Journal of Artificial Intelligence and Technology, 2022, 2, 100-110
https://doi.org/10.37965/jait.2022.0110

ISTP
RESEARCH ARTICLE

Single-Image Dehazing Based on Two-Stream Convolutional Neural Network

Meng Jun,[1] Li Yuanyuan,[1] Liang HuaHua,[2] and Ma You[2]

[1]Chongqing Key laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[2]Automation College, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

(Received 25 April 2022; Revised 19 June 2022; Accepted 20 June 2022; Published online 30 June 2022)

Abstract: The haze weather environment leads to the deterioration of the visual effect of the image, and it is difficult to carry out the work of the advanced vision task. Therefore, dehazing the haze image is an important step before the execution of the advanced vision task. Traditional dehazing algorithms achieve image dehazing by improving image brightness and contrast or constructing artificial priors such as color attenuation priors and dark channel priors. However, the effect is unstable when dealing with complex scenes. In the method based on convolutional neural network, the image dehazing network of the encoding and decoding structure does not consider the difference before and after the dehazing image, and the image spatial information is lost in the encoding stage. In order to overcome these problems, this paper proposes a novel end-to-end two-stream convolutional neural network for single-image dehazing. The network model is composed of a spatial information feature stream and a high-level semantic feature stream. The spatial information feature stream retains the detailed information of the dehazing image, and the high-level semantic feature stream extracts the multi-scale structural features of the dehazing image. A spatial information auxiliary module is designed and placed between the feature streams. This module uses the attention mechanism to construct a unified expression of different types of information and realizes the gradual restoration of the clear image with the semantic information auxiliary spatial information in the dehazing network. A parallel residual twicing module is proposed, which performs dehazing on the difference information of features at different stages to improve the model's ability to discriminate haze images. The peak signal-to-noise ratio (PSNR) and structural similarity are used to quantitatively evaluate the similarity between the dehazing results of each algorithm and the original image. The structure similarity and PSNR of the method in this paper reached 0.852 and 17.557dB on the HazeRD dataset, which were higher than existing comparison algorithms. On the SOTS dataset, the indicators are 0.955 and 27.348dB, which are sub-optimal results. In experiments with real haze images, this method can also achieve excellent visual restoration effects. The experimental results show that the model proposed in this paper can restore desired visual effects without fog images, and it also has good generalization performance in real haze scenes.

Keywords: attention mechanism; image dehazing; semantic feature; spatial information; two-stream network

I. INTRODUCTION

Natural images captured in haze weather have problems with low contrast, low color saturation, and high brightness. Haze images as input will significantly increase the difficulty of processing advanced vision tasks. Therefore, dehazing such images is an important step before performing advanced vision tasks [1].

Traditional methods [2–5] improve the visual effect of hazy images by increasing the contrast and enhancing the detailed features. However, due to the lack of physical model support, improving contrast and color information only cannot achieve true dehazing [6]. Therefore, some studies [7–10] have extrapolated unknown parameters in atmospheric scattering models [11] to restore clear images by constructing a priori knowledge. But fixed prior knowledge cannot adapt to changing scenarios, resulting in low robustness in complex real-world scenarios [12].

With the rise of deep learning, many methods [13] began to employ neural networks to estimate unknown parameters in physical models. But the estimation accuracy of unknown parameters affects the quality of dehazed images [14]. Therefore, many methods begin to directly dehaze images with end-to-end deep learning models. End-to-end deep learning methods [12,15–18] directly regress haze-free images from hazy images, without relying on unreliable physical models parameter estimation, achieving excellent dehazing performance. However, the dehazing backbone network of these methods [12,16,17] uses standard encoder–decoder structures. The cascaded codecs do not take into account the differences before and after image dehazing, resulting in incomplete dehazing results or color deviations. In addition, the downsampling in the encoding stage loses a lot of spatial information, which degrades the quality of the reconstructed image in the decoding stage. Recently, two-stream convolutional neural network [19,20], as a novel convolutional neural network architecture, has been successfully applied to image classification tasks. The model computes two-stream features by constructing two sets of feature extraction branches with different structures and then fuses the two-stream features for output prediction. The performance of the model is improved relative to that of the single-feature stream. However, in the image dehazing task, only a small number of

Corresponding author: Yuanyuan Li (email: liyy@cqupt.edu.dn).

classification output layers cannot effectively regress clear images from the fused two-stream features. Although there are already models [12] that fully regress the features by adding enhancement modules in the output stage, it also brings additional computational effort.

In response to the above problems, this paper proposes a two-stream convolutional neural network for single-image dehazing. Two dehazing feature streams are constructed in this network: spatial information feature stream and high-level semantic features stream. The spatial information feature stream retains the dehazed image detail information, and the high-level semantic feature stream extracts the abstract multi-scale semantic information in the dehazed image. This paper designs a spatial information auxiliary module, which uses the attention mechanism to eliminate the attention difference between different feature streams, so that the multi-scale semantic information of the hazy image can effectively assist the spatial information to restore the haze-free image. A parallel residual twicing module is embedded in the high-level semantic feature stream, which performs dehazing on the difference information of the features at different stages in the encoding and decoding process, so that the model focuses on the change area of the features during the reconstruction process. By cascading multiple parallel residual twicing modules, the model gradually recovers the information during the feature decoding process. Comparative experiments show that the dehaze model proposed in this paper has better dehazing ability than the previous dehazing methods.

The contributions of our work are as follows:

- An end-to-end two-stream multi-scale feature dehazing network with parallel extraction of spatial information features and multi-scale high-level semantic features is proposed. The ability to use a unified objective function to train the network to extract two-stream features avoids the problem of spatial information loss.
- The spatial information auxiliary module is proposed, which uses the attention mechanism to merge the semantic information and spatial information and construct a unified representation of different types of information, so that the semantic information can assist the spatial information to gradually restore the clear image, which improves the dehazing ability of the model.
- A parallel residual twicing module is proposed. By dehazing the feature difference information at different stages instead of directly dehazing the features, the model can focus on the feature change area during the reconstruction process, and the model's ability to discriminate hazy images is enhanced.

II. RELATED WORK

A. SINGLE-IMAGE DEHAZING

In the traditional dehazing method, the image enhancement method starts from the image itself, improves the image contrast, and strengthens the image details. Reference [2] used a multi-scale Laplacian scheme to mix a set of artificially underexposed haze images to synthesize a haze-free image. Reference [3] obtained a set of underexposed images through gamma correction. In order to preserve the overall structure and local details, both global and local components were first decomposed to construct an effective pixel-by-pixel weight map, and then the weight map was used to guide pixel-level fusion and finally output the dehazed image after adjusting the saturation. Due to the lack of physical model support, such methods only improve contrast and color information, so they cannot achieve dehazing in the true sense. Using the atmospheric scattering model, many methods based on image restoration estimate the unknown parameters through effective priors and reasonable assumptions to restore a clear image. Reference [9] first trained a scene depth estimation model with a differentiable function and then restored a haze-free image based on the atmospheric light and transmission map estimated. Reference [10] established a linear model of the hazy image with prior information and estimated the transmission function with the help of the depth map of the hazy image. But for white object areas or haze images, this linear model may not hold.

