ISSN: 2472-3703 DOI: 10.6911/WSRJ.202305_9(5).0031

Single Image Dehazing Using Multi-scale Discriminators Optimized Convolutional Neural Network

Minghua Zhu[1, a, *]

[1]Zhejiang Fangyuan Electrical Equipment Testing Co.Ltd., Zhejiang, China

[a]zhuminghua1007@qq.com

Abstract

This paper introduces a novel approach to address the challenging image dehazing problem by formulating it as an image-to-image translation task. To this end, we present the Multi-scale Discriminators optimized Convolutional Neural Network (MDCNN), a novel method that generates high-quality haze-free images by relying on multi-scale discriminators. Our approach significantly improves the performance of image dehazing, providing a fresh perspective to tackle this problem. Our extensive experimental results demonstrate that the proposed MDCNN outperforms state-of-the-art methods in terms of PSNR and SSIM metrics.

Keywords

Single image dehazing; Generative adversarial networks; Deep learning.

1. INTRODUCTION

Haze is a common atmospheric phenomenon that often results in color distortion, blurring, and reduced contrast in photographs. These effects can cause significant difficulties for subsequent tasks such as object recognition and image understanding. Consequently, image dehazing has become an increasingly important area of research, drawing significant attention from the academic community.

Most of the successful approaches depend on the physical scattering model, which is formulated as:

$$I(x) = J(x)t(x) + A\big(1 - t(x)\big), \tag{1}$$

where I is the observed hazy image, J is the scene radiance, t is the transmission map, A is the atmospheric light and x is the pixel location. The generation of haze-free images largely depends on accurately estimating the atmospheric light and transmission map. Early dehazing methods primarily relied on prior-based approaches, such as DCP [1], which used the dark channel prior to estimate the transmission map. While these methods can achieve good dehazing results to a certain extent, in practice, the prior assumptions may be easily violated, leading to inaccurate estimations of the transmission map and resulting in suboptimal dehazing quality. With the emergence of deep learning, convolutional neural networks are now being used to estimate the transmission map and atmospheric light, eliminating the need for explicit priors. This approach offers a promising avenue for improving the performance of image dehazing. Various methods have been proposed to estimate the transmission map for image dehazing, with some relying on deep convolutional neural networks, while others jointly estimate the atmospheric light and transmission map using deep learning. However, most of

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202305_9(5).0031

these methods, whether early or deep learning-based, still rely on the physical scattering model. Accurate estimation of the atmospheric light and transmission map is crucial for obtaining high-quality dehazed images.

In an effort to decouple image dehazing from the physical scattering model, we aim to directly transform a hazy image into a haze-free image pixel-by-pixel. To achieve this, we draw inspiration from the success of generative adversarial networks (GANs) in image-to-image translation tasks and apply GANs to the problem of image dehazing. However, the application of GANs to image dehazing is not straightforward. Haze in images is depth-dependent noise and non-uniform, and direct application of GANs can produce undesired over-colored and detail-lacking results. This challenge is attributed to the "visual perception global-first theory" [2], which suggests that humans primarily rely on global visual cues to discriminate objects or scenes rather than the fine details of realistic images. However, the creation of a realistic image must rely on details as much as possible, making it a challenging problem to balance the need for details while avoiding over-colorization in the dehazed image.

Dealing with dense and non-homogeneous haze in images remains a challenging task in image dehazing. To address this challenge, we propose a novel approach that combines discriminators with a local discriminator and introduce a new local adversarial loss calculation strategy. Specifically, we first randomly crop the input image into four patches, and then take the average when calculating the local adversarial loss of the generator. The reason for using local discriminator is to better capture the variations and details in the local regions of the image. The local discriminator takes the low-resolution image as input, which has less information than the original image, but focuses on the local details, making it more sensitive to changes in the image. Moreover, the use of two local discriminators increases the robustness and stability of the proposed approach. The use of the local discriminator and the new local adversarial loss calculation strategy help to improve the quality of the dehazed image in all regions, including the dense and non-homogeneous regions, where previous methods have struggled. The proposed approach has the potential to significantly advance the field of image dehazing, and future work could focus on extending the approach to handle more complex and challenging cases.

