th}}$  residual block, and  $j$  represents the  $j^{\text{th}}$  feature map of this residual block; (3) Training the random weights; (4) Outputting the weight of each layer and seeing the distribution.



**Fig. 2.** Inter-layer weight distribution discrimination method



**Fig. 3.** Weight distribution results in Shanghai dataset.



**Fig. 4.** Different residual block number and train method results in Shanghai dataset

### 2.2 Shallow ResNet

Applying deep learning methods to images based PM<sub>2.5</sub> index Estimation is a challenging task. The existing CNN models are suitable for object recognition tasks, but our task is to explore whether the edge of the object is clear and whether the image texture is clear. The existing deep and large networks are difficult to train and easy to get over-fitting for PM<sub>2.5</sub> index estimation task. Therefore, we proposed a shallow ResNet with fewer layers compared to existing architecture. This architecture is presented in Fig. 1, which takes a square 224 \* 224 pixels RGB image as input and composed of one convolutional layers, one pooling layers, and seven residual blocks selected from ResNet-50 [11], the select of residual blocks number is from experience, the result is shown in Fig. 4.

Shallow architecture tends to learn low level features such as edges, lines, texture and colors. As the number of model layers deepens, the edges extracted by layers tend to be semantic and gradually change to the shape of objects. In our PM<sub>2.5</sub> estimation



Fig. 5. Layer enhancement samples

task, the focus is not to identify the object itself, but to identify whether the edges or lines of the object is clear, the shallow architecture is more suitable for our task.

### 2.3 Layer Enhancement

The Attention mechanism [13] was previously used in the task of text classification, and recently was widely used in object detection of images. In the object detection task, the initially selected ROI (region of interest) is given a higher weight value for more attention. Inspired by this, we proposed an enhance method, multiplying each weighted probability value learned by the convolutional layer by itself, so that the effects of activation and suppression are doubled, it also means image enhancement. At a convolution layer, the previous layer’s feature maps are convolved with learnable kernels and put through the activation function to form the output feature map. Based on Sect. 2.1 weight distribution results shown in Fig. 3, we add enhancements after residual block six and block seven. That is, each output map may combine convolutions with multiple input maps except that the output after residual block six and seven combine input map multiple with itself. In general, we have that

$$x_j^l = \begin{cases} f\left(\sum_{j \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right), & l = \text{others} \\ x_{ij}^{l-1} * x_{ij}^{l-1}, & l = 7, 9 \end{cases} \quad (1)$$

Where  $x^l$  represents the pixel of  $l^{\text{th}}$  block,  $M_j$  represents a selection of input maps. According to these observations, the proposed PMIE with enhancement suppresses the features that have little bit relationship with the output and strengthens those features with greater concern. Several examples of the enhancement are depicted in Fig. 5.

## 2.4 Training and Testing

The PMIE-model is trained by the back-propagation algorithm with batch stochastic gradient descend such that the mean squared error loss is minimized. RGB images from the training dataset are resized to  $224 * 224 * 3$  and fed to the PMIE-model for training. The observed PM2.5 index of each training image is used to calculate MSE loss. Fine-tuning was adopted due to the limited dataset was not large enough to train the full CNNs. There are two possible approaches of performing fine-tuning in a pretrained network: The first one is to fine-tune all the layers of the CNNs, the other approach is to keep some of the earlier layers fixed and fine-tune higher level layers of the network. In this paper, we fine-tuned all layers using the parameters learned from ImageNet datasets.

Training is done per epoch, the CNN parameters updated based on the best results on the validation set. After training, the testing dataset were fed to the trained model and get the predicted PM2.5 index.

## 3 Experiments

### 3.1 Dataset

We present the PM2.5 prediction task on two images datasets: Shanghai dataset and Beijing dataset. (1) Shanghai dataset is a single-scene dataset [6]. This dataset contains 1885 pictures captured at the Oriental Pearl Tower in Shanghai city, China, and contains different capture times from May to December of 2014. (2) Beijing dataset is a non-single scene dataset that contains 1514 pictures collect from Beijing tourist website by ourselves [14]. These pictures were captured at diverse locations in Beijing City, China.

The U.S. consulate in Beijing and a Shanghai provided PM2.5 indices hourly, we used these to retrieve the PM2.5 indices of two datasets. Figure 6 shows the histogram distribution of the PM2.5 index of two datasets.

