J. Meas. Sci. Instrum., 2026, 17(2): 232-242. http://jmsi.publish.founderss.cn
DOI: 10.62756/jmsi.1674-8042.2026020 jmsi@nuc.edu.cn

Parallel feature fusion framework for image dehazing driven by haze-detail collaboration

YU Xuan, MA Jiaqing, YU Leiyi*

School of Computer Science and Engineering, Chongqing College of Humanities Science & Technology, Chongqing 401524, China

***Corresponding author:** YU Leiyi (yuleiyi2008@163.com)

Received: December 16, 2025 **Revised:** January 15, 2026 **Accepted:** January 23, 2026

Abstract: Image dehazing remains a highly ill-posed problem in low-level computer vision, primarily due to the complex coupling of spatially variant haze and high-frequency background details. Existing deep learning-based methods often struggle to balance the trade-off between aggressive haze removal and the preservation of fine textures, frequently resulting in color distortion, halos, or detail loss. To address these limitations and achieve high-fidelity restoration, this paper proposes a parallel feature fusion framework driven by a haze-detail collaboration mechanism. Specifically, the proposed framework adopts a dual-branch parallel architecture to disentangle the dehazing process. The upper branch functions as a haze layer extraction network. It employs a Res2Net-based encoder to capture multi-scale semantic features and integrates a novel deformable convolution-residual hybrid attention module. By dynamically adjusting the receptive fields, this module precisely characterizes non-uniform haze distributions and models long-range dependencies. Simultaneously, the lower branch serves as a detail compensation network, leveraging context detail information blocks with multi-scale dilated convolutions to aggregate contextual cues and reinforce the representation of high-frequency textural details. Subsequently, a fusion network performs adaptive feature integration, effectively merging the extracted haze features with the enhanced detail information to reconstruct the haze-free image. To ensure robust training, a dual-supervision mechanism is introduced, combining a feature regularization loss to align feature distributions in the latent space and a reconstruction loss to constrain pixel-level content fidelity. Extensive quantitative and qualitative experiments are conducted on both synthetic benchmarks and real-world datasets. The results demonstrated that the proposed algorithm delivered superior performance, achieving higher peak signal-to-noise ratio and structural similarity scores compared to state-of-the-art methods. Visual comparisons further confirmed that our method effectively removed dense haze while recovering vivid colors and sharp structural details without introducing artifacts.

Key words: image dehazing; Res2Net; attention mechanism; deformable convolution; texture preservation; feature supervision

0 Introduction

Images captured under hazy weather conditions suffer from severe quality degradation, characterized by a significant loss of structural information and textural details. This visual impairment adversely affects the performance of downstream high-level vision tasks, such as satellite remote sensing, object detection, and environmental monitoring. Consequently, effective image dehazing to restore visibility is of critical importance.

Existing image dehazing algorithms are primarily categorized into two paradigms: traditional prior-based methods and deep learning-based approaches[1]. Traditional methods typically rely on the physical atmospheric scattering model, leveraging hand-crafted priors to estimate atmospheric light and transmission maps for haze removal. He et al.[2] introduced the dark channel prior (DCP), which estimates the transmission map by employing the dark channel of the image in conjunction with the atmospheric scattering model, thereby enabling image dehazing. Ling et al.[3] proposed the saturation line prior, which exploits pixel-wise correlation to construct saturation lines for transmission estimation and subsequent image reconstruction. Similarly, to address the inaccurate estimation of atmospheric light and transmission caused by white objects or bright regions in the dark channel prior, Li et al.[4] proposed a superpixel-based dehazing algorithm, by utilizing simple linear iterative clustering (SLIC) to estimate local atmospheric light and obtaining the final transmission map. Although these prior-based methods yield satisfactory results in specific scenes, they often suffer from performance degradation when their

© The Author(s) 2026. Published by Publishing Center of North University of China. The articles published in this open access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

underlying assumptions are violated. In complex real-world scenarios, this limitation frequently leads to artifacts such as color distortion and incomplete haze removal.

With the rapid advancement of deep learning, methods based on deep neural networks have emerged as a prominent research direction in the field of image dehazing. Cai et al.[5] proposed DehazeNet, a dehazing network that learns the global transmission map through multi-scale feature extraction and restores haze-free images based on the atmospheric scattering model. Chen et al.[6] proposed GCA-Net, a gated contextual aggregation network that employs smooth dilated convolutions to suppress gridding artifacts and introduces a gating subnetwork to fuse multi-level features, thereby enhancing perceptual quality. Qin et al.[7] proposed FFA-Net, an end-to-end network that incorporates both channel and pixel attention mechanisms to construct a feature attention module, enabling the model to adaptively attend to informative regions of the image. Luo et al.[8] designed a novel large-kernel convolution dehazing block, which comprises deep large-kernel convolution components and a channel-enhanced feed-forward network. In this design, the deep large-kernel convolution is decomposed into shallow-depth convolution and dilated convolution, while channel attention is introduced into the feed-forward network to identify important channels and enhance dehazing robustness. Chen et al.[9] proposed a detail-enhanced attention block composed of a detail enhancement convolution and content-guided attention to strengthen feature learning and improve dehazing performance.

Although existing deep learning-based methods have achieved remarkable performance, issues such as over-removal or residual haze persist when processing images with uneven haze distribution. This primarily arises from the inadequate characterization of haze features during training. To address these challenges, we propose a parallel feature fusion framework driven by haze-detail collaboration. Specifically, the proposed framework integrates a haze layer extraction network, a detail compensation network, and a fusion network, and it is supervised by a feature regularization loss. This effectively alleviates the issue of uneven haze removal in images with non-uniformly distributed haze. The main contributions are summarized as follows.

1) To effectively resolve the trade-off between haze-layer feature removal and detail restoration, we propose a parallel feature fusion framework driven by haze–detail collaboration.

