ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

DSC-Net: Single Image Haze Removal based on Depthwise Separable Convolution

Can Leng[a], Gang Liu[b]

College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610025, China

[a]1252903096@qq.com, [b]liugang@cuit.edu.cn

Abstract

Single image haze removal is an extremely difficult problem. As a hot topic in the field of computer vision, deep learning has been used to get plausible dehazing solutions. Dehaze-Net is one of the popular deep learning models proposed for single image dehazing, with excellent performance while maintaining efficiency and ease of use. However, in complex scenes, Dehaze-Net can't fully extract the boundary information. In this paper, we propose a new model called DSC-Net, which can retain the advantages of Dehaze-Net while overcoming its shortcomings to some extent. DSC-Net uses multiple groups of convolutions and maxout unit which can generate almost all haze-relevant features. And the depthwise separable convolution in DSC-Net effectively reduces the number of parameters. The numerical experimental results show that DSC-Net performs better than other similar models, e.g., Dehaze-Net, PD-Net, especially in the large sky area of the image and detail processing.

Keywords

DSC-Net; Image Dehazing; Deep CNN; Dehaze-Net; Depthwise Separable Convolution.

1. Introduction

The information in the outdoor image is usually covered by the turbid medium in the atmosphere. Haze, smoke, and fog [1] are such phenomena caused by atmospheric absorption and scattering. Due to the existence of smoke, dust, and other floating particles in the atmosphere, images taken in such an environment are often subject to low contrast, color distortion, and other visible quality degradation. And the hazy image input will make it difficult to solve visual tasks. Therefore, haze removal has great significance in photography and computer vision applications.

Haze removal is a challenging problem because haze transmission depends on unknown depths at different locations. According to the constraints of haze removal, many methods have been proposed by using additional information or multiple images, including the adaptive histogram equalizationmethod [2] and the saturation based method [3]. In [4], the author proposes the method of image homomorphic filtering. The single-scale Retinex image enhancement algorithm is proposed in [5]. The methods based on image enhancement have the characteristics of high contrast, uniform brightness, and great visual effects, but generally have poor effects on detail processing.

To better describe the constraint of a hazy image, the atmospheric scattering model was first proposed by McCartney [6], and further developed by Narasimhan [7]. The formula of the atmospheric scattering model can be written as:

$$I(x) = J(x)t(x) + \alpha(1 - t(x)) \tag{1}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

where $I(x)$ is the observed hazy image, $J(x)$ is the real scene to be recovered, $t(x)$ is the medium transmission, α is the global atmospheric light, and x indexes pixels in the observed hazy image I.

Most of the papers are applied to atmospheric scattering (1). He et al. [8] propose the dark channel prior (DCP) to calculate the transmission matrix. Tang et al. [9] use four types of haze-relevant features to estimate the transmission. Matan Sulami et al. [10] propose an accelerated method for the automatic recovery of atmospheric light. Deep learning has received more and more attention in single image dehazing. Cai et al. [11] propose Dehaze-Net to estimate the value of α through the dark channel prior [8], so only the value of $t(x)$ needs to be solved. Song et al. [12] introduce the pyramid pooling module [13] in the Dehaze-Net model and add a residual block. Golts et al. [14] propose a deep neural network that uses DCP loss for unsupervised training. Other convolutional neural network models also achieve good dehazing effects [15, 16, 17, 18]. In addition, the generative adversarial network (GAN). in recent years has achieved remarkable success in image restoration [19, 20].

