Leveraging Unsupervised Haze Region Segmentation and Class-Correlated Deep Learning for Visual Haze Severity Estimation

Michael Shell, *Member, IEEE,* John Doe, *Fellow, OSA,* and Jane Doe, *Life Fellow, IEEE*

***Abstract*—Visual haze severity estimation via camera photos has gained much attention recently as a low-cost solution for air quality evaluation. This task requires efforts in terms of collecting photos of haze weather annotated with air quality measures, extracting representative visual features of haze, and developing statistical and machine learning algorithms for haze severity level/value prediction. Existing efforts typically focus on leveraging pixel-level features and depth information of photos and state-of-the-art supervised learning algorithms, including deep learning. However, two challenges have not been touched. First, existing approaches usually achieve downgrade predictive performance on light-haze photos, because light haze can hardly hide scene objects and thus the extracted features may be similar to those from non-haze photos due to the diverse scene objects.**

the small-scale training data, and the cost-expensive feature extraction for haze. In this paper, we figure out that the downgrade performance is mainly resulted by the diverse scene objects in the light-haze photos. To address these problems, we propose a fast and fully unsupervised algorithm, called intensity-invariant contrast-based haze region segmentation (IC-HRS), to detect haze/cloud regions for extracting effective visual features of haze avoiding the noise from background scenes. More importantly, IC-HRS simplifies the prediction of haze severities as two subtasks, i.e. a binary haze/cloud classification and a haze level classification. It also makes the classifiers be effectively trained on rich patches of cloud/haze regions, instead of a small-scale photo set. Moreover, we further propose a degree-progression multi-task convolutional neural network (DM-CNN) designed for robust haze level classification by incorporating the increasing degrees of haze between haze level classes as task relatedness. Experiments on a publicly available dataset and our collected real-world phone photo dataset demonstrate the impressive performance of our proposed approach on haze severity estimation.

***Index Terms*—Unsupervised haze region segmentation, Class-correlated deep learning, Haze/cloud bi-classification, Patch-based sample enrichment.**

I. Introduction

Particulate matters in the air is now a severe global issue challenging people' health, especially children, and the aging population [1]. According to the statement of World Health Organization (WHO) updated by September, 2016[1], 92% of the world population was living in places where the WHO air quality guidelines levels were not met; particle matter (PM), notably the $PM_{2.5}$ and PM_{10}, is the major pollutant, having a close relationship with mortality or morbidity. Therefore, knowing the ambient air quality becomes an important step in people's daily activity arrangement. This raises the need for easy-access mobile air quality measurements, allowing people to check the ambient air quality, especially the haze severity levels, by themselves at any time and place in their daily life.

Besides evaluating local air quality from the nearest air quality measure tower or purchasing mobile sensor devices [2], [3], an emerging approach is to explore the use of digital photos of daily living captured by phones or surveillance cameras for estimating air quality, or more specifically the haze severity, considering that high PM levels result in observable haze in the sky [4], [5], [6], [7], [8], [9], [10]. As an example, HackAIR[2] [11], an EU-funded project starting from 2016, is one of recent actions towards this goal. This solution has been deemed to be easy access by everyone and low cost in the resource, taking advantages of crowdsourcing platform, rich haze photos on the Web, and advanced machine learning techniques.

Existing related studies in the literature typically follow three key procedures, including acquiring adequate non-haze and haze photos with accurate air quality measure values/levels, such as AQI and $PM_{2.5}$, for preparing training and testing data, developing algorithms for extracting visual features of haze information from photos, and developing predictive algorithms for air quality estimation. Despite different photo resources and algorithms adopted, these studies usually face three issues:

1) **Downgrade performance on light-haze photos** is usually observed, which are easy to be predicted to have lighter or no haze [10], [7], [8]. This is mainly because light haze can hardly hide scenes objects behind while commonly-used features in the literature, including dark channel, light transmission map, deep neural network features, and image content features (color/texture/contrast), are sensitive to scene objects. It makes the photos similar to the non-haze ones at the feature level.
2) **No dataset of real-world haze photos** has been published for haze severity estimation. Therefore, existing studies [4], [5], [6], [7], [8], [9], [10] usually conduct experiments on small-scale datasets of less than 2,000 photos either collected online or taken by the authors. These datasets can hardly cover the whole value range of

M. Shell was with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA e-mail: (see http://www.michaelshell.org/contact.html).

J. Doe and J. Doe are with Anonymous University.

Manuscript received April 19, 2005; revised August 26, 2015.

[1]http://www.who.int/mediacentre/factsheets/fs313/en/

[2]http://www.hackair.eu/

air quality measures, such as AQI and PSI, thus usually leading to the unsatisfied performance.

3) **Cost-expensive feature extraction** has always be applied for haze severity prediction. The most widely-recognized effective features either are computationally expensive, such as dark channel and light transmission map [9], [5], [10], [6], [7], or require preparing external training data and classifiers, such as depth map [9], [5] and features from sky region [10], [7].