2) In the haze-layer extraction network, we leverage Res2Net to more effectively capture image features at different scales, and propose a deformable convolution–residual hybrid attention block to enhance the network's sensitivity to the distribution of haze features during training.

3) In the detail compensation network, we design a detail contextual enhancement module that employs dilated convolutions with multi-scale dilation rates to aggregate contextual information, thereby expanding the receptive field while enhancing the network's ability to preserve fine-grained structural details.

4) In the fusion network, we introduce an attention mechanism to generate weight masks to disentangle the haze layers from background structures. Additionally, we design a haze–background fusion module to facilitate haze removal while further enhancing texture details.

5) We formulate a composite loss function by integrating feature regularization loss, mean squared error loss, and perceptual loss. By feeding paired hazy and clean images into the network and extracting their feature maps, the feature regularization loss imposes constraints on the training process. This not only enhances feature alignment and consistency but also improves the model's robustness and generalization capability.

1 Theoretical background

In the field of computer vision, atmospheric scattering models are widely adopted to model the haze formation process within imaging systems[10]. The mathematical expression can be formulated as

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

where $\boldsymbol{I}(x)$ denotes the haze image, $\boldsymbol{J}(x)$ denotes the scene radiance (haze-free image), $\boldsymbol{t}(x)$ is the transmission map, and $\boldsymbol{A}$ refers to the global atmospheric light.

In practice, image dehazing is commonly achieved by leveraging priors to estimate $\boldsymbol{t}(x)$ and $\boldsymbol{A}$, thereby recovering the haze-free image. To simplify the dehazing model, we reformulate the physical model into an additive form[11]. By defining the background image as $\boldsymbol{B}(x)=\boldsymbol{J}(x)$, Eq. (1) can be rewritten as

$$\boldsymbol{I}(x)=\boldsymbol{B}(x)+(\boldsymbol{A}-\boldsymbol{B}(x))(1-\boldsymbol{t}(x)), \tag{2}$$

where the second term represents the haze-induced degradation. If we denote the haze-layer features as $\boldsymbol{D}(x)=(\boldsymbol{A}-\boldsymbol{B}(x))(1-\boldsymbol{t}(x))$, the haze image can be simplified as

$$\boldsymbol{I}(x)=\boldsymbol{B}(x)+\boldsymbol{D}(x), \tag{3}$$

where $\boldsymbol{I}(x)$ denotes the haze image, $\boldsymbol{B}(x)$ denotes the

background image, and $D(x)$ denotes the haze-layer features. However, as indicated by the derivation $D(x)=(A-B(x))(1-t(x))$, the haze component $D(x)$ is intrinsically dependent on the background information $B(x)$. Consequently, a simple additive assumption oversimplifies the complex coupling between $B(x)$ and $D(x)$ caused by the transmission medium[12]. As a result, Eq. (3) cannot accurately characterize the dehazing process in complex scenes. To address this, we reformulate the model to explicitly account for their correlation.

$$I(x)=B(x)+D(x)+F(B(x),D(x)), \quad (4)$$

where $F(B(x),D(x))$ denotes the coupling relationship between the background image and the haze-layer features, capturing latent correlations and redundancy shared between $B(x)$ and $D(x)$.

2 Proposed algorithm

Fig.1 illustrates the overall architecture of the proposed haze-detail synergistic parallel feature fusion framework. This framework comprises four main components: the haze layer extraction network, the detail compensation network, the fusion network, and the feature supervision mechanism. Specifically, the haze layer extraction network is structured as an encoder-decoder architecture. The encoder takes hazy images as input and employs stacked Res2Net blocks to enhance the representational capability of haze layer features across different spatial scales. Additionally, we introduce a deformable convolution-residual hybrid attention module to better adapt to diverse haze distribution characteristics, thereby increasing the network's sensitivity to haze-related features.

To further restore structural fidelity, we design a detail compensation network that employs a dilated convolution architecture with multi-scale dilation rates to aggregate contextual cues. This design enhances the network's ability to capture fine texture details while expanding the receptive field. Additionally, we design a fusion network to achieve effective feature integration between the haze layer and background, thereby obtaining richer and more precise feature representations.

Complementing the architectural design, this study employs a feature supervision mechanism to guide network optimization. As shown in Fig.1, the supervision includes two parts. 1) By computing feature regularization loss (L_F) between paired hazy and clear images, the network learns the intrinsic differences between them and aligns feature representations adaptively; 2) Simultaneously, a reconstruction loss (L_R) is imposed between the restored output and the reference image to constrain the content fidelity. This dual-supervision mechanism enhances feature compactness and consistency, enabling more precise characterization of haze-related components.

