



An end-to-end deep learning approach for real-time single image dehazing

Chi Yoon Jeong^{1,2} · KyeongDeok Moon¹ · Mooseop Kim^{1,2}

Received: 2 June 2022 / Accepted: 24 November 2022 / Published online: 8 February 2023
© The Author(s) 2023

Abstract

Image dehazing methods can restore clean images from hazy images and are popularly used as a preprocessing step to improve performance in various image analysis tasks. In recent times, deep learning-based methods have been used to sharply increase the visual quality of restored images, but they require a long computation time. The processing time of image-dehazing methods is one of the important factors to be considered in order not to affect the latency of the main image analysis tasks such as detection and segmentation. We propose an end-to-end network model for real-time image dehazing. We devised a zoomed convolution group that processes computation-intensive operations with low resolution to decrease the processing time of the network model without performance degradation. Additionally, the zoomed convolution group adopts an efficient channel attention module to improve the performance of the network model. Thus, we designed a network model using a zoomed convolution group to progressively recover haze-free images using a coarse-to-fine strategy. By adjusting the sampling ratio and the number of convolution blocks that make up the convolution group, we distributed small and large computational complexities respectively in the early and later operational stages. The experimental results with the proposed method on a public dataset showed a real-time performance comparable to that of another state-of-the-art (SOTA) method. The proposed network's peak-signal-to-noise ratio was 0.8 dB lower than that of the SOTA method, but the processing speed was 10.4 times faster.

Keywords Single-image dehazing · Lightweight dehazing networks · Image restoration · End-to-end learning · Convolutional neural networks

1 Introduction

The goal of image dehazing is to restore an unknown clean image from a source image whose quality is degraded by haze [1], as shown in Fig. 1. Image dehazing methods can lead to performance gain in several image analysis tasks including object detection and scene segmentation because

they can enhance the image quality in the preprocessing step before the main image analysis tasks are conducted [2]. Therefore, many studies on image dehazing have been conducted to enhance the visual quality of degraded images.

Previous studies on image dehazing can be classified into prior-knowledge-based and deep learning-based methods [3]. Prior-knowledge-based methods rely on domain knowledge, such as dark channel priors (DCP) [4], which suffer from generalization because prior information cannot cover all cases [5]. Because deep learning can automatically find representative features from huge amounts of data, methods using deep learning outperform traditional methods in various image analysis tasks [6–10]. Therefore, many studies [2, 5, 11–15] using deep learning approaches have been conducted to improve the performance of image dehazing. Single image dehazing methods using deep neural networks (DNNs) increase performance with regard to the peak signal-to-noise ratio (PSNR) and structural similarity

✉ Mooseop Kim
gomskim@etri.re.kr

Chi Yoon Jeong
iamready@etri.re.kr

KyeongDeok Moon
kdmoon@etri.re.kr

¹ Artificial Intelligence Research Lab., Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Daejeon 34129, Republic of Korea

² Artificial Intelligence, University of Science and Technology, 217 Gajeong-ro, Daejeon 34113, Republic of Korea

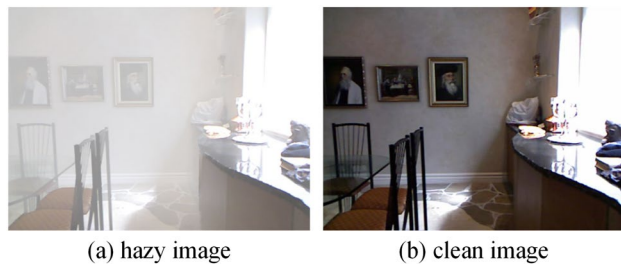


Fig. 1 Example of image dehazing

index measure (SSIM), but they require high computational complexity and computation time [14].

Recently, situational awareness methods based on image analysis have become widely used in unmanned aerial vehicles (UAVs) [16], advanced driver-assistance systems [17], and pedestrian navigation assistance systems [18], and haze can substantially decrease the performance of these methods. Furthermore, the latency of these methods is important to quickly recognize and respond to various situations. Given that dehazing operations are typically used as preprocessing steps for these methods, the dehazing processing time is an important factor that can affect the latency of the situational awareness methods.

A few studies [17, 19–22] have been conducted to decrease the latency of dehazing methods. Methods to increase the processing speed of a dehazing algorithm were proposed by replacing complex operations in the DCP algorithm with low-complexity operations. In addition, methods that utilize device properties, such as digital signal processors (DSPs) [19] or mobile devices [22], have been proposed to reduce the latency of dehazing algorithms. However, these methods mainly focus on the DCP algorithm, which is a representative prior-knowledge-based method, and have limited dehazing performance with regard to PSNR and SSIM. A recent study [23] also pointed out that when the performance of the dehazing method is low, the performance of object detection, which is one of the major image analysis tasks, does not improve. Few studies [20, 24] on lightweight dehazing networks have been conducted to speed up deep learning-based methods, but they did not use an end-to-end approach that directly generates a clean image from a source image. Recent end-to-end learning approaches [2, 13–15] for single-image dehazing have shown superior performance compared to approaches that only partially adopt deep learning networks [5, 11] for single-image haze removal.