Dehaze-Net [11] shows good performance in image dehazing, with high contrast and without obvious noise after dehazing, and it has a compact structure and requires only a small amount of memory to run. But in some outdoor pictures, Dehaze-Net is ineffective in dense haze areas with large areas of color distortion in the sky. Inspired by Dehaze-Net, we propose a haze removal model called DSC-Net (Depthwise separable convolution Dehaze-Net), which can retain the advantages of Dehaze-Net and suppress its shortcomings. DSCNet first takes haze images as input, and then extracts features from multiple scales of the image. To some extent, it can enable the network to extract image features more comprehensively. And we introduce depthwise separable convolution [21] to change the number of channels. The network structure of DSC-Net effectively reduces the number of parameters and enhances the ability of the model to extract global information from the image. At the same time, the ReLU function [22] is used as the activation function as we find it more effective than the BReLU activation proposed by [11], in our unique setting. DSC-Net outputs its medium transmission map, which is used to recover the haze-free image. After getting the medium transmission map, due to the existence of a local extremum, the blocking artifacts will appear in the transmission map obtained from DSC-Net. To refine t(x), we use guided filtering [23] to smooth the image. The smoothed image can effectively eliminate blocking artifacts and retain edge information. The numerical experimental results show that DSC-Net achieves great performance while remaining efficient and easy to use.

The rest of the paper is organized as follows: In Section 2, we provide background knowledge to understand the design of DSC-Net. In Section 3, we show the details of the proposed DSC-Net. Experiments are presented in Section 4, and conclusions are drawn in Section 5.

2. Background

According to the atmospheric scattering model, solving the medium transmission map is the most critical step. In this section, we briefly introduce some related works and present the layer designs of DSC-Net. DSC-Net first uses Maxout unit [24] in multi-scale feature extraction, then it uses convolution kernels of different sizes to conduct multi-scale mapping and select the max pooling to retain the main features. The depthwise separable convolution [21] in subsequent networks is used to reduce parameters. Rectified Linear Unit (ReLU) is proposed [22] to overcome the problem of vanishing gradients. In the last layer of DSC-Net, ReLU is used as a nonlinear activation function.

2.1 Dark Channel Prior

The dark channel prior is based on extensive observation of haze-free outdoor images. In most of the haze-free patches, at least one color channel has some pixels whose intensity is very low or close to zero. To formally describe this observa tion, the dark channel [8] is defined as the minimum of all pixel colors in a local patch:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} J^c(y) \right) \tag{2}$$

where $J(c)$ is a color channel of an arbitrary image J, and $\Omega(x)$ is a local patch centered at = x.Two minimum operators are obtained from a dark channel: $\min c \in \{r,g,b\}$ is performed on each pixel, and $\min y \in \Omega(x)$ is a minimum filter.

2.2 Maxout Unit

Maxout unit [24] is actually an excitation function. For the input feature vector $x = (x1, x2, ..., xd)$, the calculation formula of Maxout hidden layer neurons can be expressed as:

$$z_{i,j} = x^T W_{i,j} + b_{i,j}, W \in R^{d \times m \times k} \tag{3}$$

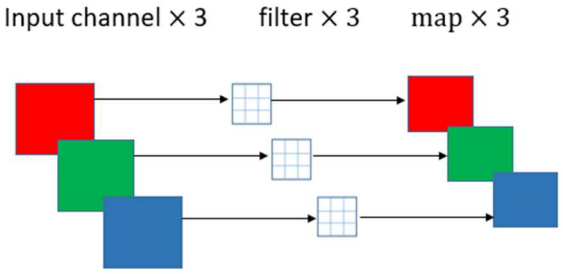
$$h_i x = \max_{j \in [1,k]} z_{i,j} \tag{4}$$

Where d represents the number of nodes in the input layer, m represents the number of nodes in the hidden layer, k represents the node corresponding to each hidden layer node, b is the parameter we need to learn, a two-dimensional matrix with the size of $m \times k$, and maxout selects the maximum activation value as the activation value of neurons in the next layer. When we use the maxout unit in our module, the number of parameters will increase, but the advantages of the maxout unit are also obvious. Its fitting ability is very strong, and it can fit arbitrary convex functions. In the process of feature extraction, we are equivalent to using an appropriate filter to convolve the input haze image and then conducting nonlinear mapping. We select convolution kernels of different sizes in the model to extract features, use the maxout unit to reduce 16 dimensions to 4 dimensions to achieve dimension reduction processing, and then splice the reduced arrays to effectively extract relevant features.