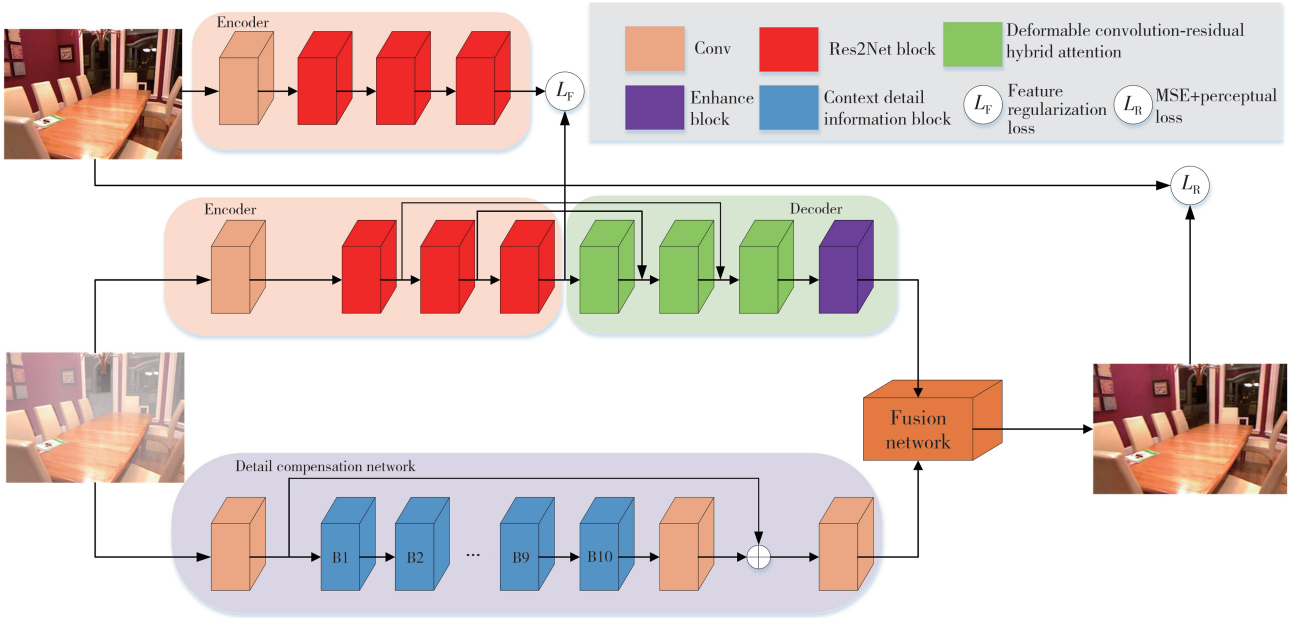


Fig. 1 Network structure

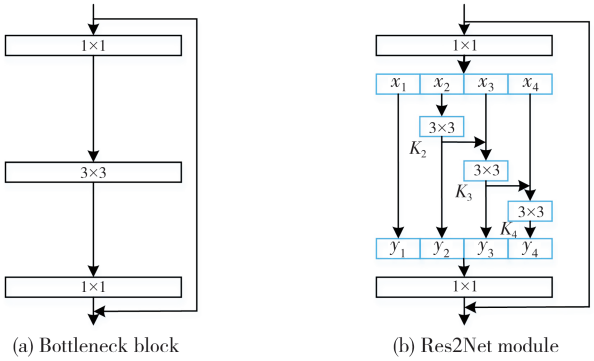
2.1 Haze-layer feature extraction network

The haze layer feature extraction network is illustrated in Fig.1. Specifically, we initialize the encoder component using ImageNet pre-trained weights. This initialization strategy incorporates rich semantic prior information, significantly enhancing the network's feature representation capabilities and accelerating training convergence. The decoder integrates a deformable convolution-residual hybrid attention module, enabling the network to

selectively emphasize meaningful channel responses and spatial structures associated with haze. Additionally, an enhancement module is appended after the DC-RHA module to further enrich the representational capacity of the extracted features.

2.1.1 Encoder module

Res2Net enhances the performance of the ResNet architecture by introducing a novel hierarchical residual-like structure designed to handle multi-scale feature maps effectively[13]. Unlike the standard ResNet block, where convolutional layers are restricted to a fixed receptive field size, Res2Net addresses this limitation by partitioning input feature maps into multiple subsets. These subsets are processed through a cascade of convolutions, effectively decomposing the standard 3×3 convolution into a series of multi-scale representations. By introducing a scale dimension parameter, denoted as "s", to control the number of channel groups, Res2Net enables a single convolutional layer to represent varied receptive fields. This mechanism significantly improves the network's ability to capture fine-grained multi-scale features. Fig.2(a) illustrates the basic bottleneck block of ResNet. Fig.2(b) shows the basic module of Res2Net. Specifically, Res2Net first adopts a 1×1 convolution to split the input feature map into s subsets, denoted as $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_s\}$.



(a) Bottleneck block (b) Res2Net module

Fig. 2 Comparison between bottleneck block and Res2Net module

The first subset $\boldsymbol{x}_1$ serves as an identity mapping to produce $\boldsymbol{y}_1$, while the second subset is processed by a 3×3 convolution to produce the output $\boldsymbol{y}_2$. For the remaining subsets, each current subset is added to the output of the previous subset and then fed into a 3×3 convolution. Let the output of the i-th subset be denoted as $\boldsymbol{y}_i$. This process can be formulated as

$$\boldsymbol{y}_i = \begin{cases} \boldsymbol{x}_i, & i = 1; \\ \boldsymbol{K}_i(\boldsymbol{x}_i), & i = 2; \\ \boldsymbol{K}_i(\boldsymbol{x}_i + \boldsymbol{y}_{i-1}), & 3 \leqslant i \leqslant s. \end{cases} \tag{5}$$

2.1.2 Decoder module

To effectively characterize regional haze features, we design the DC-RHA module, as shown in Fig.3.