Therefore, in this study, we propose an end-to-end network method for real-time haze removal. Because convolutional neural networks (CNNs) have shown the best performance among various neural network architectures for single-image haze removal, the proposed method adopts a CNN for the end-to-end network. CNNs are generally

constructed by stacking a convolution group consisting of multiple convolution operations, activation functions, and residual connections, and convolution groups incur extensive computational costs. To reduce the computation time of convolution groups, the proposed method utilizes a zoomed convolution group, inspired by zoomed convolution in [7], which effectively reduces the processing time of convolution. Furthermore, the proposed network adopts an efficient channel attention method [25] because the attention mechanism can increase network performance without heavy computational burden. The experimental results on a public dataset confirmed that the proposed end-to-end dehazing network can process in real time with a comparable performance to that of a state-of-the-art (SOTA) method. The contributions of this study are as follows.

- We propose an end-to-end dehazing network that can restore haze-free images with high visual quality in real time.
- To decrease the processing time of the network model without degrading the quality of restored images, we devise a zoomed convolution group that processes computation-intensive operations with low resolution. Also, we adopt efficient channel attention mechanisms that are extremely lightweight in comparison with the various existing channel attention mechanism.
- Finally, we verify that the proposed network can restore haze-free images in real time with a performance comparable to that of a SOTA method.

The remainder of this paper is structured as follows. Section 2 presents related works on single-image haze removal and methods to improve the speed of dehazing methods. Section 3 describes the proposed network for real-time image dehazing, while performance evaluation experiments and results are described in Sect. 4. Finally, Sect. 5 describes our concluding remarks and indicates further research directions.

2 Related works

A haze is a physical phenomenon associated with atmospheric light scattering produced by smoke, dust, and other floating particles [1]. Thus, most dehazing methods use the atmospheric scattering model [26] as defined below:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

where x denotes the pixel coordinates of an image, and $I(x)$ and $J(x)$ denote the hazy image acquired by the camera and clear scene radiance, respectively. A denotes the global atmospheric light, and $t(x)$ denotes the transmission map

defined by the atmospheric scattering parameter and the distance between the camera and the object.

Traditional approaches called prior knowledge-based methods use various statistical properties observed in haze-free images, and one of the representative features of haze-free images is the DCP proposed by Hu et al. [4]. DCP assumes that there exists a dark channel with very low intensity at some pixels in most non-sky patches. Dehazing methods using DCP have shown superior performance to other prior knowledge-based methods but have high computational complexity.

Therefore, many studies [17, 19–22] have been conducted to decrease the computational complexity of the DCP algorithm. Moreover, an empirical study [27] to compare the performance of the efficient dehazing methods has been conducted. Lu and Dong [19] used a mean filter instead of guided filtering or soft matting, which is the most complex operation in the DCP algorithm, and then accelerated the operation of the mean filter using the features of DSP. Tsai et al. [17] substituted a fast guided filter for the soft matting method and used them in the sky estimation step. In [21], morphological reconstruction was applied to estimate the transmission map rapidly. A recent study [20] proposed a method for estimating atmospheric light based on a weighting scheme and substituted morphological filtering for the soft matting method to reduce the computation time of the DCP algorithm. Cimtay [22] proposed a method to increase the processing speed of the dehazing method by reusing previously calculated atmospheric light information after analyzing changes in the orientation of mobile devices. Although these methods decrease the processing time of the DCP algorithm, DCP has limited dehazing performance with regard to PSNR and SSIM.

Due to the success of deep learning in various image analysis tasks, much research [2, 5, 11–15] has been conducted to use deep learning in single image dehazing. Pioneer [5, 11, 12] introduced CNNs to calculate the transmission map and utilized it to restore haze-free images with atmospheric light computed using a traditional method. However, these methods, which consist of multiple steps, suffer from performance limitations because errors in each step amplify the degradation of the visible image quality. To solve this problem, end-to-end approaches [2, 13–15] using various network architectures have been proposed.

Attention-based multi-scale network, inspired by GridNet in [28], was proposed by Liu et al. [14]. Their backbone network consisted of three rows representing different scales, and each row had five residual dense blocks. By enabling flexible information interchange between the image scales and the residual dense blocks using a grid network and attention module, the visible quality of restored images can be increased. Dual residual networks (DuRN) [13] was proposed to solve image restoration tasks, including dehazing.