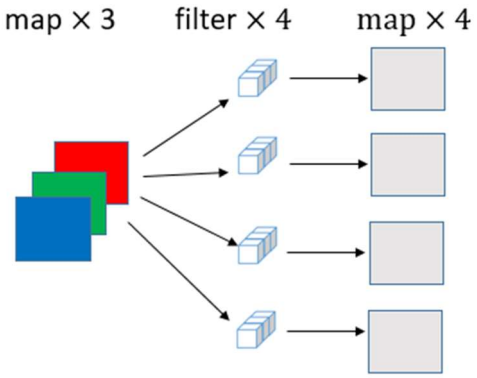
2.3 Depthwise Separable Convolution

Depthwise separable convolution (DSC) [21] is mainly divided into two processes, namely, depthwise convolution and pointwise convolution. The architecture of depthwise separable convolution is shown in Figure 1. For depthwise convolution, one convolution kernel of the depth convolution is responsible for one channel, and one channel is convolved by only one convolution kernel. In this process, the number of feature map channels is exactly the same as the number of input channels. The size of the convolution kernel of pointwise convolution is $1 \times 1 \times n$, n is the number of channels on the upper layer. Therefore, the convolution operation will combine the maps in the depth direction to generate a new feature map. And the number of convolution kernels is equal to the number of output feature maps. We chose depthwise separable convolution because it can significantly improve efficiency without sacrificing the performance of the model. Compared with ordinary convolution, it changes the consideration of both channels and regions to only consider regions first and then consider channels. The characteristics of depthwise separable convolution enable it to effectively reduce parameters.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031



(a) Depthwise Convolution



(b) Pointwise Convolution

Figure 1. The architecture of Depthwise separable convolution.

3. The Proposed DSC-NET

3.1 Layer Designs of DSC-NET

The proposed DSC-Net is composed of convolutional and pooling layers, with appropriate nonlinear activation functions used after these layers. Figure 2 shows the architecture of DSC-Net. We can easily recover the haze-free image $J(x)$ while the values of the medium transmission $t(x)$ and the atmospheric light α are given. Equation (1) can be written as follows:

$$J(x)=\frac{I(x)-\alpha(1-t(x))}{t(x)} \tag{5}$$

Densely extracting haze-relevant features is equivalent to convolving an input hazy image with appropriate filters, followed by nonlinear mappings. In order to extract as many image-related features as possible, we chose five different convolution kernels for feature extraction. The size of any convolution filter is among 1×1, 3×3, 5×5, 7×7, 9×9, and we select replicate padding to enhance boundary information extraction. Then the Maxout unit [24] maps each of kn1- dimensional vectors into an n1-dimensional one and extracts the haze-relevant features by automatic learning rather than heuristic ways in existing methods.

After feature extraction, we concatenate features extracted by convolution kernels of different sizes according to dimensions and conduct multi-scale mapping and maxpooling. Multi-scale features have been proven effective for haze removal in [9], which densely compute features of an input image at multiple spatial scales. Multi-scale feature extraction is also effective to achieve scale invariance. We use parallel convolutional operations where the size of any convolution filter is among 3×3, 5×5 and 7×7, and the number of filters for these three scales is the same. Then we use maxpooling to retain the main features while reducing the parameters and computation. At the same time, it can also prevent overfitting and improve the generalization ability of the model.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

The added depthwise separable convolution considers only the region first and then the channel, effectively reducing the parameters required by the model. And the parameters of DSC-Net are shown in Table 1.