With the rise of deep learning, many methods began to use neural networks to directly estimate the transfer function and atmospheric light in atmospheric scattering models. Reference [13] proposed a dehazing model named DehazeNet. The model estimates its transmission map directly from a given hazy image and then feeds the hazy image and the estimated transmission map into an atmospheric scattering model to obtain a clear image. The activation function of the network is a bilateral nonlinear rectification function, which constrains the numerical space of the feature map between [0,1] to improve the dehazing quality. Reference [21] proposed the AOD-Net (All-in-one dehazing network) dehazing model. The model does not estimate the transfer function and atmospheric light separately but combines the two into one parameter, using a lightweight CNN performs regression to get the final clear image.

The estimation accuracy of the unknown parameters by the deep learning network will affect the quality of the dehazed image, so in recent years, many methods have begun to apply end-to-end deep learning models to directly regress the clear image from the hazy image. Reference [15] proposed the GridDehazeNet model, in which a trainable preprocessing module was designed to solve the shortcomings of insufficient diversity and low pertinence of manually selected preprocessing methods. An attention-based multi-scale network is used for the dehazing backbone of the model, which effectively alleviates the inflexibility of traditional multi-scale methods. Reference [22] constructed an image dehazing network with an encoder–decoder structure using octave convolution. In addition, this work designs self-attention modules for features at different stages to enhance the dehazing effect of the model. Reference [16] proposed an image restoration network constrained by a physical model. The restoration network was supervised by a discriminator constructed based on the physical model during training to ensure the final quality of the restored image. Reference [12] based on generative adversarial networks used the discriminator to guide the image reconstruction network to generate rough dehazed images and then used the enhancement network to enhance the color and detail effects of the images, thus proposing an enhanced image-to-image dehazing network. The Cycle GAN model [23] proposes a cycle consistency loss for image-to-image translation, which can accomplish image-to-image translation without relying on paired data. Based on this, [17] proposed a domain-adaptive framework for single-image dehazing. The method first bidirectionally transforms the real hazy image and the synthetic hazy image to reduce the domain shift between the synthetic domain and the real domain; then the transformed image and its original image are used as the input of the dehazing network. Although these methods have achieved excellent image dehazing results so far, the dehazing backbone of these methods adopts a standard encoder–decoder structure. The cascaded codecs do not take into account the differences before and after image dehazing, resulting in incomplete dehazing results or color deviations.

In addition, this structural at encoding stage loses a lot of spatial information due to downsampling, which destroys the quality of the reconstructed image.

B. TWO-STREAM NETWORK

Two-stream network, as a novel convolutional neural network structure, has been used for image classification tasks. By constructing two feature extraction branches with different structures, two heterogeneous features are calculated to solve the limitation of model performance brought by single-image information. For example, [19] proposed a two-stream network model, which can simultaneously extract local and global spatial features in the input data. The two-stream model proposed by [20] could simultaneously extract spatial and transformed features. The advantages of the two sets of features complement each other, which improved the performance of the model compared to a single-feature stream. At present, there are few works using two-stream network for image dehazing. In addition, if the image reconstruction model is constructed according to the mode of classification network in the above method, only relying on a small number of output layers cannot make full use of the combined two-stream features. Although there is already model [12] by adding augmentation modules to fully learn the features output, this also brings additional storage and computational resource consumption.

Aiming at the above problems, this paper proposes a two-stream multi-scale feature dehazing network. The main body of the model adopts a two-stream structure so that the model can maintain the spatial information of the input image. A spatial information auxiliary module is proposed, which uses the attention mechanism to eliminate the attention difference between the two feature streams, so that the semantic information of the hazy image can effectively assist the spatial information to restore the clear image. The encoding and decoding part of the dehazing network is embedded with a parallel residual twicing module, which learns the difference between features and gradually recovers the information in the image during the feature decoding process, so that the model can extract more effective image features.

III. THE PROPOSED SOLUTION

A. THE OVERALL NETWORK STRUCTURE

As shown in Fig. 1, this model contains two feature streams: a spatial information feature stream that extracts spatial features of images and a high-level semantic feature stream that extracts multi-scale semantic information of images. Specifically, the high-level semantic feature stream is an encoder–decoder network based on a parallel residual twicing module. The feature encoding part consists of a feature extraction module and a convolutional layer with stride 2 concatenated, and the feature decoding part consists of parallel residual push-pull. The pull module and the deconvolution layer with stride 2 are concatenated. The spatial information feature flow consists of a small full-resolution convolutional neural network. The features calculated by the spatial information auxiliary module are input into the feature extraction module so that the feature stream can also use the semantic information for reconstruction tasks. The final spatial information feature stream and high-level semantic feature stream are spliced according to the feature channel direction and then enter the output module to obtain a haze-free image.

B. SPATIAL INFORMATION AUXILIARY MODULE

The architecture of the spatial information auxiliary module is shown in Fig. 2. The method first extracts the effective structural information in the high-level semantic feature stream. High-level

