IMAGE-BASED AIR QUALITY ANALYSIS USING DEEP CONVOLUTIONAL NEURAL NETWORK

Avijoy Chakma[1], Ben Vizena[1], Tingting Cao[2], Jerry Lin[2], Jing Zhang[1]

[1]Computer Science Department, Lamar University, TX, 77710, USA
[2]Center of Advances in Water and Air Quality, Lamar University, TX, 77710, USA

ABSTRACT

Air pollution may cause many severe diseases. An efficient air quality monitoring system is of great benefit for human health and air pollution control. In this paper, we study image-based air quality analysis, in particular, the concentration estimation of particulate matter with diameters less than 2.5 micrometers ($PM_{2.5}$). The proposed method uses a deep Convolutional Neural Network (CNN) to classify natural images into different categories based on their $PM_{2.5}$ concentrations. In order to evaluate the proposed method, we created a dataset that contains total 591 images taken in Beijing with corresponding $PM_{2.5}$ concentrations. The experimental results demonstrate that our method are valid for image-based $PM_{2.5}$ concentration estimation.

***Index Terms*—** Particulate Matter, Air Quality Analysis, Deep Convolutional Neural Networks, Image Classification

1. INTRODUCTION

Air pollution has become an alarming environmental issue globally due to rapid urbanization and industrialization. Among different air pollutants, airborne particulate matter (PM) with diameters less than 2.5 micrometers ($PM_{2.5}$) has significant harmful effects on the human body as these particles are capable of transmitting hazardous chemicals into the human lung and blood and cause cardiovascular, respiratory and cerebrovascular diseases, reduced lung functions, and heart attacks. Therefore, $PM_{2.5}$ concentration has been used as a worldwide major air quality metric. Currently, air quality monitoring methods are mainly based on monitoring stations, which are not available to the majority of regions because of the high setup cost and expensive sophisticated sensors.

With the increasing availability of portable cameras and smart phones, images play a more and more important role in information representation and description. If air quality metrics, such as $PM_{2.5}$ and Air Quality Index (AQI), can be estimated by analyzing photo images, it will provide an efficient and affordable way to monitor air quality. For example, smartphone users can take photos and estimate the real-time local air quality by themselves.

In recent years, using computer vision and machine learning techniques to analyze the air quality (e.g., haze level and $PM_{2.5}$ concentration) has begun to receive increasing attention and a few papers can be found in the literature.

The existing image-based haze level analysis methods are mainly inspired by the dehazing algorithms [1, 2]. In computer vision and computer graphics areas, the haze image can be modeled as:

$$I(x) = J(x)t(x) + A(1 - t(x)), \quad (1)$$

where I is the observed image, J is the haze-free scene image, A is the sky luminance that indicates the lighting condition, and t is transmission matrix that describes the portion of the light that reach the camera. The transmission matrix *t(x)* can be expressed as

$$t(x) = e^{-\beta d(x)} \quad (2)$$

where *d(x)* is the scene depth map and β is the atmosphere scattering coefficient, which is related to the air quality (e.g., haze level) [1]. Therefore, the transmission matrix *t(x)* and the depth map *d(x)* are important image features for air quality analysis.

Li *et al.* [3] proposed a method to estimate haze level using different pooling and transformation functions based on depth and transmission information extracted from haze images. Firstly, the depth map and transmission matrix of the haze image are generated by a Deep Convolutional Neural Fields-based approach proposed in [4] and a Dark Channel Prior-based dehazing algorithm proposed in [1], respectively. After that, different transformation, bivariate, and pooling functions are applied on both transmission and depth matrices to estimate haze level statistically. The performance is evaluated using absolute Spearman correlation coefficients. Although this method performs well with synthetic images, the correlation on a real image dataset with thousands images is 40.83%.

Liu *et al.* proposed a method based on support vector regression after extracting 6 different image features [5]. The authors adopted Dark Channel Prior for image transmission computation and analyzed image contrast and entropy that may be affected by $PM_{2.5}$ significantly. They also observed the effect of $PM_{2.5}$ by analyzing weather conditions, the color of sky region, and the location of sun during the image taken. A support vector regression model was developed to predict the $PM_{2.5}$ using all of these features. The method was assessed using clear and cloudy photos taken at three fix locations in Beijing, Shanghai, and Phoenix with manually marked reference regions with different depths. The proposed method achieves good performance for the images from Beijing and Shanghai but fails for the images from Phoenix due to the narrow region of $PM_{2.5}$ values.