In this paper, we propose a cascaded classification framework (Figure 1) with two key techniques, including the intensity-invariant contrast-based haze region segmentation (IC-HRS) algorithm and the degree-progression multi-task convolutional neural network (DM-CNN) classification model, for a practical solution to fast haze severity estimation from a single photo of daily living.

IC-HRS is the core of this framework, which incorporates the smoothness and intensity-invariant properties of haze to detect the cloud and haze regions of a photo by using solely the photo in a fully unsupervised manner. Besides, IC-HRS can even stop after scanning through the haze regions without traversing the entire photo by incorporating our proposed heuristic scanning strategies. These characteristics make IC-HRS different from existing approaches that extract features either from the entire photo [9], [5], [10], [6], [7] or using external resources to detect depth map or sky region [9], [5], [10], [7].

More importantly, patches of the segmented cloud and haze regions make the task of haze severity estimation much simplified. That is, we could differentiate the non-haze from haze photos through the difference in cloud and haze, and distinguish haze in different severity levels through their fine-grained characteristics. To achieve this goal, we design a seven-layer convolutional neural network (CNN) architecture and build two classifiers to perform the bi-class cloud/haze classification and the multi-class haze severity level classification, respectively. This two-step classification method also has advantages of simplifying the learning complexity of the second classifier by making it focus on recognizing fine-grained features among different haze severity levels rather than that of different types of cloud, such as cloudy and overcast.

Moreover, we further improve the CNN classifier for multi-class haze severity level classification by the proposed DM-CNN, which considers the increasing degrees of haze between haze level classes as task relatedness and incorporates a pairwise constraint in the objective function to guide the prediction process so that classes farther from the true class have lower prediction values.

We conducted experiments on a published dataset of 3,024 synthetic images of different haze levels used in [4], [5] and a self-collected dataset of 1,811 real-world phone photos. The effectiveness and efficiency of the IC-HRS algorithm, the cloud/haze patch-based training method, the DM-CNN classifier, and the entire cascaded classification model has been evaluated and compared with existing approaches. We observed an impressive performance of IC-HRS in both segmentation results and time cost compared with existing supervised methods and found a significant improvement in classification performance using the rich patches of cloud/haze regions than using the photos. Besides, DM-CNN consistently improves the performance of CNN in terms of classification performance. The overall performance of the proposed cascaded classification framework achieves superior performance than existing methods.

To summarize, the contribution of this paper includes:

1) We propose the first unsupervised algorithm, i.e. the IC-HRS algorithm, for haze region segmentation. IC-HRS can detect haze region of a photo using just the photo and achieve results and time cost similar to existing supervised methods.
2) We demonstrate the feasibility of using rich cloud/haze patches segmented by IC-HRS rather than small-scale photos for effective haze severity estimation, which benefits from the elimination of the side-effect of diverse background scenes, the two-step classification method, and the significantly increased training samples.
3) We identify the degree-progression prior in haze level prediction and formulate it as a pair-wise constraint in the loss function of CNN, which improves the prediction of CNN by making classes farther from the true class have lower prediction values. The formulate of this prior can be extended to similar classification tasks.

II. Related Work

Generally speaking, our study is related to the ambient air quality estimation with digital photos. To summarize, existing work on pure photo-based air quality estimation can be categorized into three classes in terms of different estimation approaches:

- **Direct calculation**: directly computes a parameter value with a function obtained from experimental observation to reflect the severity levels of haze. Mao et al. [4] model such haze parameter with a function leveraging the lowest/highest pixel values in RGB channels across the image. Li et al. [5] calculate the parameter by leveraging the medium transmission map [12] and the depth map of the original photo through operations, such as logarithm, multiplication, division, and pooling.
- **Statistical regression**: uses statistical methods to model the parameter of air quality measure as a function of related factors. Wang et al. [9] use depth map and dark channel of original photos to model $PM_{2.5}$ values as a function of air light, distance, and dark channel. Liu et al. [10] predict the $PM_{2.5}$ value of a photo with support vector regression (SVR) and a range of features, including medium transmission map, image contrast and that of the manually selected region of interest (ROI), image entropy, sky color and smoothness, solar zenith angle, and local humidity.
- **Handcrafted feature-based classification**: extracts hand-crafted visual features that describe the image as a whole, such as color histogram, based on which classifiers are trained to perform air quality level classification. Zhang et al. [6] use four types of visual features, including medium transmission map, power spectrum slope,