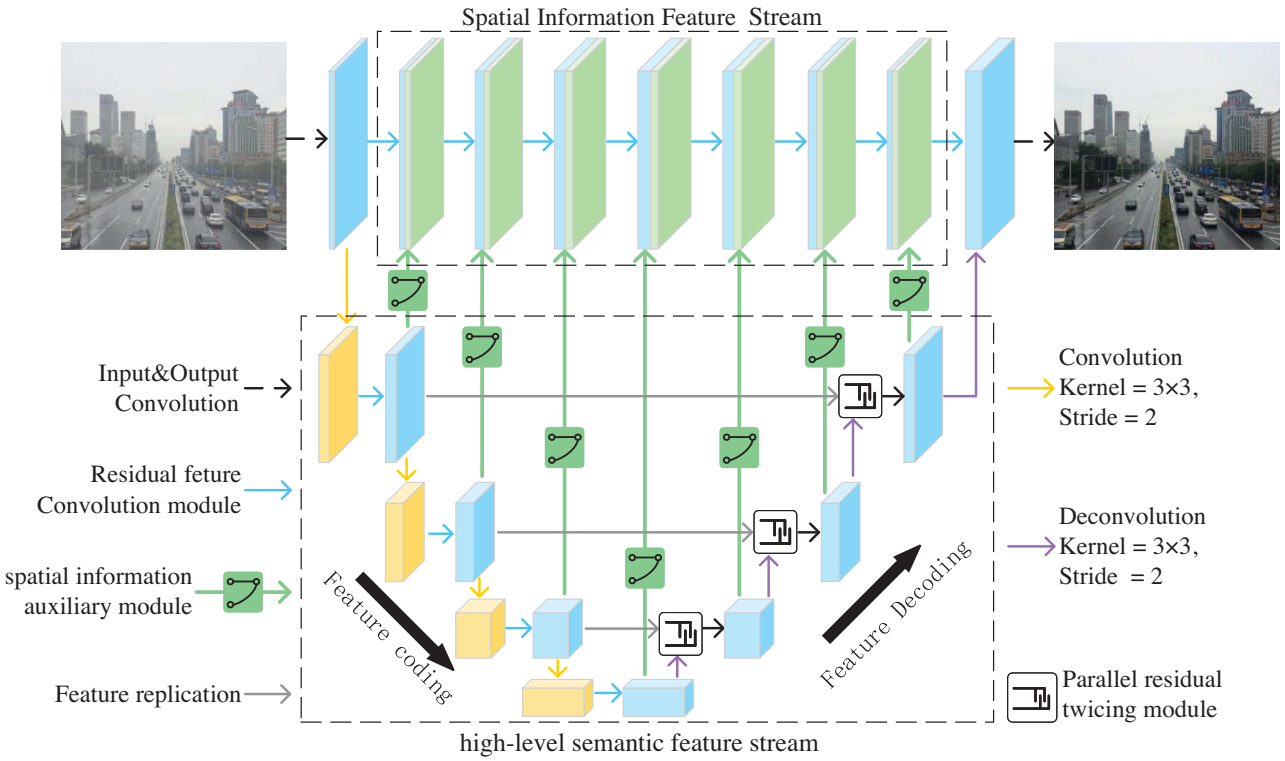

Fig. 1. The overall network structure.

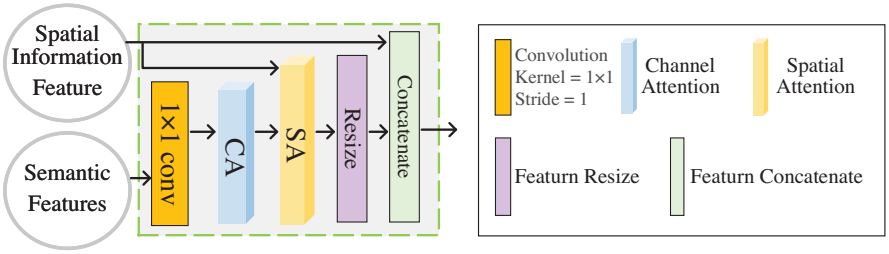

Fig. 2. Architecture of spatial information auxiliary module.

semantic feature stream attention (HSFSA) is then computed to highlight semantic feature maps that are effective for the dehazing task. Two-stream mixed attention (TSMA) is then computed to emphasize regions of interest in the spatial information features stream. Finally, the feature upsampling of attention calculation and spatial information feature splicing are completed. Using the spatial information auxiliary module, the image structure information in the high-level semantic feature stream is gradually integrated into the spatial information feature stream, and the full-resolution network can utilize the image structure features to improve the dehazing effect of the model.

In the high-level semantic features of convolutional neural networks, some feature maps have strong correlation or no feature information. Therefore, in order to extract the effective information in the features, the method uses convolution to merge the feature channels to achieve the feature extraction.

Attention mechanism [24] has been widely used in convolutional neural networks to emphasize the region of interest of the model and enhance the adaptability of the model. In the attention model, a specific algorithm is used to calculate the weight information in a specific dimension for the feature map. By assigning weights, the feature information that deserves more attention is highlighted, and the feature information that is not important or even unfavorable for prediction is ignored. Therefore, in HSFSA, attention features are used to treat different feature channels in high-level semantic features unequally, providing additional flexibility for the network model to handle different types of fog. As shown in Fig. 3, for the input feature F, the semantic information of the feature map is first aggregated using average pooling and max pooling to generate two feature vectors with different semantics: F^c_{avg} and F^c_{max}. In order to obtain the weights of different channels, these two sets of features are first spliced and then pass through a convolutional layer and Sigmoid function to obtain the channel attention feature F^c_a:

$$F^c_a = \sigma(\text{Conv}(\text{Cat}(F^c_{\text{avg}}, F^c_{\text{max}}))) \tag{1}$$

where σ represents the Sigmoid activation function, Conv represents the two-dimensional convolution, and Cat represents the splicing of tensors along the number of channels. Finally, multiply each element of the channel attention feature with each feature map of the input feature:

$$F^* = F \otimes F^c_a \tag{2}$$

Considering the uneven distribution of haze on different image pixels and the different processing procedures of the full-resolution network and the encoding–decoding network, the regions of interest for the features of these two parts are not exactly the same. If they are directly fused, it will hinder the expression of spatial information features. Therefore, the TSMA is utilized in the spatial information auxiliary module to unify the focus of the two sets of features.

As shown in Fig. 4, in the TSMA, for the feature map F from the high-level semantic feature stream, the maximum spatial information feature F^s_{max} and the average spatial information feature F^s_{avg} are first calculated along the channel direction:

$$\begin{cases} F^s_{\text{max}_{w,h}} = \text{Max}(F_{1,w,h}, F_{2,w,h} \cdots, F_{ch,w,h}) \\ F^s_{\text{avg}_{w,h}} = \frac{\sum_{ch=1}^{CH} F_{ch,w,h}}{CH} \end{cases} \tag{3}$$

where w,h represents the abscissa and ordinate on the feature map, ch represents the number of the feature map, and CH represents the total number of feature maps.

Then the convolution module is used to extract a self-convolution spatial feature F^s_{selfconv} of size $1 \times \text{H} \times \text{W}$ from the $\text{C} \times \text{W} \times \text{H}$ feature map F:

$$F^s_{\text{selfconv}} = \tau(\text{Conv}(F)) \tag{4}$$

where τ represents the nonlinear rectification activation function ReLU. In order to enable high-level semantic features to better focus on the area of interest for spatial information features, the spatial information feature map F^h of the corresponding stage is downsampled to $\text{W} \times \text{H}$ and then input to the convolution module to obtain a size of $1 \times \text{H} \times \text{W}$ for advanced Semantic convolution spatial features F^s_{hconv}:

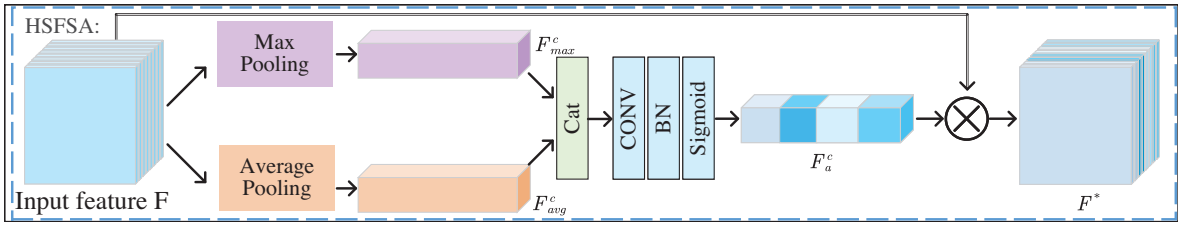

Fig. 3. The structure of the high-level semantic feature stream attention (HSFSA).