DuRN is made by stacking residual blocks that consist of dual residual connections that use the promise of paired operations, such as convolution with kernels of different sizes. They pointed out that a dual residual connection can increase the potential interaction between paired operations, leading to a performance increase in the image restoration tasks. The feature fusion attention network (FFA-Net) [15], which is based on a CNN architecture, adopts a novel attention mechanism that considers both channel features and image pixels. FFA-Net composites a convolution group compromising several convolutions, skip connections, and a feature attention module, and then constructs the dehazing network using multiple convolution groups and the feature attention module. FFA-Net has shown excellent performance in the visual quality of the restored image, but it requires a high computational complexity. A dehazing network using an encoder-decoder architecture was proposed by Dong et al. [2]. They exploited the dense feature fusion module to use the information of nonadjacent features and to prevent the loss of spatial information, which is one of the issues in the encoder-decoder architecture. In addition, they applied a boosting strategy to the decoder of the network to progressively refine the feature map. These methods have shown a significant performance gain compared with traditional methods, and they require high computational complexity and computation time. Considering that image dehazing is adopted as a preprocessing step for image analysis tasks, studies on lightweight dehazing networks are necessary.

Recently, studies [20, 24] have been conducted to speed up the CNNs used to calculate the transmission map. In [20], a fully convolutional network that replaces dense layers with convolution layers was used. Furthermore, a lightweight CNN architecture [24] consisting of a few layers was proposed for fast transmission map estimation. However, a study on accelerating the dehazing network for an end-to-end approach is required because end-to-end learning approaches have superior performance.

3 Proposed method

The proposed end-to-end image-dehazing network is designed based on the CNN architecture [15], which has shown superior performance among neural networks. However, we modified its CNN structure to decrease the computational complexity for real-time image-dehazing, which is the main goal of our study. As illustrated in Fig. 2, the proposed network comprises three parts: preprocessing, post-processing, and a backbone network. The preprocessing part is used to provide valuable image enhancement compared to manually designed methods and consists of convolution layers. The postprocessing part also consists of convolution layers and is used to reduce artifacts to enhance the quality

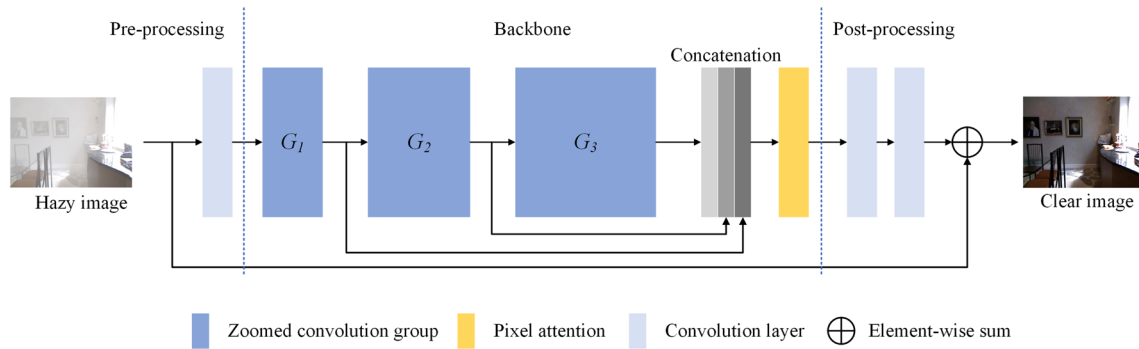


Fig. 2 Overview of the proposed method for real-time dehazing

of restored images. In the backbone network, we propose a zoomed convolution group to effectively decrease the computational complexity of the network layers and exploit skip connections to effectively convey the information of the earlier layers to the later layers. As shown in Fig. 2, we used three zoomed convolution groups, G_1 , G_2 , and G_3 , with different sampling ratios for the zoomed operation and different numbers of layers. The output feature maps of each zoomed convolution group were concatenated and passed onto the network’s efficient channel attention mechanism, without loss of performance.

Many variants of standard convolution exist, depending on the purpose of the network models. For example, group convolution [29] can decrease the number of parameters by half compared to standard convolution. Dilated convolution [30] focuses on enlarging receptive fields; therefore, it is popularly used in dense prediction tasks. However, the latency of network models is determined by complex factors such as the number of parameters and the computational complexity [7].

Therefore, we focus on variants of standard convolution that are specifically for reducing the latency of network models. Zoomed convolution was introduced in [7]. First, it downsamples the input feature map and then processes the convolution on a low-resolution feature map. Finally, the output feature map is upsampled to the resolution of the input map. Considering that the processing time of the convolution operations is tightly coupled with the resolution of the feature map, zoomed convolution effectively reduces latency by processing the convolution on a small feature map. In comparison with a standard convolution with kernel size 3×3 , zoomed convolution can reduce the processing time of operations by 40% and has twice the receptive field [7].

