



OPEN SAD-Net: a full spectral self-attention detail enhancement network for single image dehazing

Qingjun Niu^{1,2}, Kun Wu^{1,2}, Jialu Zhang^{1,2}, Zhenqi Han^{1,2}✉ & Lizhuang Liu^{1,2}✉

Single-image dehazing technology plays a significant role in video surveillance and intelligent transportation. However, existing dehazing methods using vanilla convolution only extract features in the temporal domain and lack the ability to capture multi-directional information. To address the aforementioned issues, we design a new full spectral attention-based detail enhancement dehazing network, named SAD-Net. SAD-Net adopts a U-Net-like structure and integrates Spectral Detail Enhancement Convolution (SDEC) and Frequency-Guided Attention (FGA). SDEC combines wavelet transform and difference convolution (DC) to enhance high-frequency features while preserving low-frequency information. FGA detects haze-induced discrepancies and fine-tunes feature modulation. Experimental results show that SAD-Net outperforms six other dehazing networks on the Dense-Haze, NH-Haze, RESIDE and I-Haze datasets. Specifically, it increases the peak signal-to-noise ratio (PSNR) to 17.16 dB on the Dense-Haze dataset, surpassing the current state-of-the-art (SOTA) methods. Additionally, SAD-Net achieves excellent dehazing performance on an external dataset without any prior training.

Image dehazing plays a significant role in computer vision systems, such as autonomous driving, video surveillance, and intelligent transportation¹. Therefore, restoring clear scenes from hazy images has become the core issue in single-image dehazing and has drawn significant attention from academia and industry over the past five years. With the development of deep learning, convolution neural networks (CNNs) have made significant progress in image dehazing^{2–7}. Early CNN-based methods^{8–10}, such as AOD-Net¹¹, DehazeNet¹², and GCA-Net¹³, are based on the dark channel prior theory. They output dehazed images by estimating the transmission map and atmospheric light, combine with the atmospheric scattering model¹⁴. However, these methods heavily rely on the accurate estimation of the transmission map and atmospheric light. To address this issue, subsequent research^{15,16} propose end-to-end image restoration models. Examples include FFA-Net¹⁷, GridDehaze-Net¹⁸, and MB-TaylorFormer¹⁹, which directly generate dehazed images and significantly improve performance. However, convolutional networks or transformer architectures often result in the loss of high-frequency texture details in foggy images during end-to-end image restoration. To address this issue, methods such as Gaussian filtering and the Laplacian operator are used to extract high-frequency texture information from images, which is then integrated into the network model to enhance the high-frequency details in the generated image. For example, WSAMF-Net²⁰ introduces an additional high-frequency information extraction branch in the end-to-end network.

Despite these advancements, there are still issues with the current dehazing networks. (1) Ignore high-frequency information. AOD-Net and FFA-Net use convolution to extract features and in the attention mechanism of FFA-Net, attention weights are computed solely through global average pooling²¹ and standard convolutions. These two methods tend to focus on processing low-frequency information, neglecting high-frequency details such as edges and noise, which limits the dehazing performance in terms of detail recovery and prevents the full restoration of high-frequency features in the image. In contrast, WSAMF-Net introduces an additional branch for extracting high-frequency information. (2) Convolution operations are insufficiently sensitive to multi-scale and directional information²². The MSBND network tackles this limitation by deepening the network architecture. However, MSBND approach significantly increases the model's computational cost and complexity. (3) The dehazing problem involves spatial and channel-level non-uniformity. To handle spatial non-uniformity, FFA-Net generates a spatial importance map through pixel attention, adaptively processing regions with varying fog densities. DEA-Net²³ combines spatial and channel attention, generating unique importance maps for each channel. FFA-Net and DEA-Net are limited by the size of their receptive fields, making

¹Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. ✉email: hanzq@sari.ac.cn; liulz@sari.ac.cn

it challenging to capture interactions between global features effectively. As a result, FFA-Net and DEA-Net lack the ability to model details across large-scale regions.

To solve the three aforementioned issues, a U-Net-like^{15,16,19} frequency-domain attention dehazing network named SAD-Net is developed. The network is divided into three parts: encoder, feature transform, and decoder. The feature transform uses a spectral detail enhancement attention block (SDEAB) constructed by combining Spectral Detail Enhancement Convolution (SDEC) and Frequency-Guided Attention (FGA). The encoder-decoder part uses SDEB. SAD-Net first extracts image features at different scales through the encoder and enhances the details in the feature transformation section. The decoder then gradually restores the spatial resolution of the image, ultimately outputting the dehazed image. SDEC integrates wavelet transform²⁴, inverse wavelet reconstruction, and five convolution layers (four difference convolutions (DCs) and one vanilla convolution). It separates and extracts useful information from both high and low frequencies, and reintegrates them through inverse wavelet transform to produce enhanced detail feature maps. FGA separates high and low-frequency information using wavelet transform, generating feature maps for different frequency directions and channels. It focuses on the key areas affected by haze while suppressing irrelevant interference. SAD-Net is tested on the public datasets Dense-Haze²⁵, I-Haze²⁶, RESIDE²⁷ and NH-Haze²⁸, achieving excellent results. Specifically, on the Dense-Haze dataset, SAD-Net reaches a Peak Signal-to-Noise Ratio (PSNR) of 17.16dB, surpassing the current state-of-the-art (SOTA) methods. The results on other datasets also show that SAD-Net has outstanding dehazing performance in handling complex foggy scenes. Additionally, SAD-Net demonstrates strong generalization capabilities. An external dataset is constructed with 45 images from O-Haze to test SAD-Net. Results show that SAD-Net effectively restores hazy images in unseen scenes. All dehazed images achieve a PSNR above 14 dB. This demonstrates its strong generalization capability.

In summary, our main contributions are as follows:

1. We design a novel frequency-domain attention-based dehazing network, named SAD-Net. SAD-Net integrates SDEC and FGA. SDEC innovatively combines wavelet transform with DC, while FGA further integrates wavelet transform with spatial and channel attention.
2. Extensive model comparisons and ablation experiments are conducted, demonstrating the feasibility of SAD-Net's convolution network in frequency-domain dehazing. Our approach has shown superior performance compared to SOTA dehazing networks.
3. SAD-Net is tested directly on external datasets without prior training and successfully restores most scene information. This demonstrates its strong generalization capability.

Methods

The SAD-Net network consists of three parts: an encoder, an adaptive feature transformation module, and a decoder. The network structure is shown in Fig. 1. We built Spectral Detail Enhancement Block (SDEB) and SDEAB as basic modules by integrating SDEC, FGA, and GCA modules. SDEB is used in the encoder and decoder, while SDEAB is used in the feature transformation part. The network adopts a three-layer U-Net-like encoder-decoder architecture to achieve efficient image dehazing.