CNN techniques have been extensively used in image processing and computer vision research areas since 2012 and have achieved many breakthroughs and plausible performance in several challenging tasks, such as object detection and recognition, semantic segmentation, and image classification [6, 7]. Recently, CNNs have been applied to weather classification. In [8], Karayev *et al.* used CNN to recognize image styles using Flickr style dataset, which has 80,000 images labeled with 20 different visual styles, including two atmosphere styles: sunny and hazy. In [9], Elhoseiny *et al.* adopted CNN to classify images into sunny weather and cloudy weather using 10,000 images. For air pollution ($PM_{2.5}$) analysis, Li *et al.* [10] proposed a spatiotemporal deep learning method driven on data from 12 air quality monitoring stations in Beijing city. The created model preserves both spatial correlation and temporal correlation by using the current data collected and the past interval air quality data from all stations as the input. The stacked autoencoder is used to extract inherent air quality feature in combination with logistic regression on top of it to provide real-value air quality. However, the proposed method is dependent on the air quality data from stations equipped with Thermo Fisher detectors and the data are from non-image resources.

In this paper, motivated by the remarkable success of CNNs, we propose the first method that applies CNN for air quality analysis, in particular, image-based $PM_{2.5}$ concentration estimation for any locations. The RGB color channels of the image are input to the deep learning network to estimate $PM_{2.5}$ concentration directly, which overcomes the restrictions of having sensors, auxiliary data, fixed location, or sophisticated image features. In addition to the proposed method, we also created a dataset of 591 images from Beijing city with associated $PM_{2.5}$ data.

2. METHODOLOGY

As discussed in Section 1, the existing image-based haze level analysis methods usually need to use sophisticated algorithms to extract the transmission matrix and the depth map from the single input image for haze level estimation, which not only increases the computation complexity but also requires accurate transmission and depth information. As a simple and explicit end-to-end architecture, the CNN can extract both low-level and high-level image features automatically and greatly simplify image analysis process. In this section we will introduce the proposed CNN transfer learning-based $PM_{2.5}$ concentration estimation method.

2.1. CNN Architecture

The adopted CNN is *imagenet-matconvnet-verydeep* model. This model achieved very high performance on ImageNet dataset, which has over 15 million high resolution images from 1000 categories. This CNN model has 8 convolutional layers with max pooling and Relu and the last three are full connection layers (FC6, FC7 and FC8). The output of the last FC layer is input to the softmax to produce the classification results. The input to the CNN is the RGB three color channels of an image with size 224×224.

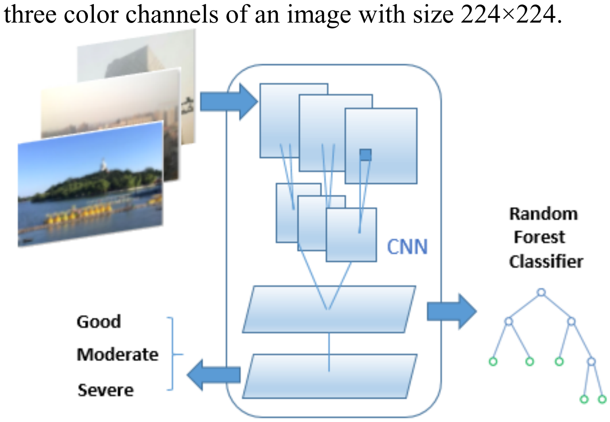


Fig. 1. CNN for $PM_{2.5}$ concentration estimation and two transfer learning methods, CNN fine-tuning and random forest classification using image features extracted by CNN.

2.1. Transfer Learning

One impressive phenomenon of many deep neural networks is that the specific trained CNNs are applicable to quite different datasets and tasks by transferring knowledge learned in one area to another area and can achieve surprisingly good results. This transferability enables the applications of CNNs to research tasks without very large dataset. Typically, there are two major strategies for transfer learning: (1) Fine-tuning the CNN. The weights of the pre-trained CNN can be fine-tuned using new datasets for new tasks. (2) Using features extracted by CNNs. The CNN is considered as a feature extractor. The features extracted by the CNN are used as the inputs to other classifiers.

In our study, we downloaded the *imagenet-matconvnet-verydeep* model pre-trained using ImageNet dataset from MatConvnet[1] and used both transfer-learning strategies in the proposed method. Figure 1 illustrates the CNN architecture and two transfer learning methods for $PM_{2.5}$ concentration estimation.

[1] http://www.vlfeat.org/matconvnet/pretrained/

(1) For CNN fine-tuning, the first fifteen layers of CNN architecture used in this study follow the CNN-layer specifications proposed in [11]. To reduce overfitting, one dropout layer is added before the last full connection layer. The 1000 nodes in the FC8 are replaced with three nodes for three difference $PM_{2.5}$ concentration levels (Good, Moderate, and Severe). The parameters of the first fifteen layers are initialized using pre-trained weights and the parameters of the FC8 are initialized randomly. During the training, the parameters of all fully connected layers are adjusted by the back-propagation algorithm with batch stochastic gradient descend. The multinomial logistic regression objective is used as the loss function. The softmax loss of a training image x is defined as $loss(x, l)$ with $l \in$ {Good, Moderate, Severe}.

