ISSN: 2472-3703 DOI: 10.6911/WSRJ.202305_9(5).0029

Single Image Dehazing Using Attention Enhanced Net

Minghua Zhu[1, a, *]

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

[a]zhuminghua1007@qq.com

Abstract

Producing clear and haze-free images is crucial for many computer vision systems and algorithms. Therefore, single image dehazing is a vital area of research in this field. In the past, prior-based methods have shown promising results. However, these methods tend to produce unwanted artifacts in their outputs since their priors cannot account for all possible scenarios. Conversely, learning-based methods have emerged as a more natural and effective approach. In this paper, we proposed an attention enhanced net for single image dehazing, which incorporates a channel attention branch and a spatial attention branch to enable the network to focus on the most informative parts of the high-dimensional feature map, thereby improving the performance of the subsequent layers in our neural network.

Keywords

Single image dehazing; Attention mechanism; Deep learning.

1. INTRODUCTION

Under hazy conditions, the visibility of images is significantly reduced due to the scattering of atmospheric aerosol particles. This poses a challenge for many computer vision applications, including object detection, recognition, and Advanced Driver Assistance Systems (ADAS), as it becomes difficult to perceive and understand visual information. Hence, haze removal, particularly single-image dehazing, has gained significant attention in the past decade, owing to its immense value in enhancing image quality and facilitating computer vision tasks.

Dehazing methods can be broadly categorized into two groups: prior-based and learning-based. The former relies on the well-established physical model of atmospheric scattering [1] which can be defined as follows:

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

where x denotes the pixel position, while $I(x)$ and $J(x)$ represent the apparent luminance (i.e., the hazy image) and intrinsic luminance (i.e., the clear scene), respectively. A denotes the global skylight, which represents the ambient light in the atmosphere. The transmission of intrinsic luminance in the atmosphere is denoted by $t(x)$ and can be modeled as follows:

$$t(x) = e^{-\beta d(x)}. \tag{2}$$

Koschmieder's law contains multiple unknown variables, including the extinction coefficient (denoted as β) and the scene depth ($d(x)$). Therefore, it is impossible to determine these variables solely based on the input hazy image. To address this challenge, researchers of prior-based methods have proposed incorporating various priors as additional constraints to obtain an appropriate solution for $J(x)$. The goal of these priors is typically to enhance the contrast of objects relative to the ambient light, which is a key factor in determining visibility. By effectively restoring contrast, prior-based methods can produce dehazed images with improved visibility. However, such priors may not be universally applicable and can lead to over-enhancement of contrast, causing unwanted artifacts like halos and color blockages in certain circumstances.

Learning-based dehazing methods differ from prior-based approaches in that they use convolutional neural networks (CNNs) to estimate A and $t(x)$ or directly recover $J(x)$ from the input hazy image via supervised learning. As CNNs are adept at generating images with minimal artifacts [2], these methods can produce dehazed images that are visually realistic. However, the training process requires a significant number of clear and hazy image pairs from the same scene, which can be challenging to acquire in real-world conditions. To address this limitation, learning-based methods often use synthetic hazy images generated by applying Koschmieder's law to indoor scenes where depth information is available. Although this approach allows for more accessible data collection, there can be discrepancies between indoor synthetic and real-world outdoor images, potentially causing learning-based methods to overfit to synthetic data and limiting their ability to remove real-world haze.

In this paper, we propose a novel attention mechanism, it including a spatial attention branch and a channel branch to enhance the performance of the model, it is beneficial to assign distinct weights to the feature maps of different channels and pixels at varying positions in the feature map. By assigning varying weights to different parts of the feature map, the model can more effectively identify and utilize the most salient features, leading to improved accuracy, reduced overfitting, and faster training and inference times.