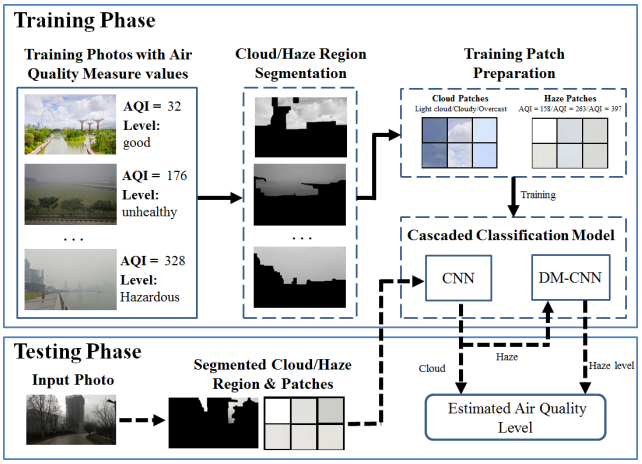

Fig. 1: Use case of the cascaded framework for air quality estimation. The procedures of training and testing phases are drawn using arrowlines and the dashed ones, respectively.

contrast, and saturation, and train five SVMs as binary classifiers to categorize air quality index (AQI) into six levels. In their follow-up study [7], features from the dark channel maps and sky regions are included, and cascaded SVMs with multiple kernel learning is used for air quality level classification.

- **End-to-end classification**: uses images instead of hand-crafted features as inputs for air quality level classification. Zhang et al. [8] propose to use a 10-layer CNN with a new activation function and negative Log-log ordinal classifier for classifying original photos of daily living into six severity levels of air quality measures, in this case, the $PM_{2.5}$ and PM_{10}.

It is worth mentioning that besides the photos of daily living, other data sources, such as satellite images [13], social media [14], [15], [16], and official records on the web [17], [11], also have been investigated to find their links with air quality. However, studies using photos of daily living are the majority, because of its easy accessibility and the strongest association with ambient air quality.

III. The Cascaded Classification Framework

A. Overview

Our proposed cascaded classification framework includes two key components: 1) the cloud/haze region segmentation, and 2) the cascaded classification model. As shown in Figure 1, in training phase, training photos with air quality measure values, in this case, the AQI levels, are first fed to the IC-HRS algorithm for segmenting patches of cloud/haze regions. Subsequently, the training patches are organized to train the cascaded classification model, where CNN is trained using both haze and cloud patches for bi-class cloud/haze classification while DM-CNN is trained using haze patches for haze severity level classification.

In the testing phase, the input photo goes through similar procedures to obtain the cloud/haze patches. Note that if the HRS algorithm returns a nil, it indicates that the input photo has a pure blue sky and a decision of good air quality is directly made without going through the classification model. Otherwise, the obtained patches subsequently are fed to the bi-class CNN for cloud/haze classification. The final decision on air quality is determined by a major voting of decisions of patches. A voting of cloud leads to the final decision of good air quality; otherwise, the patches are fed to DM-CNN for fine-grained classification of haze severity levels and the final decision is also made by major voting.

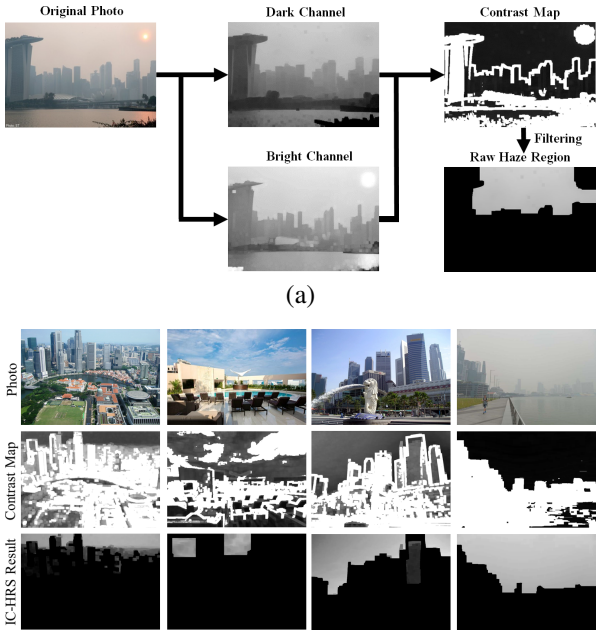

Fig. 2: (a) Procedures of IC-HRS algorithm. (b) More examples of HC-HRS results.

The design of this framework stands on two principles: 1) learning from cloud/haze patches rather than the original photos can reduce the possible noise from diverse scenes, thus alleviating the training complexity and improving the performance; and 2) Separating the classification of cloud patches from haze patches can help the learning model to be more focused, i.e. the first classifier can focus on the difference between cloud and haze and the second one can focus on fine-grained characteristics of haze in different severity levels. The second point has been supported by observations in [8], where CNN trained with both haze-free and haze photos performs well in classifying non-haze photos while not impressive on differentiating haze levels.