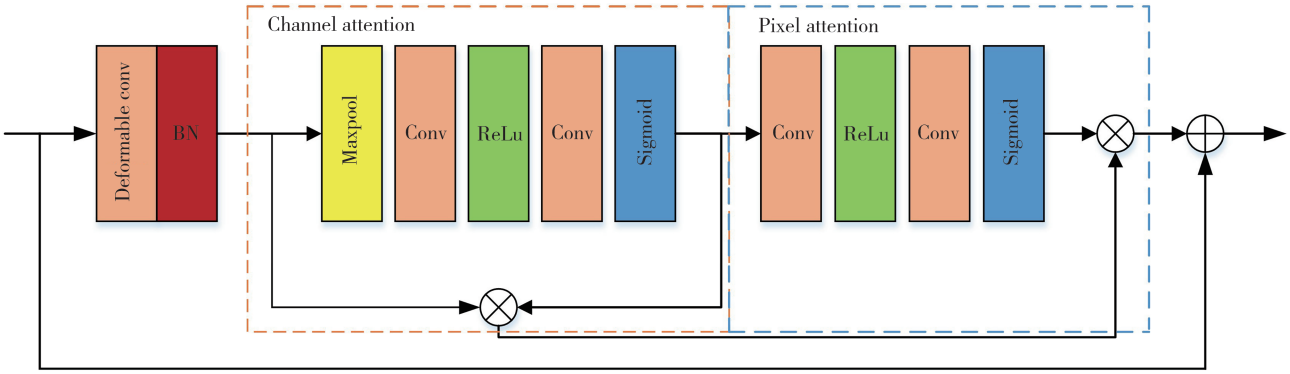


Fig. 3 DC-RHA structure diagram

This module consists of a deformable convolution[14] and a hybrid attention mechanism. Specifically, the deformable convolution dynamically adjusts the spatial sampling positions of convolutional kernels to accommodate the spatially non-uniform distribution of haze. Subsequently, the module employs a hybrid attention mechanism to enhance the focus on informative features. The channel attention[15] adaptively reweights different feature channels to enhance the representation of crucial channels, while the pixel attention[16] assigns weights to each spatial position, further refining the feature representation. By sequentially combining channel and pixel attention, the proposed module enhances feature representation along both channel and spatial dimensions, thereby bolstering the model's robustness and generalization capabilities. We define the mathematical formulation of this process as

$$F_{\text{ca}} = \sigma(\text{Conv}(\delta(\text{Conv}(\text{Max}(\text{DC}(F_x)))))), \tag{6}$$

$$F_{\text{pa}} = \sigma(\text{Conv}(\delta(\text{Conv}(F_{\text{ca}})))), \tag{7}$$

$$F_{\text{DC-RHA}} = ((F_{\text{ca}} \otimes \text{DC}(F_x)) \otimes F_{\text{pa}}) \oplus F_x, \tag{8}$$

where F_x denotes the feature map input at the x-th layer; and DC and Max denote the deformable convolution layer and the max-pooling layer, respectively; Conv denotes a 3×3 convolution operation; δ and σ denote the ReLU and Sigmoid activation functions, respectively; F_{ca} and F_{pa} denote the feature responses after channel attention and

pixel attention, while F_{DC-RHA} denotes the final output feature map of the deformable convolution-residual hybrid attention block. Notably, during the channel attention stage, the scalar weights generated by the Sigmoid function are multiplied by the feature map output from the batch normalization (BN) layer. This operation effectively recalibrates the normalized features, enabling the network to adaptively focus on important channels.

2.1.3 Feature representation enhancement module

To bolster the extraction of haze-layer features and expand the model's receptive field, we design a feature enhancement module, as illustrated in Fig.4.

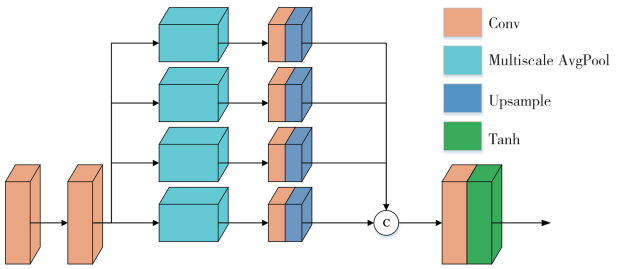


Fig. 4 Enhancement module

This module captures features across multiple spatial scales through a multi-scale pooling strategy. First, feature maps undergo preliminary feature extraction via a convolutional layer. Subsequently, we apply average pooling operations with kernel sizes of 32×32, 16×16, 8×8, and 4×4 to generate multi-scale feature representations, enriching the semantic diversity of the extracted features. The pooled feature maps are then upsampled to restore their spatial resolution, enabling seamless concatenation with the original features. Finally, we apply a convolutional layer with a Tanh activation function to further enhance nonlinear representation capabilities. The specific enhancement process is

$$F_{en}\begin{cases} F_{en1}=\text{Upsample}(\text{Conv}(\text{Avg}_{32\times32}(F_i))),\\ F_{en2}=\text{Upsample}(\text{Conv}(\text{Avg}_{16\times16}(F_i))),\\ F_{en3}=\text{Upsample}(\text{Conv}(\text{Avg}_{8\times8}(F_i))),\\ F_{en4}=\text{Upsample}(\text{Conv}(\text{Avg}_{4\times4}(F_i))),\end{cases} \tag{9}$$

$$F_{en}=\alpha(\text{Conv}(\text{cat}(F_{en1},F_{en2},F_{en3},F_{en4}))), \tag{10}$$

where F_i denotes the output feature map of the i-th layer; Avg and Upsample denote the average pooling layer and the upsampling layer, respectively; the operator cat denotes the feature concatenation operation; α indicates the Tanh activation function; F_{en} denotes the final output feature map of the enhancement module.

2.2 Detail compensation network

To compensate for the structural details compromised during the haze removal process, we design a dedicated detail compensation branch. By leveraging a context detail information block (CDIB), this branch effectively expands the receptive field and aggregates multi-scale contextual cues, thereby facilitating the precise restoration of fine-grained structures.

Fig.5 illustrates the structure of the CDIB, which is built upon a contextual dilated convolutional layer (CDCL) with residual connections. In the CDCL, three dilated convolutions with dilation rates of 1, 3, and 5 are deployed to capture detailed features at different spatial scales. 1×1 convolution is applied to fuse the multi-scale features and restore the channel dimensionality. The mathematical formulation of the CDCL is expressed as

$$\begin{aligned}\text{CDCL}=\text{Conv}_{1\times1}(\text{cat}[&\text{Conv}_{3\times3,p1}(\boldsymbol{x}),\text{Conv}_{3\times3,p2}(\boldsymbol{x}),\\&\text{Conv}_{3\times3,p3}(\boldsymbol{x})]),\end{aligned} \tag{11}$$

where $\boldsymbol{x}$ denotes the input feature map, $\text{Conv}_{x\times x,pi}$ denotes a dilated convolution with kernel size k and dilation rate pi. Formally, we define the CDIB as

$$\text{CDIB}=\varepsilon(\text{CDCL}(\varepsilon,\delta(\text{CDCL}(\boldsymbol{x}))))\oplus\boldsymbol{x}, \tag{12}$$

where $\boldsymbol{x}$ denotes the input feature map, ε and δ denote the batch normalization and ReLU activation operations, respectively.

