IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 12, 2019, accepted September 29, 2019, date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945805

Air Quality Measurement Based on Double-Channel Convolutional Neural Network Ensemble Learning

ZHENYU WANG[1], WEI ZHENG[1], CHUNFENG SONG[2], (Student Member, IEEE), ZHAOXIANG ZHANG[2], (Senior Member, IEEE), JIE LIAN[1], SHAOLONG YUE[1], AND SENRONG JI[1]

[1]School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
[2]National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing 100190, China

Corresponding author: Zhenyu Wang (zywang@ncepu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976090 and Grant 61573139, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018ZD05.

ABSTRACT Environmental air quality affects people's lives and has a profound guiding significance for the development of social activities. At present, environmental air quality measurement mainly adopts the method that setting air quality detectors at specific monitoring points in cities with fix-time sampling and slow analysis, which is severely restricted by the time and location. To address this problem, recognizing air quality with mobile cameras is a natural idea. Some air quality measurement algorithms related to deep learning mostly adopt a single convolutional neural network to directly train the whole image, which will ignore the difference of each part of the image. In this paper, in order to learn the combined feature extracted from different parts of the environmental image, we propose the double-channel weighted convolutional network (DCWCN) ensemble learning algorithm. This mainly includes two aspects: ensemble learning of DCWCN and self-learning weighted feature fusion. Firstly, we construct a double-channel convolutional neural network, which uses each channel to train different parts of the environment images for feature extraction. Secondly, we propose a feature weights self-learning method, which weights and concatenates the extracted feature vectors to measure the air quality. Moreover, we build an environmental image dataset with random sampling time and locations to evaluate our method. The experiments show that our method can achieve over 87% accuracy on the newly built dataset. At the same time, through comparative experiments, we proved that the proposed method achieves considerable improvement in terms of performance compared with existing CNN based methods.

INDEX TERMS AQI measurement, deep learning, CNN, image recognition.

I. INTRODUCTION

Environmental air quality is closely related to human production and life. The decline of air quality is likely to cause ecological damage and induce human diseases. At present, air quality monitoring mainly adopts the method of setting up monitoring stations in several specific locations in the city, using the air quality detector to regularly sample and measure air pollutants, and finally obtaining the air quality index (AQI) through calculation and analysis. This method is severely be limited by time and locations, and it can only obtain air quality at specific monitoring points at the regular time. It is difficult to obtain the air quality information of the random location in real-time, and the measurement cost is high. How to obtain the AQI in real time and accurately is a subject worth studying.

Image-based air quality measurement is a method using image processing algorithms to extract environmental image features and estimate AQI based on image features. In recent years, with the rapid development of deep learning technology, using deep learning technology to complete recognition, detection and other tasks is efficient. Environmental images under different air quality grades are often different to some extent, therefore, it is feasible and valuable to use the

The associate editor coordinating the review of this manuscript and approving it for publication was Tossapon Boongoen.

deep convolutional neural network (CNN) to extract features of environmental images and complete the measurement of real-time AQI at random site. Compared with the traditional air quality measurement method, the former can obtain the air quality at any time and any location, which has the advantages of real-time and low cost, it has been widely concerned by the academic circle in recent years.

At present, the existing air quality measurement methods related to image or deep learning are mainly divided into two types: based on traditional image processing or deep learning. The methods based on traditional image processing [1], [7] are using traditional image processing algorithms for feature extraction, such as edge detection, histogram of oriented gradient (HOG), etc. The extracted features are analyzed and calculated to get air quality measurement values. The image-based deep learning methods [2]–[5] generally train the deep convolutional neural network model, extract the environmental image features, and calculate the air quality.

Most of the above image-related air quality measurement algorithms adopt convolutional neural network to extract features of the whole image. However, due to the complexity of the environmental image contents, the change rule of air pollution between the sky part and the building part is different. Under the condition of using the same convolutional neural network for feature extraction, these differences will be ignored. Considering the particularity of the problem, we proposed a double-channel convolutional weighted neural network (DCWCN) to extract different parts of the image features separately and measure the AQI shown in the environmental image. Compared with other methods, our method is more targeted to extract different parts of environmental image features, focuses on the selection of excellent features.

This paper proposes an air quality measurement algorithm based on double-channel convolutional neural network ensemble learning, we first construct a kind of suitable double-channel convolutional neural network architecture for air quality measurement. From the idea of ensemble learning, we use two single-channel convolutional neural networks to extract the feature in the different parts of the environmental image respectively. Secondly, considering that different parts of the image have different effect weights on the final recognition result, we propose a weighted feature fusion method to fuse the features extracted from the two channels. Finally, the integrated global feature is used to measure the air quality of the environment image. In addition, we also propose a weights self-learning method to find the appropriate feature fusion weights. In terms of training and evaluating, we established an environmental image dataset with different locations, times and air quality by means of the manual camera capturing and collection. The experiment shows our method can achieve considerable accuracy and a small MAE in our dataset. The algorithm pipeline is shown in Figure1.

The main contributions of this paper:

- We construct a double-channel weighted convolutional neural network architecture called DCWCN used to perform feature extraction for different parts of the environmental image.
- We propose a weighted feature fusion method and a feature weights self-learning method to select excellent features.
- We apply the DCWCN and weighted feature fusion to the classification and regression tasks, complete the tasks of air quality grade measurement and air quality index measurement. Through experiments, we prove the effectiveness of the proposed method.
- We preliminarily explore the effect of different image parts on air quality measurement, obtain a preliminary conjecture that the building parts have greater influence than sky parts.
- We establish a dataset of environmental images, which contains various locations and times, and annotated by both air quality grade and air quality index labels.

The rest of this paper is organized as follows. Section 2 is related work. Section 3 introduces the air quality measurement algorithm based on double-channel convolutional neural network ensemble learning. Section 4 is the experiment. Section 5 is the discussion. Section 6 is the conclusion.

II. RELATED WORK

Image-based air quality measurement methods are mostly based on traditional image processing algorithms at first. With the rapid development of deep learning, deep learning-related air quality measurement methods are attracting more and more attention.

A. DEEP LEARNING AND DOUBLE CNN NETWORK

The research on deep learning can be traced back to 1989 when LeCun applied the BP algorithm to a multi-layer neural network [8]. With the LeNet-5 model proposed by LeCun in 1998 [9], the basic architecture of deep neural network was formed. In 2006, Geoffrey Hinton, a professor at the University of Toronto, formally proposed the concept of deep learning [10]. Alex proposed AlexNet [11] in 2012, built the first big convolutional neural network, and adopted the ReLu activation function instead of Sigmod, managed to avoid the problem of gradient disappeared in neural network training, the performance of image recognition is much better than traditional methods. VGG [12] used multiple small convolution kernels instead of large convolution kernels in AlexNet [11], which also achieved the performance of large convolution kernels, and greatly deepened the network depth, further improving image recognition accuracy. GoogleNet [13] from the perspective of extending the width of CNN, proposed a method of using multiple inception modules in series by widely used 1D convolution kernels, which improved the accuracy and reduced the parameter amount. ResNet [14] proposed using the short-cut connection to transform the learning convolutional layers output task into the learning convolutional layers residual task, which effectively avoids the gradient disappearance problem when

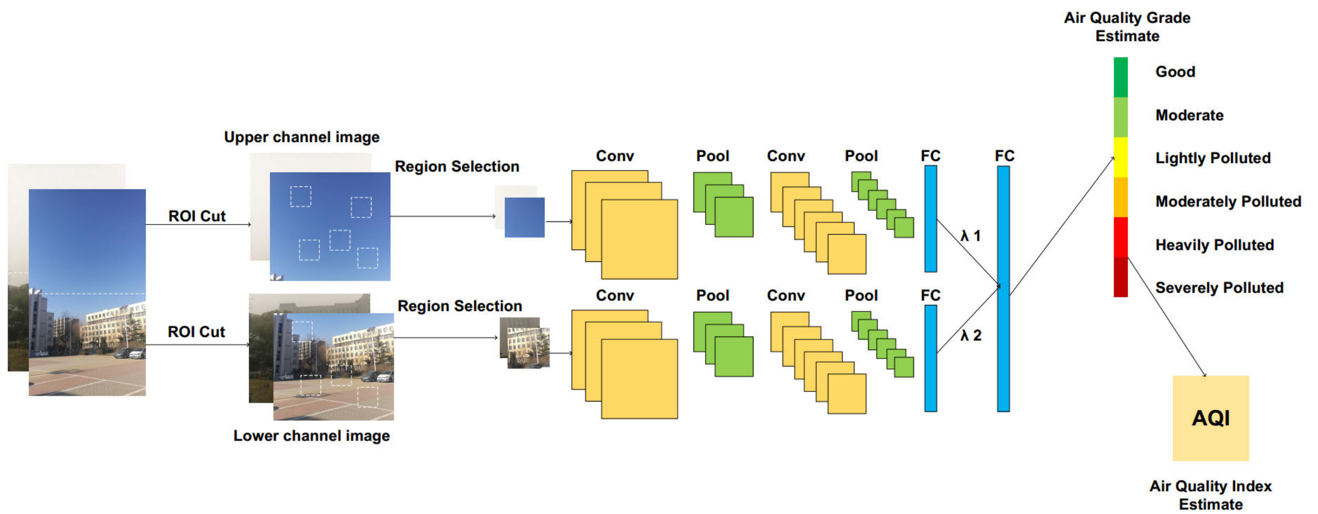


FIGURE 1. **The pipeline of air quality measurement based on the DCWCN algorithm. When estimating AQI, we add an additional layer after the last FC layer for value regression.**