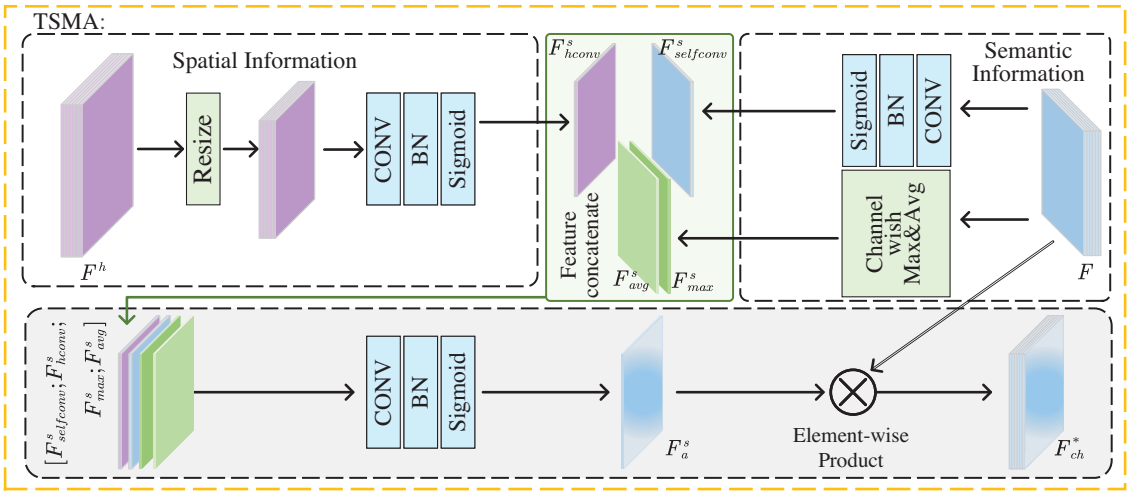

Fig. 4. The structure of two-stream mixed attention (TSMA).

$$F^s_{\text{hconv}} = \tau(\text{Conv}(\text{Resize}(F^h))) \tag{5}$$

After all the spatial features are spliced, the convolution module completes the calculation of the TSMA tensor F^s_a:

$$F^s_a = \sigma(\text{Conv}([F^s_{\max}; F^s_{\text{avg}}; F^s_{\text{selfconv}}; F^s_{\text{hconv}}])) \tag{6}$$

Finally, perform point-to-point multiplication of the TSMA tensor with each feature map to complete the weighted calculation of the input features:

$$F^*_{ch} = F_{ch} * F^s_a \tag{7}$$

C. PARALLEL RESIDUAL TWICING MODULE

As shown in Fig. 5a, in the U Net [25] structure, the horizontal transfer process of features adopts the direct splicing:

$$i^n = G([i^{n+1}; j^n]) \tag{8}$$

Among them, i^n represents the feature of the n-th layer of the decoding part, j^n represents the feature of the n-th layer of the encoding part, and G represents the computing unit used for dehazing, which represents the feature extraction module in the model proposed in this paper. In this U Net structure, feature splicing is performed first, and then convolution is performed to extract features. For image segmentation tasks, the feature extraction module can simultaneously receive high-level semantic features and low-level fine-grained information to improve classification accuracy. However, the U-Net structure is not a fusion method specially designed for image dehazing tasks, and the structure does not consider the difference between before and after reconstructed features.

The twicing technique has been shown to be effective for image restoration tasks [26]. In this technology, the iterative way of features is improved, specifically filtered version of the data residual was added back to the inital estimate. The twicing structure shown in Fig. 5b is as follows:

$$i^n = G(i^{n+1} - j^n) + i^{n+1} \tag{9}$$

For the image dehazing task, the residual is defined as the difference between a hazy image and its dehazed image. Since a single dehazing computing unit cannot recover the perfect haze-free image features from the features of the hazy image, the model needs to integrate multiple dehazing computing units to gradually reconstruct the haze-free image, so the residual between the features at different stages will not be different zero. By extracting features from the residuals and then adding them back to the estimated features, the dehazing module can pay more attention to the hazy parts of the features.

In order to enable the shallow layers in the network to also receive the residual signal and further improve the performance of the network dehazing, as shown in Fig. 5c, this paper uses a parallel twicing method to complete this process. Parallel twicing is represented by the following formula:

$$i^n = G(i^{n+1} - j^n) + i^{n+1} + j^n \tag{10}$$

Compared with the ordinary twicing structure, in the parallel twicing structure, the estimated haze-free image features are added to the image semantic features generated by the previous layer and the corresponding shallow image semantic features so that both parts can be A residual signal is received.

D. LOSS FUNCTION IN TRAINING STAGE

In this paper, the mean squared error (MSE) is used as the loss function to calculate the difference between the network output and the corresponding real clear image. The loss function is expressed as:

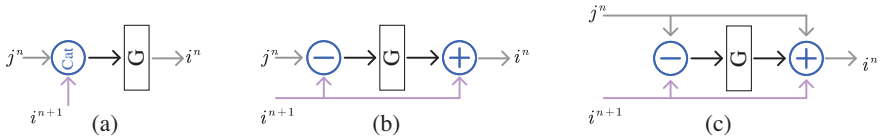

Fig. 5. Different ways to transfer features, (a) feature concatenate, (b) twicing technology, and (c) the method proposed in this article.