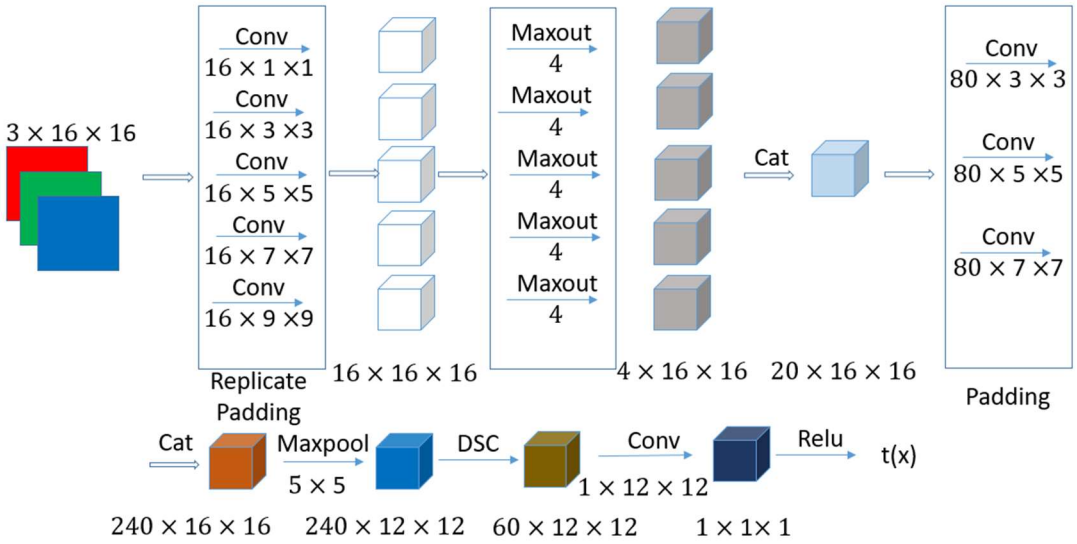


Figure 2. The architecture of DSC-Net.

Table 1. The parameter of DSC-Net

Type	Input size	num	Kernel size	padding	Output size
conv1	3 × 16 × 16	16	3 × 1 × 1	0	16 × 16 × 16
conv2	3 × 16 × 16	16	3 × 3 × 3	1	16 × 16 × 16
conv3	3 × 16 × 16	16	3 × 5 × 5	2	16 × 16 × 16
conv4	3 × 16 × 16	16	3 × 7 × 7	3	16 × 16 × 16
conv5	3 × 16 × 16	16	3 × 9 × 9	4	16 × 16 × 16
Maxout	16 × 16 × 16	4	/	0	4 × 16 × 16
cat1	4 × 16 × 16	5	/	/	20 × 16 × 16
conv6	20 × 16 × 16	80	20 × 3 × 3	1	80 × 16 × 16
conv7	20 × 16 × 16	80	20 × 5 × 5	2	80 × 16 × 16
conv8	20 × 16 × 16	80	20 × 7 × 7	3	80 × 16 × 16
cat2	80 × 16 × 16	3	/	/	240 × 16 × 16
Maxpool	240 × 16 × 16	/	5 × 5	0	240 × 16 × 16
DSC	240 × 16 × 16	60	240 × 3 × 3	1	60 × 12 × 12
conv9	60 × 12 × 12	1	60 × 12 × 12	0	1 × 1 × 1

Layers of DSC-Net are cascaded together to form a trainable system, where biases and filters associated with convolutional layers are network parameters to be learned. The first layer of DSC-Net is designed for feature extraction. Maxout activation functions can be considered a piecewise linear approximation of any convex function. In the second layer, DSC-Net concatenates features from the previous five convolutional layers and uses three convolutional layers to form multi-scale features by fusing filters of varied sizes. And then DSC-Net concatenates features again before the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

maxpooling layer. Such a multi-scale design captures features at different scales and also compensates for the information loss during convolutions. In the last layer, the introduced depthwise separable convolution and ordinary convolution can effectively reduce parameters. We utilize the relu activation function to get $t(x)$ as we find it more effective than others in our model.

