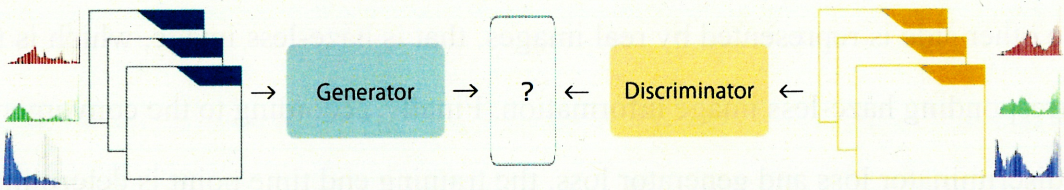
decoder structure uses this potential feature to represent the image content of the highlighted area by restoring the original resolution through the transposition convolution operation. Unlike the original GAN model, which starts directly from the noise vector, the hidden representation obtained from the encoder captures more changes and relationships between the unknown region and the known region, and then inputs the decoder to generate content. The expansion convolution is used in the middle layer, which allows the use of a larger input region to calculate each output pixel, and the field of vision is also expanded. The expansion convolution network model can calculate each output pixel under the influence of a larger pixel region of the input image. The generator uses a standard encoder network, on which an extended convolution layer is added.

4.2.2. Discriminator

Discriminator is through training the generator-model, it can use small reconstruction loss to removal the corresponding pixels in the highlighted area. However, only using the generator can't ensure that the removed area is visually true and consistent, so the target image is added, the multi-scale discriminator is used as the discriminator to distinguish whether the information is true or false, so as to distinguish whether the image is removed. The discriminator helps the network to improve the quality of removal results, and the trained discriminator will not be fooled by unrealistic images. Based on convolutional neural network, these discriminators compress the image into small eigenvectors. The prediction corresponds to the probability that the image is real.

4.3. Combining GAN with Dark Channel Prior

The operation principle of generator and discriminator is shown in the figure below (Figure. 4-1). The input is the bright area bug which can't be solved well by DCP processing. Through the CGAN designed in this study, RGB training with haze-free image is realized, and the ideal RGB value of recovery image is obtained, so as to restore the bright area recovery of dehaze image that DCP can't correspond it.



DCP dehaze image RGB information b) haze-less image RGB information

Figure. 4-1 Diagrammatic sketch for solving RGB-Model using Conditional GAN

The new concept of target map, which is embedded in the process of counter training, pays more attention to the information of discriminator recognition generator, especially the atomization area, which is helpful to the final feature extraction. The target map and its position in the training network are shown in the following figure (Figure. 4-2).

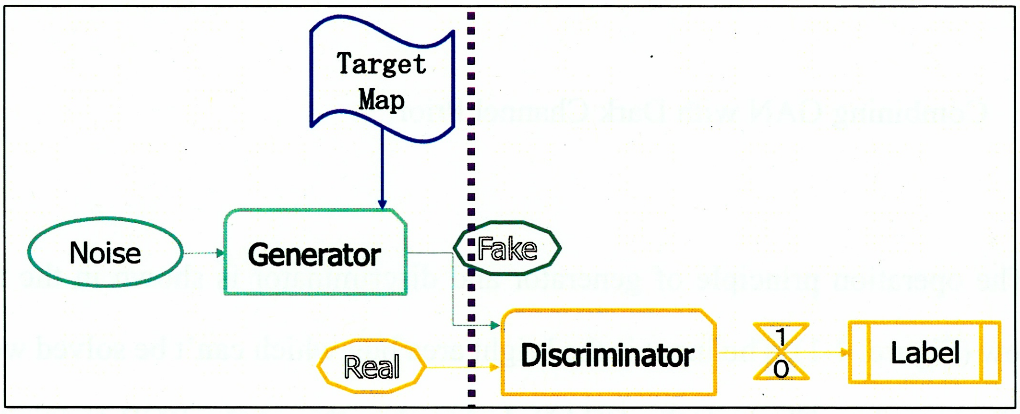


Figure. 4-2 Image of DbGAN composition

Random input is the image after DCP haze removal, but there is a problem of light processing, it is used as a side of GAN to carry out confrontation training. The other side is represented by real images, that is haze-less image, which is the corresponding haze-less image information. Finally, according to the convergence of discriminator loss and generator loss, the training end time point is determined, and the training model is obtained as the basis for processing the unprepared area of the bright area of the image again (Figure. 4-2).

The innovation of this work is to use the RGB information data after dehazing of DCP and use the RGB information data of haze-less image as the counter training, which effectively solves the unavoidable confusion of dehazing distortion of bright part of DCP.

In order to solve the problem of bright distortion caused by DCP, the generation of network and output discrimination network are used to evaluate, so as to ensure that the output looks like the real image. In this process, the target map is generated first. Target map is an important part of the network, because it will guide the network to focus on the haze distortion area. Target map is generated by the cyclic network.