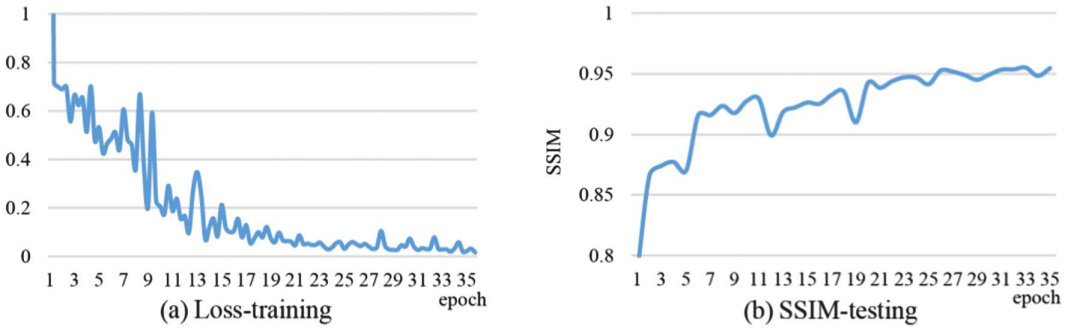

Fig. 6. The dehazing results of different methods on the HazeRD dataset.

$$L = \frac{1}{n_b}\sum_{i=1}^{n_b}(Y^i - \phi(X^i, w))^2 \tag{11}$$

Among them, n_b represents the batch size of the input data, Y represents the real clear image, X represents the hazy image, ϕ represents the proposed two-stream multi-scale feature dehazing network in this paper, and w represents the parameters in the model. The training process of the model is summarized as Algorithm 1.

Algorithm 1. Two-stream multi-scale convolutional neural network for image dehazing training algorithm

Input:
$n_b \leftarrow$Batch size
$t \leftarrow$Training times of model
$w \leftarrow$ Parameters of untrained two-stream multi-scale feature dehazing network
Output:
The parameters of the trained two-stream multi-scale feature dehazing network $\rightarrow w$;
1: **for** $num = 1$; $num \leq t$ **do**
2: haze image sample $X = \{x^{(1)}, \ldots, x^{(n)}\}$
3: clear free image sample $Y = \{y^{(1)}, \ldots y^{(n)}\}$
4: $X \leftarrow \phi(X, w)$, The haze image is input into the model to obtain the prediction image, ϕ represents the proposed two-stream multi-scale feature dehazing network in this paper
5: $L \leftarrow \frac{1}{n_b}\sum_{i=1}^{n_b}(Y - X)^2$, The difference between the predicted image and the real image is calculated according to equation (11)
6: Update the parameters w of the two-stream multiscale feature dehazing network using gradient descent for L
7: **End for**

IV. Experiment

In this section, the specific training settings of the network model are first explained, and then the dehazing effect of the model is shown. In the experiment, two synthetic public data [27,28] were used for training and testing, and the Structural Similarity Index (SSIM) and the peak signal-to-noise ratio (PSNR) are a total of two full reference indicators to evaluate the dehazing effect of the image. The larger the value, the closer the restoration result is to the original image. Finally, this paper tests the image dehazing effect of all methods in real foggy conditions and performs ablation analysis on the proposed module.

A. DATASET

ITS (Indoor Training Set) and OTS (Outdoor Training Set) are two large-scale synthetic haze image datasets for training dehazing models [27]. ITS contains 10,000 clear images that can be used for training. Each clear image corresponds to 10 haze images with atmospheric light A between [0.7, 1.0] and transmittance beta between [0.6, 1.8]. The OTS contains 2061 real outdoor images from Beijing's real-time weather. Using the estimated image depth information, as well as the set atmospheric light and transmittance, each clear image corresponds to 35 haze images. SOTS (Synthetic Object Testing Set) [27] and HazeRD [28] datasets are used to evaluate the dehazing performance of the model. SOTS is a test set in RESIDE that contains 500 pairs of indoor haze/clear images and 500 pairs of outdoor haze/clear images; HazeRD is the result of Zhang et al. Synthetic haze dataset contains 75 pairs of synthetic clear haze/clear images.

B. TRAINING SETTINGS

The network model is implemented in Pytorch and trained using the Adam optimization method [29], where β_1 and β_2 are set to 0.9 and 0.999, respectively. Five thousand indoor clear images and 5,000 outdoor clear images were randomly selected from ITS and OTS, for a total of 10,000 clear images to train the model. In order to increase the generalization ability of the model to dehaze and to simulate the process of haze formation from shallow to deep in the real world for each clear image, three corresponding synthetic haze images with haze density from shallow to deep are selected. During the training process, all image sizes are uniformly scaled to 256×256, and the pixel values are normalized to between $[-1, 1]$ after being read in RGB format. The model was trained with a batch size of 8, a learning rate of 1×10^{-4}, and trained on an RTX Titan GPU for 35 epochs. The loss change of the model in the training phase and the change process of the SSIM indicator on the test set are shown in Fig. 6. It can be observed from the figure that in the first 20 epochs of the training process, the loss value of the model rapidly drops below 0.2, and the SSIM indicator on SOTS rapidly approaches 0.95 at the same time. After that, the model loss slowly decreased and tended to stabilize. The model successfully converged, and the SSIM indicator finally reached 0.955.

C. RESULT

1) COMPOSITE DATASET. In this section, the method in this paper and the method based on image enhancement: AMEF [2]; the

method based on image restoration: CO [7], CAP [10], DEFADE [8]; methods based on convolutional neural networks: MSBDN [14], RDN [30], GDN [15], EPDN [12], DA [17], and PBG [16] for comparison. All methods use SSIM and PSNR metrics to objectively evaluate their performance on SOTS and HazeRD datasets, and the test image size is uniformly 512×512. The dehazing images of all methods are shown in Figs. 7–9.

From Figs. 8, 9, it can be observed that the images of CO, CAP, AMEF, and DEFADE have limited ability to dehaze, and there is still obvious fog remaining in the image after dehazing. End-to-end models for direct estimation of sharp images, RDN, GDN, EPDN, DA, MSBDN, and PBG, yield better results than other indirect methods. However, there is still a certain amount of haze remaining in the dehazing results of GDN. In some cases,

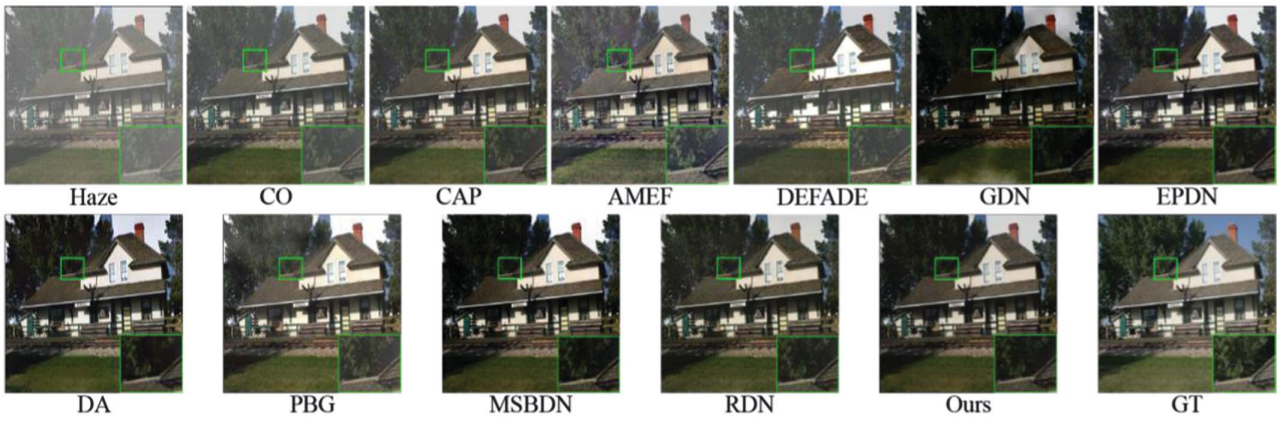

Fig. 7. The dehazing results of different methods on the HazeRD dataset.