The encoder compresses the spatial dimensions of the feature maps through two downsampling steps, preserving more global semantic information and effectively extracting high-frequency details using the SDEC module. The feature transformation part consists of multiple SDEABs, which use the FGA attention mechanism to enhance fine-grained processing in uneven fog areas, improving feature recognition. The decoder restores the spatial resolution of the image through two upsampling steps and combines the detailed features retained by the encoder. This ensures that the dehazed image maintains a good global structure while preserving rich local details. The skip connections and GCA module's adaptive weighting mechanism further optimize the fusion of low-level and high-level features, enhancing image clarity. Through this encoder-decoder structure, combined with the collaborative work of the SDEC, FGA, and GCA modules, SAD-Net can efficiently handle features at different scales. It achieves robust dehazing performance and excellent image quality.

Spectral detail enhancement convolution (SDEC)

Before introducing our proposed SDEC, we first review DC²⁹. DC, inspired by Local Binary Patterns (LBP)³⁰, enhances feature representation by capturing the differences between a pixel and its surrounding pixels in specific directions. This method extracts the variations between each pixel and its neighbors, generating richer features for subsequent image processing tasks. DC has achieved excellent results in face anti-spoofing tasks^{31–33}, proving its effectiveness. However, relying solely on DC has limitations when handling complex dehazing tasks. High-frequency details, such as edge information in hazy images, are often hard to capture accurately. Therefore, spatial-domain convolution alone is not enough to extract these critical details effectively. To further enhance the ability to capture high-frequency information, we introduce wavelet transform²⁴, which is widely used in image processing before CNN^{34–37}. It can decompose signals into different sub-bands, enhancing the high-frequency sub-bands to restore image details and compressing the low-frequency sub-bands to reduce the impact of haze. This improves image contrast and clarity. This method is widely applied in the dehazing field, especially excelling in detail recovery and edge enhancement. We combine wavelet transform with DC. Wavelet transform extracts high and low-frequency information from feature maps, and different DC methods are applied to different sub-bands. This better handles fine-grained information in the feature maps and restores haze-free images.

In SDEC, five convolutions are used (four DCs and one vanilla convolution). First, the input feature map is decomposed into four sub-bands using wavelet transform: one LL and three high-frequency detail sub-bands (HL, LH, HH). The LL sub-band retains the overall contour information of the image, reflecting smoother areas, while the high-frequency sub-bands focus on capturing edge features in the image. Specifically, the HL sub-band mainly extracts horizontal edge information, capturing horizontal details in the image. LH focuses on

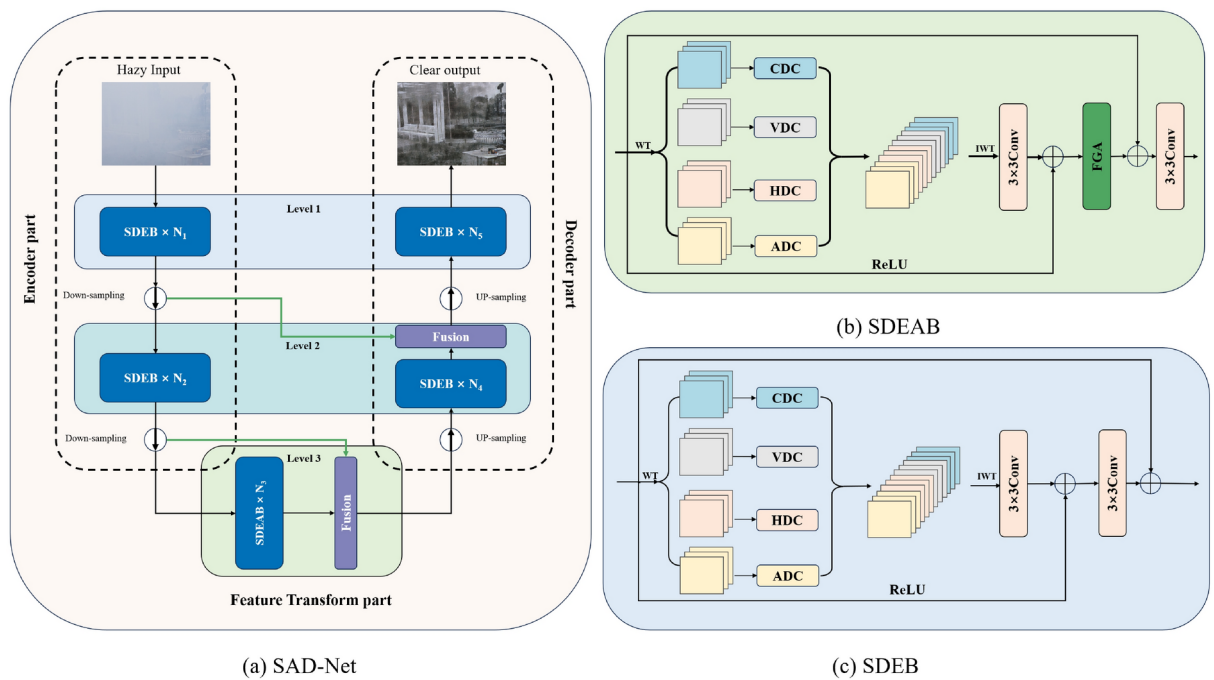


Fig. 1. The overall structure of SAD-Net and its block diagrams. **(a)** Shows the structure of SAD-Net, including the encoding part, feature transformation part, and decoding part. SDEAB is used in the feature transformation part, while SDEB is used in the other parts. **(b)** Shows the structure of SDEAB, including SDEC, FGA, and a vanilla convolution. **(c)** Shows the structure of SDEB, including SDEC and a vanilla convolution.

vertical details, while HH captures diagonal edge features, helping to recognize complex diagonal structures. After the wavelet transform, SDEC uses four different types of DCs to process these sub-bands, enhancing feature extraction. The central difference convolution (CDC) is used for LL, preserving the overall features by calculating the difference between the center pixel and its surrounding pixels while enhancing image details. For the high-frequency sub-bands, diagonal difference convolution (ADC) is applied, along with horizontal difference convolution (HDC) and vertical difference convolution (VDC)²³, which incorporate traditional local descriptors like Sobel, Prewitt, or Scharr. These convolutions process features in the horizontal, vertical, and diagonal directions, strengthening the high-frequency sub-bands in different directions and frequencies. After the DCs, the sub-band features are concatenated to form a feature representation containing multi-frequency information. Then, the inverse wavelet transform is used to reconstruct the concatenated features back to the same size as the original input feature map.