Inspired by a zoomed convolution study in [7], we propose a zoomed convolution group to reduce the latency of the network model, without visual quality degradation of restored images. As illustrated in Fig. 3, the proposed

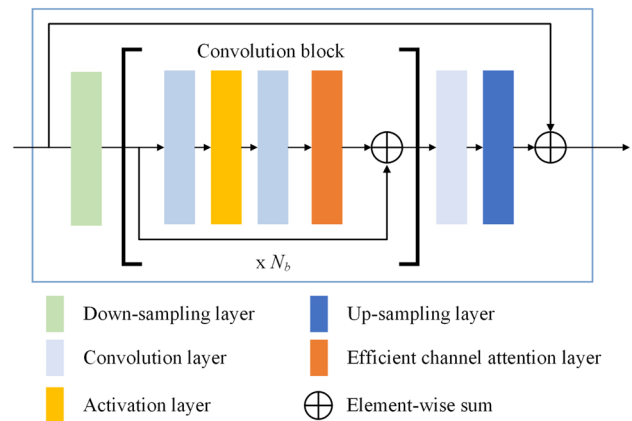


Fig. 3 Structure of zoomed convolution group

zoomed convolution group comprises multiple convolution blocks, an additional convolutional layer, a downsampling layer, and an upsampling layer. The convolution block consists of a stacked neural network that includes two convolution layers, an activation layer, an attention layer, and a skip connection. Bilinear interpolation was used for downsampling and upsampling of the feature maps, and the sampling ratio (Z_s) was varied by the convolution group. G_1 , G_2 , and G_3 have 11, 13, and 17 convolution blocks, respectively. Furthermore, they have a sampling ratio of two, four, and four for the zoomed operation, respectively. Similar to zoomed convolution, the processing time of the convolution group can be effectively reduced by performing various operations at a low resolution. In addition, the zoomed convolution group has a significant gain in processing speed because it performs upsampling and downsampling only once for many convolution blocks.

The attention mechanism can increase the performance of a network without a high amount of computational burdens; therefore, many attention modules including efficient channel attention (ECA) [25], squeeze-and-excitation networks (SENet) [31], and convolutional block attention modules

(CBAMs) [32] have been proposed and widely used in various networks. Among the conventional attention modules, ECA provides a lightweight method to learn effective channel attention through global average pooling (GAP) and fast 1D convolution. Therefore, we adopted the ECA module for the convolutional block that performs the attention module multiple times. Given a feature map $\chi \in \mathcal{R}^{W \times H \times C}$, ECA module first performs GAP, defined as follows:

$$G(\chi) = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W \chi_{ij}, \quad (2)$$

where C denotes the number of filters or channel dimensions and W and H are the width and height of the feature map, respectively. Channel attention \mathbf{w}_c is obtained as follows [25]:

$$\mathbf{w}_c = \sigma(\text{C1D}_k(G(\chi))), \quad (3)$$

where σ represents sigmoid function, C1D represents the fast 1D convolution, and k is the kernel size for the fast 1D convolution.

Since previous models [15] consisted of convolution groups with the same number of convolution blocks regardless of their position, there exist redundant operations. Thus, the structures of the convolution groups needed to be adjusted to decrease the latency of the dehazing network. Similar to the decoder in the encoder-decoder architecture [2], we designed a network model to progressively recover the haze-free images using a coarse-to-fine strategy. Therefore, we reduced the number of convolution blocks of the convolution group in the early stage and increased it toward a later stage. Additionally, we set the sampling ratio to two in the first zoomed convolution group and four in the other zoomed convolution groups. The proposed network can reduce the latency of the network by efficiently distributing computational complexity using a coarse-to-fine strategy.

The output of each zoomed convolution group is concatenated into a feature map using weights calculated from channel attention. In addition, pixel attention proposed by [15] is executed on the concatenated feature map because the haze is non-uniformly distributed on the image pixels. By adopting a pixel attention module, the proposed dehazing network adaptively learns the informative features of image pixels. Given the concatenated feature map F , pixel attention \mathbf{w}_p is calculated using the following equation [15]:

$$\mathbf{w}_p = \sigma(\text{C2D}(\delta(\text{C2D}(F))))), \quad (4)$$

where δ and σ denote the ReLU activation function and sigmoid function, respectively, and C2D represents the 2D convolution.

The output feature map of the pixel attention module passes through two convolution layers in what is a

postprocessing step. Then, a haze-free image is restored by summing the elements of the input image and the result of the postprocessing step. Residual learning can also help to improve the performance of network models by effectively conveying the information of shallow layers to deep layers [6]. Therefore, skip connections for residual learning are also used in convolution blocks, convolution groups, and entire networks.

4 Experimental results

The proposed dehazing network was compared with SOTA methods on a public dataset. For the experiments, the proposed network consisted of three zoomed convolution groups with 11, 13, and 17 convolutional blocks, respectively. The sampling ratio of the first zoomed convolution group was set to 2 and that of the others to 4. All convolution layers used a 3×3 kernel, and the channel dimensions of all convolution layers was set to 64. The kernel size of the 1D convolution, which is used in the ECA module, was set to 3.

