Single Sea Surface Image Dehazing via Multi-scale Concatenated Attention Network

Xin He\textsuperscript{1,a,*}, Wanfeng Ji\textsuperscript{1,b}, Zheng Wang\textsuperscript{3,c}, Yaoqing Zhang\textsuperscript{4,d}

\textsuperscript{1}School of Basic Sciences for Aviation, Naval Aviation University, Yantai, China
\textsuperscript{2}School of Basic Sciences for Aviation, Naval Aviation University, Yantai, China
\textsuperscript{3}Zhenjiang Campus, Army Military Transportation University, Zhenjiang, China
\textsuperscript{4}School of Basic Sciences for Aviation, Naval Aviation University, Yantai, China

\textsuperscript{a,*} hexin6770@163.com, \textsuperscript{b}3368993237@qq.com, \textsuperscript{c}1204597545@qq.com, \textsuperscript{d}abbychn@126.com

Abstract—Haze degrades obscures and content information of sea surface images, which can negatively impact the navigation safety of the intelligent ship in real-time systems. Therefore, it is imperative to develop algorithms that effectively remove haze and make the intelligent ship have higher environmental adaptability. In this paper, we propose a novel Multi-scale Concatenated Attention Network (MCA-Net) to address this problem. Specifically, we use the concatenation operation to connect multi-scale residual blocks (MRB), while introducing the channel attention mechanism for guiding the hazy region feature extracting. Extensive experiments on both synthetic datasets and real sea surface images, which demonstrate our proposed method outperforms are better than other recent dehazing approaches.

1. INTRODUCTION

Intelligent perception and analysis of navigational environment is widely used in technologies of intelligent navigation, which enables intelligent ships to realize autonomous navigation and obstacle avoidance through target detection, target identification and target tracking. The intelligent ship monitors the near scene through the surrounding environment information obtained from the visual imaging system, which mainly consists of visible light cameras. Due to the influence of bad weather factors, sea fog is often accompanied during sea voyages. The scenes taken by the camera are scattered by suspended particles, which greatly reduces the visibility and contrast and seriously affects the performance of the visual system of intelligent ships. Improving the clarity of images obtained in foggy environment is an important prerequisite for intelligent ships to realize technical means such as intelligent perception of surrounding targets and obstacles [1]. Thus, the research of single image dehazing technology for intelligent ships in sea fog environment is of great practical significance.

Mathematically, existing works [2, 3] usually model a hazy image by

\[ I(x) = J(x)T(x) + A[1 - T(x)] \]  

where \( I(x) \) and \( J(x) \) denote the observed clean image and hazy image at location \( x \), while \( A \) describes the global atmospheric light. When the haze is homogeneous, the medium transmission map \( T(x) \) can be expressed as...
\[ T(x) = e^{-\beta d(x)} \quad (2) \]

where \( d(x) \) is the scene depth and \( \beta \) indicates the scattering coefficient of the atmosphere. As only the hazy image \( I(x) \) is available, the haze-free scene \( J(x) \) is an ill-posed problem.

To make the problem well posed, the early priori-based methods attempted to use the statistical characteristics of images, such as dark channel prior (DCP) to estimate the transmission map and atmospheric light \[4\]. Although the physical methods work well in some cases, there are still unable to precisely estimating parameters, particularly in the case of unsuitable modeling.

In recent decades, learning-based methods is introduced by employing convolutional neural networks (CNN) to estimate transmissions or predict clear images. Cai et al. \[5\] proposed a simple but powerful network, named as DehazeNet, added feature extraction and non-linear regression layers to recover the image. Ren et al. \[6\] trained a Multi-Scale CNN, consists coarse and fine scale networks, to reconstruct the transmission map. AOD-Net \[7\] developed a K-estimation module that can combine the transmission map’s and atmospheric scattering model’s variables through a single convolutional network. Afterwards, the gated fusion network (GFN) \[8\] was presented.

\[ F_0 = H_{FE} (I_{\text{haze}}) \quad (3) \]

Figure 1. The framework of our proposed multi-scale concatenated attention network.

In this paper, we propose an end-to-end Multi-scale Concatenated Attention Network (MCA-Net) for single image dehazing, which can directly output dehazed image without the estimation of intermediate parameters. Specifically, we use the concatenation operation to connect multi-scale residual blocks, while integrating the channel attention mechanism for guiding the hazy region feature extracting into the learning procedure. Compare with the related methods above, our methods produce the better dehazing results in the synthetic dataset as well as the real-world ones.