Frequency-guided attention (FGA)

In real haze scenarios, the fog is usually unevenly distributed, and the blocking of non-uniform haze leads to varying degrees of loss or reduction of edge information in different regions of the image. To restore the dehazed image more accurately, the neural network needs to learn the features of clear, lightly hazy, and heavily hazy areas. The network should adaptively remove the haze according to the degree of haze obstruction while enhancing the high-frequency edges and texture information that are obscured by the haze. To this end, both FFA-Net and DEA-Net utilize attention mechanisms that combine spatial and channel dimensions to help the network distinguish between lightly and heavily hazy areas. However, the attention mechanism that generates attention weights solely through average pooling can only retain low-frequency information and is unable to effectively capture high-frequency edge details³⁸. To enhance the obscured high-frequency edges and texture information, we propose a multi-frequency spatial-channel collaborative attention mechanism. This method not only allows the network to learn the haze distribution and assign corresponding weights to different regions, but also adaptively enhances the obscured high-frequency edge and texture information, thereby significantly improving the dehazing performance. To verify that the FGA has this capability, we remove the entire FGA module from SAD-Net to construct a new model, SD-Net, and compare the results on the non-uniform haze dataset, i.e., the NH-Haze dataset. The results are shown in Fig. 2. It is clearly evident that without the FGA module, the network leaves significant haze residue on many images, indicating that it cannot adaptively remove haze based on varying levels of haze obstruction. Additionally, the dehazed areas appear as if they have been smeared, with the texture details of the image almost completely disappearing.

The specific process of the multi-frequency attention mechanism is as follows: First, the input feature map undergoes a wavelet transform, such as using Daubechies or Haar wavelets. This process decomposes the image into four subband feature maps: LL, LH, HL, and HH. The core purpose of the wavelet transform is to decompose the different frequency components of the image, allowing the model to process frequency-specific information

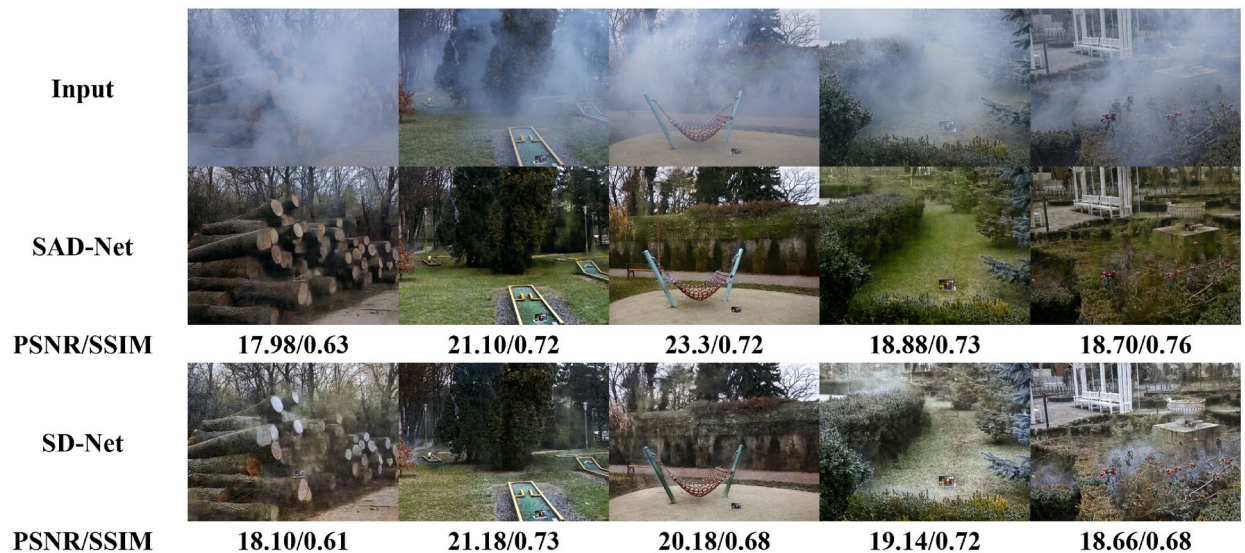


Fig. 2. Comparison results between SAD-Net and SD-Net on the NH-Haze dataset. PSNR/SSIM are used as evaluation metrics.

separately, such as global low-frequency features and local high-frequency detail features. Next, the average is calculated along the height (H) and width (W) directions for each subband feature map, generating two single 1D sequence structures, h_{atten} and w_{atten} . This step helps compress spatial information while significantly reducing the spatial dimensions of the feature maps, thereby reducing the computational load in subsequent steps. Then, 1×1 convolutions are applied to reduce the channel dimensions of each subband feature map, compressing the number of channels and reducing the computational complexity of the subsequent channel attention mechanism. After dimensionality reduction, the feature maps from different subbands are concatenated along the channel dimension. The concatenated feature map is then normalized using group normalization³⁹. This step is designed to better distinguish the information differences between spatial features and reduce the interference of spatial information. The goal is to enhance the salient information in the feature map and suppress background noise. Next, by applying the Sigmoid activation function, a spatial attention map is generated. These spatial attention maps represent the salient regions along the height and width dimensions, reflecting the more important areas in the image. The generated spatial attention map is then element-wise multiplied with the four subband feature maps to produce a spatially enhanced feature map. The low-frequency subband (e.g., LL) is multiplied by the generated spatial attention map, as shown in formula 1:

$$LL_{\text{atten}} = LL \cdot h_{\text{atten}} \cdot w_{\text{atten}} \quad (1)$$

These enhanced feature maps are adjusted in the spatial dimension, allowing the model to better capture the features of the regions obscured by haze, while reducing the influence of other irrelevant areas. Next, to reduce the computational complexity, the spatially enhanced feature maps are further downsampled using average pooling⁴⁰. Next, a channel self-attention mechanism is used to generate three matrices: the query matrix (Q), key matrix (K), and value matrix (V). These matrices are used for self-attention calculations between channels. By performing matrix multiplication between the query and key matrices, the attention weight matrix between channels is computed and normalized using Softmax. Then, the attention weight matrix is multiplied by the value matrix to generate the weighted channel features⁴¹.

$$\text{Attention}(q, k, v) = \text{Softmax} \left(\frac{q \cdot k^T}{\sqrt{d}} \right) \cdot v \quad (2)$$

This process enhances the expression of relevant features and helps reinforce the effective features for haze removal. Finally, an inverse wavelet transform is applied to restore the feature maps, which have undergone self-attention and spatial enhancement, back to the time domain, maintaining the same size as the input feature maps. The inverse wavelet transform process allows the feature maps to be restored to the original image size, while retaining the enhanced low-frequency and high-frequency information, ultimately achieving an effective dehazing effect. The detailed structure is shown in Fig. 3.