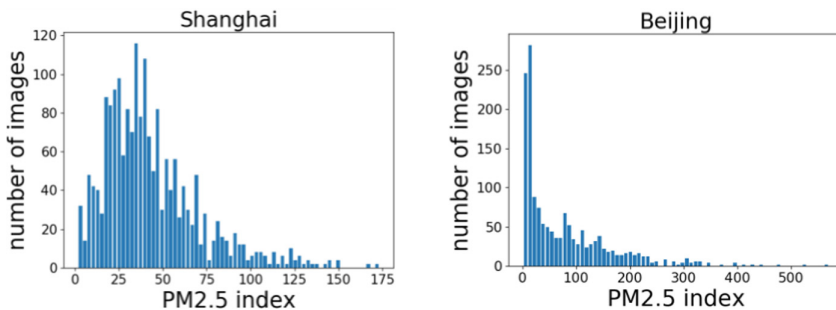


Fig. 6. The histograms of PM2.5 of Shanghai dataset and Beijing dataset.

### 3.2 Evaluation Protocols

We use mean squared error (RSME) and R-squared to evaluate the error of prediction. RSME is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y'_i)^2}, \quad (2)$$

Where  $y'_i$  is the  $i^{\text{th}}$  forecast value, and  $y_i$  is the  $i^{\text{th}}$  observed value,  $i = 1, 2, \dots, N$ . And R-squared is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - y'_i)^2}{\sum_{i=1}^N (y_i - \text{avg}(y))^2}, \quad (3)$$

Where  $y'_i$  is the  $i^{\text{th}}$  forecast value,  $\text{avg}(y)$  is the average forecast value, and  $y_i$  is the  $i^{\text{th}}$  observed value,  $i = 1, 2, \dots$ , R-squared increases with the agreement between the observed value and the forecast value with a maximum value of 1, which indicates the best match.

### 3.3 Experiment Setting

In order to train and evaluate our PMIE network, we randomly select 80% images as training set and tune the CNN, 10% images as validation set, and 10% are used for testing. To fine-tune the model, we loaded the first convolutional and the earlier seven residual block weights from ResNet-50. After loading the pretrained weights, we adjusted all the parameters in PMIE networks to fit our goal.

For the PMIE network training parameters, we ran the code for 500 iterations and set the learning rate to a very small variations of 0.001. The batch size is 64 and the momentum is assigned to 0.9. The program was implemented using keras 2.05 in Ubuntu 16.04.

### 3.4 Experiments Result

In this study, we measure the performance of the proposed method using evaluation protocol described in Sect. 3.2. Table 1 presents the results of PMIE on two datasets with the RSME and R-squared values with comparison results in [6, 8]. In order to better compare our experiments, we also joined the VGG16 network as a comparison.

It is clear that the deep learning method such as ResNet and VGGnet achieved better result than traditional image feature extraction method like literate [6]. Since our network is improved on ResNet50, our focus is on comparing our PMIE methods with ResNet50 reported literate [8]. We can see that our method performed better in the same dataset of Beijing and Shanghai.

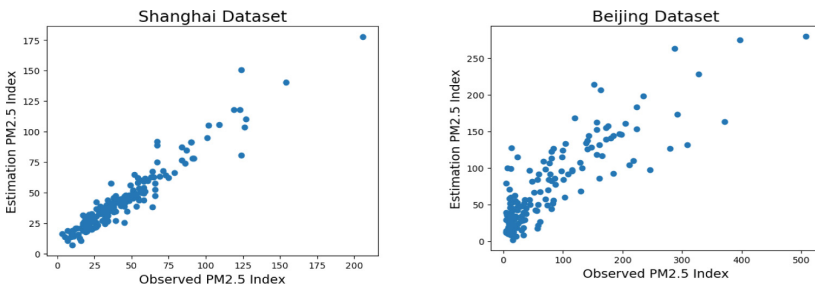
For Shanghai dataset, the RMSE of our PMIE method is reduced by 11.86% and R-squared increased by 4.59% in Shanghai dataset with ResNet50 from literate [8]. For Beijing dataset, the RMSE and R-squared values of PMIE are 50.64 and 0.68, reduced by 14.38% and increase by 23.63% respectively. Besides, compared with the traditional resnet50 network, the training time of PMIE is greatly reduced.