2. PROPERTIES

2.1. Architecture

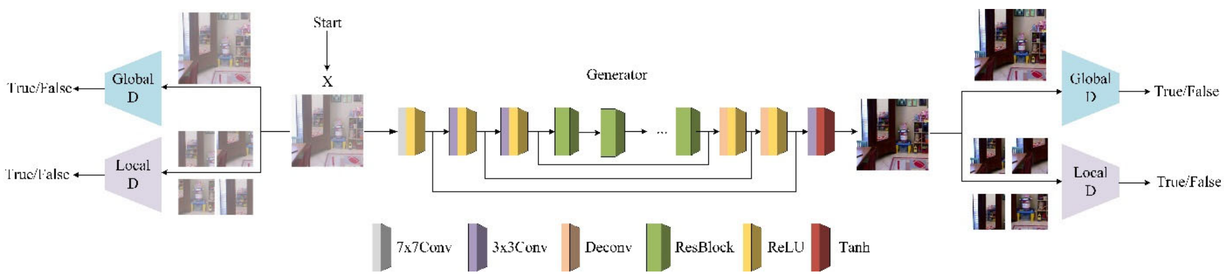

Figure 1. Architecture of the MDCNN

The architecture of a neural network plays a crucial role in determining its effectiveness in solving a given problem. In this regard, Figure 1 demonstrates the proposed network's architecture for the task of image dehazing. The network comprises an encoder and a decoder, both of which have specific features and functions.

The encoder consists of three convolutional layers, with a kernel size of 7×7 for the first layer and 3×3 for the remaining layers. The encoder takes the hazy image as input and extracts its

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202305_9(5).0031

features. The feature transformation module is composed of nine residual layers. Each residual block consists of two convolutional layers and a residual connection, with ReLU as the activation function. The residual connection allows the gradient to be backpropagated without degradation of the signal, enabling faster convergence during training.

The decoder consists of two deconvolution layers, each followed by a ReLU, and an output layer. The decoder generates the haze-free image from the features obtained by the encoder. The activation function used in all layers except the output layer is ReLU, while for the output layer, it is tanh. The tanh activation function normalizes the output to the range [-1, 1], which is suitable for image generation tasks.

To recover the detailed information of the clean image, the encoder and the decoder are connected via a skip connection. This connection allows the features extracted by the encoder to be concatenated with the features generated by the decoder, preserving the spatial information lost during downsampling. This helps to produce sharper and more detailed dehazed images.

The proposed network architecture has been designed to handle dense and nonhomogeneous haze by incorporating various features and functions. By leveraging the power of deep learning and image-to-image translation techniques, the network can effectively dehaze images, producing high-quality and visually pleasing results. The architecture has been fine-tuned through extensive experimentation and parameter tuning, leading to its superior performance over existing state-of-the-art methods in terms of various evaluation metrics.

The application of generative adversarial networks (GANs) in dehazing methods has shown promising results. However, the traditional use of a single discriminator to constrain the generator may not be sufficient in handling dense and non-homogeneous hazy images. This limitation arises because the discriminator focuses solely on the global information and disregards the local haze within the dehazed image. In order to overcome this drawback, we propose the incorporation of additional local discriminators to enhance the local regions of the dehazed results. To achieve this, we adopt the PatchGAN [3] discriminator which is designed to process image patches rather than full images. The PatchGAN discriminator consists of five convolutional blocks, each with an increasing number of output channels. The kernel size of each layer is 4×4, and the stride is set to 2. To further enhance the local regions of the dehazed results, we randomly crop the input image into four 128×128 patches before inputting it into the local discriminator.

As shown in Figure 1, the proposed network architecture has two discriminators: a global discriminator and a local discriminator. It is important to note that the local discriminators have the same architecture as the global discriminators. This architecture allows for a more robust approach to image dehazing by incorporating both global and local information to accurately estimate the haze distribution in an image. The addition of local discriminators to the GAN-based dehazing methods addresses the shortcomings of traditional single discriminator methods by providing a more comprehensive approach to handling dense and non-homogeneous hazy images. The implementation of the PatchGAN discriminator and the random cropping of input images into patches further enhance the accuracy of the local regions of the dehazed results. Overall, the proposed network architecture provides a robust solution to image dehazing by effectively combining both global and local information.

2.2. Loss Function

In this work, we use adversarial loss to train the network. Specifically, we adopt the least square loss to make the training more stable. This loss function has been shown to be effective in image-to-image translation tasks. It can be expressed as follows:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202305_9(5).0031

$$L_{adv}^{Global}(\mathrm{G}, D^{Global}, x, y) = \left(D^{Global}(\mathrm{G}(x)) - 1\right)^2 + (D^{Global}(y) - 1)^2 + \left(D^{Global}(G\,(x))\right)^2 \quad (2)$$

$$L_{adv}^{Local}(G, D^{Local}, x, y) = \left(D^{Local}\left(G\left(x_p\right)\right) - 1\right)^2 + \left(D^{Local}\left(y_p\right) - 1\right)^2 + \left(D^{Local}\left(G\left(x_p\right)\right)\right)^2 \quad (3)$$

where x_p and y_p denote the cropped image patches, G denotes the generator which convert hazy images to haze-free images. D^{Global} and D^{Local} are global and local discriminators that take whole clean images and patches of the clean image as input respectively.