training large deep convolutional neural networks. The propose of these network architectures successively enhances the feature extraction ability of deep convolutional neural networks. In terms of lightweight networks, SqueezeNet [15], MobileNet [16], ShuffleNet [17], etc., make it easy to deploy image recognition to the mobile terminal. Using deep convolutional neural network to extract image features and complete recognition task has become the primary choice and important research direction for more and more researchers. The use of double-channel CNN has also received some attention on recognition tasks. Among many video-related recognition tasks, two-stream CNN has been widely used. Ye *et al.* [18] has conducted in-depth evaluation of two-stream CNN video classification, he constructed a two-channel network to receive static frame and optical flow information in each channel CNN respectively, and used more kinds of network architectures, a variety of model fusion methods, and tested on the behavior recognition dataset and the scene recognition dataset, achieved the performance similar to state-of-the-art. Chung *et al.* [19] proposed A two-stream CNN which contains SpatialNet and TemporalNet to process static frame and optical flow information respectively. Each CNN is a siamese network for processing images captured from the different cameras of the same person for pedestrian recognition task, with a weighted objective function proposed. Compared with the previous method, the precision has improved. In addition, two-channel CNN is widely used in the research of hyperspectral images. Yang *et al.* [20] applied the two-branch CNN to the hyperspectral classification, including a spectral CNN branch and a spatial CNN branch, which are used to extract spectral information and spatial information of the image respectively. Experiments show that compared with the existing methods, the method has a strong classification ability. Yang *et al.* [21] used two-branch CNN for image super-resolution reconstruction, the two-branch CNN is used to extract spectral features from low-resolution hyperspectral images, while the multispectral image is used for spatial information extraction, then the two features are combined for high-resolution hyperspectral image reconstruction.

B. AIR QUALITY MEASUREMENT

In recent years, the measurement of air quality using deep learning method has attracted much attention in academic circles. In the study of air quality measurement related to deep learning, Zhang *et al.* [2] built a convolutional neural network, improved the convolutional layer activation function and classification layer activation function, proposed an improved activation function of convolutional neural network EPAPL, and used a Negative Log-Log Ordinal Classifier to replace softmax classifier in the classification layer, used the environment image to train their network model for classification prediction, completed measurement task of $PM_{2.5}$ and PM_{10} in six grades; Chakma *et al.* [3] used convolutional neural network training images for feature extraction, combined with random forest classification, and classified the air quality shown in the images into three grades of good, middle, and bad. Rijal *et al.* [4] adopts a method of neural network ensemble learning, they used three different convolutional neural networks, VGG16 [12], InceptionV3 [13] and ResNet50 [14] to respectively conduct regression training on image $PM_{2.5}$ values. The predicted values of the input $PM_{2.5}$ of the three networks were input as features into a feedforward network for training to predict the image $PM_{2.5}$ values. Ma *et al.* [5] combined the dark channel prior theory [6], firstly extracted the dark channel from the image, trained two convolutional neural networks respectively with the original image and the dark channel images, and identified the good and bad air quality of the image in three grades. Chen *et al.* [7] proposes a traditional image processing algorithms and deep learning combining approach. First, they counted distribution characteristics of image pixel values, statistics the proportion of high brightness points (pixel value $>$ 128) to all pixel points of each image, use edge detector statistics all image

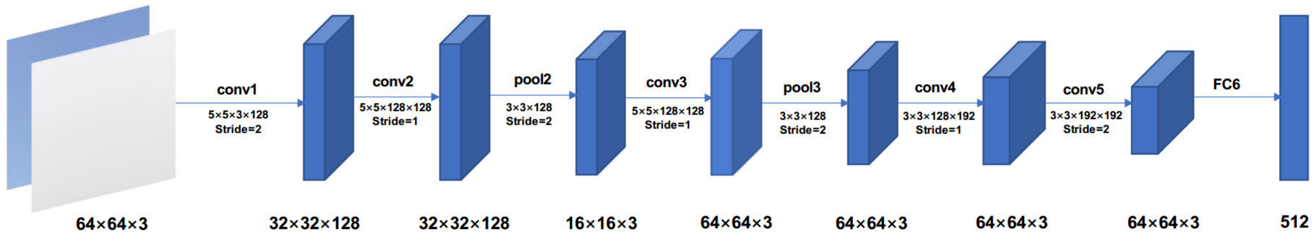


FIGURE 2. Basic single-channel convolutional neural network.

TABLE 1. Architecture of DCWCN, where 'C' is the channel number of convolutional kernel and 'S' is the stride.

Upper channel 64×64×3	Lower channel 64×64×3
Conv1a 5×5 c=128 s=2	Conv1b 5×5 c=128 s=2
Conv2a 5×5 c=128 s=1	Conv2b 5×5 c=128 s=1
Max_pool2a 3×3 s=2	Max_pool2b 3×3 s=2
Conv3a 5×5 c=128 s=1	Conv3b 5×5 c=128 s=1
Max_pool3a 3×3 s=2	Max_pool3b 3×3 s=2
Conv4a 3×3 c=192 s=1	Conv4b 3×3 c=192 s=1
Conv5a 3×3 c=192 s=2	Conv5b 3×3 c=192 s=2
FC6a 512	FC6b 512
Concatenate FC6a×λ1+FC6b×λ2	
FC7 1024	
FC8 6	

the proportion of edge points to all pixels. The two values of each image as input features to train the BP neural network, to predict AQI value.

Considering that the different parts of the environmental image have different information, we construct a double-channel convolutional neural network based on the method of deep convolutional neural network and the idea of ensemble learning to extract features from different parts of the image.

III. PROPOSED METHOD

In view of the air quality measurement is image recognition task essentially, therefore, similar to AlexNet [11], we constructed a double-channel convolutional neural network to extract features of the sky and building parts respectively, with proposed a double-channel weighted convolutional neural network (DCWCN) ensemble learning algorithm for air quality measurement by weighted and fused the extracted features. It is composed of two feature extraction convolutional neural networks, a weighted feature fusion layer, and a classification layer, as shown in Table1.

A. DCWCN ARCHITECTURE

The architecture of the DCWCN is shown in Table1. It is composed of upper and lower channel sub-convolutional neural networks, each subchannel is a basic single-channel convolutional neural network shown as Figure2, it contains 5 convolution layers, 2 pooling layers, and 1 fully connected layer. The first three convolution layers adopt 5 × 5 convolution kernel, and the last two convolution layers adopt 3 × 3 convolution kernel for feature extraction of image; Maximum pooling is used in each pooling layer to extract important features from downsampling; The 512-nodes fully connected layer is used to output feature vectors extracted from each network for feature fusion and prediction.

For different components of the environmental image, DCWCN adopts the strategy of ensemble learning to receive different parts of the image simultaneously in the upper and lower channels for training. Before inputting the environment image into DCWCN, the environmental image should be preprocessed first, the part of the sky image and the part of the building image should be segmented. As shown in Figure3, for each image, the horizontal central axis average segmentation method is adopted, to divide the image into the image mainly containing the upper half of the sky and the image mainly containing the lower half of the buildings. At the same time, we crop each part of the image to feed into DCWCN with the random flip. Among them, the upper channel convolutional neural network focuses on feature extraction of the sky. In each round of iterative training, the images of the upper half with more sky elements are input into the upper channel convolutional neural network for training; The lower channel convolutional neural network focuses on feature extraction of the building part. In each round of iterative training, the images of the lower half with more building elements are input into the lower channel convolutional neural network for training. After feature extraction at the fully connected layer of the last layer of each subnetwork, the feature vectors of the upper and lower parts were weighted and fused. The feature vectors containing complete features of the upper and lower channels were used for recognition.

B. WEIGHTED FEATURE FUSION AND WEIGHTS SELF-LEARNING

We found in our observation that, considering the input images of two channels, the content of the sky is relatively simple, generally simple sky and clouds, the image complexity is relatively low; The content of the buildings is relatively rich, there are a variety of different buildings, pedestrians, roads, and plants, etc., and the image complexity is relatively high. Due to the different complexity of the two images, the complexity of features extracted by the two channels is also different, and the influence on the final measurement result is different.