The input size of the network was set to 240×240 , as in existing methods. We used the adaptive moment estimation (ADAM) optimizer [33] and cosine annealing strategy [34]. The parameters of the optimizer, β_1 and β_2 , were set to default values of 0.9 and 0.999, and the initial learning rate was set to 0.0001. The proposed network was trained during 5×10^5 steps, and the learning rate was scheduled using the cosine annealing strategy every 5000 steps. A previous study [35] pointed out that the training of a dehazing network using L1 loss has a superior performance to that using L2 loss. Therefore, we used the L1 loss for model training. The batch size was set to 4 for model training, and 1 for the inference step.

To train the network model and test its dehazing performance, we used the RESIDE dataset [36], which is popularly used in the dehazing field. Images of the RESIDE dataset were generated by synthesizing images using an atmospheric scattering model and depth map of the images. We used the indoor training set of the RESIDE dataset for model training and the indoor images of the synthetic objective testing set (SOTS), which is a subset of the RESIDE dataset for performance evaluation. The number of training and test images was 13,990 and 500, respectively, and the resolution of the images was 620×460 .

The proposed method and SOTA methods were implemented using PyTorch v1.7, and the pretrained models provided by the authors of each method were used in the experiment. All experiments were conducted on a PC equipped with an Intel Xeon W-2295 CPU, 384 GB system memory, RTX 3090 GPU, and Ubuntu 20.04 operating system. The versions of CUDA and cuDNN were v11.0 and v8.0, respectively.

4.1 Quantitative evaluation

We evaluated the proposed and SOTA methods using PSNR and SSIM, which are full reference metrics. In addition, two no-reference metrics such as blind image integrity notator using DCT-statics (BLIINDS-II) [37] and spatial-spectral entropy-based quality (SSEQ) [38] were used.

To compare performance, we used four end-to-end dehazing networks: MSBDN [2], GDN [28], DuRN [13], FFA-Net [15] and DCP [4] which is a representative prior-knowledge-based method. In addition, we used two deep learning-based methods, AOD-Net [11] and DehazeNet [12], which partially adopted DNNs for single-image dehazing.

Table 1 summarizes the quantitative evaluation results of the proposed method and SOTA methods. As shown in Table 1, the end-end-end dehazing networks showed significant performance improvements compared to traditional methods such as DCP, AOD-Net, and DehazeNet. Among the end-to-end network models, FFA-Net exhibited the best performance with regard to full-reference metrics such as PSNR and SSIM and the proposed method exhibited the second-best performance. The difference with regard to PSNR and SSIM between FFA-Net and proposed methods was 0.8 dB and 0.003, respectively. This means that the performance gap between FFA-Net and the proposed network was only 2.2 % and 0.3 % with regard to PSNR and SSIM. Considering the no-reference metrics, our methods exhibited the third-best performance with regard to BLIINDS-II and SSEQ. The experimental results on no-reference evaluation were inconsistent with full-reference evaluation, and these results align with those of previous works [36] in that the no-reference metrics are limited in their ability to provide a suitable quality assessment of dehazed images. Considering that the purpose of this work was to develop a network model that can restore hazy-free images in real time, the experimental results

Table 1 Quantitative results comparison of SOTA methods on the SOTS dataset

Methods	PSNR	SSIM	BLIINDS-II	SSEQ
DCP [4]	16.62	0.8179	74.41	64.94
AOD-Net [11]	19.06	0.8504	79.02	67.65
DehazeNet [12]	21.14	0.8472	71.71	65.46
DuRN [13]	32.12	0.9803	75.17	69.23
GDN [14]	32.16	0.9836	76.53	70.25
FFA-Net [15]	36.39	0.9886	76.08	68.30
MSBDN [2]	32.77	0.9813	77.62	71.01
Ours	35.59	0.9854	76.88	69.64

confirmed that the proposed method minimizes the performance degradation of the SOTA method.

4.2 Qualitative evaluation

We compared the visible quality of the restored images obtained using the proposed method and conventional methods. Examples of images restored using the various methods are illustrated in Fig. 4. In Fig. 4, dehazing methods using end-to-end networks showed better visual quality than the DCP method as with the qualitative results. Overall, the end-to-end networks showed competent results in restoring clear images from hazy ones. However, Fig. 4 showed that the end-to-end network with relatively low PSNR sometimes fails to uniformly restore the homogeneous regions and restores some dark regions relatively brightly.

Figure 5 shows the resulting images after haze removal according to the degree of haze using FFA-Net and the proposed method. This confirms that the proposed dehazing network can recover clean images regardless of the haze degree. When comparing the visual quality of the restored images obtained using the proposed method and FFA-Net, Fig. 5 shows that the visual quality of the restored images is almost identical to the proposed method and FFA-Net.