Results

Datasets and metrics

Datasets

In this experiment, we use real-world foggy weather datasets to train and test SAD-Net. These datasets include Dense-Haze, NH-Haze, RESIDE, and I-Haze. Except for RESIDE, all datasets contain real foggy images generated

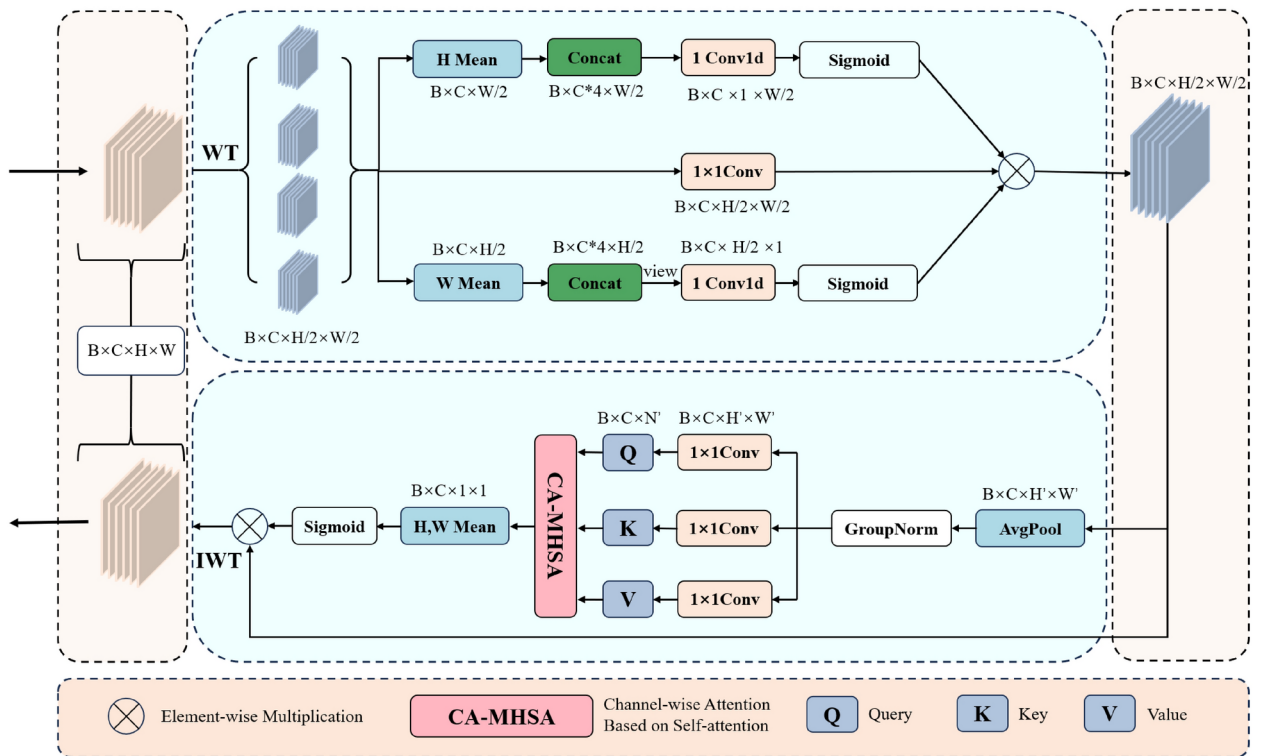


Fig. 3. Schematic diagram of the FGA mechanism. WT denotes wavelet transform, and IWT denotes inverse wavelet transform. B represents batch size, C represents the number of channels, and H and W represent the height and width of the input features, respectively. The upper part shows the frequency domain spatial attention section of FGA, while the lower part shows the channel attention section.

by professional fog machines and are widely used as benchmarks for dehazing performance evaluation. The Dense-Haze dataset includes 55 pairs of real dense fog images and their corresponding fog-free images. NH-Haze is the first dehazing dataset to contain non-homogeneous fog and also includes 55 pairs of real dense fog and fog-free images. The I-Haze dataset contains 35 pairs of real dense fog images and their corresponding fog-free images. In the NH-Haze and Dense-Haze datasets, the first 50 images are used for training, while images 51 to 55 are used for validation and testing. In the I-Haze dataset, the first 30 images are used for training, with images 31 to 35 used for validation and testing. The RESIDE dataset is a synthetic dataset, and due to the large number of images in it, the first 2000 images of the SOTS-indoor and SOTS-outdoor subsets are selected as the training set, while images 2001 to 2500 are used for validation. These four datasets are employed to comprehensively assess the dehazing performance of the proposed SAD-Net.

Evaluation metrics

To evaluate the dehazing performance, we use commonly applied image quality metrics in computer vision, such as PSNR and SSIM⁴². PSNR measures the noise level in the image, while SSIM assesses the structural similarity between the dehazed image and the original image. Additionally, we introduce the ratio of high- to low-frequency energy distribution to further validate the effectiveness of the network. This metric quantifies the network’s ability to restore high-frequency image details while preserving the overall low-frequency structure. First, we apply a Fourier transform to the image $I(x,y)$ to obtain the frequency domain image $F(u,v)$, where x and y represent spatial domain coordinates, and u and v represent frequency domain coordinates. The result of the Fourier transform is a complex matrix, with the energy represented by its magnitude. The magnitude $M(u,v)$ is defined as:

$$M(u, v) = |\mathcal{F}(I(x, y))| \tag{3}$$

Low-frequency energy is usually concentrated in the central region of the spectrum and represents the overall structural information of the image. Image details are typically represented by high-frequency energy, which is distributed in the outer regions of the spectrum. The low-frequency energy, E_{low} , represents the total energy in the central region, while the high-frequency energy, E_{high} , represents the total energy in the non-central regions of the spectrum. Finally, the ratio of high-frequency to low-frequency energy, $R_{high/low}$, is calculated as shown in Eq. (4):

Method	Dense-Haze		NH-Haze		I-Haze		SOTS-Indoor		SOTS-Outdoor	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DehazeNet	13.84	0.43	16.62	0.52	–	–	19.82	0.821	24.75	0.927
AOD-Net	13.14	0.41	15.40	0.57	17.28	0.52	20.51	0.816	24.14	0.919
GridDehazeNet	14.96	0.53	13.80	0.54	18.73	0.77	32.16	0.983	30.86	0.981
MSBDN	15.37	0.49	19.23	0.71	19.62	0.62	32.77	0.981	33.48	0.982
FFA-Net	14.39	0.45	19.87	0.69	19.72	0.73	36.39	0.988	33.57	0.984
DeHamer	16.62	0.56	20.66	0.68	–	–	36.63	0.988	35.18	0.986
AECR-Net	15.80	0.46	19.88	0.71	–	–	37.17	0.990	–	–
MB-TaylorFormer	16.66	0.56	–	–	–	–	42.64	0.994	38.09	0.991
SAD-Net	17.16	0.55	20.48	0.72	21.65	0.84	41.71	0.993	37.78	0.990

Table 1. Quantitative comparison of different dehazing methods on Dense-Haze, NH-Haze, and I-Haze datasets. We compare PSNR and SSIM, where higher values of PSNR and SSIM indicate better performance. The symbol “–” denotes that the number is not available. Bold indicates the best results.