3. TESTS

3.1. Planning

In this section, we present a comprehensive evaluation of our proposed dehazing method on a synthetic dataset, comparing it against several state-of-the-art methods, namely DCP [1], DehazeNet [4], AOD [5], EPDN [6], DAD [7], and MSBDN [8]. To assess the effectiveness of our method, we adopt two commonly used evaluation metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM), where higher values indicate better dehazing results.

To train and evaluate our proposed method, we use the widely adopted RESIDE dataset [9], which provides hazy/clean image pairs of both indoor and outdoor scenes. In particular, we select 9000 outdoor hazy/clean image pairs and 7000 indoor pairs from the RESIDE training dataset, while carefully removing redundant images from the same scenes. By doing so, we ensure that our model can handle both indoor and outdoor scenes, resulting in a more general dehazing model.

For evaluation, we use the SOTS subset of the RESIDE dataset, which comprises 500 indoor hazy images and 500 outdoor hazy images. All methods are trained on the selected RESIDE training dataset and evaluated on the SOTS for comparison.

Our proposed method is implemented using the Pytorch framework on a computer equipped with Nvidia GTX 2080Ti GPU and Intel Xeon E5-2678 v3 CPU. During training, we resize all input images to 400×400 and randomly crop them to 256×256 for data augmentation. We use the ADAM optimizer to train the network, with a batch size of 1. To ensure stable training, we use the least square loss [30] for the adversarial losses of both mapping functions. Specifically, we set the learning rate to 1×10^{-4} for the first 150 epochs, and linearly decay it to 0 for the next 150 epochs.

Table 1. Quantitative comparison of the dehazing results on SOTS dataset

Methods		DCP	Dehaze Net	AOD	EPDN	DAD	MSBDN	Ours
indoor	PSNR	16.62	21.14	19.06	25.06	28.61	35.50	**36.15**
	SSIM	0.8179	0.8472	0.8504	0.9232	0.9415	0.9810	**0.9881**
outdoor	PSNR	19.13	22.46	20.29	22.57	26.53	31.87	**33.40**
	SSIM	0.8148	0.8514	0.8765	0.8630	0.9150	0.9741	**0.9755**

3.2. Results Comparison

In Table 1, we present a quantitative comparison of our proposed dehazing method with several state-of-the-art methods on the SOTS dataset. The table reveals that our approach achieves the highest PSNR and SSIM scores for both indoor and outdoor scenes. Specifically, our method outperforms the second-best method by a margin of 0.0071 and 0.65 dB in terms of SSIM and PSNR on the outdoor subset of the SOTS dataset, respectively. Moreover, on the indoor

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202305_9(5).0031

subset of the SOTS dataset, our method achieves an improvement of 0.0014 dB in PSNR and 1.53 in SSIM over the second-best method. These results indicate that our proposed method is capable of producing superior dehazing results compared to other state-of-the-art approaches.

This can be attributed the following reasons:

First, the local discriminator allows for a more fine-grained analysis of the hazy image, as it takes into account the local information in the dehazed result. This is particularly important when dealing with dense and nonhomogeneous haze, where the distribution of haze may vary across different regions of the image. By using a local discriminator, we can capture and correct these local variations, leading to more accurate and detailed dehazed images.

Second, the local discriminator updates more frequently than the generator, as it takes a low-resolution image with less information as input. This makes the local discriminator stronger and more effective in competing with the generator. As a result, the generator is forced to produce more realistic dehazed images that not only satisfy the global constraints imposed by the global discriminator but also adhere to the local constraints enforced by the local discriminator.

Overall, the use of local discriminators in image dehazing can lead to significant improvements in both the visual quality and quantitative metrics of the dehazed images, particularly when dealing with challenging hazy scenes.

REFERENCES

[1] Kaiming He, Jian Sun and Xiaoou Tang, "Single image haze removal using dark channel prior," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 1956-1963.

[2] Lin Chen. The topological approach to perceptual organization. Visual Cognition, 12(4):553–637, 2005.

[3] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967-5976.

[4] B. Cai, X. Xu, K. Jia, C. Qing and D. Tao, "DehazeNet: An End-to-End System for Single Image Haze Removal," in IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187-5198, Nov. 2016.

[5] B. Li, X. Peng, Z. Wang, J. Xu and D. Feng, "AOD-Net: All-in-One Dehazing Network," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 4780-4788.

[6] Y. Qu, Y. Chen, J. Huang and Y. Xie, "Enhanced Pix2pix Dehazing Network," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 8152-8160.

[7] Y. Shao, L. Li, W. Ren, C. Gao and N. Sang, "Domain Adaptation for Image Dehazing," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 2805-2814.

[8] Dong H, Pan J, Xiang L, et al. Multi-scale boosted dehazing network with dense feature fusion[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 2157-2167.