4.3 Run-time analysis

We evaluated the processing times of the end-to-end dehazing networks because the processing time of the dehazing methods is an important factor to be considered to avoid affecting the latency of the main tasks. The selected existing methods were evaluated using the source codes and pre-trained models provided by the authors of each method, and all experiments were conducted in the same environments. The time taken to process all indoor images of the SOTS was measured five times, and then the average processing time per image was calculated.

The results for the SOTS dataset including PSNR and average processing time are presented in Table 2. As shown in Table 2, the proposed method outperformed all other methods with regard to latency and is the only method that can restore haze-free images in real time. Table 2 also shows that methods using CNNs such as DuRN and FFA-Net require a relatively long processing time. The proposed method can process images 10.4 times faster than FFA-Net. Considering that the difference with regard to PSNR between the proposed method and FFA-Net was only 2.2 %, the experimental results showed that the proposed method successfully decreased the latency using the zoomed convolution group and coarse-to-fine strategy.



Fig. 4 Visual results of the proposed method and the SOTA methods. **a** hazy images, **b** DCP [4], **c** DuRN [13], **d** GDN [14], **e** FFA-Net [15], **f** MSBDN [2], **g** ours, **h** ground-truth images



Fig. 5 Visual results of the proposed method and FFA-Net according to the degree of the haze. The extent of haze becomes gradually thick from left to right. **a** hazy images **b** FFA-Net [15] **c** ours

Table 2 Performance comparison to SOTA methods on SOTS dataset

Methods	PSNR (\uparrow)	Avg. processing time (s) (\downarrow)
DuRN [13]	32.12	0.091
GDN [14]	32.16	0.040
FFA-Net [15]	36.39	0.208
MSBDN [2]	32.77	0.044
Ours	35.59	0.020

Table 3 Performance comparison for different configurations of the proposed network on the SOTS dataset

Network	PSNR (\uparrow)	FPS (\uparrow)
Baseline	36.15	7.5
+ Zoomed convolution group	34.63	25.0
+ ECA module	35.87	22.7
+ Coarse-to-fine strategy	35.59	49.5

4.4 Ablation study

To analyze the effectiveness of each component in our network model, including the attention module, zoomed convolution group, and coarse-to-fine strategy, we conducted ablation studies as shown in Table 3. From the baseline network that consists of convolution groups with the same number of blocks ($N_b = 17$) but no efficient attention module, we evaluated the performance of the dehazing network with regard to PSNR and frames per second (FPS) by adding each component. Table 3 confirms that the zoomed convolution group with the same sampling ratio of two can effectively decrease the latency of the dehazing network. In addition, Table 3 confirms that the efficient attention module increases the quantitative results of the dehazing network with a small burden for the processing time. The coarse-to-fine strategy that efficiently distributes the computational complexity by

adjusting the number of convolution blocks and the sampling ratio also effectively reduces the processing time while minimizing performance degradation.

In addition, we assessed the performance of the proposed method according to the various configurations of zoomed convolution groups, as shown in Table 4. The experimental results showed that a high sampling ratio (Z_s) in the first convolution group can decrease the performance compared to the other cases. Furthermore, the higher number of convolution blocks (N_b) in the later convolution group can help improve the performance than vice versa. The proposed configuration is composed of three zoomed convolution groups consisting of 11, 13, and 17 convolutional blocks with sampling ratios of 2, 4, and 4, respectively, and it exhibited the highest performance in terms of PSNR among the configurations that achieved a processing speed of over 30 FPS.

5 Conclusion

In this study, we proposed an end-to-end network for real-time image dehazing because the latency of a dehazing network is an important factor to be considered to avoid affecting the speed of main image analysis tasks. The proposed method uses a CNN architecture, which has been shown to outperform other neural network architectures. To reduce the computation time of our model's CNN, we used a zoomed convolution group, inspired by an existing study on zoomed convolution, which effectively reduces the processing time of convolution. Zoomed convolution groups can significantly reduce their processing time of the convolution group by performing various operations such as convolution, activation, and attention at low resolution. Additionally, the zoomed convolution group adopts an efficient channel attention module because the attention mechanism is one of the important factors for improving the performance of deep learning networks. Through the foregoing, we designed a network model to progressively recover haze-free images using a coarse-to-fine strategy and efficiently distribute the

Table 4 Performance comparison for different configurations of the zoomed convolution groups on the SOTS dataset

G_1		G_2		G_3		PSNR (\uparrow)	FPS (\uparrow)
N_b	Z_s	N_b	Z_s	N_b	Z_s		
17	2	17	2	17	2	35.87	22.7
11	2	13	2	17	2	35.49	27.8
11	4	13	4	17	4	33.03	59.3
11	2	13	4	17	4	35.59	49.5
11	4	13	4	17	2	33.63	40.9
11	4	13	2	17	4	35.38	45.7
17	4	13	4	11	2	33.40	45.5
17	2	13	4	11	4	35.31	43.0
17	4	13	2	11	4	34.73	44.7

computational complexity by adjusting the number of convolution blocks and sampling ratio. The experimental results on a public dataset showed that the visual quality of the images restored by the proposed method is almost identical to that of FFA-Net, which had the highest performance in the quantitative evaluation results. The performance of FFA-Net differed from the proposed network by only 2.2 % and 0.3 % in terms of PSNR and SSIM, respectively. The experimental results further showed that our method outperformed all compared methods with regard to latency. Moreover, the proposed network is the only method for restoring haze-free images in real time. The proposed method can process images 10.4 times faster than FFA-Net and enables a frame rate of 49.5 FPS, with minimal performance degradation.