Fig. 8. The dehazing results of different methods in the indoor scene of the SOTS dataset.

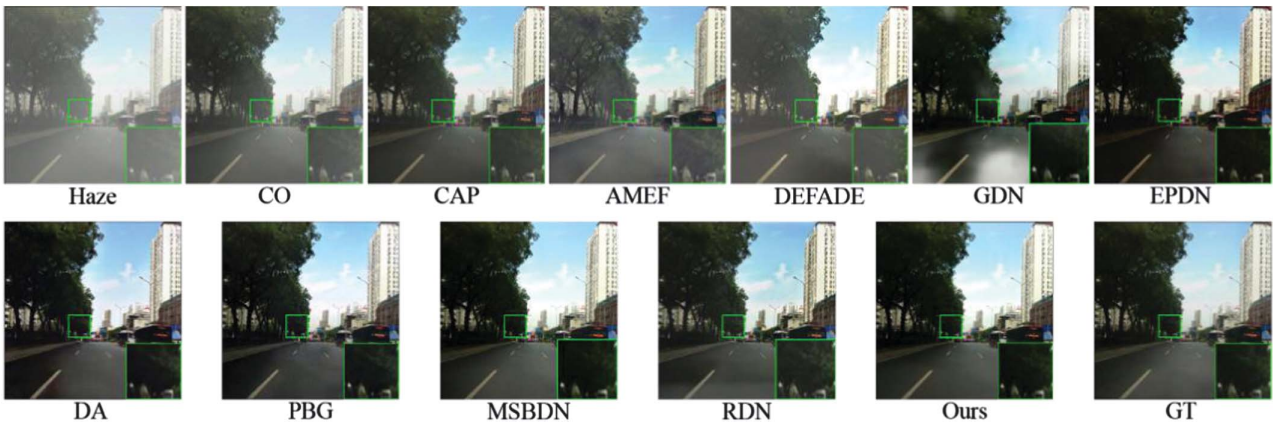

Fig. 9. The dehazing results of different methods in the outdoor scene of the SOTS dataset.

Table I Quantitative comparisons of different methods on two data sets. Two best results are marked bold. (1) the best result and (2) the second best result

	SOTS		HazeRD	
Test Set	**SSIM**	**PSNR**	**SSIM**	**PSNR**
CO	0.841	19.923	0.769	14.879
CAP	0.857	20.662	0.804	15.803
AMEF	0.821	18.604	0.806	17.027
DEFADE	0.766	17.911	0.735	15.598
GDN	0.897	20.760	0.680	12.528
EPDN	0.898	22.747	0.796	15.717
DA	0.906	22.872	0.814	17.151
PBG	0.937	26.539	**0.850(2)**	**16.705(2)**
MSBDN	**0.982(1)**	**33.520(1)**	0.808	14.885
RDN	0.943	24.230	0.845	17.050
ours	**0.955(2)**	**27.348(2)**	**0.852(1)**	**17.557(1)**

EPDN dehaze results in outdoor scenes will be darker than real images. In addition, DA is accompanied by some color distortion and poor results for high-frequency details such as edges and blue sky. RDN has poor results for outdoor scenes, as shown in Fig. 9, where white bands appear on the road surface. MSBDN achieves excellent dehazing results on SOTS, but on the HazeRD dataset, the excessive contrast of the scene causes some dark details to be lost, such as the shadowed parts in the foliage. Compared with these methods, the method in this paper can effectively restore the structure and details of the image, and the reconstructed image is closer to the original image.

The objective evaluation metrics of the 11 methods on the datasets are given in Table I and Fig. 10. The method proposed in this paper obtains the suboptimal PSNR and SSIM values on the SOTS dataset and the optimal PSNR and SSIM values on the Hazard dataset. Compared with the latest PBG, the PSNR and SSIM of the method in this paper are improved by 0.809 dB and 0.018, respectively, on the SOTS dataset. For the HazeRD dataset, the PSNR and SSIM metrics of the dehazing results of the method proposed in this paper are 0.852 dB and 0.02, respectively, higher than the PBG.

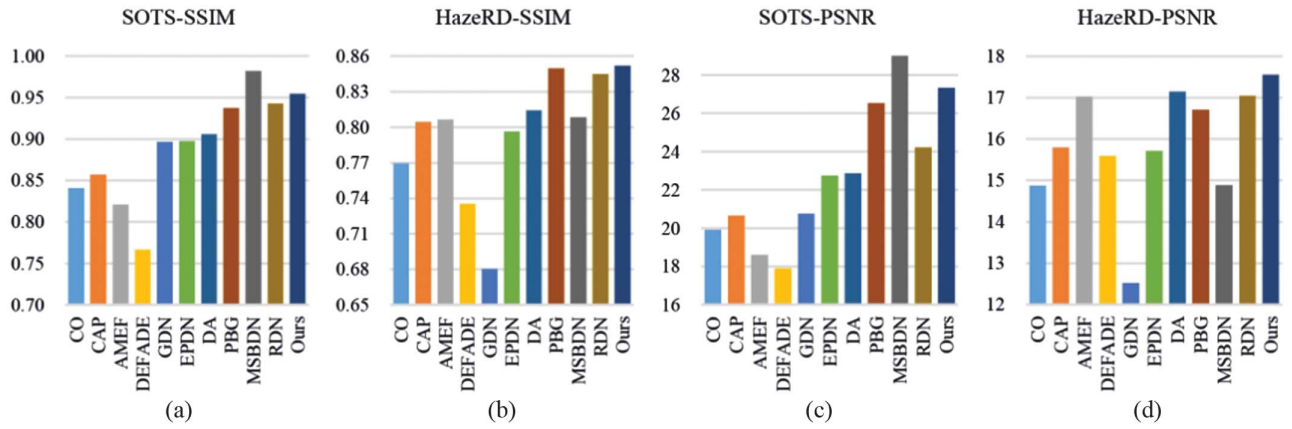

Fig. 10. Objective evaluation index results of different methods on SOTS and HazeRD datasets. The higher the SSIM and PSNR values, the better the effect of image reconstruction. (a) SSIM of the SOTS dataset, (b) SSIM of the HazeRD dataset, (c) PSNR of the SOTS dataset, and (d) PSNR of the HazeRD dataset.