B. The Intensity-Invariant Contrast-Based Haze Region Segmentation (IC-HRS) Algorithm

The proposed Intensity-invariant contrast-based haze region segmentation (IC-HRS) algorithm is based on the smoothness and intensity-invariant properties of haze. We observe that pixels in haze regions usually have little contrast difference in RGB channels and compared with neighboring pixels. In contrast, scene objects usually are colorful, thus having neighboring pixels of low-intensity values in at least one of the RGB channels [12]. As such, we proposed IC-HRS to detect haze regions from background scenes.

As shown in Figure 2(a), IC-HRS has three key steps: 1) the calculation of the dark and bright channel images; 2) the calculation of the contrast map using the dark and bright channel images; and 3) the filtering over contrast map to obtain haze regions. The first step aims to select the lowest and highest intensity values for each pixel in RGB channels, while second steps aim to identify candidate haze regions that have invariant intensity values under a threshold. The last step adopts our proposed heuristic filtering strategies to segment out qualified haze regions from the contrast map. The technical details of IC-HRS are illustrated in the following sections, and the pseudocode is given in Algorithm 1.

1) The Dark/Bright Channel Images: As shown in Figure 2(a), given an input photo I, HRS first computes two images, namely the dark channel I_{dark} [12] and our proposed bright channel I_{bright}, defined as:

$$I_{dark}(x) = \min_{\hat{x} \in neighbor(x)} (\min_{k \in \{R,G,B\}} I^k(\hat{x})), \tag{1}$$

$$I_{bright}(x) = \max_{\hat{x} \in neighbor(x)} (\max_{k \in \{R,G,B\}} I^k(\hat{x})), \tag{2}$$

where x is a pixel, $I^k(.)$ is the k-th channel of image I, and $\hat{x} \in neighbor(x)$ is a neighboring pixel of x.

We observe that I_{bright} is a counterpart of I_{dark} choosing all brightest pixel values. We propose the so-called bright channel to maximize the contrast difference.

2) The Contrast Map: The contrast map $I_{contrast}$ is essentially a mask image marking intensity-invariant pixels, defined as:

$$I_{contrast}(x) = \begin{cases} 0 & if\ v_{contrast} < v_{threshold} \\ 1 & otherwise \end{cases}, \tag{3}$$

where $v_{contrast} = I_{bright}(x) - I_{dark}(x)$ and $v_{threshold} = \frac{1}{2}(\text{mean}(v_{contrast}) + \text{median}(v_{contrast}))$. The contrast map is further smoothed by a mode filter F_{mode} (See Algorithm 1 for details).

Similar to the contrast map shown in Figure 2(a), we consistently observe that the contrast maps of most photos in our experiment can perfectly draw a "white line" separating cloud/haze regions from scene objects. This demonstrates that IC-HRS can effectively make use of the contrast difference between pixels at the same boundary positions of I_{bright} and I_{dark} for haze region segmentation. The noise, i.e. the black areas below the boundary, are introduced by the scene objects in black and white so that their contrast different is also small. However, we will show that such noise can be effectively filtered using our proposed filtering algorithm described in the following section.

3) The Heuristic Filtering Algorithm: Our proposed filtering algorithm aims to effectively and efficiently detect haze regions from the contrast map $I_{contrast}$. More specifically, we expect the developed algorithm can perfectly detect the "white line" between cloud/haze regions and background scenes in the contrast map so as to segment out the haze regions without scanning through the entire photo. To achieve this goal, we introduce two types of prior knowledge:

Algorithm 1 IC-HRS Algorithm

Input: RGB channels of original photo I of size $(m \times n)$

1: Calculate the dark/bright channels $I_{dark}(x)$ and $I_{bright}(x)$ according to Equations (1) and (2) with the square neighborhood mask of side length = $\frac{1}{40}\min(m, n)$;
2: Calculate the contrast map $I_{contrast}(x)$ according to Equation (3);
3: Smooth $I_{contrast}(x)$ using a square mode filter F_{mode} with side length = $\frac{1}{20}\min(m, n)$;
4: Set a filter F_{raw} of size $(\frac{1}{10}m \times \frac{1}{10}n)$
5: Set $Haze\_region\_in\_row = 1$
6: Set $No\_haze\_region = 1$
7: Set a zero matrix I_{mask} of size $(m \times n)$
8: **For** each row $\in [\frac{1}{20}m, m - \frac{1}{20}m]$ of $I_{contrast}(x)$
9: **If** $Haze\_region\_in\_row = 0$ and $No\_haze\_region = 0$
10: break;
11: **End If**
12: Set $Haze\_region\_in\_row = 0$
13: **For** each column $\in [\frac{1}{20}n, n - \frac{1}{20}n]$ of $I_{contrast}(x)$
14: Scan this local patch with filter F_{raw}
15: **If** this patch contains no zero entry
16: Set $Haze\_region\_in\_row = 1$
17: Set $No\_haze\_region = 0$
18: Set the corresponding pixels in I_{mask} to be 1's
19: **End for**
20: **End for**

Output: The mask image I_{mask} with pixel values equal to 1 for haze regions and otherwise 0

- Haze/sky regions usually appear in the upper positions of photos;
- Haze/sky regions usually are connected large regions rather than the scattered small ones.