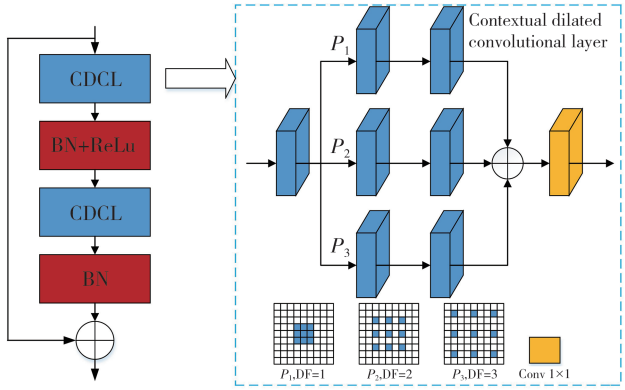


Fig. 5 Context detail information block

2.3 Haze-background feature fusion module

To facilitate effective feature separation and interaction, we design a feature fusion network to integrate the information from the dual branches, as illustrated in Fig.6.Let $\boldsymbol{f}_h$ denote the haze-layer features extracted by the haze layer extraction branch (upper branch), and $\boldsymbol{f}_b$ denote the detail features generated by the detail compensation branch (lower branch). Specifically, the low-level feature $\boldsymbol{f}_h$ is processed by a global channel attention mechanism to learn a weighting matrix $\boldsymbol{Q}_h$, enabling the network to selectively emphasize haze-related characteristics. In contrast, the high-level feature $\boldsymbol{f}_b$ is fed into a local pixel attention mechanism to learn a weighting matrix $\boldsymbol{Q}_b$, allowing the

network to effectively attend to background information. Both $\boldsymbol{Q}_{\mathrm{h}}$ and $\boldsymbol{Q}_{\mathrm{b}}$ are subsequently normalized through a Softmax function to generate complementary weight masks, ensuring that pixels at different spatial locations are adaptively assigned to different layers (background or haze layer). We formulate the fusion process as

$$F_{\mathrm{h}}=\boldsymbol{Q}_{\mathrm{h}}\otimes\boldsymbol{f}_{\mathrm{h}}+\mathrm{Softmax}(\boldsymbol{Q}_{\mathrm{b}})\otimes\boldsymbol{f}_{\mathrm{b}}, \tag{13}$$

$$F_{\mathrm{b}}=\boldsymbol{Q}_{\mathrm{b}}\otimes\boldsymbol{f}_{\mathrm{b}}+\mathrm{Softmax}(\boldsymbol{Q}_{\mathrm{h}})\otimes\boldsymbol{f}_{\mathrm{h}}, \tag{14}$$

where $\boldsymbol{Q}_{\mathrm{h}}\otimes\boldsymbol{f}_{\mathrm{h}}$ denotes the haze components contained in the haze-layer feature map, $\mathrm{Softmax}(\boldsymbol{Q}_{\mathrm{b}})\otimes\boldsymbol{f}_{\mathrm{b}}$ represents the residual haze components that remain in the background feature map, $\boldsymbol{Q}_{\mathrm{b}}\otimes\boldsymbol{f}_{\mathrm{b}}$ denotes the detail information contained in the detail information that appears within the haze-layer feature map.

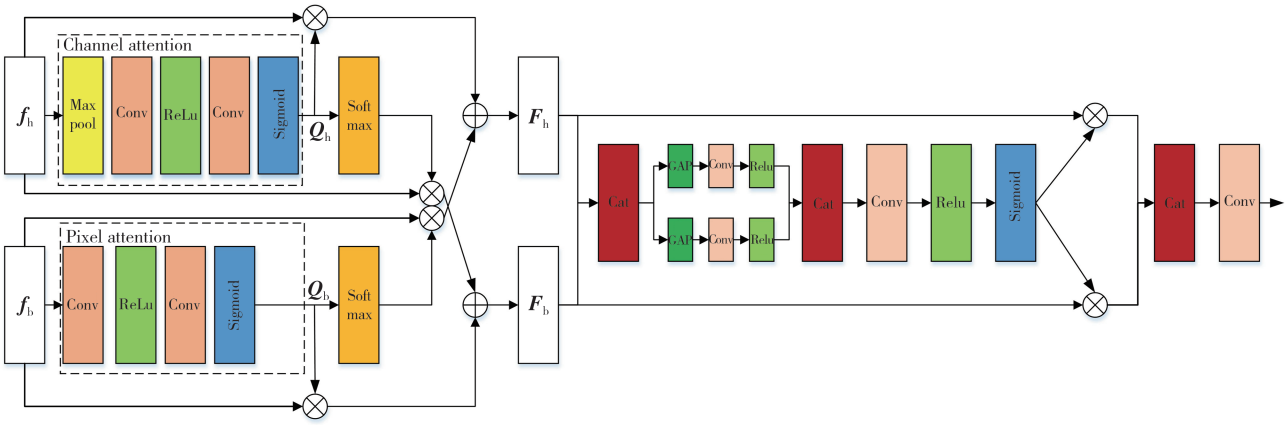


Fig. 6 Fusion network

Building upon this, we design a dual-stage fusion module. In the first stage, global average pooling (GAP) and global max pooling (GMP) are applied to extract multi-scale detail information. The resulting features are concatenated and passed through a convolution layer followed by a Sigmoid activation function to compute the corresponding fusion weights. In the second stage, the learned weights are applied to F_{h} and F_{b}, and the two are subsequently combined through weighted fusion to produce the final output. We define the detailed fusion process as