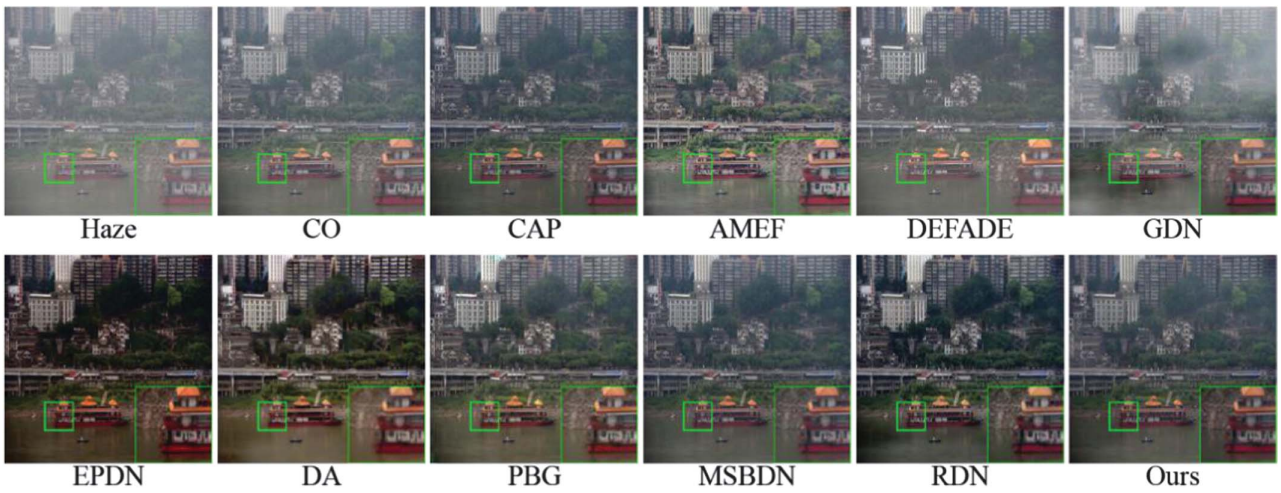

Fig. 11. The dehazing result in real scenes.

Fig. 12. The dehazing result in real scenes.

2) REAL SCENE. The dehazing effects of all 11 methods on the real fog image are shown in Figs. 11, 12. It can be seen from the result images that EPDN defogs most thoroughly, but the overall image brightness is dark leading to excessive loss of dark details, and the image color is also distorted, with an overall bias toward red-yellow. The MSBDN model with the best index in the synthetic data has limited actual dehazing ability, as shown in the local zoom results in Fig. 11, there are obvious haze on the river bank and the stern part after dehazing. The dehazing ability of CO, CAP, DEFADE, and GDN for the real image is weak, and there are obvious haze left in the image. Color distortion and detail loss occurred as a result of dehazing by DA, for example, the texture on the river bank was blurred in Fig. 11, and the lawn of the course turned yellow in Fig. 12. The image brightness of RDN was unnatural, for example, part of the water surface turned black in Fig. 11. Compared with these methods, the realistic dehazing effect obtained in this paper is more complete, and the color of the image can be accurately restored.

D. ABLATION STUDIES

In order to verify the validity of each module proposed in this paper and the way it affects the model, after removing and replacing the corresponding modules, this paper uses the same training method as before to obtain the ablated model, which is analyzed quantitatively using evaluation metrics, and the results are shown in Table II and Fig. 13.

The results from the ablation show that each module plays an important role in the network performance. The model performance is severely limited by using the spatial information feature

Table II Comparisons on SOTS for different configurations, a–f total six different combinations

Scheme	a	b	c	d	e	f (ours)
High-level semantic feature stream	✓		✓	✓	✓	✓
Spatial information feature stream		✓	✓	✓	✓	✓
Parallel residual twicing module				✓	✓	✓
Spatial information auxiliary module with no attention mechanism					✓	
Spatial information auxiliary module						✓
SSIM	0.699	0.869	0.934	0.951	0.952	0.955
PSNR	19.917	22.121	25.158	25.699	25.453	27.348

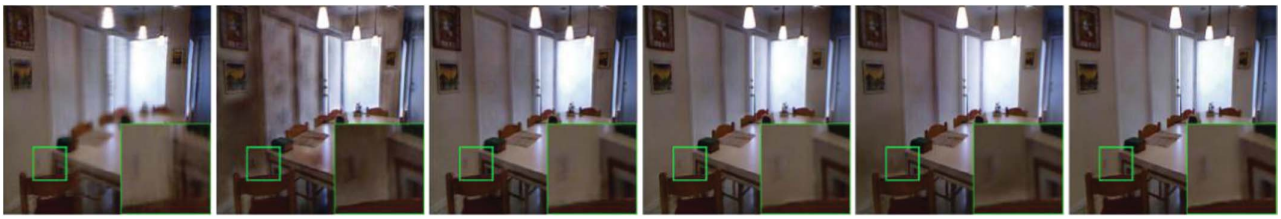

Fig. 13. The result of ablation studies.

stream (scheme a) or the high-level semantic feature stream (scheme b) alone. When using only the high-level semantic feature stream (scheme a) for the dehazing task, although the network output has a relatively obvious dehazing effect, the loss of advanced spatial features of the image is severe. In the case of using spatial information feature streams alone (scheme b), the feature extraction capability is limited due to the limited size of the full-resolution network, which leads to poor model dehazing performance. After the combination of the two feature streams (scheme c), the model performance is improved substantially for the first time, with SSIM and PSNR reaching 0.934 and 25.16 db, respectively.

The horizontal transfer of features in the model using parallel residual twicing modules (scheme d) allows the model to enhance the performance of the model without introducing any additional modules. Compared with the direct stitching of high-level semantic features at different stages (scheme c), the parallel residual twicing module enhances the ability of the model to represent the image structure in the dehazing task, and the SSIM metrics have a substantial improvement. Since the two features have different concerns, the direct fusion of high-level semantic features with enlarged resolution and spatial information features with the attention mechanism removed (scheme e) makes the expressions between the features interfere with each other, and the PSNR metrics show a decrease. By using the spatial information auxiliary module to build the model (scheme f), the attention mechanism is used to unify the focus between different features, which enables the semantic information to effectively assist the spatial information to recover the clear image, so that the model's ability to maintain color and details is improved and the PSNR index achieves a considerable improvement. The union of all modules finally enables the model in this paper to achieve a better dehazing effect.