2. PROPOSED METHOD

In this section, we introduce concatenation framework firstly, describe each part as well as the loss function secondly.

2.1 Concatenation Framework

The overall architecture of MCA-Net is shown in Fig. 1, which mainly consists of N multi-scale residual blocks (MRB), and skip connections. MCA-Net employ similar components to step-wise reconstruct the dehazed image. Each component consists of one MRB block and a channel-wise attention layer. The transfer of features between each component is accomplished by local and global skip connections. We set \( I_{\text{haze}} \) for the hazy images and \( I_{\text{clear}} \) for the haze-free images. The first convolution layer extracts feature \( F_0 \):

\[ F_0 = H_{FE} (I_{\text{haze}}) \]
where $H_{fe}(\cdot)$ represents the operation transformation from pixel space to feature space. Then we use $I_{haze}$ and $F_0$ for global residual learning by skip connection, and further have

$$F_i = H_{B_i}(F_i)$$

(4)

where $H_{B_1}(\cdot)$ indicates the first MRB. $F_i$ is further transmit to the global residual learning, and have

$$F_n = H_{B_{n}}(H_{CA}(concatenate(F_{n-1}, F_{n-2})))$$

(5)

where $F_n$ denotes the output of n-th MRB and $H_{CA}$ is the attention layer. Each output of the MRB is transmit to the attention layer and has a local skip connection to the output of the next block. $F_{n-1}$ and $F_{n-2}$ are concatenated for the local skip connection. Finally, we have

$$O_p = H_{ff}(concatenate(F'_0, \cdots, F'_n))$$

(6)

and

$$O_{clear} = conv(O_p) + I_{haze}$$

(7)

Here, $H_{ff}$ is the composite function, and $O_{clear}$ denotes a clear image of the final predicted output.

2.2 Multi-scale Residual Block

Multi-scale features are widely applied in all kinds of computer vision tasks. Multi-scale combining features have better representation in objects and their surroundings [9]. Therefore, a multi-scale residual block (MRB) is proposed, which is the connection between different scales of the feature map and the residual blocks, as shown in Fig. 2.

Mathematically, we first adopt pooling operation with different convolution kernels and strides to form the multi-scale features:

$$y_i = Pooling(x), i \in \{1, 2, 4\}$$

(8)

Then, all scales are blend into three convolution layers. Lastly, we use jump connections to add input signals to learn residuals:

$$z = H(Cat(Up_1(y_1), \cdots, Up_4(y_4))) + x$$

(9)

where $Cat$ is concatenation operation and $Up$ is upsampling operation. $H$ represents a series of operations consisting of two 3×3 convolution operations and one 1×1 convolution operation.

![Figure 2. The architecture of Multi-scale Residual Block (MRB).](image)

2.3 Channel-wise Attention

Attention can improve the focus of salient parts in visual system, which is significance of capturing vision. Since diverse channel features include diverse weighted information, channel-wise attention [10] is introduced after the MRBs to obtain useful local features. Firstly, feature maps are fed into a channel descriptor based on global average pooling.
\[ y_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j) \]  

where \( x_c(i, j) \) is the value of \( c \)-th channel at the location \((i, j)\). The global average pool operator collects the channel global information and changes the feature map. To obtain the weights of various channels, the Sigmod and ReLU activation functions are selected following with two convolution layers, which is given by:

\[ r_c = \alpha(\text{conv}_2(\delta(\text{conv}_1(y_c)))) \] (11)

Lastly, we element-wise multiply the weights information and the input, we have

\[ f_c = r_c \times x_c \] (12)

By adding the channel attention layer after MRB, it can guide the hazy region feature and extract useful information in different channels.

2.4 Loss Function
For our network optimization, we choose the simple L1 loss by default.

\[ L(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \| I_{gt} - MCA(I_{haze}) \| \] (13)

where \( \Theta \) denotes the network parameters, \( I_{gt} \) and \( I_{haze} \) stand for ground truth and original input image.

3. EXPERIMENTAL RESULTS
In this section, we first introduce datasets and experimental details. Then we quantitatively and qualitatively evaluate our method against other dehazing algorithms.