$$\begin{aligned}F_1=&\mathrm{cat}(\delta(\mathrm{Conv}(\mathrm{GAP}(\mathrm{cat}(F_{\mathrm{h}},F_{b})))),\\&\delta(\mathrm{Conv}(\mathrm{GMP}(\mathrm{cat}(F_{\mathrm{h}},F_{\mathrm{b}}))))),\end{aligned} \tag{15}$$

$$\begin{aligned}F_2=&\mathrm{Conv}(\mathrm{cat}(\sigma(\delta(\mathrm{Conv}(F_1)))\otimes F_{\mathrm{h}},\\&\sigma(\delta(\mathrm{Conv}(F_1)))\otimes F_{\mathrm{b}})),\end{aligned} \tag{16}$$

where F_1 denotes the output of the first stage, F_2 denotes the output of the second stage, cat represents the concatenation operation, and Conv denotes the convolution operation. In addition, GAP and GMP represent the global average pooling and global max pooling operations, respectively.

2.4 Loss function

During network training, the mean squared error (MSE) loss, perceptual loss, and feature regularization loss are employed to optimize the model parameters. The MSE loss quantifies the pixel-wise discrepancy between the dehazed image and the corresponding clear image, and is defined as

$$L_2=\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{J}_i-J_i\right\|^2, \tag{17}$$

where $\tilde{J}_i$ denotes the dehazed image, J_i denotes the corresponding clear image, and N represents the number of restored image samples.

The perceptual loss is adopted to enhance the visual quality of the reconstructed image, ensuring that the dehazed image is more similar to the clear image at a semantic level. Based on the VGG16 feature extraction network pretrained on ImageNet[17], the perceptual loss is defined as

$$L_{\mathrm{p}}=\sum_{j=1}^{3}\frac{1}{C_jH_jW_j}\left\|\phi_j(\tilde{J}_i)-\phi_j(J_i)\right\|_2^2, \tag{18}$$

where $\tilde{J}_i$ denotes the dehazed image, J_i denotes the clear image, and ϕ_j represents the j-th layer of the VGG16 network. In addition, C_j, H_j, and W_j denote the number of channels, height, and width of the feature map at layer j.

The feature regularization loss is introduced to ensure that the encoder captures robust semantic and structural information, impervious to haze interference. As illustrated in Fig.1, during training, both the hazy image and the clear reference image are fed into the shared Res2Net encoder. We enforce consistency between their deep feature representations to compel the network to learn haze-invariant features. Specifically, the feature maps are extracted from the 4th stages of the Res2Net encoder, as these deep layers contain rich semantic context crucial for disentangling background structures from haze components. The regularization loss is formulated using the $L1$ norm.

$$L_{\mathrm{F}}=\frac{1}{N}\sum_{i=1}^{N}(\left\|E_k(J_i)-E_k(J_h)\right\|_1), \tag{19}$$

where $E_k(J_i)$ and $E_k(J_h)$ represent the feature maps extracted from the clean and hazy images at the k-th stage, and N denotes the number of restored image samples. By minimizing L_F, the encoder is constrained to align the latent representations of hazy and clean images, thereby preserving structural fidelity. Accordingly, the composite loss function is expressed as

$$L = L_2 + \lambda_1 L_p + \lambda_2 L_F, \tag{20}$$

where L denotes the composite loss function, and λ_1 and λ_2 are weighting factors, which are set to 0.1 and 0.05.

3 Experimental results and analysis

3.1 Experimental setup

In this study, both synthetic and real-world datasets were utilized to validate the efficacy of the proposed network. Regarding synthetic data, we employed the indoor training set (ITS) and the synthetic objective testing set (SOTS) from the RESIDE benchmark dataset[18].

To further assess generalization capabilities under realistic haze conditions, the O-HAZE[19] and NH-HAZE[20] datasets were leveraged. The proposed method was implemented using the PyTorch framework, and all experiments were conducted on a workstation powered by an NVIDIA GeForce RTX 3060 GPU. To ensure a rigorous evaluation, the datasets were randomly partitioned into training and validation sets at a ratio of 8∶2. The network was trained for 100 epochs using the Adam optimizer with a batch size of 8. The initial learning rate was set to 1×10^{-4} and was annealed by a factor of 0.1 upon validation loss saturation.

3.2 Comparative experiments

To empirically validate the effectiveness of the proposed algorithm, we compared it against state-of-the-art methods, including DCP[1], GCA-Net[6], FFA-Net[7], LKD-Net[8], and DEA-Net[9], across different datasets. The evaluation was conducted from both qualitative and quantitative perspectives.

Fig.7 presents the visual results on the RESIDE dataset. As observed, while the DCP algorithm effectively removes haze, it introduces pronounced halo artifacts. Both GCA-Net and FFA-Net exhibit limited dehazing capabilities, leaving residual haze in various regions. LKD-Net tends to generate oversaturation sky regions, whereas DEA-Net is plagued by local distortions and color shifts. In contrast, the proposed method achieves superior dehazing completeness while preserving natural color fidelity and optimal brightness levels.



Fig. 7 Comparison of different algorithms on RESIDE dataset

Fig.8 illustrates the comparative results on the O-HAZE dataset. Although DCP manages to remove haze, the restored images suffers from a significant loss of luminance, appearing noticeably darker. GCA-Net, FFA-Net, and LKD-Net all exhibit varying degrees of residual haze. DEA-Net produces uneven dehazing effects and introduces significant distortions. Conversely, our method achieves remarkable performance, striking a balance between color fidelity and brightness preservation.