Then, the generated network uses the designed automatic encoder to input the RGB information of the image as the input, referring to the target map. In order to obtain more extensive context information, multi-scale loss is used in the decoder side of the automatic encoder. Each loss compares the output of the convolution layer with the corresponding ground truth. The input of convolution layer contains the characteristics of decoder layer. In addition to these losses, for the final output of the automatic encoder, a perceptual loss is used to obtain a more comprehensive similarity with the ground truth. The final output is also the output of the generated network. As following formula 4-1 , the f (x) and g (x) are the feature vector of the front layer, and the number of the N is the front hidden layer features. The model that synthesizes the No.j area integrates the degree of participation at the No.i position, local information and all stations are integrated.

$$o_j = v\left(\sum_{i=1}^{N}\left(\frac{\exp(f(x_i)^T g(x_j))}{\sum_{i=1}^{N}\exp(f(x_i)^T g(x_j))} h(x_i)\right)\right) \tag{4-1}$$

About the learn and output target layer, using the adjust proportional parameter, it can be learned γ (0, ∞) with scalar, and put TM into GAN model gradually, as following formula 4-2.

$$y_i = \partial o_j + x_i \tag{4-2}$$

After obtaining the output of the generated image, the discrimination network will check whether it is true or not. In this study, target map is used to guide the discrimination network to the local target area. In general, target map is introduced into the generation network and discrimination network, which is a new method. It can effectively achieve haze removal which is impossible for feature extraction and dehazing DCP.

The algorithm of DbGAN as following, $p_g(z)$ is the distribution of haze image sets, and $p_{data}(x)$ is the vector of haze-free image sets.

Table 4-1 the algorithm of DbGAN

Algorithm : DbGAN

for (number of training iterations){
    for (hyperparameter){
        minibatch hazy images $\{z^{(m)}\}$ with $p_g(z)$;
        $T(m)=f_m(C_{m-1},T(m-1))$; // introducing the target map
        minibatch haze-less image $x^{(m)}$ with $p_{data}(x)$;
        update the discriminator;
    }
    sample minibatch of m hazy samples $p_g(x)$ with $p_g(z)$;
    update the generator;
}

It must be update discriminator and generator for $p_g(z)$ approaches $p_{data}(x)$,

$P_G(x;\theta)$, now θ is the parameter.

4.4. Experimental Results

In the generation countermeasure network, the generation network and the discrimination network. Receiving and inputting an image information with incomplete processing, the generating network tries to generate a bright area for processing as much as possible to achieve the effect of bright area dehazing. The discrimination network will verify whether the image generated by the generated network looks real. Its loss can be expressed as:

$$\min_{G} \max_{D} \ E_{R\sim Pciean}[\log(D(R))] + E_{I\sim Praindrop}[\log(1 - D(G(I)))] \quad (4\text{-}3)$$

This kind of haze removal algorithm based on RGB after DCP. Using the generated countermeasure network, the generated network generates the target map through the attention recurrent network, generates the highlighted area repair image through the contextual autoencoder together with the input image. Then, the global and local validity of the output generated by the network evaluation is judged. In order to be able to locally verify, we inject target map into the network, which is also the innovation of this method, that is, we use target map in generating network and discriminating network. For haze-less image and hazy image of the same object, it is hoped that the output of this image with haze will approach haze-less image infinitely after the network processing.

The model was trained in this study, and the data fitting of the results is shown that the accuracy of training and validation data tend to be 1, then tend to be stable in a certain period of time, and rather than rising. Therefore, the approximate value of epoch can be determined.

4.5. Simulation and Conclusions

The new algorithm was introducing the target map to GAN, it is a model that mimics the neural circuits of the human brain. The visualization and accuracy improvement of gaze areas obtained by visual explanation, and it can add the receptive field of feature per layer spreads. It amount to increase the depth of the net. For example, as following figure 4-3, In original image, areas marked with red, yellow and blue are considered as sensitive parts of human eyes. Through TM model, we can notice these areas and realize centralized processing, which improves efficiency. As shown in Figure 4-3, the visible area in the original image will be displayed in the layer. These conspicuous areas are often the location of DCP miss restored area.