2. PROPERTIES

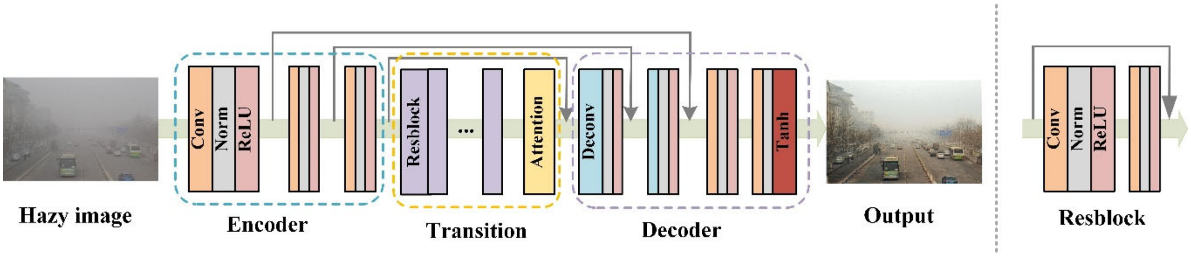

Figure 1. Architecture of the network

Figure 1 demonstrates the proposed network's architecture, which consists of an encoder with three convolutional layers and a feature transformation module with nine residual layers. The kernel size for the first layer is 7×7, while for the rest of the layers in the encoder, it is 3×3. The residual block comprises two convolutional layers and a residual connection with ReLU as the activation function. The decoder comprises two deconvolution layers, each followed by a 1×1 convolutional kernel, five convolutional layers, and an output layer. The activation function used in all layers except the output layer is ReLU, and for the output layer, it is tanh. To recover the detailed information of the clean image, the encoder and the decoder are connected via a skip connection.

Spatial attention refers to a mechanism that selectively focuses on certain regions or parts of an image, based on their relevance to the task at hand. By selectively attending to relevant regions, spatial attention allows the model to ignore irrelevant or noisy regions of the image, which can improve the accuracy and efficiency of the model. Some of the advantages of spatial attention include:

(1) Improved accuracy: By focusing on relevant regions of an image, spatial attention can help the model identify important features and patterns that are crucial for accurate classification or recognition.

(2) Faster training: Spatial attention can help reduce the amount of noise and irrelevant information that the model needs to process, which can speed up training and inference times.

(3) Better interpretability: Spatial attention can provide insights into which parts of an image the model is focusing on, which can help with model interpretation and debugging.

Channel attention, on the other hand, refers to a mechanism that selectively weights different channels of an input feature map, based on their importance to the task at hand. By giving more weight to important channels and less weight to unimportant channels, channel attention can help the model focus on the most relevant information, while ignoring irrelevant or noisy channels. Some of the advantages of channel attention include:

(1) Improved feature representation: Channel attention can help the model identify and highlight the most important features in the input, which can improve the quality of the feature representation and lead to better performance.

(2) Reduced overfitting: Channel attention can help reduce overfitting by suppressing irrelevant or noisy channels, which can lead to better generalization performance.

(3) Faster training: Channel attention can reduce the dimensionality of the input feature map, which can speed up training and inference times.