Qualitative results on the NH-HAZE dataset are depicted in Fig. 9. Images restored by DCP suffer from local oversaturation and overall darkening. While GCA-Net removes haze, it introduces noticeable color distortions. FFA-Net, LKD-Net, and DEA-Net fail to effectively penetrate dense haze regions, resulting in images with insufficient brightness. By comparison, the proposed method not only eliminates dense haze more effectively but also yields restored images with vivid details and natural color reproduction.

Fig.10 demonstrates the generalization capability of the algorithms on real-world hazy images. DCP, while effective in haze removal, introduces halo artifacts around bright objects and sky regions. GCA-Net and

FFA-Net struggle with incomplete haze removal in distant sky areas. LKD-Net and DEA-Net produce noticeable distortions in sky regions. In contrast, the proposed method effectively removes haze while preserving rich details and sharp edge structures. Moreover, for scenes containing large sky regions, our method successfully corrects color deviations while maintaining visually pleasing brightness levels.



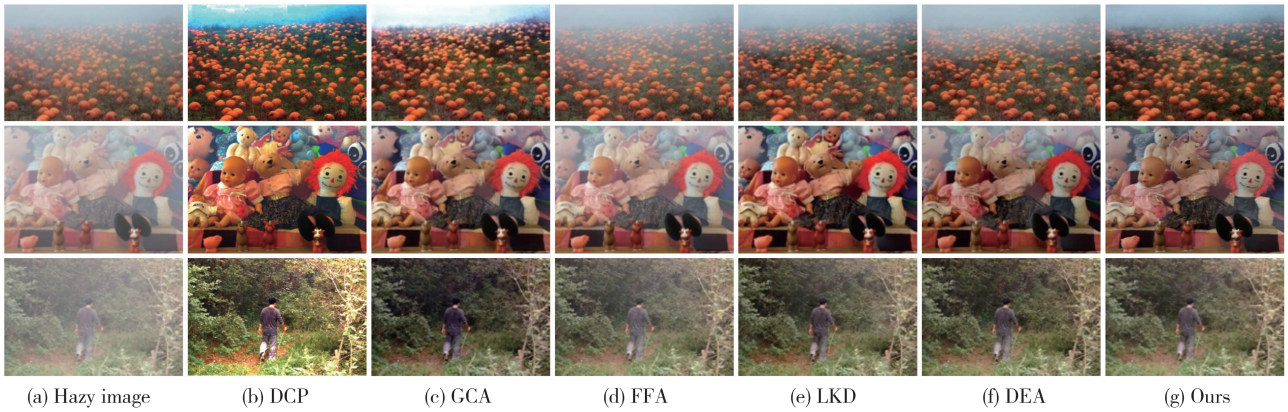
(a) Hazy image (b) DCP (c) GCA (d) FFA (e) LKD (f) DEA (g) Ours (h) GT

Fig. 8 Comparison of different algorithms on O-HAZE dataset



(a) Hazy image (b) DCP (c) GCA (d) FFA (e) LKD (f) DEA (g) Ours (h) GT

Fig. 9 Comparison of different algorithms on NH-HAZE dataset



(a) Hazy image (b) DCP (c) GCA (d) FFA (e) LKD (f) DEA (g) Ours

Fig. 10 Comparison of different algorithms in real images

To quantitatively assess the performance of the proposed method against competing algorithms, we employed two widely adopted full-reference metrics: peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). Specifically, PSNR quantifies pixel-level reconstruction fidelity, whereas SSIM assesses the preservation of structural integrity. The quantitative comparisons are summarized in Table 1. As evidenced by the tabulated data, the proposed method consistently outperforms state-of-the-art approaches, achieving the highest PSNR and SSIM scores across the tested datasets. These quantitative indicators, corroborating the visual comparisons, validated the robustness and superior efficacy of our proposed algorithm.

Table 1 Quantitative comparison on synthetic and real-world datasets

Method	RESIDE		O-HAZE		NH-HAZE	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
DCP	15.467	0.764	15.787	0.639	12.298	0.574
GCA-Net	21.172	0.896	16.281	0.639	18.762	0.752
FFA-Net	28.301	0.937	24.078	0.813	19.396	0.761
LKD-Net	26.762	0.897	21.117	0.704	18.736	0.692
DEA-Net	28.301	0.906	22.465	0.778	18.169	0.747
Ours	30.099	0.953	25.709	0.890	22.242	0.766

3.3 Ablation study

To validate the efficacy of the proposed architectural components, we conducted comprehensive ablation studies focusing on three key aspects: the parallel feature fusion framework, the composite loss function, and the internal mechanisms of the DC-RHA block. The detailed procedures are described as follows.

1) To quantify the contribution of each component, we performed a progressive ablation study. Specifically, we defined four core components. A denotes the Res2Net encoder for haze-layer extraction; B represents the DC-RHA decoder for detail compensation; C corresponds to the detail sub-network (DSN); and D signifies the proposed parallel feature fusion module. We constructed baseline variants by systematically integrating these components to analyze their individual impact on the overall performance. The quantitative results of this analysis are summarized in Table 2. The progressive incorporation of the Res2Net, DC-RHA, DSN, and the fusion module yields consistent gains in both PSNR and SSIM metrics, thereby validating the individual contribution of each component to the overall restoration performance. The corresponding qualitative comparisons across different ablation settings are presented in Fig. 11, where Fig. 11(a) depicts the hazy input and Fig. 11(b)—11(g) illustrate the results corresponding to 1—6. Relative to the baseline, the inclusion of the Res2Net module enhances feature extraction but fails to address non-uniform haze distributions effectively. While the DC-RHA module suppresses haze, it tends to introduce chromatic distortions. Although combining Res2Net with DC-RHA improves dehazing capacity, the restored images still exhibit textural blurring and grid-like artifacts. The parallel module further boosts performance but struggles to fully recover high-frequency structural details. In contrast, the proposed parallel feature fusion network successfully eliminates haze, yielding high-quality images characterized by natural color fidelity and sharp structural details.