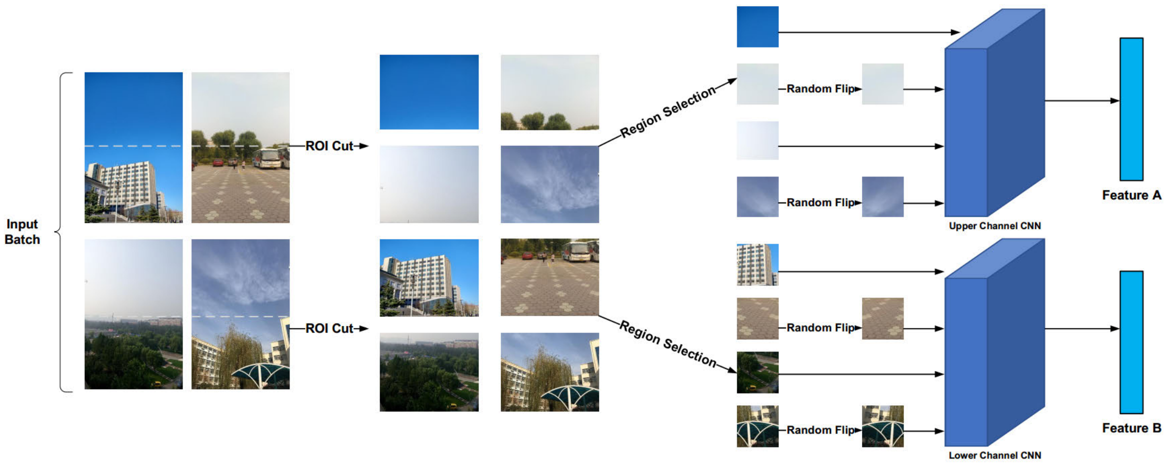


FIGURE 3. **The images preprocessing before training.**

Therefore, considering that the image features of the upper and lower channels may have different influences on the measurement results, we propose a method of weighted feature fusion. Before feeding the output features of the two feature layers into the classification layer, the weighted feature fusion is carried out first. The weights are multiplied by the output feature vectors of the upper and lower channels by two constants, and then the two feature vectors are concatenated. The formula of feature fusion is as equation (1):

$$feat = \lambda_1 * feat_a \oplus \lambda_2 * feat_b \tag{1}$$

where λ_1 and λ_2 are weight values of upper and lower channels respectively; $\oplus$ represents concatenate operation; $feat_a$ and $feat_b$ are feature vectors extracted by upper and lower channels respectively; $feat$ is global features after feature fusion.

As an improvement of artificially assigning feature weights, we propose a self-learning method for feature weights, in which the fusion weights also are calculated by learning with training. In the initial stage, two weights λ_1 and λ_2 are set to 0.5 by adopting the strategy of balancing weights. In the training of the DCWCN, we only train other network parameters of the DCWCN with the weights are frozen. After the training of the network, other parameters of the network should be frozen and the two weights are trained to find the appropriate feature fusion ratio. In feature weights training stage, considering the proportional relationship between the two weights, we propose a weight loss constraint function to limit the training of weight values. The weight loss constraint function is defined as equation (2):

$$Loss_w = [1 - (\lambda_1 + \lambda_2)]^2 \tag{2}$$

When training the feature weight, the objective loss and the weight loss constraint function are combined to form the joint loss function, we optimize joint loss function to adjust the weight parameter values. Finally, we multiply the two weights obtained after training by the feature vectors extracted from each channel and concatenate two weighted features.

C. AIR QUALITY MEASUREMENT

For air quality measurement tasks, start from the two directions of classification and regression, we consider applying our DCWCN in two aspects, air quality grade measurement and air quality index measurement.

1) AIR QUALITY GRADE MEASUREMENT

Air quality grade measurement is a classification task essentially. According to the 6 grades of air quality, the corresponding environmental images are divided into 6 categories and classified in the fully connected layer. Softmax was used for the activation function to conduct the one-hot operation on all kinds of labels to obtain the predicted probability value of each grade, and the maximum probability index of grades was taken as the measurement result. At the same time, we put forward a calculation method of AQI according to the prediction probability of each grade, like equation (3):

$$AQI = AQI_L + (AQI_H - AQI_L)(1 - P) \tag{3}$$

where AQI_H and AQI_L are the upper and lower limits of the predicted grade air quality index respectively, and P is the predicted probability of the predicted grade. According to the calculation, we can get the calculated value of AQI.

2) AIR QUALITY INDEX MEASUREMENT

Based on the idea of regression, we consider the direct measurement of AQI. Therefore, we add a 1-node fully connection layer after the above DCWCN, the AQI value corresponding to the environment image was used as the training label to conduct regression training. With the direct measurement of AQI, we can calculate the air quality grade according to the AQI value. The loss function adopts the mean square error between the predicted value and the labeled

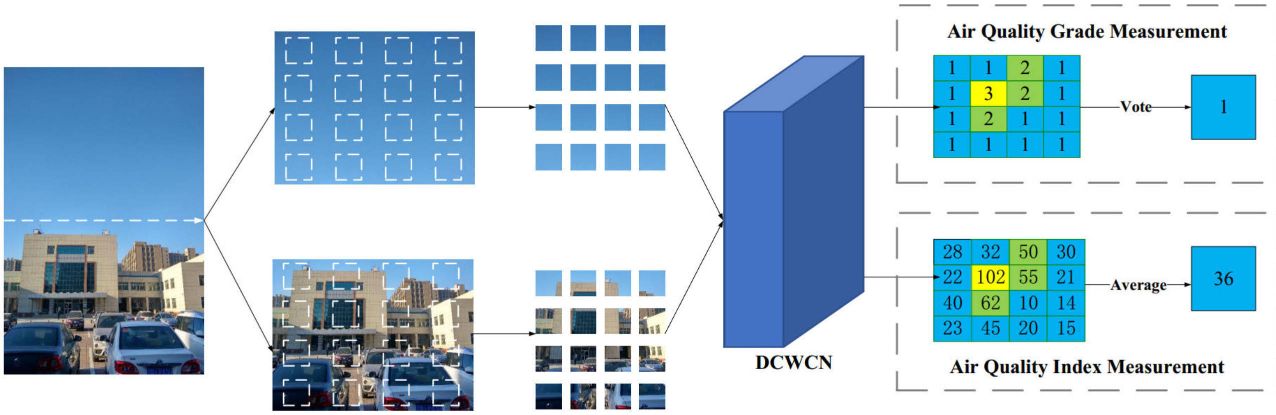


FIGURE 4. The inference process of DCWCN.

value:

$$Loss_R = \frac{1}{M}\sum_{i=1}^{M}(y_i - f(x_i))^2 \tag{4}$$

where M is the number of training images, y_i is the AQI labeled value of the i-th image, and $f(x_i)$ is the predicted value.

3) TEST METHODS AND EVALUATION CRITERIA

At the testing stage, we used test-time augmentation, that is, we also cropped the image in the test, and introduced the voting mechanism and the averaging mechanism to obtain the final result. As shown in Figure4, for each image, we first cut it into the upper part of the sky and the lower part of the buildings, next perform 16 times specific position crop for each part. The crop size is 64∗64, which is the same as the training stage. Each picture gets 16 pairs of image patch as the test data of this image, with fed into the network to use for inferencing. Finally, 16 prediction results will be obtained for each image. For the classification task, the voting mechanism is used for regarding the majority of the 16 prediction results as the final prediction category. For the AQI measurement task, the average of the 16 predicted values is regarded as the measurement result.

Two evaluation criteria, mean accuracy and mean absolute error (MAE), were used to evaluate the classification accuracy. The mean accuracy is shown in the following equation (5), that is, the ratio of the predicted correct sample number to the total sample number. where N is the total number of test samples, y_i^p is the i-th sample predicted grade, y_i^t is the i-th sample labeled grade.