V. CONCLUSION

In this paper, a novel two-stream convolutional neural network for image dehazing tasks was proposed. The network is built on a dual-stream network structure of high-level semantic feature streams and spatial information feature streams and included a two-stream feature fusion module and a parallel residual twicing module. The spatial information auxiliary module is designed based on the structural features of the two sets of feature streams, which constructs a unified representation of different dehazing information by the feature extraction module, so that the semantic information extracted from the dehazed image assists the spatial information to reconstruct the clear image step by step, allowing the semantic and spatial information to be fully learned before the regression output, reducing the complexity of the network. The parallel residual twicing module is designed based on the twicing technique to gradually extract effective reconstructed features in the process of feature decoding by learning the differences between the features before and after dehazing. The results of the ablation experiments show that the proposed module is effective for the dehazing problem. A large number of comparison tests also show that the proposed model in this paper performs well on practical image dehazing tasks.

In future research work, the general applicability of the proposed solution will be further explored and applied to tasks such as target detection, instance segmentation, and pedestrian re-identification in haze weather environments. The model structure will continue to be optimized to achieve further improvements in dehazing effect and computational speed.

ACKNOWLEDGMENTS

This work is jointly supported by the National Natural Science Foundation of China under Grant No. 61803061, 61906026; Innovation research group of universities in Chongqing; the Chongqing Natural Science Foundation under Grant cstc2020jcyj-msxmX0577, cstc2020jcyj-msxmX0634; "Chengdu-Chongqing Economic Circle" innovation funding of Chongqing Municipal Education Commission KJCXZD2020028; the Science and Technology Research Program of Chongqing Municipal Education Commission grants KJQN202000602; Ministry of Education China Mobile Research Fund (MCM 20180404); Special key project of Chongqing technology innovation and application development: cstc2019jscx-zdztzx0068.

REFERENCES

[1] W. U. Di and Q. S. Zhu, "The latest research progress of image dehazing," *Acta Autom. Sinica*, vol. 41, no. (2015-2-221), p. 221, 2015.

[2] A. Galdran, "Image dehazing by artificial multiple-exposure image fusion," *Signal Process.*, vol. 149, no. (AUG.), pp. 135–147, 2018.

[3] Z. Zhu, H. Wei, G. Hu, Y. Li, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," in IEEE Trans. Instrum. Meas., vol. 70, pp. 1–23, 2021, Art no. 5001523. doi: 10.1109/TIM.2020.3024335

[4] M. Zheng, G. Qi, Z. Zhu, Y. Li, H. Wei, and Y. Liu, "Image dehazing by an artificial image fusion method based on adaptive structure decomposition," *IEEE Sens. J.*, vol. 20, no. 14, pp. 8062–8072, 2020.

[5] W. Li, H. Wei, G. Qi, H. Ding, and K. Li, "A fast image dehazing algorithm for highway tunnel based on artificial multi-exposure image fusion," *IOP Conf. Series: Mater. Sci. Eng.*, vol. 741, no. 1, p. 012038, Jan 2020.

[6] Y. Yang, G. Chen, and J. Zhou, "Iterative optimization defogging algorithm using Gaussian weight decay," *Acta Autom. Sin.*, vol. 45, no. 4, pp. 819–828, 2019.

[7] J. He, C. Zhang, Y. Ran, and Z. Kai, "Convex optimization for fast image dehazing," in *2016 IEEE Int. Conf. Image Process. (ICIP)*, pp. 2246–2250, 2016. doi: 10.1109/ICIP.2016.7532758

[8] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, 2015.

[9] Z. Zhu, Y. Luo, H. Wei, Y. Li, and P. Li, "Atmospheric light estimation based remote sensing image dehazing," *Remote Sens.*, vol. 13, no. 13, p. 2432, 2021.

[10] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, 2015.

[11] E. J. McCartney and F. F. Hall, "Optics of the atmosphere: Scattering by molecules and particles," *Phys. Today*, vol. 30, no. 5, p. 76, 1977.

[12] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *2019 IEEE/CVF Conf. Comput. Vis. Pattern Recogn. (CVPR)*, pp. 8152–8160, 2019. doi: 10.1109/CVPR.2019.00835

[13] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, 2016.

[14] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2154–2164, 2020. doi: 10.1109/CVPR42600.2020.00223

[15] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 7313–7322, 2019. doi: 10.1109/ICCV.2019.00741

[16] J. Pan, J. Dong, Y. Liu, J. Zhang, J. Ren, J. Tang, Y.-W. Tai, and M.-H. Yang, "Physics-based generative adversarial models for image restoration and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2449–2462, 2021.

[17] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," pp. 2805–2814, 2020. doi: 10.1109/CVPR42600.2020.00288

[18] Y. Li, J. Meng, Z. Zhu, X. Huang, G. Qi, and Y. Luo, "Context convolution dehazing network with channel attention," in *2021 5th Asian Conf. Artif. Intell. Technol. (ACAIT)*, pp. 259–265, 2021. doi: 10.1109/ACAIT53529.2021.9731215

[19] X. Li, M. Ding, and A. Piurica, "Deep feature fusion via two-stream convolutional neural network for hyperspectral image classification." *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2615–2629, 2020.

[20] X. Liao, K. Li, X. Zhu, and K Liu Jr, "Robust detection of image operator chain with two-stream convolutional neural network," *IEEE J. Sel. Top. Signal Process.*, vol. PP, no 99, 1, 2020.

[21] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4780–4788, 2017. doi: 10.1109/ICCV.2017.511

[22] Z. Zhu, Y. Luo, G. Qi, J. Meng, and N. Mazur, "Remote sensing image defogging networks based on dual self-attention boost residual octave convolution," *Remote Sens.*, vol. 13, no. 16, p. 3104, 2021.

[23] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE Inter. Conf. Comp. Vision (ICCV)*, pp. 2242–2251, 2017. doi: 10.1109/ICCV.2017.244

[24] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. Lecture Notes in Computer Science, vol 11211. Springer, Cham. pp. 3–19, 2018. doi: 10.1007/978-3-030-01234-2_1

[25] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation," In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham., 2015. doi: 10.1007/978-3-319-24574-4_28

[26] M. R. Charest, M. Elad, and P. Milanfar, "A general iterative regularization framework for image denoising," in *40th Annual Conf. Inf. Sci. Syst.*, pp. 452–457, 2006. doi: 10.1109/CISS.2006.286510

[27] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, 2019.

[28] Y. Zhang, D. Li, and G. Sharma, "Hazerd: An outdoor scene dataset and benchmark for single image dehazing," in *2017 IEEE Int. Conf. Image Process. (ICIP)*, pp. 3205–3209, 2017. doi: 10.1109/ICIP.2017.8296874

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[30] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, "Refinednet: A weakly supervised refinement framework for single image dehazing," *IEEE Trans. Image Process.*, vol. 30, pp. 3391–3404, 2021.