Based on the first fact, our filtering algorithm performs a top-down scan of rows of the contrast map $I_{contrast}$ for haze regions, while based on the second one, our algorithm will search for the connected regions and stop the process when no connected haze regions are found in a row scan - this helps to filter the noise introduced by scene objects, as mentioned in Section III-B2, and can greatly save computation for high-resolution photos.

The procedures of the heuristic filtering algorithm are presented in Steps 4-20 of Algorithm 1. Note that $Haze\_region\_in\_row = 1$ indicates that a haze region has been found in a row, and $No\_haze\_region = 1$ indicates that no haze region has been found in $I_{contrast}$. The two conditions are signs indicating the complete scanning of the haze region in photos. The images of raw haze regions shown in Figure 2 are results of the heuristic filtering algorithm. We observe that the proposed filtering strategies work well on diverse photos of daily living to filter scene objects. This will greatly simplify the haze/non-haze weather classification task as cloud/haze classification.

4) Computational Complexity Analysis: IC-HRS is the first unsupervised algorithm for haze region segmentation, which is effective, the light resource required, and much faster. As shown in Algorithm 1, it has a linear time complexity of $O(n)$, where n is the number of pixels of the input image. IC-HRS uses just the original photo and is able to complete the extraction and database storage of an arbitrary number of

TABLE I: Statistics of datasets.

[5em] Dataset Statistics	# of class	# of training data	# of testing data	Image size
Synthetic dataset	9	2700	324	(640×480)
Real-world dataset	4	1500	372	(640×360), (1080×810), (1200×800),..., (4032×3024)

haze patches in a single pass of the filter. More importantly, benefiting from our developed heuristic filtering algorithm, IC-HRS may complete after detecting the "white line" without traversing through the entire photo. This may save a lot of computation especially for non- or light haze photos.

C. Degree-Progression Multi-Task CNN

The degree-progression multi-task CNN (DM-CNN) is motivated by the observation that the commonly used objective functions of CNN, such as the cross entropy, do not consider a case of the inter-related classes during prediction, i.e. they solely increase the prediction response to the true class while depressing those for other classes. When applied to the haze level evaluation, it may results in the case that the classifier successfully predicts the air quality in the photo as very unhealthy, but the second highest response is good weather. This is clearly unreasonable.

To address this problem, we incorporate the degree-progression prior, i.e. the increasing degrees of haze between classes, as pairwise constraints into the CNN objective function, defined as:

$$Loss_{total} = Loss_{cnn} + Loss_{pairwise}, \tag{4}$$

$$Loss_{cnn} = -\sum_{i} y_i \log \hat{y}_i, \tag{5}$$

$$Loss_{pairwise} = \sum_{c_j \in \{C \setminus c_J\}} Loss_{pairwise}(c_j), \tag{6}$$

$$Loss_{pairwise}(c_j) = \begin{cases} R & if\ j < y\ \&\ R > 0 \\ -R & if\ j >= y\ \&\ R < 0 \\ 0 & otherwise \end{cases}, \tag{7}$$

where $Loss_{total}$, $Loss_{cnn}$, $Loss_{pairwise}$ are the total loss, the cross entropy loss of CNN, and our proposed pairwise constraints, respectively. y and $\hat{y}$ are the true and predicted labels, respectively. $Loss_{pairwise}(c_i)$ is the pairwise loss for the i-th class c_i that we want to minimize. J is the number of classes so that the last class c_J is excluded from the loss calculation. $R = r_{c_j} - r_{c_{j+1}}$, and r_{c_j} is the prediction response of classifier for c_j.

This pairwise constraint shows that DM-CNN penalizes the cases that the classes far away from the true class have competitively high prediction responses and aims to achieve the objective that the farther the class is away from the true class, the lower the prediction response should be.

Table 2 illustrates our designed architecture for the bi-class CNN and DM-CNN. From our experimental observations, this architecture can be easily optimized by the Tensorflow optimizer in terms of convergence in the training phase and achieve good performance-efficiency tradeoff comparing with those with fewer or more layers.

TABLE II: The designed Architecture for our bi-class CNN and DM-CNN. m and n are the height and width of input image.

Layer #	Input size	Operation (Parameter Details)
1	$m \times n$	cov1 (3×3, 32) & max_pooling (3×3)
2	32×32	cov2 (3×3, 64) & max_pooling (3×3)
3	16×16	cov3 (3×3, 64)
4	16×16	cov4 (3×3, 64) & max_pooling (3×3)
5	8×8×64	fc1 (512)
6	512	fc2 (512) & dropout (0.6)

Fig. 3: Photos of real-world dataset. Left to right: light cloud, cloudy, overcast, AQI=158, AQI=263, AQI = 397.