Method	DehazeNet	AOD-Net	GridDehazeNet	MSBDN	SAD-Net
#Param (M)	0.008	0.0018	0.96	31.35	3.15
#FLOPs (G)	0.5409	0.1146	21.43	41.54	20.21
Method	FFA-Net	DeHamer	AECR-Net	MB-TaylorFormer	SAD-Net
#Param (M)	4.456	132.4	2.611	7.43	3.15
#FLOPs (G)	287.5	48.93	52.20	88.1	20.21

Table 2. The comparison of parameter count and FLOPs between SAD-Net and other methods. We tested the models’ floating-point operations and parameter count using tensors of size $1 \times 3 \times 256 \times 256$. The unit for floating-point operations is Giga (G), and the unit for parameter count is Million (M).

$$E = \sum_{(u,v)} M(u,v), R_{\text{high/low}} = \frac{E_{\text{high}}}{E_{\text{low}}} \quad (4)$$

This ratio reflects the network’s ability to balance restoring high-frequency details while maintaining the low-frequency structure, further validating its effectiveness in processing different frequency components. To ensure fairness, all evaluations are conducted on uncropped RGB color images.

Implementation details

We implement the SAD-Net model on the PyTorch deep learning platform using a single NVIDIA RTX3090 GPU. In the first and second stages, we use SDEB, and in the third stage, we use SDEAB. The number of modules in each stage, denoted as [N1, N2, N3, N4, N5], is set to [4, 4, 8, 4, 4]. The network optimization is carried out using the Adam optimizer, with parameters β_1 , β_2 , and ϵ set to the default values of 0.9, 0.999, and 1×10^{-8} , respectively. The initial learning rate is set to 1×10^{-4} , with a batch size of 2. We use a cosine annealing strategy to smoothly adjust the learning rate from the initial value to 1×10^{-6} . To enhance training diversity, we randomly crop 640×640 image patches from the original images and apply two data augmentation techniques: 90° , 180° , or 270° rotations and vertical or horizontal flips. The model is trained for a total of 120,000 iterations. Training SAD-Net on the Dense-HAZE dataset takes approximately 2 days.

Comparisons with SOTA methods

We compare our SAD-Net with six SOTA single image dehazing methods, including DehazeNet, MSBDN¹⁶, AOD-Net, GridDehazeNet, FFA-Net, DeHamer⁴³, and MB-TaylorFormer, conduct a comprehensive evaluation on the Dense-Haze and NH-Haze datasets.

Table 1 presents the quantitative evaluation results of SAD-Net on five datasets (PSNR and SSIM metrics). For ease of comparison, the results for the Dense-Haze, NH-Haze, and I-Haze datasets are rounded to two decimal places. To better highlight the subtle differences between methods on the SOTS datasets, the SSIM results for the SOTS-Indoor and SOTS-Outdoor datasets are rounded to three decimal places. As shown in the table, SAD-Net achieves a PSNR of 17.16 dB and an SSIM of 0.53 on the Dense-Haze dataset, ranking first compared to other SOTA dehazing methods. On the SOTS-Indoor and SOTS-Outdoor datasets, SAD-Net’s SSIM performance is close to that of MB-TaylorFormer. On the NH-Haze and I-Haze datasets, SAD-Net outperforms most existing dehazing methods, demonstrating excellent dehazing performance. Table 2 compares the FLOPs and the parameter counts of SAD-Net with other methods. It can be seen that SAD-Net surpasses the performance of DeHamer using only half of its floating-point operations. It achieves comparable results with MB-TaylorFormer using only about one-quarter of its floating-point operations.

Figure 4 shows the visual comparison between SAD-Net and other SOTA methods on the Dense-Haze dataset. From the figure, it can be seen that SAD-Net is capable of generating clearer images, preserving more object contour information, and effectively reducing haze residue. In the visual comparison on the NH-Haze dataset, we further observe that SAD-Net restores color information more effectively in light haze regions, while achieving clearer object contours in heavy haze regions. To verify SAD-Net's ability to preserve image details and structures in the frequency domain, we performed a Fourier transform on the dehazed images. By analyzing the distribution of high-frequency and low-frequency energy in the frequency domain images, we further quantified the network's balance between restoring high-frequency details (such as object contours and textures) and retaining low-frequency information (such as the overall image structure). Specifically, high-frequency energy reflects the richness of image details, while low-frequency energy corresponds to the overall structure of the image. The ratio of high-frequency to low-frequency energy distribution indicates that SAD-Net significantly enhances the restoration of high-frequency details while maintaining the stability of low-frequency components. However, relying solely on high- and low-frequency information or the distribution of high- and low-frequency energy as evaluation criteria for image restoration is not sufficient, as clearly demonstrated in Fig. 4. For instance, on the Dense-Haze dataset, although the R-value of the image restored by MSBDN is close to that of the original clear image, its high- and low-frequency components are significantly lower than those of the clear image. Therefore, it is not sufficient to conclude that MSBDN outperforms SAD-Net in image restoration. Similarly, on the NH-Haze dataset, although the low-frequency components of images restored by AECR-Net, FFA-Net, and Dehamer are closer to those of the clear image, Dehamer's R-value is the closest to the clear image, indicating that Dehamer achieves the highest image quality. The purpose of introducing the R-value is to intuitively demonstrate how SAD-Net directly performs dehazing in the frequency domain while preserving more high-frequency information. Compared to other methods, SAD-Net's high-to-low-frequency energy ratio is closer to that of the original haze-free image, and the overall distribution of high- and low-frequency energy is also more similar to the original image. This proves that the network effectively preserves the natural appearance and detail information of the image while avoiding artifacts caused by excessive sharpening.



Fig. 4. Visual and frequency domain comparison of different dehazing methods on Dense-Haze and NH-Haze datasets. The images above are from Dense-Haze and their corresponding spectral plots. The images below are from NH-Haze and their corresponding spectral plots. PSNR and SSIM refer to the values obtained by comparing the images above with the original images. E_{low} and E_{high} represent the low and high-frequency energy of the spectral plots above, respectively. The closer these two values are to the low and high-frequency energy of the clear image, the better. $R_{high/low}$, when the low and high-frequency energy is close to that of the clear image, should be as close to the clear image as possible.

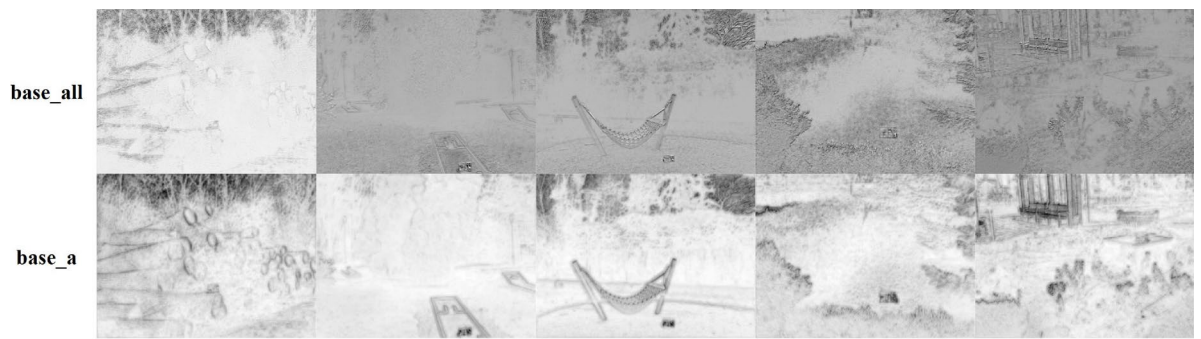


Fig. 5. The average feature mappings of base_all and base_a after different attention mechanisms on the NH-Haze dataset.