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} y_i^p = y_i^t \tag{5}$$

At the same time, because of the particularity of the problem, the label information collected from the nearest location of stations as well as the time is the most closed to the hour of the AQI value. It is difficult to obtain the location accurate AQI value, so the annotation information has a little error; In addition to the limitation of time and location, the measuring error of the measuring instrument itself makes the images of different grades at the critical point of air quality grade also have the problem of inaccuracy. Therefore, we use the MAE as the second evaluation standard, that is, to calculate the mean value of the absolute value of the difference between the predicted grade and the true grade of each sample. The formula is as equation (6):

$$MAE = \frac{1}{N}\sum\nolimits_{i=1}^{N} |y_i^p - y_i^t| \tag{6}$$

For the AQI regression problem, MAE is also used as the primary evaluation criteria to measure the mean deviation between the measured value and the real AQI value. At the same time, we introduce the mean deviation rate (MDR) as another evaluation criteria of the index. The formula of MDR is as equation (7). Where $f(x_i)$ is the i-th a sample AQI predicted value, y_i is the i-th a sample AQI of true value.

$$MDR = \frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{|f(x_i) - y_i|}{y_i} \tag{7}$$

IV. EXPERIMENT

A. ENVIRONMENTAL IMAGES DATASET

In order to establish an effective dataset, we used the method of manual collection to randomly capture environmental images at different times of each day in the Beijing area. Images content is basically similar, including the buildings and the sky two parts. At the same time, according to [27], we collected the AQI value of the current period for each image. According to [26], we divided each image into corresponding grades, as the image label. The corresponding grade and AQI relations as shown in Table2. In this way, we built a dataset containing about 2500 environmental images under various air quality conditions.

In the study of this task, we screened the dataset. Due to the quality problem of the image itself, we manually removed the images with poor image quality. Such as poor weather conditions, inappropriate shooting time, sunset time which light is

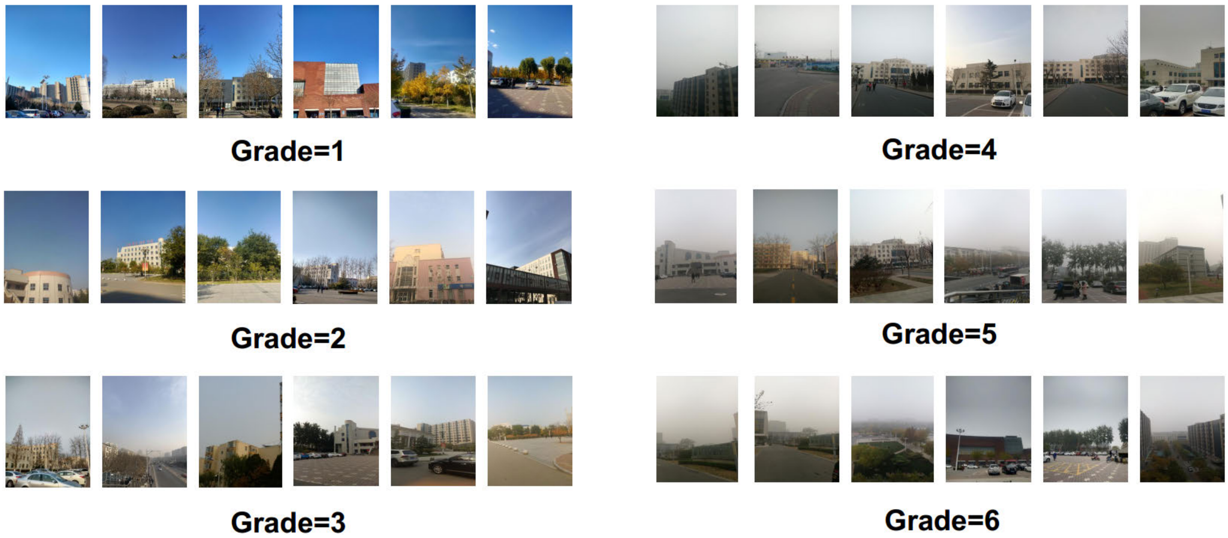


FIGURE 5. Samples of environmental images dataset.

TABLE 2. Corresponding relationship of air quality index and grade.

AQI	0-50	51-100	101-150	151-200	201-300	>300
Grade	1	2	3	4	5	6

TABLE 3. Air quality grade distribution of dataset.

Grade	1	2	3	4	5	6	Total
Train Set	98	94	70	90	79	34	465
Test Set	28	13	18	15	23	5	102
Whole Set	126	107	88	105	102	39	567

poor, night by the influence of street lights. Considering the unbalanced samples problem, images with low AQI are much more than images with high AQI, this will lead to the poor ability of the model to predict high AQI. We chose a relatively evenly number of each grade, finally formed a dataset with 567 images, some sample images are shown in Figure5. The screened dataset contains 465 training images and 102 testing images. The number of images of each category is shown in Table3, and the distribution of AQI is shown in Figure6.

B. IMPLEMENTAL DETAILS

1) IMAGE PREPROCESSING

Before training, we first resized each image and scaled all images to the size of the short side 500 pixels and the long side with the same scaling ratio. In each iteration of training, we selected mini-batch images, for each image, cut evenly based on the horizontal central axis, and each image was divided into upper and lower parts. Secondly, considering insufficient training data, we used random crop and random horizontal flip as data augmentation methods. The random crop was performed on two partitioned images, with a 64×64 image patch cropped respectively. At the same time,

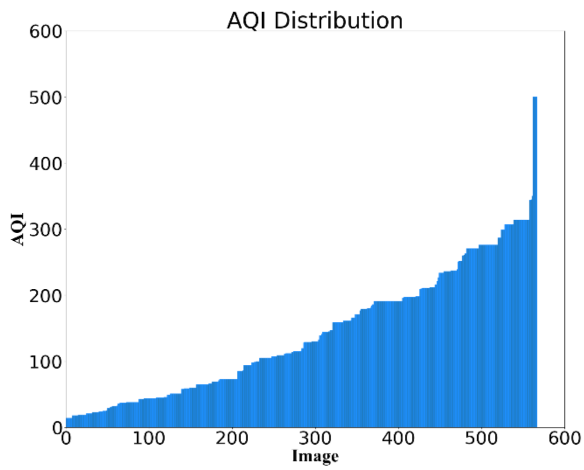


FIGURE 6. Air quality index distribution of environmental images dataset. the x-axis is the image, and the y-axis is the corresponding AQI value.

the horizontal flip probability of 0.5 was used to carry out the random horizontal transformation on the two images. Finally, for each mini-batch, we obtained 64×64 images of batch-size $\times 2$ as the training data of this iteration, and the two cropped images on the same image were taken as a group, which were fed into the upper or lower channel convolutional neural network respectively.

2) DCWCN TRAINING

At the end of image preprocessing, the obtained training data are simultaneously fed into the upper and lower channel sub-convolutional neural network for training. The upper part containing more sky elements is fed into the upper channel convolutional neural network. The lower part containing more building elements is fed into the lower channel convolutional neural network. The loss function is calculated at the output FC layer. For the grade measurement problem, the loss function is the mean of the cross entropy between the predicted grades and the labeled grades. For the AQI measurement problem, the loss function is the mean square error (MSE). The fusion weights λ_1 and λ_2 of feature fusion layer are frozen, and the Adam optimization algorithm [22] is adopted to optimize other network parameters of the network. At the same time, we adopted dropout [23] with a probability of 0.5 to prevent network overfitting at the last convolution layer of each channel convolutional neural network.

3) FEATURE FUSION WEIGHTS OPTIMIZATION

When the training of the DCWCN meets the requirements and the loss value no longer decreases significantly, we stopped the first step of training and froze the network parameters. Next, only two fusion weights of feature fusion layer were optimized, and Adam optimization algorithm [22] was used to update λ_1 and λ_2. After a certain number of iterations, the training was completed.

Training Environment Configuration: We adopt Intel Xeon e5-2650 v3@2.30GHz CPU, NVIDIA Tesla K40c GPU hardware environment; In terms of software, we adopt TensorFlow 1.10.0 deep learning framework and Python3.5 programming language. In terms of training settings, batch-size was 128, the learning rate was 1-e4, the training period was about 3300 epochs, and the number of iterations was 11,000. The first 10,000 times are used to train the parameters of the DCWCN. The last 1000 iterations are used to update the feature fusion weights.

C. EVALUATE THE EFFECTIVENESS OF THE PROPOSED METHOD

Based on the proposed DCWCN, from the point of view of classification and regression, we evaluate our method at two aspects: air quality grade recognition and air quality index measurement.

1) AIR QUALITY GRADE RECOGNITION PERFORMANCE ANALYSIS

We evaluated our proposed method with two feature fusion approaches: double-channel equal weights (DCEW-C) and double channel self-learning weights (DCSLW-C) for classification. In addition, we used classical CNN architecture, such as AlexNet [11], VGG [12], ResNet [14], InceptionV3 [24], and traditional classification algorithms such as SVM [25], MLP, to compare with our method. The testing results of the above methods are shown in Table4. Due to the large similarity between adjacent air quality grade images, we also introduced the neighbor accuracy as another reference evaluation criterion for grade recognition.

As shown in Table4, both our method DCEW-C and DCSLW-C achieved better performance than another approach. Compare with the best performance of other approaches, AlexNet [11], our method improved on all above three criteria, and this improvement shows greater advantages with the use of self-learning weights fusion.

TABLE 4. Results of air quality grade recognition of DCWCN.