(2) For CNN-based feature extraction, we use the image features extracted by FC8 of CNN to train a random forest classifier. The output of FC8 is 4096-dimensional feature vector extracted from each image by the multiple convolutional filters. The designed random forest classifier contains 500 trees and each tree is trained using 20 randomly selected features.

4. EXPERIMENTS

In this section, we will introduce the image dataset, training, and testing process, and discuss the experimental results.

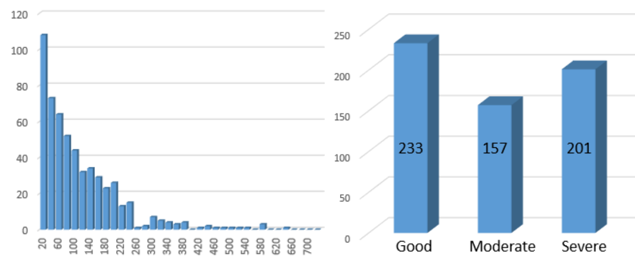
4.1. Dataset

Because no $PM_{2.5}$ concentration image dataset publicly available, we created a dataset that has 591 images collected from a Beijing tourist website[2], which manages archives of Beijing real-time weather photos and other relevant information, including the snapshot time, location, and weather condition. The corresponding $PM_{2.5}$ value of each image was retrieved using the hourly $PM_{2.5}$ historical data provided by the U.S. embassy in Beijing[3] based on the image taken time and date. The collected images were taken under different weather conditions at different times, different locations, and in different seasons from the year 2013 to 2017. All image resolutions are higher than 500×500. Meanwhile, each selected image contains sky region and objects with different depths to ensure that the image has enough information to reflect the air quality (e.g., $PM_{2.5}$ values) accurately. The range of $PM_{2.5}$ concentrations is from 2 to 634 and we separated the images into three class: Good (<75), Moderate (75~115), and Severe (>115). Figure 2-a shows the histogram of the $PM_{2.5}$ concentrations of 591 images and Figure 2-b shows the number of images in each class. This dataset can be publicly downloaded from the website of our image-based air quality analysis project[4].

[2] http://www.tour-beijing.com/real_time_weather_photo/

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

[4] galaxy.cs.lamar.edu/~jingz/air_quality



(a) (b)

Fig. 2. (a) Histogram of $PM_{2.5}$ concentrations and (b) The numbers of images in Good, Moderate and Severe classes.

4.2. Training and Testing

In order to augment the training images, we divided the image into two equal sub-images and flip each sub-image in left-right direction. Note that we did not flip image in up-down direction because the sky region should be at the top part of the image. This augmentation step produces total 591×4=2364 images for CNN training and testing. Each image is rescaled to 224×224 to match the input size of the 1st layer of CNN.

We randomly select 20% images from each class as test set and the other 80% images are further divided into training set (90%) that is used to train and tune the CNN and validation set (10%) that is used to optimize the setting of the CNN. The CNN settings used are as follows: The batch size is 100 images and the number of epochs is 60. The learning rate decay is set in the log space from -3 to -5.5. The momentum is assigned to 0.9.

The program was implemented using MATLAB R2016b. *Matconvnet-1.0-beta22* package was used for CNN-based classification and *TreeBagger* and *predict* functions were used for random forest-based classification.

4.3. Experimental Results

To comprehensively evaluate the performance of the CNN, we tested the CNN and the random forest classifier five times with randomly split training, validation, and test sets.

In this study, we measure the performance of the proposed method using the evaluation metric 'Accuracy', which is defined as the number of correctly classified images divided by the number of all images.

The average classification accuracies of the CNN and the random forest classifier are list in Table 1.

Table 1. The performance of CNN and Random Forest

Methods	Classification Accuracy (Average)
Random Forest	63.62%
CNN	68.74%

We can see that the classification accuracy of CNN is approximately 5% better than that of the random forest classifier. Both method achieved similar performance, because the features fed to the random forest are from the last full connection layer of CNN.

Figure 3 shows some images that were correctly classified by the CNN. Each column illustrates 10 images from each class, Good, Moderate, and Severe.