Method	Dense-Haze		NH-Haze		I-Haze	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
base_all	15.80	0.51	18.87	0.65	16.89	0.68
base_a	16.60	0.55	19.72	0.67	19.67	0.80
base_c	16.20	0.51	18.94	0.64	19.76	0.71
SAD-Net	17.16	0.55	20.48	0.72	21.65	0.84

Table 3. A quantitative comparison of different models on the Dense-Haze, NH-Haze, and I-Haze datasets is conducted. We compare PSNR and SSIM, where higher values indicate better performance. The best result is highlighted in bold.

Discussion

To verify the effectiveness of the proposed FGA and SDEC modules, we design two experiments: FGA capability analysis and SDEC capability analysis. Both experiments run on the Dense-Haze dataset to analyze the contributions and functions of the SDEC and FGA components. To further validate the generalization performance of SAD-Net, we create an external dataset with 45 images selected from the O-Haze dataset. We then apply the untrained SAD-Net directly to this external dataset for dehazing to evaluate its generalization ability.

FGA capability analysis

First, all SDEC modules in SAD-Net are replaced with vanilla convolutions, and FGA is replaced with CGA²³ modules to build a baseline model. This resulting model is called “base_all”. To verify the independent contribution of FGA, the CGA attention module in base_all is replaced with FGA. The resulting model is named “base_a”. This model is used to evaluate the dehazing performance improvement brought by FGA. PSNR and SSIM are used as evaluation metrics. The results are shown in Table 2. From the table, it can be seen that base_a outperforms base_all in PSNR and SSIM across different datasets. Especially in the I-Haze dataset, PSNR is improved by 2.78dB, and SSIM is improved by 0.12. This indicates that FGA can fully utilize high and low-frequency information when reassigning weights to the spectrum, has some noise reduction capability, and can pay attention to important foggy areas in the image, which helps restore image structure. This suggests that FGA makes an important independent contribution to dehazing.

Traditional temporal self-attention focuses on low-frequency information, whereas FGA attention can emphasize low-frequency information while also attending to high-frequency details. To substantiate this conclusion, we compare the output features of the first DEABlock at level 3. Specifically, we feed the test set of the NH-Haze dataset into the model and average the feature maps along the channel dimension to approximate the overall representation of the feature map. As shown in Fig. 5, the base_a network using FGA exhibits a brighter image, indicating that, compared to base_all, FGA retains more low-frequency information. Additionally, the edges and textures of the base_a image are overall clearer, while base_all loses nearly all edges and textures in dense fog regions. This demonstrates that FGA retains more low-frequency information while also preserving as much high-frequency information as possible.

SDEC capability analysis

To demonstrate the independent contribution of SDEC, we replace the vanilla convolution modules in base_all with SDEC. The resulting model, named base_c, is used to evaluate the improvement in dehazing performance. The results are shown in Table 3. In the Dense-Haze and NH-Haze datasets, which represent uniform and non-uniform dense fog, respectively, there is no improvement in SSIM for base_c compared to base_all, but PSNR increased by 0.39 dB and 0.2 dB, respectively. Thus, base_c show no significant restoration effect on uniform dense fog and non-uniform dense fog, but it could fully extract and utilize high and low-frequency information,



Fig. 6. The average feature mapping of base_all and base_c on the NH-Haze dataset.

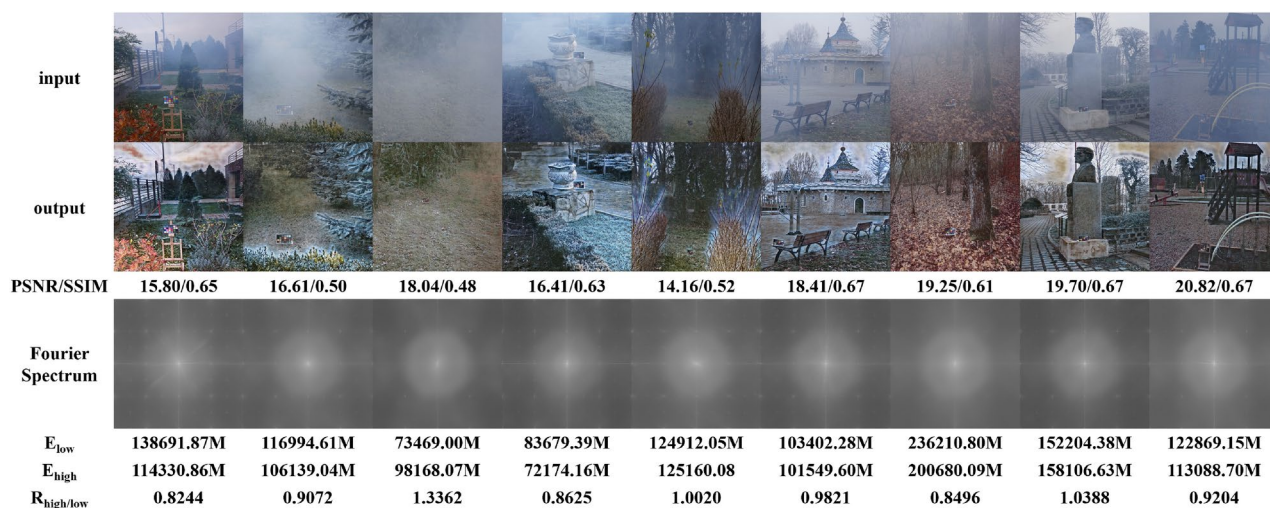


Fig. 7. The generalization test results of SAD-Net on external datasets. The first row shows the foggy images from the dataset. The second row shows the images after defogging using SAD-Net. PSNR and SSIM refer to the PSNR and SSIM of the defogged images compared to the original images.

reducing the presence of noise. In the I-Haze dataset, which represents uniform indoor fog and is lighter than the fog in the other two datasets, both SSIM and PSNR showed a significant improvement for base_c over base_all. This indicates that base_c has a certain restorative effect on indoor uniform light fog. All these data suggest that SDEC makes an important independent contribution to dehazing.

Compare to vanilla convolution, SDEC preserves more high-frequency details, resulting in a richer visual texture. To verify this, we compare the encoder's level 2 final layer output features in both base_all and base_c models, specifically the output from the last SDEBlock in encoder level 2, which serves as the input to level3 before downsampling. Up to this point, the network primarily focuses on feature extraction. We input the NH-Haze dataset's test set into the model and average the feature maps along the channel dimension to approximate the overall feature representation. As shown in Fig. 6, vanilla convolution produces a blurred average feature map in regions with edges and textures. In contrast, base_c, which uses SDEC as its feature extraction module, provides sharper boundary contours and richer high-frequency details.