Method	Accuracy	Neighbor accuracy	MAE
AlexNet [11]	80.49%	95.58%	0.2441
VGG16 [12]	78.14%	93.63%	0.2863
ResNet50 [14]	76.86%	93.33%	0.3167
InceptionV3 [24]	78.82%	94.61%	0.2794
MLP	38.92%	80.39%	0.8647
SVM [25]	28.43%	40.20%	2.0588
DCEW-C	**80.88%**	**95.58%**	**0.2353**
DCSLW-C	**81.86%**	**96.17%**	**0.2225**

2) AIR QUALITY INDEX MEASUREMENT PERFORMANCE ANALYSIS

In addition, from the perspective of regression, we implemented experiments on the direct measurement of the AQI with DCWCN and classified the images to the corresponding grade according to the predicted results. Similarly, we used double-channel self-learning weights for regression (DCSLW-R), double-channel equal weights for regression (DCEW-R), single-channel for regression (SC-R) and single-channel (AlexNet [11]) for regression (SC-R with AlexNet) to accomplish AQI measurement experiments. The experimental results of AQI prediction are shown in Figure7. For the task of AQI measurement, the difference between the DCWCN with equal weights and self-learning weights is small, and they have respective advantages in MAE and MDR. The prediction of lower air quality index is more accurate, and the accuracy decreases with the increase of the index. Compared with the single-channel convolutional neural network, the measuring error of DCWCN is obviously lower. Some AQI prediction results of DCSLW-R are shown in Figure8.

D. PERFORMANCE OVERVIEW AND PROBLEM ANALYSIS

Based on the above experiments, we compare with DCSLW-C, DCEW-C, SC-C, DCSLW-R, DCEW-R, SC-R methods performance of both grade measurement and index measurement, as shown in Figure9. On the whole, the classification method using DCWCN shows obvious advantages in terms of the accuracy of grade measurement, and the regression method using DCWCN obtains lower AQI MAE. This advantage is further enhanced with the adoption of self-learning weights.

Meanwhile, in the experiment, we also analyzed the samples that testing failed. The partial measurement wrong samples of DCSLW-C and DCSLW-R as shown in Figure10. We found that most wrong samples were high air quality grade samples, and the difference between these images with different grades was small, and there was a similarity with the error category. At the same time, due to the label limitations mentioned above, the labels were approximately collected at the nearest time and place monitoring points, so there are certain errors in the label itself, which are also part of the reasons for the measurement errors.

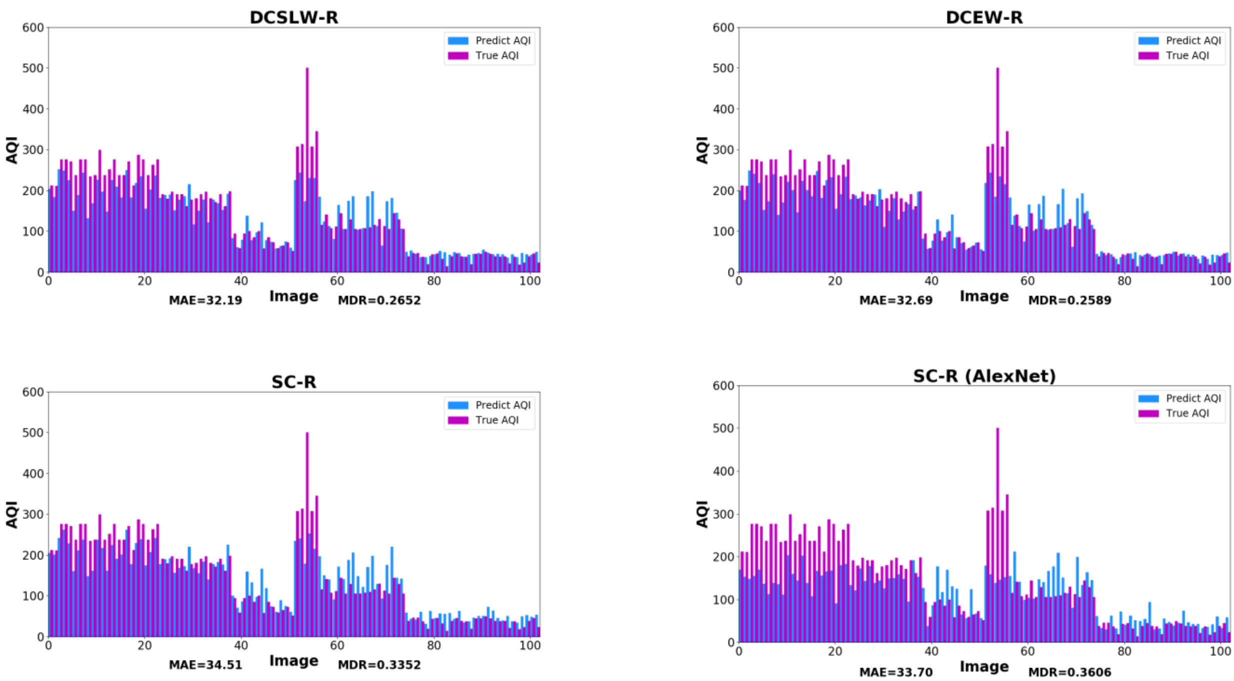


FIGURE 7. Results of air quality index regression. The top parts are method with DCWCN and the bottom parts are with single-channel CNN.

FIGURE 8. Samples of DCSLW-R AQI measurement.

TABLE 5. Experimental results of different feature selection channel.

Method	Accuracy	Neighbor accuracy	MAE
Single channel	67.25%	92.45%	0.4245
Single upper channel	69.31%	91.37%	0.4039
Single lower channel	74.80%	89.80%	0.4108
DCEW-C	**80.88%**	**95.58%**	**0.2353**
DCSLW-C	**81.86%**	**96.17%**	**0.2225**

V. DISCUSSION

A. PERFORMANCE EFFECTS OF COMPLEX CONDITIONS

In the experiment, we found the sky and buildings parts have different effects on AQI measurement, and different feature selection channels and different feature fusion weights would have a significant effect on the algorithm performance. Therefore, we discussed the different performance of the system under the above conditions.

1) PERFORMANCE ANALYSIS OF DCWCN WITH DIFFERENT FEATURE SELECTION CHANNELS

As we can see, double channels could extract features from different parts of environmental images, but what will happen to the performance when we use only a single channel? so we use the following methods for performance comparison: (1) single channel convolutional neural network is used for training and evaluating on the whole image; (2) only the upper channel convolutional neural network is used for training and evaluating the upper part of the image; (3) only the lower channel convolutional neural network is used for training and evaluating the lower part of the image. Each single-channel CNN is our base single channel CNN shown in Figure2.

It can be seen from Table5 that the method using DCWCN is much better than the method only applied a single convolutional neural network at accuracy, neighbor accuracy, and MAE. For the proposed DCEW-C, because it takes different parts of the image information into account, and adopts the strategy of separately extracting ensemble learning, it has achieved a great performance improvement compared with the single-channel convolutional neural network. Compared with the optimal performance of each single channel convolutional neural network, the accuracy is improved by more than 6 percentage points, the neighbor accuracy is improved by more than 3 percentage points, and the MAE is reduced by

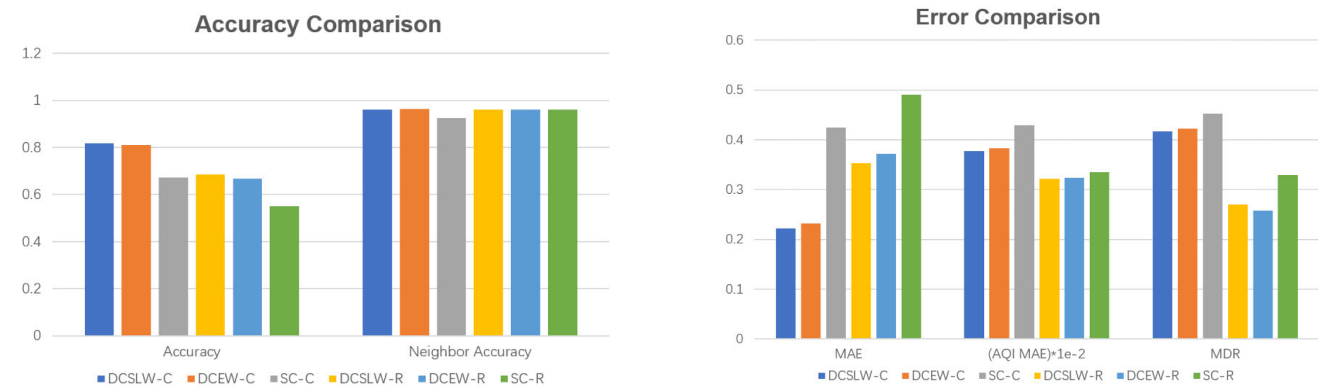


FIGURE 9. Performance comparison of algorithms. Which contains two kinds of criterion and experimental method: classification, regression.