3.2 Training and Dehazing of DSC-NET

Pairs of hazy and haze-free images of natural scenes are not widely available. Instead, we synthesize training data based on the physical haze formation model [9]. We collect haze free images from the Internet and randomly sample from them patches of size 16×16. More than 14,800 patches are obtained from the sampling process. Since the media transmission is locally invariant, we can assume arbitrary transmission for individual image patches. Given a haze-free patch $J_{P(x)}$, the atmospheric light α, we divide (0, 1) into ten intervals on average, and randomly select $t(x)$ from the ten intervals for each haze addition. We can synthesize a hazy patch as:

$$I_{P(x)} = J_{P(x)}t(x) + \alpha(1-t(x)) \tag{6}$$

The atmospheric light α is set to 1 to reduce the uncertainty in variable learning. Therefore, more than 148,000 synthetic patches are generated by the process of adding haze. We select the Adma optimizer and use Mean Squared Error (MSE) asthe loss function. The final training data is used in the model. The parameters of the model are saved.

We use the trained model on the test set. Guided filter [23] is used to refine $t(x)$ for the blocking artifacts that appear in it. According to the atmospheric scattering model (1), when $t(x)$ approaches 0, the input image is approximately equal to atmospheric light. And when $t(x)$ approaches 1, the input image is approximately equal to a clean image. Obviously, the above two situations are not reasonable when we remove haze from the input hazy image to restore a clean image. Therefore, we limit the range of $t(x)$ by taking the lower limit when $t(x)$ approaches 0 and the upper limit when $t(x)$ approaches 1. Relying on the dark channel prior (DCP), we select the pixel with the largest pixel value in the top one percent of the figure. We also select the pixel with the highest pixel value as atmospheric light α. Based on $t(x)$ estimated by DSC-Net and α, haze-free images are restored through the atmospheric scattering model.

4. Experiments

To verify the effectiveness of complete images, we collect haze-free images from the RESIDE data set and use haze of different concentrations to randomly synthesize the haze image. All the algorithms and models are implemented on a PC with an Nvidia GeForce RTX 3060 GPU, 12GB of video memory, and a Windows 10 x64 system. And we compile our code using third-party libraries such as PyTorch, OpenCV, Numpy, etc.

Table 2. Quantitative results of quality evaluation indicators on RESIDE data set

Evaluation indicators	SSR [5]	DCP [8]	Dehaze-Net [11]	PD-Net [12]	DSC-Net
PSNR	9.9564	21.6628	18.9989	20.4336	20.6801
SSIM	0.6918	0.9198	0.8476	0.8700	0.8753

To compare the advantages and disadvantages of each algorithm, we compare DSC-Net with other algorithms, including SSR [5], DCP [8], Dehaze-Net [11] and PD-Net [12]. Among them, SSR is mainly based on image enhancement, and DCP is mainly based on the physical model. DSC-Net is a

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

method based on deep learning that has a similar network structure to Dehaze-Net and PD-Net, which can better see the advantages of DSC-Net. We select Adam optimizer in the model and set the batch size value to 128.