In this study, we focused on improving the processing speed of dehazing networks while maintaining a high visual quality of restored images. However, the high visual quality of the output image in the preprocessing step does not always guarantee performance improvement in the main image analysis task. In future work, we will analyze the relationship between the performance of the image-dehazing network and that of the main tasks using the image-dehazing network as a preprocessor. Using such analyses results, we plan to develop several lightweight image-dehazing networks, each optimized for a specific main image analysis task. Also, we plan to analyze the performance on various datasets that contain real-world environments including low-light conditions.

Acknowledgements This work was supported by Electronics and Telecommunications Research Institute(ETRI) grant funded by the Korean government [22ZS1200, Fundamental Technology Research for Human-Centric Autonomous Intelligent Systems].

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Mei, K., Jiang, A., Li, J., Wang, M.: Progressive feature fusion network for realistic image dehazing. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) Computer vision—ACCV 2018, pp. 203–215. Springer, Cham (2019)
- Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.-H.: Multi-scale boosted dehazing network with dense feature fusion. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR) (2020)
- Wang, C., Meng, Z., Xie, R., Jiang, X.: A single image dehazing algorithm based on cycle-gan. In: Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence. RICAI 2019. Association for Computing Machinery, New York, NY, USA, pp. 247–251 (2019). <https://doi.org/10.1145/3366194.3366237>
- He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2011). <https://doi.org/10.1109/TPAMI.2010.168>
- Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.-H.: Single image dehazing via multi-scale convolutional neural networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer vision—ECCV 2016, pp. 154–169. Springer, Cham (2016)
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
- Chen, W., Gong, X., Liu, X., Zhang, Q., Li, Y., Wang, Z.: Fast-erseg: Searching for faster real-time semantic segmentation. In: International Conference on Learning Representations (2020)
- Jeong, C., Yang, H.S., Moon, K.: A novel approach for detecting the horizon using a convolutional neural network and multi-scale edge detection. *Multidimens. Syst. Signal Process.* **30**(3), 1187–1204 (2019). <https://doi.org/10.1007/s11045-018-0602-4>
- Redmon, J., Divvala, S.K., Girshick, R.B., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016). <https://doi.org/10.1109/CVPR.2016.91>
- Chung, S., Jeong, C.Y., Lim, J.M., Lim, J., Noh, K.J., Kim, G., Jeong, H.: Real-world multimodal lifelog dataset for human behavior study. *ETRI J.* **44**(3), 426–437 (2022). <https://doi.org/10.4218/etrij.2020-0446>
- Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 4780–4788 (2017). <https://doi.org/10.1109/ICCV.2017.511>
- Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **25**(11), 5187–5198 (2016). <https://doi.org/10.1109/TIP.2016.2598681>
- Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
- Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2019)
- Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: feature fusion attention network for single image dehazing. *Proc. AAAI Conf. Artif. Intell.* **34**(07), 11908–11915 (2020). <https://doi.org/10.1609/aaai.v34i07.6865>
- Gautam, S., Gandhi, T.K., Panigrahi, B.K.: A model-based dehazing scheme for unmanned aerial vehicle system using radiance boundary constraint and graph model. *J. Vis. Commun. Image Represent.* **74**, 102993 (2021). <https://doi.org/10.1016/j.jvcir.2020.102993>
- Tsai, C.-C., Lin, C.-Y., Guo, J.-I.: Dark channel prior based video dehazing algorithm with sky preservation and its embedded