3.1 Datasets
We selected 1,000 images from the VOC 2007 dataset and BSD 500 dataset [11]. The sea surface images and other scenes images are included, so that the network can better understand the features of hazy images. In our experiment, the blue channel hazing algorithm is used to generating synthetic datasets. There is a logarithmic relationship between image depth and normalized blue channel, which can be defined by

\[ d(x) = -\log(1 - B(x)) \] (14)

where \( B(x) \) can be formed by removing the details of \( B(x) \) with guide filtering. The transmission map is linear with the blue channel, according to formula (2) and (5), and it can expressed as

\[ T(x) = 1 - 0.95 \times B(x) \] (15)

When the clear image and transmission map are known, the hazy image can be synthesized by using formula (1). These image pars are randomly cut into 64×64 crops. In total, 500, 000 pairs of images were selected for training.

3.2 Training Settings
We perform the training process using PyTorch on an NVIDIA GTX 1080 GPU. For the accelerated training purpose, the Adam optimizer is used with a batch size of 8, where \( \beta_1 \) and \( \beta_2 \) take the default values of 0.9 and 0.999 respectively. The network is trained for 100 epochs and the learning rate is initially set to 0.001 and then is divided by 10 after each 25 epochs.

3.3 Result Evaluation
The proposed method is compared with three state dehazing methods: DCP [4], AOD-Net [7] and GFN [8]. As the availability of ground truth in synthetic data, we enable to evaluate the results using PSNR and SSIM metrics [12]. Table 1 shows the evaluation criteria between each pairs of haze-free image and dehazing result. Though the table, our method obtains the highest value, indicates that is more similar to the ground truth one. Besides, some corresponding pictures directly show visually difference,
particularly seen in Fig. 3. Furthermore, the generalization ability of our model on real images is assessed, as shown in Fig. 4. Overall, our proposed method restores more details and obtains visually pleasing images, which is more suitable for marine application.

![Dehazed results on the synthetic dataset.](image)

![Dehazed results on real hazy images.](image)

**TABLE 1. QUANTITATIVE EVALUATIONS ON SYNTHETIC DATASET**

| Method   | DCP     | AOD-Net | GFN     | Ours    |
|----------|---------|---------|---------|---------|
| PSNR/SSIM| 14.60/0.75 | 18.33/0.82 | 20.85/0.80 | **23.15/0.88** |

4. CONCLUSION

In this paper, we present multi-scale concatenated attention network (MCA-Net) for dehazing single sea surface image. Multi-scale residual block and channel attention mechanism are applied to get effective local and global features. Extensive experiment results show that, our method is superior to other dehazing methods for sea surface images.

REFERENCES

[1] DehazeZoo web site. [Online]. Available: github.com/cxtalk/DehazeZoo

[2] E. J. Mccartney and F. F. H. Jr, “Optics of the atmosphere: Scattering by molecules and particles,” Optica Acta International Journal of Optics, vol. 14, no. 9, pp. 698–699, 1977.

[3] Narasimhan, G. Srinivasa, Nayar, and K. Shree, “Vision and the atmosphere,” International Journal of Computer Vision, vol. 48, no. 3, pp. 233–254, 2002.

[4] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2341–2353, 2010.
[5] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehazenet: An end-to-end system for single image haze removal,” IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187–5198, 2016.

[6] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, “Single image dehazing via multi-scale convolutional neural networks,” in Proceedings of the European Conference on Computer Vision, pp. 154–169, 2016.

[7] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “AOD-Net: All-in-one dehazing network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4770–4778, 2017.

[8] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.H. Yang, “Gated fusion network for single image dehazing,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3253–3261, 2018.

[9] Y. Li, Q. Miao, R. Liu, J. Song, Y. Quan, and Y. Huang, “A multi-scale fusion scheme based on haze-relevant features for single image dehazing,” Neurocomputing, vol. 283, no. 29, pp. 73–86, 2018.

[10] S. Woo, J. Park, J.-Y. Lee, and I. SoKweon, “CBAM: Convolutional block attention module,” in Proceedings of the European Conference on Computer Vision, pp. 3–19, 2018.

[11] Q. Ju, C. Li and Q. Sang, "Single Sea Surface Image Dehazing Using Multi-scale Cascaded Convolutional Neural Network," in 2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI), pp. 1-5, 2018.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.