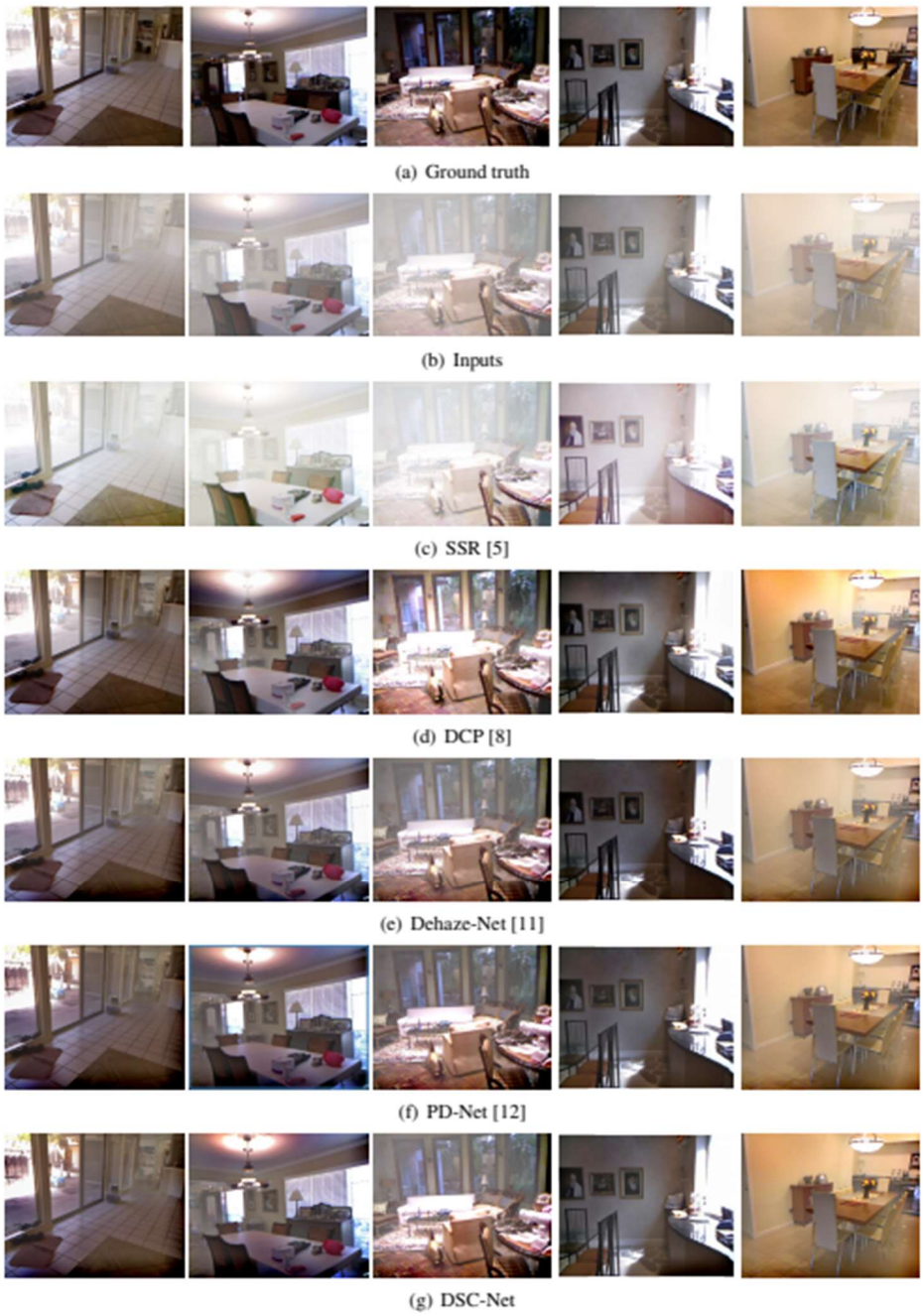


Figure 3. The haze removal results on indoor synthetic image.

For all pictures on the test set, we use the peak signal to-noise ratio (PSNR) and the structural similarity (SSIM) [25] as the full reference image quality evaluation indicators. PSNR can be expressed as:

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{(2^n - 1)^2}{MSE}\right) \tag{7}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

And the evaluation data is shown in Table 2. From the results obtained, it can be seen that the DSC-Net model has a good average SSIM value and PSNR value on the experimental data set.

Figure 4. The haze removal results on outdoor real haze image.

We chose the RESIDE data set for our experiment because it has very comprehensive images, including outdoor real hazy images and indoor simulated hazy images. We show the quantitative results on synthetic images of DSC-Net and other algorithms on the RESIDE data set in Figure 3. It can be seen from Figure 3 (c) that the picture after SSR algorithm dehazing is white for the whole image, and there is overexposure. From Figure 3 (d), we can see that the image after dehazing by the DCP algorithm has color distortion and low contrast. Figure 3 (e) and Figure 3 (f) show that the image details of them after dehazing recover well, but there is also color distortion.