Figure 7 shows the correlations between the estimated PM2.5 indices and observed PM2.5 indices of Shanghai dataset and Beijing dataset.

Figure 8 shows the training and validation loss of our PMIE and ResNet50 on two data sets. We can see that the method we proposed overall performs better on validation sets and better to overcome overfitting challenges. In addition, in non-single scene Beijing datasets, our method converges faster during training, and it performs steadier during training and validating.

**Table 1.** PMIE performance vs other networks

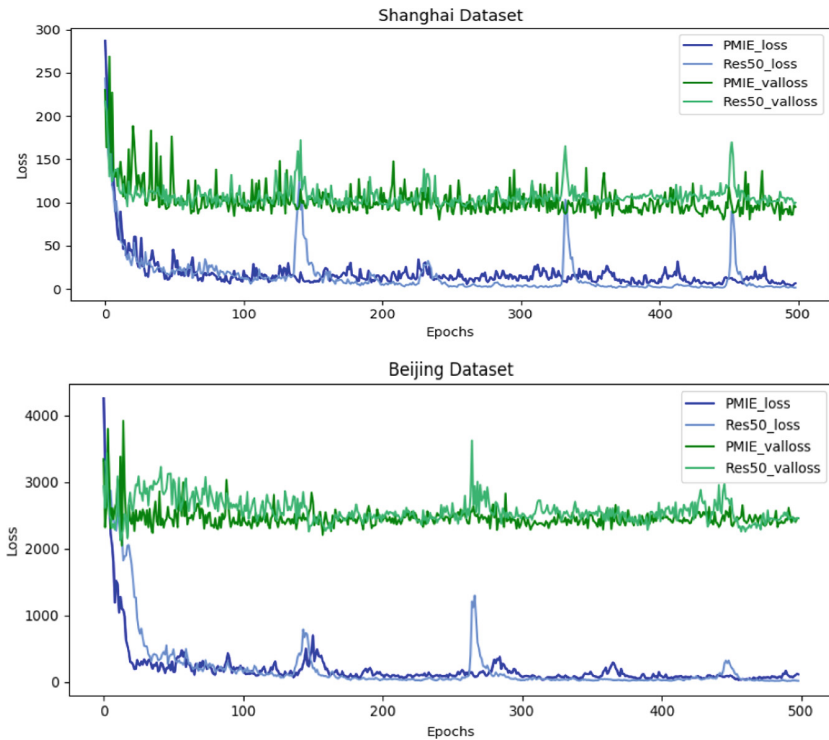
Dataset	Method	RMSE	R-squared	Run time
Shanghai	[6]	19.23	0.57	–
	VGG16	11.26	0.84	3 h 2 m 20 s
	ResNet50 [8]	10.11	0.87	3 h 8 m 50 s
	<b>PMIE</b>	<b>8.91</b>	<b>0.91</b>	<b>2 h 7 m 56 s</b>
Beijing	VGG16	60.67	0.52	2 h 28 m 23 s
	ResNet50 [8]	59.15	0.55	2 h 31 m 30 s
	<b>PMIE</b>	<b>50.64</b>	<b>0.68</b>	<b>1 h 40 m 7 s</b>



**Fig. 7.** The correlation of estimated and observed PM2.5 index.

Figure 9 shows images with their observed and estimated PM2.5 indices. The first two rows pictures are from Shanghai dataset and the last two rows are from Beijing dataset. The first row and the third row are pictures with accurate prediction, the second





**Fig. 8.** Train loss and validation loss of proposed PMIE and ResNe50 for Shanghai and Beijing datasets

row and the last row are pictures with inaccurate prediction. By analyzing the dataset, we find there are some reasons for inaccurate prediction. We can see the images from second and last looks different between human visual observation and image labels. For example, the 1<sup>st</sup> image in last row looks very clear but its actual PM2.5 index is 80, the 3<sup>rd</sup> picture in the last row has larger estimated index than observed index because of its gray hue. In addition, lack of high PM2.5 images in dataset also resulted in the inaccurate prediction on high PM2.5 images. One the other hand, it shows that our algorithm is more accurate for most image predictions except these pictures does not match subjective visual bringing lager errors.