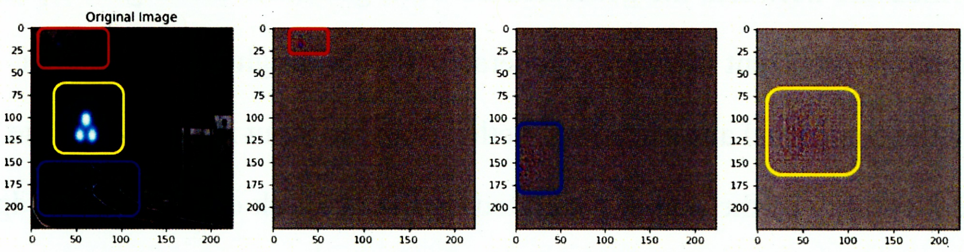
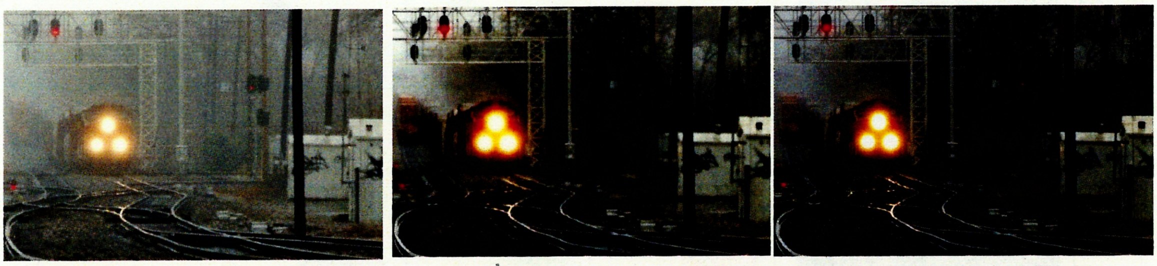


Figure. 4-3 Visualization of Target Map

In these cases, the convergence is ideal and the stable correlation values of train accuracy and test accuracy tend to be 1. The convergence is obvious, and there are local small fluctuations in them. However, the use of the generated model has little impact on the final processing results. Use an example to further illustrate, as shown below.

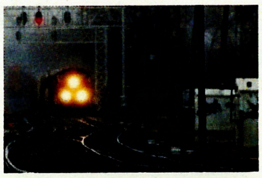
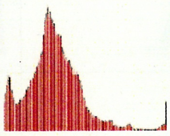
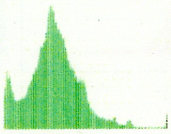
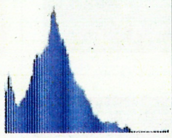
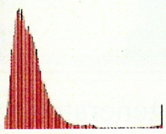
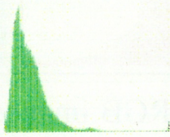
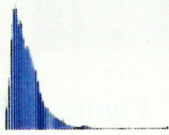
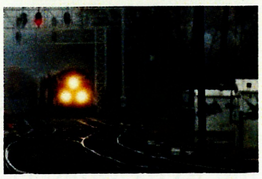
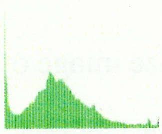


(a) Hazy image (b) DCP dehaze image (c) DbGAN dehaze image

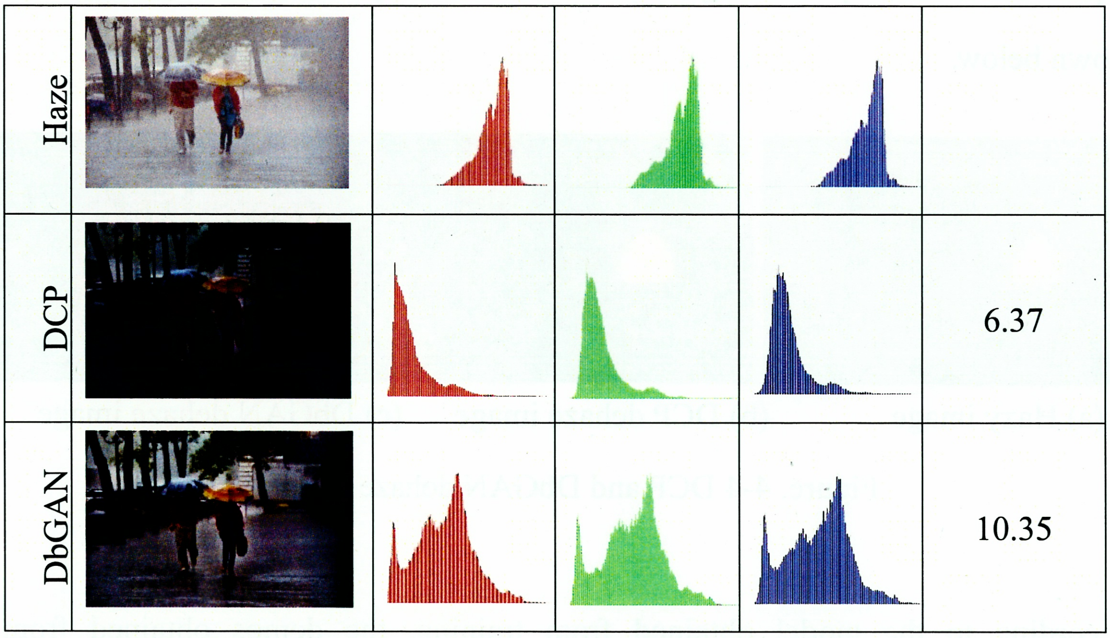
Figure. 4-4 DCP and DbGAN dehaze image

According to the model obtained from training, the demos obtained from restoration are shown in the following example, and the corresponding RGB three values, as well as PSNR(Peak Single-to-noise Ratio) , as show from demo1 to demo3.