To further analyze the dehazing effect of the DSC-Net model. The qualitative results on real-world images are shown in Figure 4. As we can see from Figure 4 (b), the SSR algorithm causes a large number of blank areas in the picture, and haze removal is incomplete. From Figure 4 (c), the DCP algorithm has color distortion in a large area of the sky. The Dehaze-Net model and PD-Net model in Figure 4 (d) and (e) have a good dehazing effect, but some sky areas in the picture also show color

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

distortion after dehazing, and the overall brightness of the pictures is a little dark. From Figure 4 (f), it can be seen that the picture after dehazing by the DSC-Net model is clear and bright, without obvious color distortion. The details of the enlarged picture of Figure 4 (f) are also very good, and the overall visual effect is better. We enlarge the lower right corner of the rightmost image in Figure 4, and we can see from Figure 5 that compared with other methods, DSC-Net has better visual quality. In general, the synthetic and real-world image results after DSC-Net processing are clear, without obvious color distortion, and the details are more consistent with human perception.

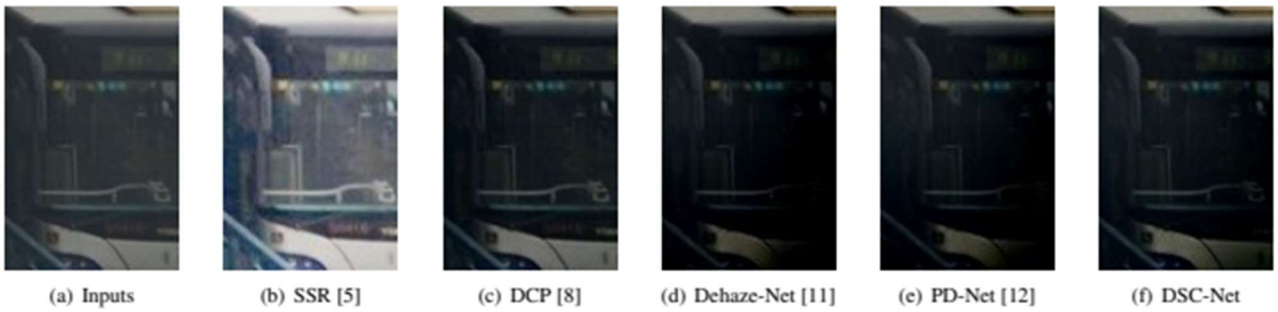


Figure 5. Visual quality comparison between DSC-Net and other methods on a hazy image.

5. Experiments

Aiming at the problem of single image haze removal, we propose DSC-Net using multiple feature extraction blocks and depthwise separable convolution. Multi-scale feature extraction effectively extracts relevant features, and the depthwise separable convolution effectively reduces parameters.

The experimental results in the RESIDE data set show that the images after the DSC-Net dehazing are clear and bright, and the processing effect is good in large areas of empty space. We compare DSC-Net with a variety of similar methods on both natural and synthetic haze images, using both subjective and objective (PSNR, SSIM) criteria. Experimental results confirm the superiority and efficiency of DSC-Net.

Acknowledgments

Thank the anonymous reviewers for their valuable comments.

References

[1] S.G. Narasimhan and S.K. Nayar:Vision and the Atmosphere, Int'l J. Computer Vision, vol. 48(2002), p. 233-254.

[2] S.M. Pizer, E.P. Amburn, J.D. Austin, et al., "Adaptive histogram equalization and its variations", Computer vision, graphics, and image processing, vol.39(1987), p. 355-368.

[3] J.E. McDonald:The saturation adjustment in numerical modelling of fog, Journal of the Atmospheric Sciences, vol.20(1963), p.476-478.