TABLE 6. Results of different feature fusion weights ratios.

Ratio（upper: lower）	Accuracy	Neighbor accuracy	MAE
3:7	79.90%	97.16%	0.2363
4:6	77.84%	93.53%	0.2990
5:5(DCEW-C)	80.88%	95.98%	0.2353
6:4	78.43%	92.05%	0.3049
7:3	80.29%	94.80%	0.2539
DCSLW-C	**81.86%**	**96.17%**	**0.2225**

0.1686. Besides, with the self-learning weights fusion method DCSLW-C, the performance has been further improved. And we can find in Table5, for the single channel, the lower channel is more efficient than others, so we speculate that the building parts are more important than sky parts.

2) PERFORMANCE ANALYSIS OF DCWCN WITH DIFFERENT FEATURE FUSION WEIGHTS

On the basis of the double-channel equal-weighted convolutional neural network for air quality grade recognition, we explore the performance of different fusion weights. Considering from two aspects, we have conducted experiments with different assigned weights and with double-channel self-learning weights for classification (DCSLW-C). For the weight assignment, we studied the system performance with the feature weights ratios of upper and lower channels at 3:7, 4:6, 5:5, 6:4 and 7:3 respectively. The experimental results are shown in Table6.

As can be seen from Table6, system performance varies for different feature weights. For the inappropriate feature weights, the performance of the system is obviously decreased compared with DCEW-C. On the other hand, some performance criterions have been improved with more appropriate feature weights. Therefore, adopting the method of weights self-learning is beneficial for the system to automatically find the appropriate feature weights. The DCSLW-C has improved its performance in terms of accuracy and MAE, compared with DCEW-C and the assigned weights feature fusion method.

FIGURE 10. Hard samples of the experiments, the top half part is the result of DCSLW-C and the bottom half part is the DCSLW-R.

B. PERFORMANCE EFFECTS OF ROI SELECTION

We found in the study that the performance of DCWCN is susceptible to the ROI's choice of two channels, because of the difference in the distribution of the sky part and the background part of the environmental image. Therefore, we deeply explored the effects of different ROI selections by the next two aspects: (1) Using the different ROI selection ratios of the image sky and background parts. (2) Increasing the number of channels, use the multi-channel convolutional neural network to subdivide the ROI.

1) PERFORMANCE ANALYSIS OF DCWCN WITH DIFFERENT ROI SELECTION RATIOS

In the experiment, we found that the division of ROI may affect the performance due to the difference between the sky and the building's position of different images. Compared with the previous method which the central axis average segmentation to distinguish the upper and lower channels ROI, we explored the effect of ROI selection on the DCWCN using the different cutting ratios for the upper and lower channels. We used five ratios of 1:3, 3:1, 2:3, 3:2 to divide the image horizontally. Each part as the input of the upper and lower channels respectively, the corresponding ROI images are cropped and fed into the upper and lower channels for training. As shown in Figure11, among the five ROI selection

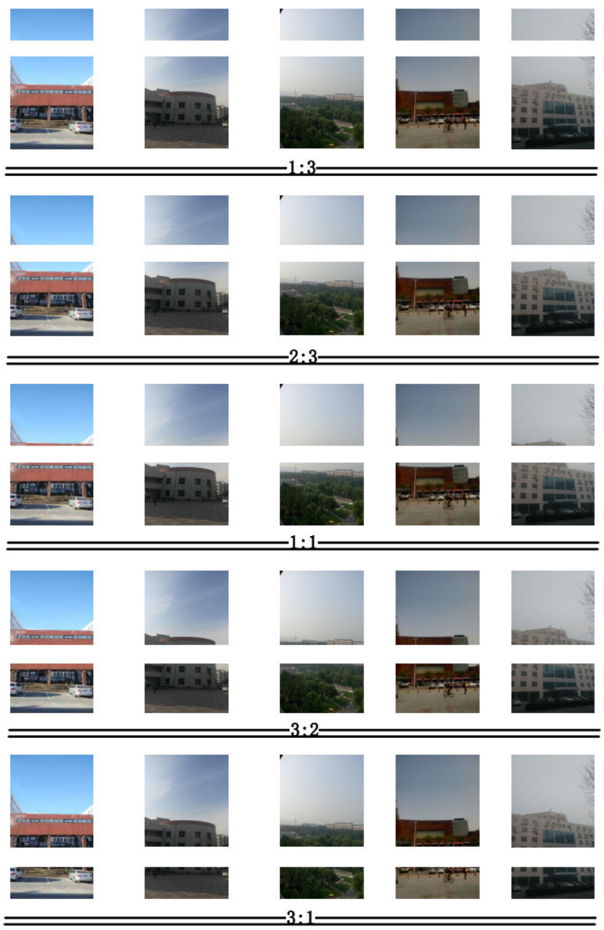


FIGURE 11. ROI partitioning examples of different selection ratios.

TABLE 7. DCWCN performance of different ROI selection ratios.

ROI selection ratio	Accuracy	Neighbor accuracy	MAE
1:3	78.43%	95.10%	0.2843
2:3	76.47%	93.14%	0.3137
1:1	80.88%	95.98%	0.2353
3:2	**83.33%**	**98.04%**	**0.1961**
3:1	76.47%	89.22%	0.3725

ratios, the 1:1 average division is the closest to separating the sky from the building part, this is mainly because of our data collection strategy, the captured image upper and lower proportion as much as possible 1:1. Among other selection ratios, it is inevitable to mix the interferential ROI content required for another channel.

The experimental results are shown in Table7. As shown in Table7, although the two-part ROI separation of the image is the clearest at a 1:1 selection ratio, we are surprised to find that performance is not degraded at all other selection ratios. The ratio of 3:2 is increased by more than 2 percentage points accuracy compared to 1:1. We analyze this may be due to the different image complexity of each channel. As shown in Figure11, under the 3:2 selection ratio, the ROI of the lower channel buildings is more purely divided, which rarely mixed with the sky elements. We suspect that there is more conducive to the lower channel focusing on the extraction of the features from the building parts, the building features may have a greater effect on system performance due to its more complex features compare to the sky part. The sky part information may be relatively simple, and the doping of the sky elements in the lower channel may cause different degrees of interference on the feature extraction of the building information. For other selection ratios, the lower channel also has different degrees of sky elements doping. But for the 3:1 division ratio, although the lower channel content is relatively pure, the excessive building elements on the upper channel make the upper channel also transform to the building feature extraction channel doped with sky elements.

Looking back at Table5, in the single-channel experiment, the performance of feature extraction using only the lower channel is significantly better than that of the upper channel for feature extraction. Therefore, we have reason to speculate that in the image-based air quality measurement, the building part provides more effective information relative to the sky part due to the different image complexity. In order to verify this conjecture, we expanded the number of feature extraction channels and subdivided the ROIs of different parts, use more channels CNN to explore the effect of ROI selection on system performance.

2) PERFORMANCE ANALYSIS WITH MORE ROI SELECTION CHANNEL

Considering that the double-channel network may have insufficient fineness for the ROI's division of different parts of the image, we expanded the number of channels of the feature extraction network to divide each image into multiple ROIs, to explore the difference of using 4,6,8,10 channels feature extraction CNN of images respectively. Each sub-channel is the base single-channel network shown in Figure2. The ROI's division of each channel adopts the average segmentation, as shown in Figure12. At the same time, we also give the weights learned by each channel after self-learning weights fusion in Figure12, which is used to explore the influence of different parts of the image. We evaluated the performance without using self-learning weight fusion and using self-learning weights. The results are shown in Table8.

As shown in Table8 and Figure12, in the case of using multi-channels for feature extraction, we can more clearly analyze the influence of each part of the ROI image on performance. The 4-channel convolutional neural network has little difference in performance compared with the double-channel because it has the most similar channel numbers to the latter than others. However, as the number of channels increases, performance begins to decline. According to the analysis of the loss in training stage, we analyzed that the system is suffered from over-fitting for the case of using more than 4 channels, because the input image content of each single channel is relatively reduced compared with double-channel and 4-channel, so the performance has a significant decline.

TABLE 8. DCWCN performance of multiple feature selection channels. The "wo" represents without self-learning weighted fusion, "w" represents with self-learning weighted fusion.