Demo 1 :

	Sample	R	G	B	PSNR
Haze					
DCP					8.46
DbGAN					15.47

Demo 2：



Demo 3：

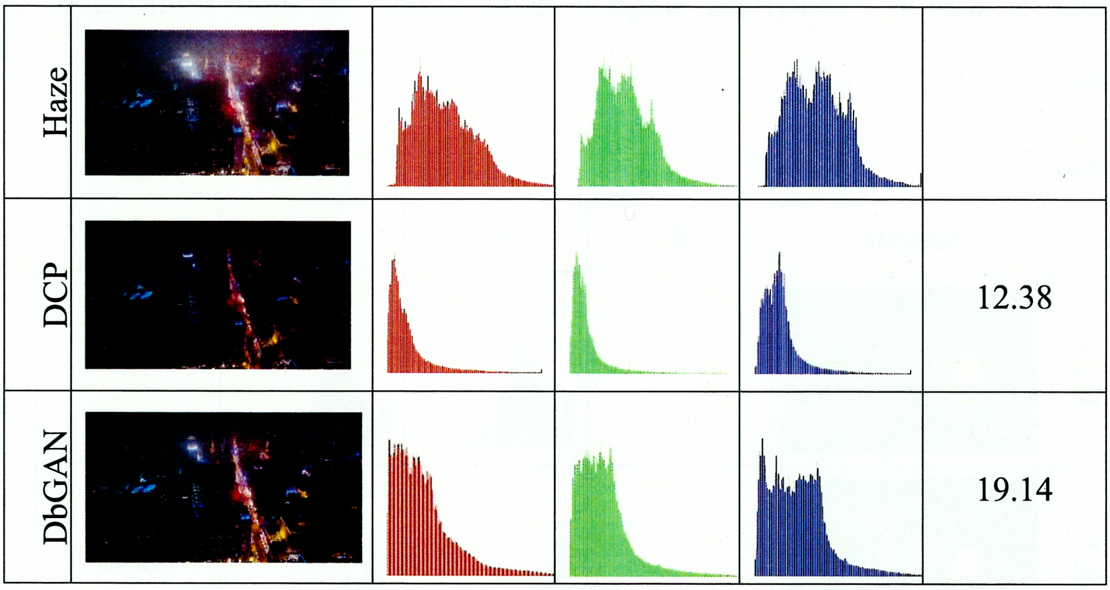


Figure. 4-5 RGB images generated by DbGAN

The RGB value of the dehaze image of DCP is calculated after the dehaze image is generated by DCP. Through DbGAN training, input the RGB value generated by a, and then GAN with haze-less image to generate an ideal model, and finally calculate the RGB value of the repaired image. As shown in the figure, the last line of the three demos is the RGB images obtained through the DbGAN confrontation training.

The PSNR test value in demo shows that, for the demo case, three different dehazing algorithms are used respectively. According to the order of DCP, DbGAN and DhNet, the corresponding PSNR value increases in turn, indicating that the image reconstruction quality is better.

In these three demo cases, the white light area has different degrees of false repair, or exaggerated restoration, and the distortion component is relatively large. After DbGAN processing, the haze removal effect of the white light area has been fully realized.

5. Deep Learning Architecture for Dehazing

Proposal 1 of this work, which has no problem for scientific research, and it is a very excellent algorithm. If it is applied to industrial production practice, portability will be weak, or even impossible to achieve. Because the conditions required by the site are too harsh, the algorithm needs to be improved.

5.1. Introduction

First of all, the premise of the implementation of the proposal is that there must be haze-free image, which can form the final model against the input information. Generally, only the defect image itself is used in hazy image processing. Therefore, it is necessary to obtain the corresponding clear image. Because the image dehaze requirements usually obtained do not correspond to haze-free image. Such as remote sensing image, automatic driving real-time processing image and so on. Secondly, considering the double dehazing characteristics of the above algorithm, the end-to-end neural network deep learning scheme is proposed [77][78][79], input hazy image, through training deep learning, directly get the dehazing image. In order to achieve this goal well, according to the homogenization characteristics of hazy image, the model of applying haze removal algorithm as end-to-end training and learning, that is designed and added, it is possible to realize end-to-end processing[80]. It is proposed that the hole convolution can keep the size of convolution kernel parameters unchanged and increase the convolution field of vision. This research is called DhNet. It has been proved by practice that the distortion problem of dehazing in bright area can be solved well, and the dehazing method is realized. The training acc-loss model results are shown in the fig. 5-1, which shows good convergence.

5.2. Contributions of our DhNet Approach

In this work, convolution neural network is used for deep learning model, and the following steps can be distinguished. One is to establish training model. The input data is mapped by function to get the required results. In this study, a convolutional neural network is a model. The second is to determine the loss function. The loss function is a constraint that forces the neural network to learn in the direction of expected results, that is, to optimize, and update the weight after back propagation (formula 5-2).