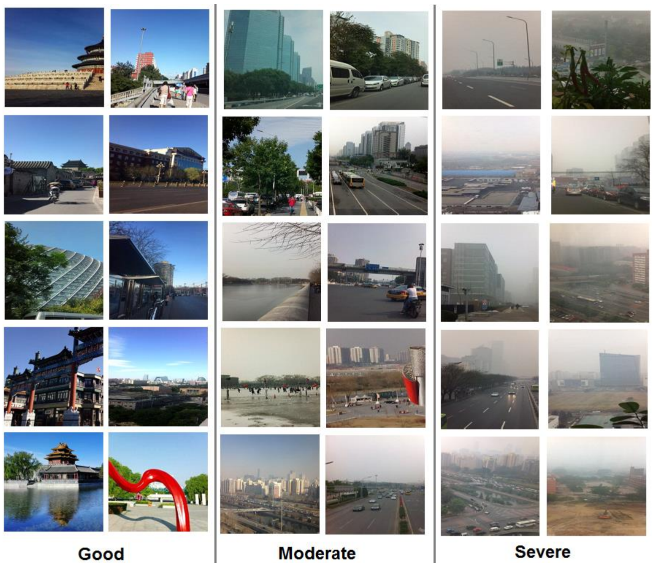


Fig. 3. Example images correctly classified by the CNN

4.4. Discussion

We compared our method with two related studies. In [3], Li *et al.* proposed a method for haze level estimation using transmission and depth matrices. Two datasets were used for evaluation purpose by using $PM_{2.5}$ as a proxy. PM25 dataset has 8,761 photos and the reported Absolute Spearman correlation is 40.83%. PM25-s dataset has three classes, NonHaze, LightHaze, and HeavyHaze and the reported correlation is 89.05%, however only 46 manually selected photos in PM25-s dataset. Compared with this state-of-the-art work [3], our study tested 2364 real images from 3 classes with the average accuracy 68.74%. Although different evaluation metrics were used due to the differences between the algorithms used, our method can achieve comparable or better performance for $PM_{2.5}$ estimation.

In [8], Karayev *et al.* used a pre-trained imageNet-CNN for weather classification. The network was fine-tuned on 10,000 images with two classes, sunny and cloudy. The best regular accuracy is 91.1%. Two main reasons that their method achieved better performance than our method are (1) weather classification task has only two classes, while our task has three classes; and (2) much more training weather images were used to fine tune the pre-trained CNN.

The $PM_{2.5}$ concentration estimation task is quite different from other tasks that classify, segment, or recognize objects based on shape, texture, or color. It needs high-level understanding of images, therefore it is a very challenging research topic. The performance demonstrated that our method is valid for $PM_{2.5}$ concentration estimation and we believe there are many promising ways to further improve the classification accuracy. For example, collecting more images with $PM_{2.5}$ concentrations and fine-tuning other CNN models, such as DAG models and RCNN models.

5. CONCLUSION

We introduce the first work in the literature that uses CNN-based method to estimate $PM_{2.5}$ concentration for natural images in this paper. Two transfer learning methods, CNN fine tuning and CNN feature-based Random Forest, are used to classify the images into 3 classes according to their $PM_{2.5}$ concentrations. The experimental results demonstrated that our method is valid for $PM_{2.5}$ concentration estimation.

Besides the proposed method, we also created a $PM_{2.5}$ concentration image dataset, which currently contains 591 images with corresponding $PM_{2.5}$ values. This dataset is publicly downloadable and other researchers working in this area can use this dataset to test and compare their methods.

Now we are collecting more images taken in 2017 to increase the dataset size. Our future work includes collecting more $PM_{2.5}$ images from different environments (e.g., cities, rural areas, national parks) and testing different CNN models to improve the classification accuracy.

6. REFERENCES

[1] He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, *33*(12), 2341-2353.

[2] Fattal, R. (2008). Single image dehazing. *ACM transactions on graphics (TOG)*, *27*(3), 72.

[3] 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* (p. 79). ACM.

[4] Liu, F., Shen, C., & Lin, G. (2015). Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE Conference on CVPR* (pp. 5162-5170).

[5] Liu, C., Tsow, F., Zou, Y., & Tao, N. (2016). Particle Pollution Estimation Based on Image Analysis. *PloS one*, *11*(2), e0145955.

[6] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional models for semantic segmentation. In *CVPR*

[7] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Torr, P. H. (2015). Conditional random fields as recurrent neural networks. In *ICCV* (pp. 1529-1537).

[8] Karayev, Sergey, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. "Recognizing image style." *arXiv preprint arXiv:1311.3715* (2013).

[9] Elhoseiny, M., Huang, S., & Elgammal, A. (2015, September). Weather classification with deep convolutional neural networks. In *Proceedings of the IEEE ICIP* (pp. 3349-3353).

[10] Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, *23*(22), 22408-22417.

[11] Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*.