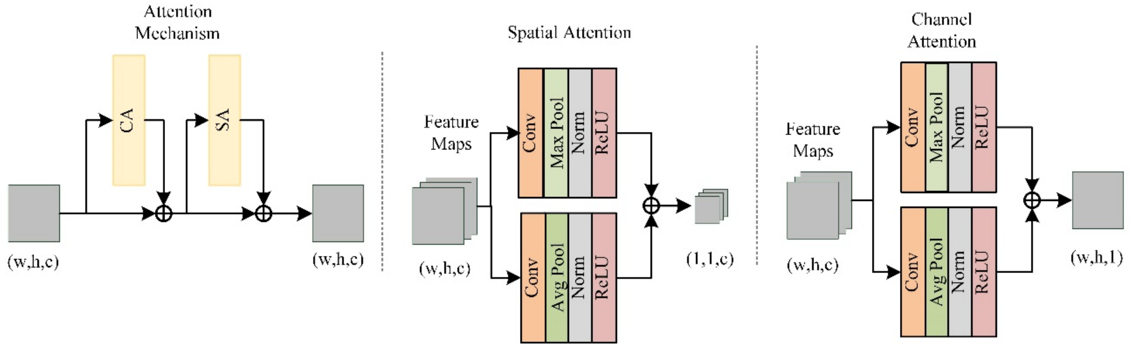

Figure 2. Diagram of attention mechanism

We incorporate an attention module at the end of the transition layer, as illustrated in Figure 2. Specifically, we pass the feature map, which has dimensions $(w,\ h,\ c)$, through a channel attention module, resulting in an output of shape $(1,\ 1,\ c)$. We then apply a broadcast mechanism to add this output to the input, assigning different weights to each channel. Here, w, h, and c correspond to the width, height, and number of channels of the feature map, respectively.

Moreover, we feed the feature map with dimensions $(w,\ h,\ c)$ into a spatial attention module, which generates an output of shape $(w,\ h,\ 1)$. We apply a broadcast mechanism to add this output to the input, which allows us to assign different weights to each spatial location. This way, we can emphasize the most relevant spatial information in the feature map.

Overall, the attention module is a powerful tool that enables network to focus on the most informative parts of the high-dimensional feature map, thereby improving the performance of the subsequent layers in our neural network.

3. TESTS

3.1. Planning

In this section, we evaluate our proposed dehazing method on synthetic dataset against several state-of-the-art methods, including DCP [3], DehazeNet [4], AOD [5], EPDN [6], DAD [7], and MSBDN [8]. We adopt two evaluation criteria, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM), where higher values indicate better dehazing results.

We use the RESIDE dataset [9], which contains hazy/clean image pairs of both indoor and outdoor scenes, for training and testing our proposed method. To train a general dehazing model that works well on both indoor and outdoor scenes, we select 9000 outdoor hazy/clean image pairs and 7000 indoor pairs from the RESIDE training dataset, while removing redundant images from the same scenes.

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

We implement our proposed method using the Pytorch framework on a computer 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. For the first 100 epochs, we set the learning rate to 2×10^{-4} and linearly decay it to 0 for the next 100 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	**36.45**
	SSIM	0.8179	0.8472	0.8504	0.9232	0.9415	0.9810	**0.9874**
outdoor	PSNR	19.13	22.46	20.29	22.57	26.53	31.87	**33.90**
	SSIM	0.8148	0.8514	0.8765	0.8630	0.9150	0.9741	**0.9761**

3.2. Results Comparison

Table 1 presents the quantitative comparisons of various dehazing methods on the SOTS dataset. It can be seen that our proposed method achieves the highest PSNR and SSIM scores for both outdoor and indoor scenes. Specifically, our method outperforms the second-best method by 0.064 and 0.95 dB in terms of SSIM and PSNR on the outdoor subset of the SOTS dataset, respectively. Similarly, on the indoor subset of the SOTS dataset, our method achieves an improvement of 0.002 dB in PSNR and 2.03 in SSIM over the second-best method.

This excellent performance can be attribute to the attention mechanism that helps neural networks focus on relevant parts of the input data when making predictions or decisions. The attention mechanism allows the network to selectively weight the importance of different parts of the input data, giving more emphasis to the most relevant features and ignoring the irrelevant ones. This can lead to improved performance, as the network can focus on the most informative features and avoid being distracted by noisy or irrelevant inputs.

REFERENCES

[1] W. E. K. Middleton and V. Twersky, “Vision through the atmosphere,” Phys. Today, vol. 7, no. 3, p. 21, Mar. 1954.

[2] Qin X, Wang Z, Bai Y, et al. FFA-Net: Feature fusion attention network for single image dehazing[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11908-11915.

[3] 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.

[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.

[9] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Reside: A benchmark for single image dehazing. IEEE Transactions on Image Processing, 28(1):492–505, 2018.