IV. Experiments

A. Experimental Setup

1) Datasets: We evaluate the performance of the proposed techniques using two datasets: 1) a synthetic dataset created by [5]; and 2) our self-collected real-world dataset of phone photos. The statistics is summarized in Table I.

The synthetic dataset expands the FRIDA datasets [18], [19] and contains 3024 images of nine haze levels. To ensure an unbiased evaluation, we randomly select 300 images from each of the nine classes as training data. Our real-world dataset includes 1872 photos either collected online or taken with our phones in different places and a diverse range of weathers. Specifically, it has 323 clear/light cloud, 437 cloudy, 328 overcasts, and 784 haze weathers. The values of air quality measure of photos, in this case, the AQI, are obtained from the corresponding official government websites (See Figure 3 for example photos). We follow the generally agreed AQI levels[3] to classify the photos to four levels of air quality, i.e. good (AQI $<$ 100), unhealthy (AQI $\in$ [100, 200]), very unhealthy (AQI $\in$ [200, 300]), and hazardous (AQI $>$ 300). In experiments, we randomly select 900 photos evenly from clear/light cloud, cloudy, and overcast weathers as the training data for good air quality, and select 200 photos from each of the three haze levels, i.e. unhealthy, very unhealthy, and hazardous, as the training data.

To facilitate the evaluation of our patch-based approach, for each of the two datasets, we extract 90k and 10k patches from the training and testing data, respectively. To further investigate the influence of patch size on performance, we extract patches of three sizes, i.e. (64×64), (128×128), and (256×256).

2) Parameter Selection: Our cascaded classification framework uses pre-defined parameters in the IC-HRS, the bi-class CNN, and the DM-CNN. Regarding the parameters used in the IC-HRS, we consistently use the settings presented

[3]https://en.wikipedia.org/wiki/Air_quality_index

in Algorithm 1. These settings consider the performance-computation tradeoff, and higher or lower values also work. Note that the size of filter for raw haze regions should have the same ratio with the input image to best preserve the haze regions in contrast map.

Regarding the bi-class CNN and DM-CNN models used in our method, we consistently use the designed network with four convolutional layers and two fully connection layers (See Table II), and use the stochastic gradient descent (SGD) with batch size $= 16$. We evaluate several deeper networks, such as VGG with 13 and 19 layers [20] and deep residual nets with 18 layers [21]. However, all these networks did not converge very well in our experiments. We reckon that these networks are overly deep to our problems. All CNN models are implemented using TensorFlow in the Ubuntu14.04 environment.

B. Performance of IC-HRS

1) Effectiveness of IC-HRS: We evaluate the effectiveness of the IC-HRS algorithm in terms of both prediction and F1 score in haze region detection. Because there is no unsupervised algorithm for haze region segmentation, we compare the performance of IC-HRS with the supervised methods, including random forest, SVM, and CNN. We follow the training method proposed in [7] to trains classifiers using haze/non-haze patches for haze region detection. Random forest and SVM are trained using SIFT and HSV features extracted from the patches while CNN is trained using the patches. We conduct experiments on both datasets, where the training of supervised methods is performed on the training patches while the testing is performed on ten testing photos with manually labeled haze regions. For efficiency purpose, experiments are conducted on patches of size (64×64) with their dark channel as features.

Table IV reports the average performance of each algorithm over the ten testing photos. The standard deviation, usually < 0.02, is omitted due to space limitation. We observe that IC-HRS achieves the best performance in both datasets. The superior performance of CNN over random forest and SVM is due to CNN's effective feature generalization power over handcrafted features. Interestingly, instead of the filtering strategy used in IC-HRS, we use CNN to detect haze region following the proposed filtering algorithm and find that there is a significant improvement than using CNN to scan the entire photos in terms of precision, outperforming IC-HRS. However, we also observe that IC-HRS obtains a much higher performance on real-world data in terms of F1 score, demonstrating our filtering algorithm fits better the real-world scenarios. It is also observed that by using the heuristic filtering strategy, there is no obvious drop of performance of CNN. It demonstrates that our proposed filtering strategy can effectively detect the boundaries separating haze regions from background scenes.

2) Efficiency of IC-HRS: We compare the time cost of IC-HRS with the feature extraction algorithms of sky region [7] and dark channel [12], both of which are widely-recognized effective features of haze. We perform the three algorithms on ten real-world photos of different sizes and compare their running time for feature extraction. All algorithms are run on the PC with Intel Core i7 CPU @3.20GHz. As shown in Figure 4, IC-HRS has the lowest time cost, much lower than extracting features of sky region. The extraction of dark channel achieves similar speed when the photo has less than 2 million pixels, but becomes much slower after that. It is because our algorithm employs the heuristic filtering algorithm so that IC-HRS may not scan over the entire photo.