Back propagation as following,

$$\begin{cases} \Delta^{(k)} = (\bar{\bar{y}} - y)^T F^{(k)} \\ \Delta^{(k)} = \Delta^{(k+1)} W^{(k+1)} F^{(k)}, k < K \end{cases} \tag{5-1}$$

Weight update,

$$W^{(k)}(s+1) = W^{(k)}(s) + \eta (x^{(k-1)} \Delta^{(k)})^T \tag{5-2}$$

Thirdly, back propagation is the application of a chain rule. In TensorFlow and python, the optimization step of updating back propagation weight has been encapsulated into a function, which can be used directly.

In this work, to sum up the proposed convolution neural network haze removal essentially repeats the above two or three steps until the weight changes to zero. At first, the neural network is not complex enough, resulting in under fitting. It is usually supposed to increase the applicability of neural network by building more complex models. However, this study want to change the dilated convolution technique, so that the network can obtain the characteristics of a broader level. However, dehazing is a low-level image processing task, which does not need to acquire the depth of the graphics features, so it is not need to increase the hidden layer, and the increase of the hidden layer is easy to cause the gradient, zero and the processing results are not available. It has been proved by practice that the hollow convolution method adopted in this research, it can make the neural network converge more easily and faster without distortion, save the training time of the model, improve the effect of dehazing, and obtain the relatively ideal evaluation results of PSNR and SSIM. It is reasonable and feasible to use this method in this work.

At the same time, there are the following relationships (as shown in formula 5-3):

$$\text{accuracy} = e^{-loss} \tag{5-3}$$

The illustration is as follows: four parameter diagrams are involved, for example. The accuracy of the training and the validation set were showed, and so on, the loss of the training set and the validation, as shown in figure 5-1.

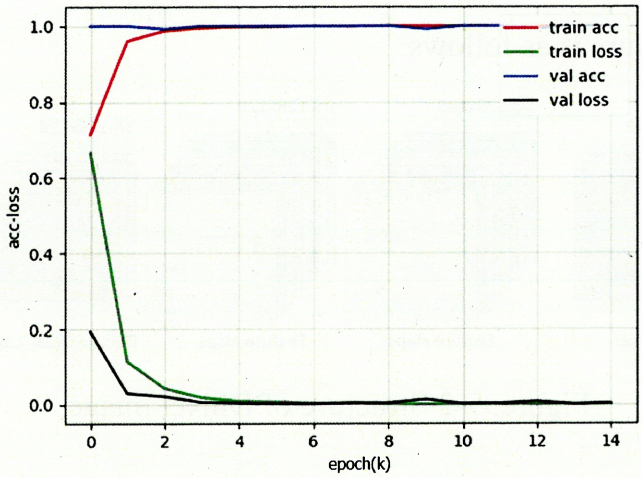


Figure. 5-1 Convergence map of DhNet model

Further analysis of the above figure shows that with the increase of epoch, the accuracy increases and the loss decreases; when the accuracy approaches 1, the loss approaches 0 and converges. Taking the fitness of training set into full consideration, a relatively ideal batch size value is obtained. loss decreased rapidly, and this learning model is available. With the increase of batch size, the loss decreases, and the error is less than 0.01.

Another innovation of this research proposal which according to the homogenization characteristics of hazy image, the convolution in end to end CNN training adopts the hollow convolution technique to realize image restoration and bright distortion area of hazy image, and the image restoration effect is significant, while the processing time of haze removal is saved.

5.3. Training and Experiments

The network structure is as follows:

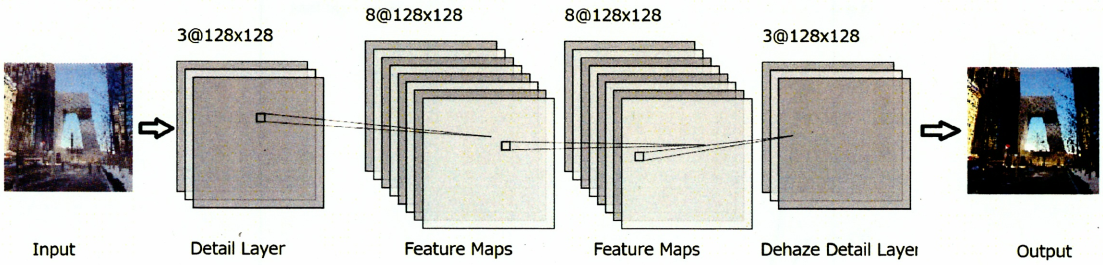


Figure. 5-2 Structure of DhNet image

Neural network structure,

$$f^l(I_{detail}) = \sigma(W^l * f^{l-1}(I_{detail}) + b^l),\ l = 1,2$$
$$f_W(I_{detail}) = W^l * f^{l-1}(I_{detail}) + b^l, \qquad l = 3 \qquad (5\text{-}4)$$

The convolution neural network is composed of three parts. The first part is the input layer, which inputs 128 × 128 size image and three channels of color image. In the second part, two groups of features are designed to extract from the combination of three groups of convolution layer and pooling layer. In the third part, a fully connected multi-layer perceptron is used to overlay the detail layer after restoration, and finally the image is restored.