Channels	Accuracy (wo/w)	Neighbor accuracy (wo/w)	MAE (wo/w)
2	80.88% / 81.86%	95.98% / 96.17%	0.2353 / 0.2225
4	81.37% / 81.37%	94.41% / 95.10%	0.2598 / 0.2460
6	75.20% / 75.78%	91.57% / 92.16%	0.3578 / 0.3480
8	76.67% / 75.20%	92.06% / 91.47%	0.3431 / 0.3608
10	73.33% / 73.92%	91.27% / 91.37%	0.3569/ 0.3578

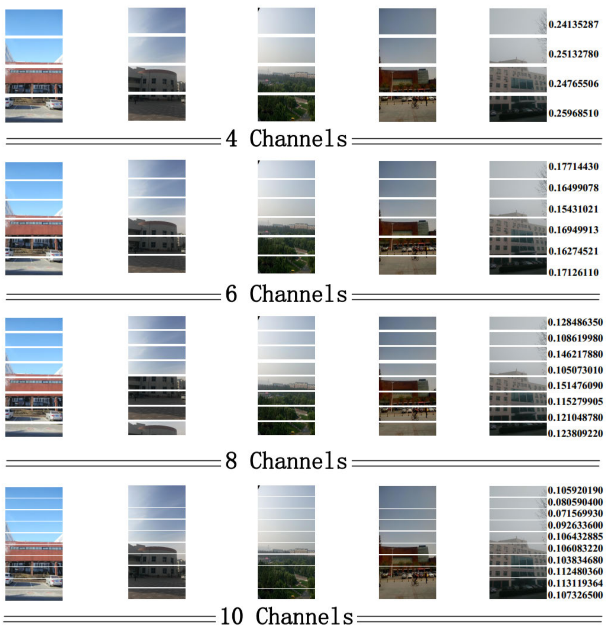


FIGURE 12. ROI partitioning examples of many kinds of feature selection channels.

It can be observed the weights of each channel in Figure12 after using the self-learning weighted feature fusion method. In general, the ROI weights of the building parts are higher than the weights of the sky part ROI, but it has fluctuated for different channels. This also confirms our view that the building features have a greater effect on air quality measurements.

C. ANALYSIS OF ALGORITHM EFFECTIVENESS IN EXPANDED DATASET AND COMPLEX WEATHER CONDITIONS

In order to verify the effectiveness of our algorithm under a large amount of data, we expanded our dataset to increase the number of samples to more than 1000 and verified the performance of DCWCN. The expanded data distribution is shown in Table9. In view of the actual situation, the image with higher air quality grade is more difficult to obtain, so we couldn't balance each category after expanding the dataset. At the same time, considering the strong feature extraction ability of classical CNN architectures such as VGG [12], ResNet [14], inceptionV3 [24] for a large amount of data,

TABLE 9. Air quality grade distribution of expanded dataset.

Grade	1	2	3	4	5	6	Total
Train Set	452	322	188	144	119	38	1263
Test Set	28	13	18	15	23	5	102
Whole Set	480	335	206	159	142	43	1365

we also used these three networks to compare performance under a single channel.

Then, considering DCWCN extensibility, our algorithm can be extended to the double-channel network of feature extraction using another basenet, not just the single-channel network architecture as shown in Figure2. Therefore, we use the above three kinds of classical CNN as two feature extraction channels respectively. On this basis, we construct the double-channel convolutional neural network to carry out experiments.

1) PERFORMANCE ANALYSIS OF SINGLE-CHANNEL NETWORKS

We used the above CNN architectures and our basenet (just single-channel CNN shown in Figure2) to implement experiments in three aspects: train the whole image, the upper part, and the lower part. The results are shown in Table10.

We were pleasantly surprised to find that although the system may be affected by sample imbalance, the performance of our DCWCN system still has been significantly improved benefit by more feature information which added from a large number of samples. In addition, as shown in Table10, in the case of single-channel networks, VGG [12], ResNet [14], inceptionV3 [24] performance is superior to use our basenet on three training methods, because of the powerful ability of feature extraction. However, their performance is still slightly less than DCWCN, we suspect that because the single-channel couldn't distinguish the sky and background features very well. In addition, we can observe that, except for VGG [12], the other three network architectures that use single-channel image have the best performance of the lower channel, which confirms our previous view: the building features have a greater effect on the performance in image-based environmental air quality measurement.

2) ANALYSIS OF SAMPLE IMBALANCE PROBLEM OF DCWCN

At the same time, because of the limitation of datasets, high-grade environmental images are far less than the low air

TABLE 10. CNN Performance with Expanded Dataset. the "W" represents train the whole image, "U" represents train the upper part, "L" represents train the lower part, S-C represents the single channel network.

CNN Architecture	Accuracy	Neighbor accuracy	MAE
VGG16 [12] (W/U/L)	85.88%/82.75%/80.20%	95.59%/94.12%/87.65%	0.2000/0.2529/0.4020
ResNet50 [14] (W/U/L)	77.08%/80.20%/81.67%	93.04%/91.76%/90.88%	0.3176/0.2912/0.3608
InceptionV3 [24] (W/U/L)	78.14%/81.76%/82.75%	94.12%/94.41%/91.96%	0.2863/0.2382/0.3029
S-C (W/U/L)	77.06%/77.65%/80.86%	89.41%/90.86%/90.67%	0.3853/0.3118/0.3078
DCEW-C	86.27%	97.06%	0.1863
DCSLW-C	**87.25%**	**96.07%**	**0.1667**

quality grade in the actual collection quantity. This problem would lead to the sample imbalance problem that may cause bias to the recognition results. Thus we further use the weight balance strategy to improve this problem. The weight balance strategy is improving the cross entropy loss function. we multiply the weight balance vector α with the label vector. The weights of different sizes according to the number of samples in different categories, aimed at increasing the weight of losses in categories with fewer samples. The improved cross entropy loss is:

$$Loss_{CE} = -\sum_{i=1}^{N} \alpha * y_i * log(f(x_i)) \tag{8}$$

where, y_i is the i-th sample label vector, $f(x_i)$ is the i-th sample predicted value, α is the weights balanced vector. In this experiment, according to the number of samples in each category, the weights α of each category is [1,1.4,2.4,3.1,3.8,11.9].

As a comparative experiment, we conducted a forced balance on the samples of the training set to reduce the training biases. The balanced dataset distribution is shown in Table11, the number of each category was balanced to be consistent with the minimum category. The experimental results are shown in Table12. As shown in Table12, the performance of the system after using the category weights balanced loss is similar to the experiment which not used in Table10, there is no big fluctuation.

Meanwhile, the precision and recall of the whole test set are shown in Table13. As can be seen from Table13, the precision and recall with weights balanced loss and balanced dataset of each grade are both considerable. We believe that the category imbalance doesn't have a too serious impact on the performance. Although the precision and recall of the higher-grade categories are slightly lower than that of the lower-grade categories, they still have certain recognition ability, and the system can effectively achieve the recognition task of all grades.

As for the balanced dataset in Table11, as shown in Table12, due to the reduction of training set samples, the ability of data fitting network is insufficient, so there is a certain decline in performance. But as shown in Table13, compared with the method which using balance loss, the method using balanced dataset has a more even distribution of all categories of precision and recall, this may be due to the fact that the balanced dataset makes the biases between each category smaller.

TABLE 11. Air quality grade distribution of balanced dataset.

Grade	1	2	3	4	5	6	Total
Train Set	43	39	42	40	44	38	246
Test Set	28	13	18	15	23	5	102
Whole Set	71	52	60	55	67	43	348

TABLE 12. DCWCN performance with weights balanced loss. The "BD" represents with the balanced dataset, "BL" represents with category weights balanced loss.

Method	Accuracy	Neighbor accuracy	MAE
DCEW-C (BL)	87.25%	96.21%	0.1856
DCEW-C (BD)	81.37%	94.12%	0.2549

3) EXPANDABILITY ANALYSIS OF DCWCN

The experimental results showed that only using a single-channel convolutional neural network to extract features from the environment image can't achieve a satisfactory result. It can't match the performance of DCWCN with the simple feature extraction network even using deeper CNN. Therefore, considering the expandability of DCWCN, the double-channel feature extraction network can adopt any CNN architecture. Therefore, we extended DCWCN to various deep CNNs and we used VGG [12], ResNet [14] and inceptionV3 [24] as the double-channel feature extraction network to carry out experiments. The results are shown in Table14.

After using the three deeper CNN architectures as the feature extraction network of DCWCN, the performance of the system is further improved thanks to the powerful feature extraction ability of deeper CNN. Relative to the above experimental results in Table10 that using the same network of the single-channel method for feature extraction, the double-channel method with deeper CNN can better separate the two parts of the sky and the building features. The system makes each channel network focus on the channel ROI part features, which makes the system performance further improved compared with the method only using a single-channel CNN to extract the features. This result fur-

TABLE 13. The test set precision and recall of DCEW-C. The "BD" represents with the balanced dataset, "BL" represents with category weights balanced loss.