Generalization ability test

To evaluate the generalization ability of the SAD-Net model, we use the O-Haze dataset⁴⁴ and construct an external test dataset for validation. We test the dehazing effect on these datasets, using PSNR and SSIM as the main evaluation metrics. The results are shown in Fig. 7. Even though the model is not trained on this dataset, SAD-Net is still able to adaptively remove haze and successfully recover most image details and information. From the PSNR and SSIM values, it is clear that the model performs well in image reconstruction and detail retention, proving its strong dehazing capability and robustness in unseen scenarios.

However, when testing on the O-Haze dataset, some of the restored images exhibit artifacts. To investigate the cause of these artifacts in unseen scenarios, we conduct a simple comparison experiment, as shown in Fig. 8. Similar to the previous ablation study, three models—base_all, base_a, and base_c—are applied for dehazing, which were not trained on the O-Haze dataset, and the results are compared. It is observed that base_all leaves a significant amount of haze and some minor artifacts. base_c, using FGA attention and standard convolutions, shows overall results similar to those processed by SAD-Net, with notable similarities, such as the brown tree



Fig. 8. The generalization test results of SAD-Net on external datasets and the comparative experiments on the effects of convolution and attention on generalization test results. The first row shows the input images to the model, the second row shows the SAD-Net test results, followed by the test results of base_all, base_c, and base_a.

branches in the first two columns of the fifth image displaying artifacts. However, SAD-Net demonstrates more severe artifacts. base_a, which uses a detail-enhancing convolution network, exhibits residual haze at the edges of objects when there are more edge details in the image, as seen in the first, second, fourth, and fifth images. Therefore, it is concluded that the cause of the artifacts lies in the model's inability to fully learn the specific haze distribution of the O-Haze dataset, particularly in edge regions, when not trained on this dataset. As a result, when processing unseen images, the model cannot effectively remove haze from the edges and may generate artifacts in these areas. Specifically, the detail-enhancing convolution module may inadvertently amplify the haze in edge areas during the process of enhancing image details, thus hindering effective haze removal. Moreover, the skip connections in the U-Net structure pass enhanced details from the encoder to the decoder, and the overlap of this edge information can exacerbate the artifacts. Additionally, the FGA attention mechanism's excessive focus on edge details further strengthens the remaining haze in these areas, contributing to the creation of artifacts in the final output, especially when the model has not learned the haze characteristics of the dataset.

Loss function

The loss function of the SAD-Net model combines the absolute error loss function and the contrastive loss function. This forms a hybrid loss function, as shown in Eq. (5):

$$Loss = \alpha L_1 + \beta \sum_{i=1}^5 w_i \cdot contrastive_i, L_1 = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (5)$$

In the contrastive loss function, we utilize the VGG19 network to extract image features and divide them into five stages. The objective of the contrastive loss function is to optimize the model by calculating the feature differences between the anchor image (a), the positive sample image (p), and the negative sample image (n). Here, a is the dehazed image output by SAD-Net, p is the fog-free image from the dataset, and n is the input image to the network. We optimize the model by calculating the L1 loss between a and p (denoted as d_{ap}) and the L1 loss between a and n (denoted as d_{an}) for each feature map. The goal is to minimize the distance between a and p , while maximizing the distance between a and n . The specific contrastive loss is given by $contrastive = \frac{d_{ap}}{d_{an}}$. For each feature map, we compute the weighted contrastive loss, and the final total loss is the weighted sum of the contrastive losses across all layers. The weight for each layer, w_i , is set according to the depth of the feature layer, specifically as $w = [\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, 1]$. In this formula, α and β are learnable parameters. w_i is the weight for the i -th VGG19 feature layer, and L_1 is the absolute error loss function, as shown in Equation 5. In this formula, N is the total number of pixels, \hat{y}_i is the i -th pixel of the model output, and y_i is the corresponding ground truth value.

The hybrid loss function aims to maximize the similarity between the dehazed image and the clear image at both the pixel and feature levels. The L_1 loss reduces pixel differences between the dehazed and ground truth images, ensuring consistency in brightness and color. Meanwhile, the contrastive loss extracts high-level information from different feature layers of the VGG19 network. This helps the model recover more details and textures, improving image clarity. The parameters α and β in the loss function are learnable and control the relative importance of L_1 loss and contrastive loss. They adjust dynamically during training. With this design,

the model achieves a balance between global and local details, generating dehazed images that are both clear and detailed.

Conclusion

In this paper, we propose SAD-Net, a single image dehazing network that combines SDEC and a FGA mechanism. SAD-Net significantly improves dehazing performance through the synergy of SDEC and FGA. SDEC effectively separates high-frequency and low-frequency information, accurately extracts and enhances edge and detail features in the image, thereby greatly improving image clarity and quality. Meanwhile, FGA, by integrating spatial and channel attention, effectively addresses the problem of uneven haze distribution, enhancing the model's focus and representation of key features. Extensive experiments conducted on public datasets such as Dense-Haze, NH-Haze, and I-Haze demonstrate the superior performance of SAD-Net. The experimental results show that SAD-Net surpasses current SOTA dehazing methods in terms of PSNR and SSIM metrics, excelling in both quantitative and qualitative analyses, proving its superior dehazing capability and robustness.

Limitations and future work

However, in the development of SAD-Net, we encounter a critical issue that needs to be addressed urgently: how to balance model efficiency and performance. Dehazing models often come with high computational complexity, which can impose a significant computational burden on devices with limited resources, such as drones or autonomous vehicles. Therefore, we believe that an important topic for future research is how to improve the inference speed and overall efficiency of the model without sacrificing the quality of dehazing. We anticipate that strategies such as lightweight network design, model pruning techniques, and model quantization may provide effective solutions for achieving high-performance dehazing models. These methods can not only optimize the computational efficiency of the model but also ensure the real-time and reliable execution of dehazing tasks in various environments.

Data availability

The data used in this study is publicly available in public repositories. The five datasets utilized in this paper, namely Dense-Haze, NH-Haze, I-Haze, RESIDE, and O-Haze, are all publicly accessible. Dense-Haze can be accessed at: <https://data.vision.ee.ethz.ch/cvl/ntire19/dense-haze/>. NH-Haze is available at: <https://data.vision.ee.ethz.ch/cvl/ntire20/nh-haze/>. I-Haze can be obtained from: <https://data.vision.ee.ethz.ch/cvl/ntire18/i-haze/>. O-Haze can be accessed at: <https://data.vision.ee.ethz.ch/cvl/ntire18/o-haze/>. The RESIDE can be downloaded via Baidu Netdisk at <https://pan.baidu.com/s/12gNfOSdn8k7bh6k9Sk0OVA?pwd=x65m> with the extraction code x65m. The training code and weights for the neural network model proposed in this paper can be found at: <https://github.com/niuqj/SAD-Net>. Regarding the RESIDE dataset, it is recommended to use a Mainland China IP when accessing Baidu Netdisk. For those without a Mainland China IP, you can visit the following URL: <https://sites.google.com/view/reside-dehaze-datasets>. Additionally, you can refer to the original RESIDE paper, which can be found at: <https://arxiv.org/abs/1712.04143>.