TABLE III: Performance comparison of IC-HRS and supervised methods in haze region detection.

[5em] Methods \ Datasets	Synthetic		Real-world	
	Precision	F1	Precision	F1
Random Forest	0.7287	0.7821	0.7766	0.7935
SVM	0.7481	0.7965	0.7113	0.8166
CNN	0.8620	**0.8922**	0.8721	0.8663
IC-HRS	**0.8781**	0.8854	**0.9026**	**0.9219**
CNN+IC-HRS	0.9137	0.8804	0.9296	0.8531

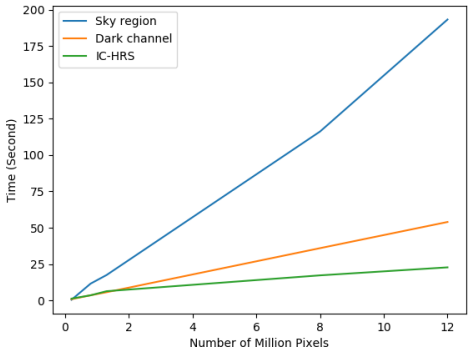

Fig. 4: Time cost comparison of IC-HRS with state-of-the-art algorithms for haze feature extraction.

C. Performance of DM-CNN

To demonstrate the effectiveness of the incorporated degree-progression prior, we evaluate the performance of DM-CNN by comparing with CNN of the same architecture. We conducted experiments on both datasets for training and testing. To achieve a comprehensive comparison, we feed the algorithms with different types of features.

Table IV reports the average precision of both algorithms in terms of precision. "O" and "D" mean that classifiers are trained using the original photos and their dark channel images, respectively, while "O+IC-HRS" means training using the haze patches of original photos extracted by IC-HRS. For a fair comparison, all training images are resized to (64×64). We observe that DM-CNN consistently performs better than CNN in most of the cases. More importantly, training using the segmented haze patches can achieve much better performance than using the original images, especially for real-world photos.

D. Performance of Cascaded Classification Framework

In this section, we conduct experiments on both datasets to evaluate the performance of the entire framework, and compare its performance with the state-of-the-art algorithms for haze severity estimation, including the two direct calculation methods [4], [5], the cascaded SVM [6] and the PAPLE [8].

TABLE IV: Performance comparison of CNN and DM-CNN with different features. (O = RGB image, D = dark channel).

[6em] Models Data	Synthetic data					Real-world data				
	O	D	O + IC-HRS	D + IC-HRS	O + Dark + IC-HRS	O	D	O + IC-HRS	D + IC-HRS	O + Dark + IC-
CNN	0.3287	0.4821	0.3766	0.5223	0.5236	0.3656	0.5625	0.5781	0.6406	0.6375
DM-CNN	0.3481	0.4965	0.3721	0.5866	0.5391	0.3318	0.5893	0.5865	0.6612	0.6971

TABLE V: Performance comparison of our proposed framework with different air quality evaluation models.

[6em] Models Datasets	Synthetic dataset		Real-word dataset	
	Precision	Spearman	Precision	Spearman
Mao's [4]	0.5463	0.7739	0.5834	0.8148
Li's [5]	0.6031	0.8902	0.6742	0.7726
Cascased SVM [6]	0.5687	0.8135	0.6218	0.8452
PAPLE [8]	0.1862	0.3270	0.4186	0.5896
Ours (64x64)	0.6187	0.8839	0.7246	0.8392
Ours (128x128)	0.7927	**0.9327**	0.8190	0.9012
Ours (256x256)	**0.8247**	0.9162	**0.8311**	**0.9289**

Mao's and Li's direct calculation methods are directly applied to the original photos. The Cascaded SVM [6] is trained on original photos with their adopted handcrafted features. The PAPLE [8] is essentially a 10-layer CNN. We do not modify its architecture and follow the paper to train it using the resized original photos of size (64×64). Our algorithm is trained using patches with the dark channel as features.

Table V reports the classification performance of our proposed framework and the algorithms in comparison. It is notable that our method consistently achieves significantly better performance than others in both datasets. This observation demonstrates the importance of our cascaded classification framework. We also observe that patches of size (128×128) achieve the best trade-off between performance and computation. PAPLE meets difficulty to converge on the synthetic dataset, resulting in the poor performance. On the real-world dataset, although converges, it achieves unsatisfactory performance in classifying haze severity levels. These observations are consistent with the authors' findings in the original paper. The SVM-based method achieves stable performance on both datasets. The adoption of a cascaded bi-class classification mode alleviates its predictive biases due to diverse non-haze weathers. However, the performance is still affected by the information loss of handcrafted features and the noisy information from scene objects.