The convolution method used in feature extraction is shown in figure 5-3. This proposal is originated from semantic segmentation, and the dilated convolution void convolution algorithm is simple and direct, it has achieved quite good results. At the same time, it can accelerate the processing speed.

As following formula 5-5, it is a standard discrete function, when l=1, it become the formula 5-6, and Dilated discrete function. In the formulas, p is the receptive field, and k is the kernel, then s is the stride.

$$(F*k)(p) = \sum_{s+t=p} F(s)k(t) \tag{5-5}$$

$$(F*_l k)(p) = \sum_{s+lt=p} F(s)k(t) \tag{5-6}$$

While performing a convolution process between the feature map and the filter in the convolution layer, the convolution process is performed on a wide range with an interval between the feature maps. It can model correlation with distant areas with few layers and few connections, as the depth increases the receptive field of view at the exponent level increases, and the amount of information increases with each convolution. So the amount of calculation does not increase even though the area of interest expands.

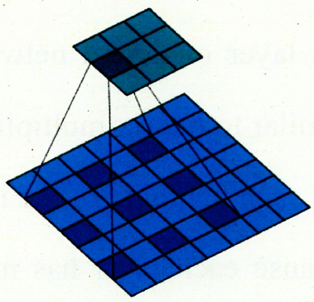


Figure. 5-3 Dilated convolution image

Using DhNet extended convolution, the data can realize from input port to output port, and receptive fields increase exponentially for expanding 1, 2, 4, 8 times. There are three layers in this structure, the dilated values are 1, 2, 4. In fact, not only pixel generation, but also generation from color channels to complex images, as following figure 5-4.

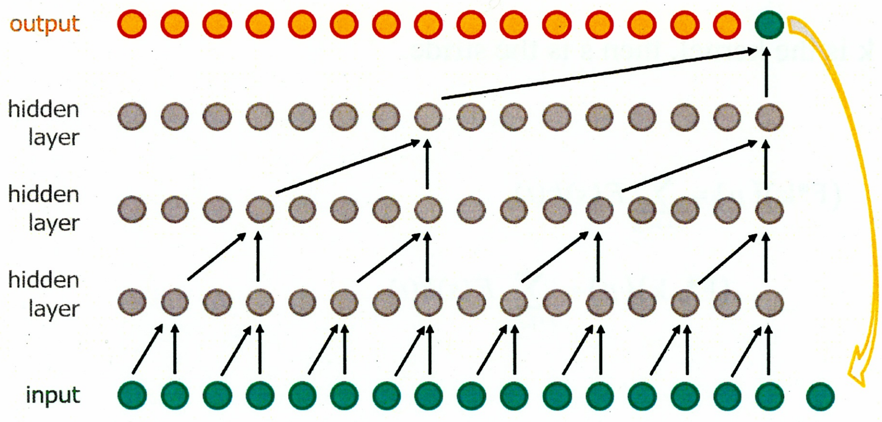


Figure. 5-4 Dilated CNN structure (one-dimensional)

Because the neural network itself can also learn the characteristics that can affect the output results. In this process, most of the features are independent of the results. However, the neural network needs to take these characteristics into account, so there is a case of poor output results. We call this phenomenon over fitting.

In the final analysis, each layer of neural network is a linear or non-linear mapping of the input data, similar to matrix multiplication, each position, or each neuron, has a weight, through such mapping of the input data, we can achieve the desired effect. However, because each layer has many neurons, and the neural network may have many layers. The convolution check image is processed by a fuzzy process, which removes the unimportant extreme data of individual parts, and filters out the feature maps in a small range through fuzzy. With the further deepening of the network structure layer, the features extracted by convolution kernel matrix will become more and more complex. It is necessary to learn the features that have the greatest impact on the results by using the machine itself. With the continuous development of convolution, the characteristic results are obtained gradually. It can be seen that at this time, the machine has ignored the useless information on the edge of the picture and retained the most unique part of the information.

5.4. Training Data

In this study, the main equipment configuration parameters used for training and learning are as follows:

・Intell CoreI i5-7300HQ CPU @ 2.5GHz

・NVIDIA GeForce GTX 1050 Ti

Software as follows:

・OpenCV4.01+Visual Studio2017 （Win10）

C/C++ & Python

・Anaconda3 (jupyter and pycharm)

The database was used, and three sets of data systems were collected or downloaded to form the data set for our research.