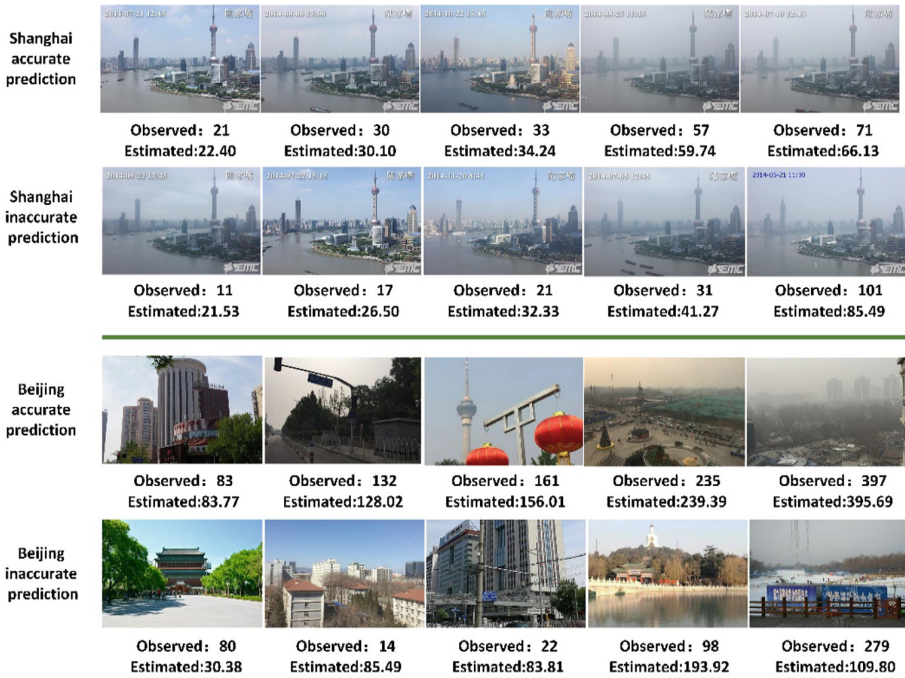


Fig. 9. Example images with their observed and estimated PM2.5

## 4 Conclusion

In this paper, we proposed a PMIE network that using residual block stacking with enhancement to estimate PM2.5 from images. Our main findings are that shallow CNN model and convolutional layer with enhancement provide better performance than typical deep CNN architecture. We also studied the performance for training and validating loss. The results on single scene Shanghai dataset and Non-single scene Beijing dataset outperforming the state-of-the-art.

**Acknowledgement.** The work in this paper is support by National Natural Science Foundation of China (No. 41601353), Shaanxi Provincial Natural Science Research Project (2017KW-010) and Shaanxi Provincial Department of Education Science Research Project (15JK1689).

## References

1. Chow, J., et al.: Health effects of fine particulate air pollution: lines that connect. *Air Repair* **56**(6), 709 (2006)
2. Mcginnis, J.M., Foegen, W.H.: Actual causes of death in the United States. *JAMA, J. Am. Med. Assoc.* **291**(10), 1238–1245 (1993)

3. Li, Y., Huang, J., Luo, J.: Using user generated online photos to estimate and monitor air pollution in major cities. In: International Conference on Internet Multimedia Computing and Service (2015)
4. Mao, J.: Detecting foggy images and estimating the haze degree factor. *J. Comput. Sci. Syst. Biol.* **7**(6), 1 (2014)
5. Liu, C., et al.: Particle pollution estimation based on image analysis. *PLoS ONE* **11**(2), e0145955 (2016)
6. Zhang, C., et al.: On estimating air pollution from photos using convolutional neural network. In: ACM on Multimedia Conference (2016)
7. Chakma, A., Vizena, B., Cao, T., Lin, J., Zhang, J.: Image-based air quality analysis using deep convolutional neural network. In: IEEE International Conference on Image Processing (2017)
8. Bo, Q., Yang, W., Rijal, N., Xie, Y., Feng, J., Zhang, J.: Particle pollution estimation from images using convolutional neural network and weather feature. In: IEEE International Conference on Image Processing (2018)
9. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(4), 640–651 (2014)
10. Szegedy, C., et al.: Rethinking the inception architecture for computer vision, pp. 2818–2826 (2015). Computer Science
11. He, K., et al.: Deep residual learning for image recognition, pp. 770–778 (2015)
12. Andrychowicz, M., Kurach, K.: Learning efficient algorithms with hierarchical attentive memory (2016)
13. Vaswani, A., et al.: Attention is all you need (2017)
14. <https://goo.gl/F1tkM4>