[4] H.G. Adelmann,:Butterworth equations for homomorphic filtering of images, Computers in Biology and Medicine, vol.28(1998), p.169-181.

[5] G.D. Finlayson, S.D. Hordley, M.S.Drew:Removing shadows from images using retinex, In Color and imaging conference , vol. 2002(2002), No. 1, p. 73-79.

[6] E.J. McCartney, Optics of the Atmosphere: Scattering by Molecules and Particles, vol.1(1976). New York, NY , USA: Wiley.

[7] S.G. Narasimhan and S.K. Nayar:Contrast restoration of weather degraded images, IEEE Trans. Pattern Anal. Mach. Learn., vol.25(2003), no.6, p.713–724.

[8] K.He, J.Sun, and X.Tang:Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell., vol. 33(2011), no. 12, p.2341–2353.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0031

[9] K. Tang, J. Yang, and J. Wang:Investigating haze-relevant features in a learning framework for image dehazing, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2014, p. 2995–3002.

[10] M. Sulami, I. Glatzer, R. Fattal, and M. Werman: Automatic recovery of the atmospheric light in hazy images, in Proc. IEEE Int. Conf. Comput. Photogr. (ICCP), May 2014, p. 1–11.

[11] B.Cai, X.Xu, K.Jia, et al.Dehazenet: An end-to-end system for single image haze removal, IEEE Transactions on Image Processing, vol.25(2016),no.11, p.5187-5198.

[12] Y.Song, G.Liu:PD-net: Improved dehaze-net based on pyramid pooling module,Advances in Applied Mathematics, vol.10(2021), p.3351.

[13] H.Zhao, J.Shi, X.Qi, et al.Pyramid scene parsing network, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp.2881-2890.

[14] A.Golts, D.Freedman, M.Elad:Unsupervised single image dehazing using dark channel prior loss, IEEE transactions on Image Processing, vol.29(2019), p.2692-2701.

[15] W.Ren, S.Liu, H.Zhang, et al.Single image dehazing via multi-scale convolutional neural networks, Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing, p.154-169.

[16] B.Li, X.Peng, Z.Wang, et al.An all-in-one network for dehazing and beyond, arXiv preprint arXiv: 1707.06543, 2017.

[17] X.Qin, Z.Wang, Y.Bai, et al.FFA-Net: Feature fusion attention network for single image dehazing, Proceedings of the AAAI conference on artificial intelligence. vol.34(2020),no.07, p.11908-11915.

[18] H.Wu, Y.Qu, S.Lin, et al.Contrastive learning for compact single image dehazing, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. p.10551-10560.

[19] X.Liu, Z.Gao, B.M.Chen:MLFcGAN: Multilevel feature fusion-based conditional GAN for underwater image color correction, IEEE Geoscience and Remote Sensing Letters, vol.17(2019),no.9, p.1488-1492.

[20] Q.Deng, Z.Huang, et al.Hardgan: A haze-aware representation distillation gan for single image dehazing, Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16. Springer International Publishing, pp.722-738.

[21] L.Sifre, S.Mallat:Rigid-motion scattering for texture classification, arXiv preprint arXiv:1403.1687, 2014.

[22] V.Nair, G.E. Hinton:Rectified linear units improve restricted boltzmann machines, Proceedings of the 27th international conference on machine learning (ICML-10), 2010, p.807-814.

[23] K. He, J. Sun, and X. Tang,:Guided image filtering, IEEE Trans.Pattern Anal. Mach. Intell., vol.35(2013), no. 6, p.1397–1409.

[24] I.Goodfellow, D.Warde-Farley, M.Mirza,et al.Maxout networks, in Proc. 30th Int. Conf. Mach. Learn. (ICML), 2013, p. 1319–1327.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli:Image quality assessment: From error visibility to structural similarity, IEEE Trans.Image Process., vol. 13(2004), no. 4, p. 600–612.