Firstly, This study uses a data set called cifar-10 (figure 5-5). This is often used as a data set for image recognition and contains 50000 learning data labelled and 10000 test data. Each image has a size of 32 x 32 pixels and is an image of a landscape, animal or vehicle. Each has 10 different label information (as for a train or an airplane).

Secondly, ImageNet project is a large-scale visual database for the research of visual object recognition software. More than 14 million image URLs are manually annotated by ImageNet to indicate objects in the image; in at least one million images, bounding boxes are also provided. ImageNet contains more than 20000 categories (figure 5-6).

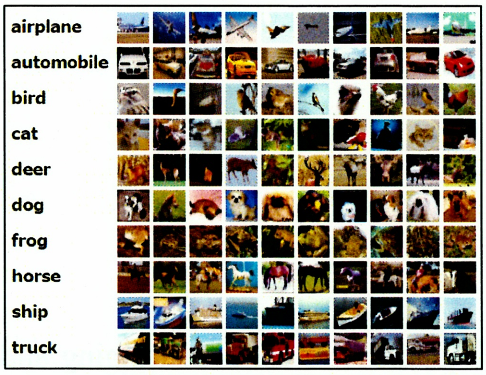


Figure. 5-5 CIFAR-10

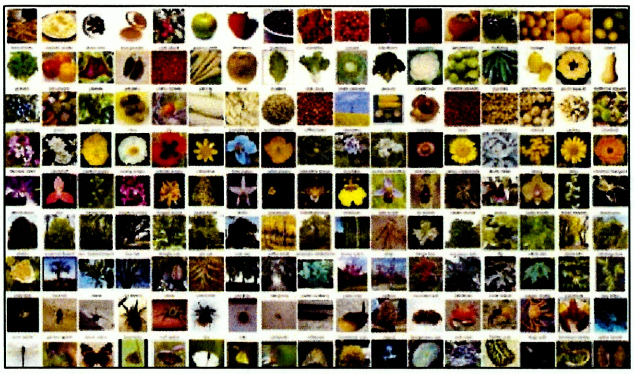


Figure. 5-6 ImageNet

Thirdly, Reside highlights a variety of data sources and image content, then it is divided into five subsets, each of them for different training or evaluation purposes. It provide a variety of evaluation criteria for defogging algorithms, from complete reference measurement, no reference measurement, to subjective evaluation and novel task driven evaluation.

Figure. 5-7 Bench Marking

Data set is composed of two parts:

Firstly, hazy image is generated by Photoshop. (Adding haze to the photos with Photoshop). In order to simulate the diversity of hazy force in real environment, 150 pairs of images with and without haze, different scenes, different backgrounds, different sizes and shapes were generated and sorted out by using self-made software. There are not only outdoor environment, but also indoor scene, day and night.

Secondly, Collect and sort out the hazy image of the real scene in the same industry, or the scene taken by oneself in haze or rainy weather, or entrust the scene taken by friends and students' mobile cameras to transmit the hazy image of their own deterioration vacuum.

5.5. Simulation and Conclusions

SSIM is based on three values between samples X and Y, namely luminance, contrast and structure. SSIM is a measure of similarity between two images, is a number between 0 and 1. The larger the value is, the smaller the gap between the repaired image and the real image is, that is, the better the image quality. When the two images are exactly the same, the value is 1. Suppose that the two input images are X and Y respectively, and the calculation formula is as follows: each time when calculating, take the $n \times n$ window from the image, then slide the window continuously for calculation, and finally take the average value as the global SSIM.

$$\mathrm{SSIM}(x,y) = [l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma}] \quad (5\text{-}7)$$

X and Y represent the average value of X and Y respectively, X and Y represent the standard deviation of X and Y respectively, XY represents the covariance of X and Y, while 1c and 2c are constants respectively, avoiding denominator 0. Set to 1 to get it,

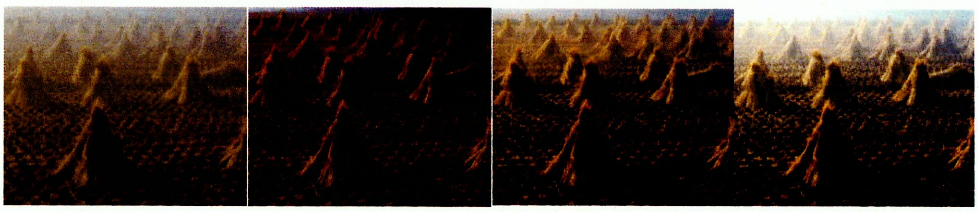
$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (5\text{-}8)$$

According to the correlation between the pixels of the image, the structural similarity between the images is calculated. The larger the value, the smaller the distortion.

The following is to calculate the PSNR of three RGB channels respectively, and then calculate the average value.