V. Conclusion

This paper presents a cascaded classification framework for fast and robust air quality estimation from single photos of daily living. We demonstrated that the proposed intensity-invariant contrast-based haze region segmentation (IC-HRS) algorithm can significantly improve the prediction performance by segmenting the haze regions for classification rather than using the entire photos. More importantly, we show that using the segmented rich patches for training can achieve much better performance than using small-scale haze photos. It benefits from both the elimination of side-effect of background scenes and the significant increase in training samples. We also observe the robustness of the proposed degree-progression multi-task convolutional neural network (DM-CNN), where the incorporating of inter-class relations in the loss function produces a consistent improvement over traditional CNN.

Despite the impressive results achieved by our proposed solution, our future work includes the investigation on the fine-grained haze features for improved predictive performance and the incorporation of multiple data sources besides photos to gain awareness on both visual and invisible pollution. Besides, we expect to extend the current version of IC-HRS to be able to detect air quality in night photos and more challenging weathers, such as rain.

References

[1] O. Raaschou-Nielsen, Z. J. Andersen, R. Beelen, E. Samoli, M. Stafoggia, G. Weinmayr, B. Hoffmann, P. Fischer, M. J. Nieuwenhuijsen, B. Brunekreef *et al.*, "Air pollution and lung cancer incidence in 17 european cohorts: prospective analyses from the european study of cohorts for air pollution effects (escape)," *The lancet oncology*, vol. 14, no. 9, pp. 813–822, 2013.

[2] A. Al-Ali, I. Zualkernan, and F. Aloul, "A mobile gprs-sensors array for air pollution monitoring," *IEEE Sensors Journal*, vol. 10, no. 10, pp. 1666–1671, 2010.

[3] X. Chen, Y. Zheng, Y. Chen, Q. Jin, W. Sun, E. Chang, and W.-Y. Ma, "Indoor air quality monitoring system for smart buildings," in *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014, pp. 471–475.

[4] J. Mao, U. Phommasak, S. Watanabe, and H. Shioya, "Detecting foggy images and estimating the haze degree factor," *Journal of Computer Science & Systems Biology*, vol. 7, no. 6, pp. 226–228, 2014.

[5] Y. Li, J. Huang, and J. Luo, "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*, 2015, p. No.79.

[6] Z. Zhang, H. Ma, H. Fu, and X. Wang, "Outdoor air quality inference from single image," in *International Conference on Multimedia Modeling*, 2015, pp. 13–25.

[7] Z. Zhang, H. Ma, H. Fu, L. Liu, and C. Zhang, "Outdoor air quality level inference via surveillance cameras," *Mobile Information Systems*, vol. 2016, pp. 1–10, 2016.

[8] C. Zhang, J. Yan, C. Li, X. Rui, L. Liu, and R. Bie, "On estimating air pollution from photos using convolutional neural network," in *Proceedings of the 2016 ACM Multimedia Conference*, 2016, pp. 297–301.

[9] H. Wang, X. Yuan, X. Wang, Y. Zhang, and Q. Dai, "Real-time air quality estimation based on color image processing," in *Visual Communications and Image Processing Conference, 2014 IEEE*, 2014, pp. 326–329.

[10] C. Liu, F. Tsow, Y. Zou, and N. Tao, "Particle pollution estimation based on image analysis," *PloS one*, vol. 11, no. 2, p. e0145955, 2016.

[11] A. Moumtzidou, S. Papadopoulos, S. Vrochidis, I. Kompatsiaris, K. Kourtidis, G. Hloupis, I. Stavrakas, K. Papachristopoulou, and C. Keratidis, "Towards air quality estimation using collected multimodal environmental data," *arXiv preprint arXiv:1610.01209*, 2016.

[12] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[13] R. Li, X. Liu, and X. Li, "Estimation of the pm2. 5 pollution levels in beijing based on nighttime light data from the defense meteorological satellite program-operational linescan system," *Atmosphere*, vol. 6, no. 5, pp. 607–622, 2015.

[14] J. Chen, H. Chen, G. Zheng, J. Z. Pan, H. Wu, and N. Zhang, "Big smog meets web science: smog disaster analysis based on social media and device data on the web," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 505–510.
[15] S. Mei, H. Li, J. Fan, X. Zhu, and C. R. Dyer, "Inferring air pollution by sniffing social media," in *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2014, pp. 534–539.
[16] B. T. Dai, K. Jayarajah, E.-P. Lim, A. Misra, and S. Nayak, "A study on singapore haze," in *Proceedings of the International Conference on Distributed Computing and Networking*, 2016, p. 44.
[17] J. Y. Zhu, C. Sun, and V. O. Li, "Granger-causality-based air quality estimation with spatio-temporal (st) heterogeneous big data," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2015, pp. 612–617.
[18] J.-P. Tarel, N. Hautiere, A. Cord, D. Gruyer, and H. Halmaoui, "Improved visibility of road scene images under heterogeneous fog," in *IEEE Intelligent Vehicles Symposium*, 2010, pp. 478–485.
[19] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, vol. 4, no. 2, pp. 6–20, 2012.
[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.