Grade	1（BL/BD）	2（BL/BD）	3（BL/BD）	4（BL/BD）	5（BL/BD）	6（BL/BD）
TP	28/28	10/9	17/13	12/11	21/19	1/3
FP	1/1	1/4	3/4	2/5	6/4	0/1
FN	0/0	3/4	1/5	3/4	2/4	4/2
Precision	96.55%/96.55%	90.91%/69.23%	85.00%/76.47%	85.71%/68.75%	77.78%/82.60%	100.00%/75.00%
Recall	100.00%/100.00%	76.92%/69.23%	94.44%/72.22%	80.00%/73.33%	91.30%/82.60%	20.00%/60.00%

TABLE 14. Variant DCWCN performance with deeper CNN architecture.

Feature Extraction CNN	Accuracy	Neighbor accuracy	MAE
VGG16 [12]	89.10%	97.06%	0.1507
ResNet50 [14]	89.22%	96.08%	0.1471
InceptionV3 [24]	88.74%	96.52%	0.1563

TABLE 15. DCWCN performance with the bad weather images.

Method	Accuracy	Neighbor accuracy	MAE
DCEW-C	80.29%	93.53%	0.2725
DCSLW-C	**81.18%**	**94.31%**	**0.2490**

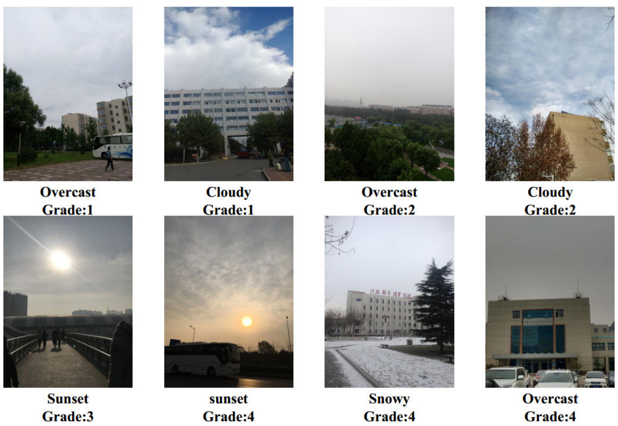


FIGURE 13. Examples of bad weather images.

ther proves that our proposed method is effective, which using the double-channel convolutional neural network to independently extract partial features of sky and buildings in environmental images and ensemble measure air quality.

4) PERFORMANCE ANALYSIS AFTER ADDING BAD WEATHER IMAGES

In addition, based on the above analysis, we also added some images of severe weather conditions to explore the impact of weather quality on system performance. some bad weather images are shown in Figure13. In the experiment, we found that the performance of DCWCN after adding the bad weather images decreased compared with the DCWCN with the expanded dataset, but still achieved good performance, which is almost similar to the performance under the use of a small dataset. This is enough to prove that in the image-based air quality measurement task, our algorithm that using the double-channel convolutional neural network to extract the features from the sky and the background separately is effective.

VI. CONCLUSION

In this paper, we have proposed an air quality measurement algorithm based on double-channel weighted convolutional network (DCWCN) ensemble learning method to measure the air quality grade and index of environmental images. Moreover, we have proposed a self-learning method of weighted feature fusion. Based on the double-channel convolutional neural network, the performance is further improved by using the weighted feature fusion method. We have established a new dataset of environmental images annotated by both air quality grades and air quality indexes. We evaluate the proposed method and compare it with existing methods on this dataset. Experimental results have shown that our method is effective and achieves considerable performance in terms of both recognizing accuracy and mean absolute error. Our method outperforms the compared methods. Meanwhile, it is found in the experiment that it is difficult to distinguish the images with adjacent grades while with similar contents. How to recognize such samples will be an important direction in our future work.

REFERENCES

[1] Z. Zhang, H. Ma, H. Fu, and X. Wang, "Outdoor air quality inference from single image," in *Proc. Int. Conf. MultiMedia Modeling*, 2015, pp. 13–15.

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

[3] A. Chakma, B. Vizena, T. Cao, J. Lin, and J. Zhang, "Image-based air quality analysis using deep convolutional neural network," in *Proc. Int. Conf. Image Process.*, Sep. 2017, pp. 3949–3952.

[4] N. Rijal, R. T. Gutta, T. Cao, J. Lin, Q. Bo, and J. Zhang, "Ensemble of deep neural networks for estimating particulate matter from images," in *Proc. IEEE 3rd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2018, pp. 733–738.

[5] J. Ma, K. Li, Y. Han, and J. Yang, "Image-based air pollution estimation using hybrid convolutional neural network," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 471–476.

[6] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[7] X. Chen, Y. Li, and D. Li, "An efficient method for air quality evaluation via ANN-based image recognition," in *Proc. Int. Conf. Artif. Intell. Ind. Eng. (AIIE)*, 2016, pp. 253–256.

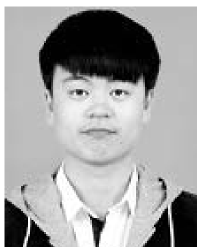
[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, ''Backpropagation applied to handwritten zip code recognition,'' *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, Nov. 1998.
[10] G. E. Hinton, S. Osindero, and Y.-W. Teh, ''A fast learning algorithm for deep belief nets,'' *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''Imagenet classification with deep convolutional neural networks,'' in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 2012, pp. 1097–1105.
[12] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, ''Going deeper with convolutions,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[14] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[15] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, ''SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size,'' 2016, *arXiv:1602.07360*. [Online]. Available: https://arxiv.org/abs/1602.07360
[16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ''MobileNets: Efficient convolutional neural networks for mobile vision applications,'' 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861
[17] X. Zhang, X. Zhou, M. Lin, and J. Sun, ''ShuffleNet: An extremely efficient convolutional neural network for mobile devices,'' 2017, *arXiv:1707.01083*. [Online]. Available: https://arxiv.org/abs/1707.01083
[18] H. Ye, Z. Wu, R.-W. Zhao, X. Wang, Y.-G. Jiang, and X. Xue, ''Evaluating two-stream CNN for video classification,'' in *Proc. ACM Int. Conf. Multimed. Retr.*, 2015, pp. 435–442.
[19] D. Chung, K. Tahboub, and E. J. Delp, ''A two stream siamese convolutional neural network for person re-identification,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1992–2000.
[20] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, ''Learning and transferring deep joint spectral-spatial features for hyperspectral classification,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.
[21] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, ''Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network,'' *Remote Sens.*, vol. 10, no. 5, p. 800, 2018.
[22] D. P. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, ''Dropout: A simple way to prevent neural networks from overfitting,'' *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
[25] C. Cortes and V. Vapnik, ''Support-vector networks,'' *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[26] *Ambient Air Quality Standards*, Standard GB3095-2012, Beijing, China, 2012.
[27] [Online]. Available: http://www.bjmemc.com.cn/

ZHENYU WANG received the B.Sc. degree from the National Defense University of Science and Technology, China, in 1997, the M.Sc. degree from Xi'an Jiaotong University, China, in 2003, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China, in 2007. He is currently an Associate Professor with the School of Control and Computer Engineering, North China Electric Power University, Beijing, China. His research interests include pattern recognition, computer vision, and machine learning.

WEI ZHENG received the B.S. degree from the School of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, China, in 2017. He is currently pursuing the M.S. degree with the School of Control and Computer Engineering, North China Electric Power University, Beijing, China. His research interests include deep learning, pattern recognition, and computer vision.

CHUNFENG SONG received the B.Sc. degree from the Qilu University of Technology, China, in 2012, and the M.Sc. degree from North China Electric Power University, China, in 2016. He is currently pursuing the Ph.D. degree with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China. His research interests include image segmentation, data clustering, person recognition, and deep learning.

ZHAOXIANG ZHANG received the bachelor's degree in circuits and systems from the University of Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2009, under the supervision of Prof. T. Tan. In 2009, he joined the School of Computer Science and Engineering, Beihang University, Beijing, where he was an Assistant Professor, from 2009 to 2011, an Associate Professor, from 2012 to 2015, and the Vice-Director of the Department of Computer Application Technology, from 2014 to 2015. In 2015, he was with the Institute of Automation, Chinese Academy of Sciences, where he is currently a Professor with the Research Center for Brain-Inspired Intelligence. He specifically focuses on brain-inspired neuronetwork and brain-inspired learning. His current research interests include computer vision, pattern recognition, and machine learning.



JIE LIAN received the B.Sc. degree from the College of Information Engineering, Southwest University of Science and Technology, Mianyang, China, in 2017. He is currently pursuing the M.S. degree with the School of Control and Computer Engineering, North China Electric Power University, Beijing, China. His research interests include deep learning, pattern recognition, and computer vision.

SHAOLONG YUE received the B.Sc. degree from the Shandong University of Technology, China, in 2018. He is currently pursuing the master's degree with North China Electric Power University, China. His research interests include pattern recognition, computer vision, and machine learning.

SENRONG JI received the B.Sc. degree from Qufu Normal University, China, in 2018. He is currently pursuing the master's degree with North China Electric Power University, China. His research interests include pattern recognition, computer vision, and deep learning.