Received: 1 November 2024; Accepted: 25 February 2025

Published online: 07 April 2025

References

1. Macaulay, M. O. & Shafiee, M. Machine learning techniques for robotic and autonomous inspection of mechanical systems and civil infrastructure. *Autonom. Intell. Syst.* **2**, 8 (2022).
2. Li, L. et al. Semi-supervised image dehazing. *IEEE Trans. Image Process.* **29**, 2766–2779 (2019).
3. Ullah, H. et al. Light-dehazenet: a novel lightweight cnn architecture for single image dehazing. *IEEE Trans. Image Process.* **30**, 8968–8982 (2021).
4. Ren, W., Pan, J., Zhang, H., Cao, X. & Yang, M.-H. Single image dehazing via multi-scale convolutional neural networks with holistic edges. *Int. J. Comput. Vision* **128**, 240–259 (2020).
5. Yang, G., Yang, H., Yu, S., Wang, J. & Nie, Z. A multi-scale dehazing network with dark channel priors. *Sensors* **23**, 5980 (2023).
6. Zhang, X., Wang, T., Luo, W. & Huang, P. Multi-level fusion and attention-guided cnn for image dehazing. *IEEE Trans. Circuits Syst. Video Technol.* **31**, 4162–4173 (2020).
7. Cui, Y., Ren, W., Cao, X. & Knoll, A. Focal network for image restoration. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* 12955–12965. <https://doi.org/10.1109/ICCV51070.2023.01195> (2023).
8. He, S., Chen, Z., Wang, F. & Wang, M. Integrated image defogging network based on improved atmospheric scattering model and attention feature fusion. *Earth Sci. Inf.* **14**, 2037–2048 (2021).
9. Qian, W., Zhou, C. & Zhang, D. Ciasm-net: a novel convolutional neural network for dehazing image. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)* 329–333 (IEEE, 2020).
10. Anan, S. et al. Image defogging framework using segmentation and the dark channel prior. *Entropy* **23**, 285 (2021).
11. Li, B., Peng, X., Wang, Z., Xu, J. & Feng, D. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision* 4770–4778 (2017).
12. 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**, 5187–5198 (2016).
13. Das, S., Islam, M. S. & Amin, M. R. Gca-net: utilizing gated context attention for improving image forgery localization and detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 81–90 (2022).
14. Narasimhan, S. G. & Nayar, S. K. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 713–724 (2003).
15. Wu, H. et al. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 10551–10560 (2021).
16. Dong, H. et al. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2157–2167 (2020).

17. 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**, 11908–11915 (2020).
18. 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* 7314–7323 (2019).
19. Qiu, Y. et al. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 12802–12813 (2023).
20. Song, X. et al. Wsamf-net: Wavelet spatial attention-based multistream feedback network for single image dehazing. *IEEE Trans. Circuits Syst. Video Technol.* **33**, 575–588 (2022).
21. Qin, Z., Zhang, P., Wu, F. & Li, X. Fcanet: Frequency channel attention networks. In *Proceedings of the IEEE/CVF international conference on computer vision* 783–792 (2021).
22. Fu, K. et al. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS J. Photogramm. Remote. Sens.* **161**, 294–308 (2020).
23. Chen, Z., He, Z. & Lu, Z.-M. Dea-net: single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Trans. Image Process.* (2024).
24. Mallat, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693 (1989).
25. Ancuti, C. O., Ancuti, C., Sbert, M. & Timofte, R. Dense-haze: a benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE International Conference on Image Processing (ICIP)* 1014–1018 (IEEE, 2019).
26. Ancuti, C., Ancuti, C. O., Timofte, R. & De Vleeschouwer, C. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *Advanced Concepts for Intelligent Vision Systems: 19th International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings* 19 620–631 (Springer, 2018).
27. Li, B. et al. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **28**, 492–505. <https://doi.org/10.1109/TI.2018.2867951> (2019).
28. Ancuti, C. O., Ancuti, C. & Timofte, R. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* 444–445 (2020).
29. Sargül, M., Ozyildirim, B. M. & Avci, M. Differential convolutional neural network. *Neural Netw.* **116**, 279–287 (2019).
30. Ojala, T., Pietikainen, M. & Maenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971–987 (2002).
31. Yu, Z. et al. Searching central difference convolutional networks for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 5295–5305 (2020).
32. Yu, Z., Qin, Y., Zhao, H., Li, X. & Zhao, G. Dual-cross central difference network for face anti-spoofing. arXiv preprint [arXiv:2105.01290](https://arxiv.org/abs/2105.01290) (2021).
33. Yu, Z. et al. Nas-fas: Static-dynamic central difference network search for face anti-spoofing. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3005–3023 (2020).
34. Rong, Z. & Jun, W. L. Improved wavelet transform algorithm for single image dehazing. *Optik* **125**, 3064–3066 (2014).
35. Fu, M., Liu, H., Yu, Y., Chen, J. & Wang, K. Dw-gan: a discrete wavelet transform gan for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 203–212 (2021).
36. Khmag, A., Al-Haddad, S., Ramli, A. R. & Kalantar, B. Single image dehazing using second-generation wavelet transforms and the mean vector l2-norm. *Vis. Comput.* **34**, 675–688 (2018).
37. Liu, X., Zhang, H., Cheung, Y.-M., You, X. & Tang, Y. Y. Efficient single image dehazing and denoising: an efficient multi-scale correlated wavelet approach. *Comput. Vis. Image Underst.* **162**, 23–33 (2017).
38. Ahmed, N., Natarajan, T. & Rao, K. R. Discrete cosine transform. *IEEE Trans. Comput.* **100**, 90–93 (1974).
39. Wu, Y. & He, K. Group normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)* 3–19 (2018).
40. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
41. Wu, K. et al. Data-balanced transformer for accelerated ionizable lipid nanoparticles screening in mrna delivery. *Brief. Bioinform.* **25**, bbae186 (2024).
42. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
43. Guo, C.-L. et al. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 5812–5820 (2022).
44. Ancuti, C. O., Ancuti, C., Timofte, R. & De Vleeschouwer, C. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 754–762 (2018).

Author contributions

Q.N., K.W., Z.H., J.Z. and L.L. conceptualized and designed the study; Q.N. conducted the network design and drafted the initial manuscript; Q.N. and K.W. participated in the creation of all tables and figures; K.W., Z.H., and J.Z. provided significant intellectual contributions and participated in key revisions of the manuscript; L.L. supervised the entire study and ensured that it adhered to the highest academic standards. All authors have read and approved the final manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.H. or L.L.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025