Comparison of research proposals:

a



b

c

d

e

f

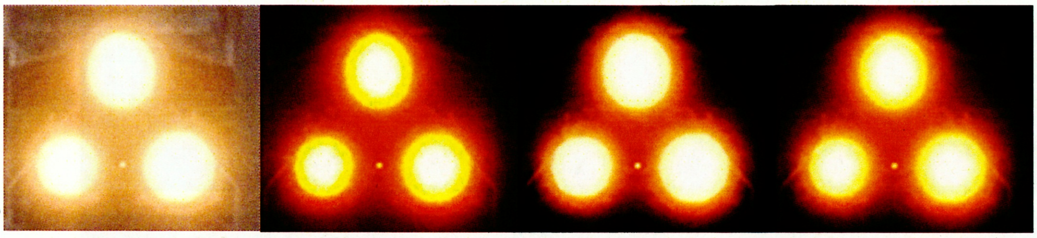
Hazy image DCP DbGAN DhNet

Figure. 5-8 Comparison of three algorithms

g

g'

h

h'


Hazy image DCP DbGAN DhNet

Figure. 5-9 Details images

In the field of digital image processing, the commonly used evaluation indexes include PSNR, SSIM and MSE. The first two categories will be used in this study as peak signal to noise ratio, PSNR values before and after image restoration indicates the ratio of the maximum power of a signal. Generally, the higher PSNR, the better image reconstruction quality. The visual characteristics of human eyes are not considered. Because the sensitivity of the human eye to the contrast difference with lower spatial frequency is higher, and the sensitivity of the human eye to the contrast difference of brightness is higher than that of chroma, the perception result of the human eye to an area will be affected by the adjacent area around it, so the evaluation result is often different from the subjective feeling of the human.

Table 5-1 Comparison of PSNR values of the image in Figure. 5-8

No.	PSNR [dB]		
	DCP	DbGAN	DhNet
a	9.93	10.58	10.73
b	12.74	17.79	16.46

c	14.35	17.76	16.94
d	11.57	13.28	13.72
e	10.85	17.50	17.60
f	9.19	10.70	11.79
g	8.46	13.27	16.39
h	9.56	15.47	19.66
Aver	10.83	14.54	15.41

From the data in Table 5-1, it can be concluded that although the DCP method has high reconstruction quality in theory, the DbGAN method is more consistent with the human visual characteristics, while the PSNR value of DhNet algorithm and DbGAN algorithm is basically close, which shows that the DhNet algorithm is close to the natural scene in terms of dehazing effect.

We have investigated mountains, field, woods, people, Tian'anmen, buildings, goose, train lights, all eight typical haze patterns, we compare the processing results of three dehazing algorithms: DCP, DbGAN and DhNet. Simply from the PSNR values of each algorithm, except for field and woods, the values of other images increase in turn, which shows that the image quality reconstruction is getting better and better. field image basically does not have the problem of special repair of the highlighted area. The hazy image is relatively uniform, the undulation is not obvious, and the background is simple, so DCP or DbGAN can well correspond. Therefore, after using DhNet algorithm, unlike other images, there are obvious numerical changes. In the same way, the processing method of woods and he represented by woods is to deepen the sense effect, specially set the threshold value, and keep some haze areas. After people see it, the sense depth is closer to the reality.

Table 5-2 Comparison of SSIM values of the image in Figure. 5-8

No.	SSIM		
	DCP	DbGAN	DhNet
a	0.77	0.73	0.75
b	0.60	0.52	0.57
c	0.56	0.46	0.51
d	0.85	0.81	0.84
e	0.78	0.80	0.84
f	0.71	0.65	0.68
g	0.82	0.71	0.76
h	0.82	0.71	0.73
Aver	0.74	0.67	0.71

As for SSIM value analysis, it is between 0 and 1. The larger the value is, the better the image quality compared with the original image is. From the SSIM value obtained in the experiment, it is basically about 0.7, that is to say, there is a gap with the original image, but after independent repair by three algorithms, the numerical differentiation is small, only a few percent difference. The specific analysis shows that the SSIM value of most hazy images, DbGAN and DhNet algorithms is generally better than that of DCP, which also proves the superiority of these two algorithms.

Table 5-3 Comparison of SSIM values of the details in Figure. 5-9

No.	Light Image SSIM	Goose Image SSIM
Haze		
DCP	0.9857	0.9916
DbGAN	0.9842	0.9901
DhNet	0.9864	0.9911

Compared with DCP, DbGAN and DhNet are obviously improved by PSNR value after dehazing, which indicates that the processing effect is better. SSIM value is between 0 and 1, This index shows that the larger the value, the greater the similarity of the image.

Table 5-4 Comparison of PSNR values of the details in Figure. 5-9

No.	Light Image PSNR[dB]	Goose Image PSNR[dB]
Haze	-	-
DCP	8.77	8.64
DbGAN	8.51	14.31
DhNet	8.81	11.49

The effect of dehazing is more obvious from the enlarged drawing and the detailed inspection. Their SSIM values are all close to 1, and the goose image is