- system realization for adas applications. *Opt. Express* **27**(9), 11877–11901 (2019). <https://doi.org/10.1364/OE.27.011877>
18. Zhang, J., Yang, K., Constantinescu, A., Peng, K., Müller, K., Stiefelhagen, R.: Trans4trans: efficient transformer for transparent object segmentation to help visually impaired people navigate in the real world. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 1760–1770 (2021). <https://doi.org/10.1109/ICCVW54120.2021.00202>
 19. Lu, J., Dong, C.: Dsp-based image real-time dehazing optimization for improved dark-channel prior algorithm. *J. Real-Time Image Process.* **17**(5), 1675–1684 (2020). <https://doi.org/10.1007/s11554-019-00933-3>
 20. Yang, G., Evans, A.N.: Improved single image dehazing methods for resource-constrained platforms. *J. Real-Time Image Process.* **18**(6), 2511–2525 (2021). <https://doi.org/10.1007/s11554-021-01143-6>
 21. Salazar-Colores, S., Cabal-Yepez, E., Ramos-Arreguin, J.M., Botella, G., Ledesma-Carrillo, L.M., Ledesma, S.: A fast image dehazing algorithm using morphological reconstruction. *IEEE Trans. Image Process.* **28**(5), 2357–2366 (2019). <https://doi.org/10.1109/TIP.2018.2885490>
 22. Cimtay, Y.: Smart and real-time image dehazing on mobile devices. *J. Real-Time Image Process.* **18**(6), 2063–2072 (2021). <https://doi.org/10.1007/s11554-021-01085-z>
 23. Jeong, C.Y., Ha, D., Moon, K., Kim, M.: An empirical study of the effect of single image dehazing on object detection. *J. Korean Inst. Intell. Syst.* **32**, 117–126 (2022). <https://doi.org/10.5391/JKIS.2022.32.2.117>
 24. Ullah, H., Muhammad, K., Irfan, M., Anwar, S., Sajjad, M., Imran, A.S., de Albuquerque, V.H.C.: Light-dehazenet: a novel lightweight CNN architecture for single image dehazing. *IEEE Trans. Image Process.* **30**, 8968–8982 (2021). <https://doi.org/10.1109/TIP.2021.3116790>
 25. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
 26. Cox, L.J.: Optics of the atmosphere-scattering by molecules and particles. *Opt. Acta: Int. J. Opt.* **24**(7), 779–779 (1977). <https://doi.org/10.1080/713819629>
 27. Girija, M.G., Shanavav, K.T., Ajith, G.S.: Image dehazing using MSRCR algorithm and morphology based algorithm: a concise review. *Mater. Today: Proc.* **24**, 1890–1897 (2020). <https://doi.org/10.1016/j.matpr.2020.03.614>
 28. Fourure, D., Emonet, R., Fromont, E., Muselet, D., Trémeau, A., Wolf, C.: Residual conv-deconv grid network for semantic segmentation. In: Proceedings of the British Machine Vision Conference, 2017 (2017)
 29. Howard, A., Zhmoginov, A., Chen, L.-C., Sandler, M., Zhu, M.: Inverted residuals and linear bottlenecks: mobile networks for classification, detection and segmentation. In: CVPR (2018)
 30. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *Computer Vision—ECCV 2018*, pp. 833–851. Springer, Cham (2018)
 31. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
 32. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
 33. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings* (2015). [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
 34. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 558–567 (2019). <https://doi.org/10.1109/CVPR.2019.00065>
 35. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2017)
 36. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **28**(1), 492–505 (2019)
 37. Saad, M.A., Bovik, A.C., Charrier, C.: Blind image quality assessment: a natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.* **21**(8), 3339–3352 (2012). <https://doi.org/10.1109/TIP.2012.2191563>
 38. Liu, L., Liu, B., Huang, H., Bovik, A.C.: No-reference image quality assessment based on spatial and spectral entropies. *Signal Process.: Image Commun.* **29**(8), 856–863 (2014). <https://doi.org/10.1016/j.image.2014.06.006>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Chi Yoon Jeong received his BS and MS degrees in electronic and electrical engineering from Pohang University of Science and Technology, Pohang, Republic of Korea, in 2002 and 2004, respectively, and his Ph.D. degree in computer science from Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea, in 2018. He is currently a principal researcher in the artificial intelligence laboratory at Electronics and Telecommunications Research Institute, Daejeon,

Republic of Korea. His main research interests include computer vision, pattern recognition, machine learning, and sensory substitution.



KyeongDeok Moon received his B.S. and M.S. degrees in computer science from Hanyang University, Korea in 1990 and 1992 respectively. He received a Ph.D. degree in information engineering from KAIST, Korea in 2005. From 1992 to 1996, he was a researcher at System Engineering Research Institute where he worked on high performance computing and clustering computing. Since 1997, he has been a principal researcher at Electronics and Telecommunications

Research Institute, where he develops the home network middleware, deep learning for video analysis, framework for autonomously navigated ship and human augmentation architecture. His research interests are in sensory substitution technology, human augmentation technology, deep learning architecture, and autonomous ship.



Mooseop Kim received his M.S. in Electrical Engineering from Kyungpook National University, Korea, and a Ph.D. degree in Computer Science and Engineering from Chungnam National University, Korea, in 1998, and 2008, respectively. He was a research engineer in the Organic LED (OLED) group in the Device and Materials Laboratory, LG Electronics Institute of Technology (LG Elite), Seoul, Korea from 1998 to 1999. He has

been with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, since 1999 and currently he is a principal researcher. His current research interests include wearable computing, activity recognition, and sensory substitution.