Table 2 Haze layer detail dual branch network ablation experiment

Index	Model	PSNR/dB	SSIM
1	Baseline	19.197	0.651
2	Baseline+A	20.245	0.685
3	Baseline+B	20.569	0.671
4	Baseline+A+B	22.054	0.748
5	Baseline+A+B+C	24.371	0.826
6	Baseline+A+B+C+D	25.709	0.890



(a) Hazy image (b) Model 1 (c) Model 2 (d) Model 3 (e) Model 4 (f) Model 5 (g) Model 6

Fig. 11 Parallel feature fusion network ablation experiment dehazing results

2) To evaluate the contribution of the proposed composite loss function, we incrementally integrated the mean squared error loss, perceptual loss, and feature regularization loss. The quantitative results are detailed in Table 3.

Table 3 Loss function ablation experiment

Loss	PSNR/dB	SSIM
L_2	20.381	0.719
L_2+L_p	22.950	0.820
$L_2+L_p+L_{FR}$	25.709	0.890

It is evident that supervising the network with the full composite loss yields the most robust dehazing performance. The corresponding visual comparisons are presented in Fig. 12.



(a) Hazy image (b) L_2 (c) L_2+L_p (d) $L_2+L_p+L_{FR}$

Fig. 12 Loss function ablation experiment dehazing results

When the network was trained solely using the L_2 loss, noticeable residual haze remains in the restored images. Although incorporating the perceptual loss improves haze removal, it tends to introduce chromatic artifacts, as evidenced by the unnatural color shifts in the white pillar region in Fig. 12(c). By further integrating the feature regularization loss, the proposed method achieves superior dehazing completeness, effectively restoring natural color fidelity and suppressing artifacts.

3) To investigate the specific contribution of the proposed DC-RHA block to haze feature extraction, we disaggregated the module and evaluated the deformable convolution (DC) and the hybrid attention mechanism independently. The quantitative results are listed in Table 4, demonstrating that the synergistic design achieves superior performance compared to employing either component in isolation.

Fig.13 visualizes the haze-layer feature maps extracted by different configurations, alongside their corresponding restored images. When employing only the channel attention or pixel attention module, the captured haze features lacked spatial continuity. Although the hybrid attention module improves the network's ability to attend to haze characteristics, the extracted features remain spatially inconsistent, leading to residual haze artifacts in the restored output (the left car region in Fig.13). In contrast, the full DC-RHA module generates spatially coherent haze feature maps, facilitating thorough haze removal and preserving sharper edge details in the reconstructed images.

Table 4 Experimental study on deformable convolution residual hybrid attention block ablation

Index	Model	PSNR/dB	SSIM
1	Baseline	24.591	0.902
2	Channel attention	26.938	0.915
3	Pixel attention	27.205	0.923
4	Hybrid attention	28.313	0.927
5	DC+Hybrid attention	30.099	0.953



(a) Hazy image (b) Model 1 (c) Model 2 (d) Model 3 (e) Model 4 (f) Model 5

Fig. 13 Experimental results of deformable convolution residual mixed attention block ablation

4 Conclusions

We presented a parallel feature fusion framework designed to synergistically address the dual challenges of haze removal and detail restoration.

We developed a fusion network to modulate the interaction between the haze layer and background components, yielding refined dehazing results. Complementing this architecture, we introduced a feature regularization loss to constrain the network to learn haze-invariant feature representations, thereby enhancing both haze elimination and detail preservation.

Extensive experimental results on the RESIDE dataset and real-world benchmarks demonstrated that the proposed algorithm achieved superior dehazing performance. The restored images exhibited natural color fidelity and well-preserved textural details. However, dehazing in scenarios with extremely dense haze remains a challenge. Future work will focus on bolstering the network's capability to characterize and eliminate dense haze, aiming to further elevate restoration quality under extreme atmospheric conditions.

Acknowledgement

This work was supported by the Scientific Research Startup Fund for Introduced Talents of Chongqing College of Humanities, Science and Technology (No. CRKRC2025002).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] ZHENG C B, YING W J, HU Q P. Comparative analysis of dehazing algorithms on real-world hazy images. Scientific Reports, 2025, 15: 10822.

[2] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior//2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 1956-1963.

[3] LING P Y, CHEN H A, TAN X, et al. Single image dehazing using saturation line prior. IEEE Transactions on Image Processing, 2023, 32: 3238-3253.

[4] LI B, HU H P, YANG Z M. Dark channel prior image dehazing algorithm based on superpixel segmentation. Journal

of Test and Measurement Technology, 2025, 39(4): 415-423.
[5] CAI B L, XU X M, JIA K, et al. DehazeNet: an end-to-end system for single image haze removal. IEEE Transactions on Image Processing, 2016, 25(11): 5187-5198.
[6] CHEN D D, HE M M, FAN Q N, et al. Gated context aggregation network for image dehazing and deraining//2019 IEEE Winter Conference on Applications of Computer Vision, January 7-11, 2019. VillageWaikoloa, HI, USA. New York: IEEE, 2019: 1375-1383.
[7] QIN X, WANG Z L, BAI Y C, et al. FFA-net: feature fusion attention network for single image dehazing. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11908-11915.
[8] LUO P J, XIAO G Q, GAO X B, et al. LKD-net: large kernel convolution network for single image dehazing//2023 IEEE International Conference on Multimedia and Expo, July 10-14, 2023, Brisbane, Australia. New York: IEEE, 2023: 1601-1606.
[9] CHEN Z X, HE Z W, LU Z M. DEA-net: single image dehazing based on detail-enhanced convolution and content-guided attention. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.
[10] HIDE R. Optics of the atmosphere: scattering by molecules and particles. Physics Bulletin, 1977, 28(11): 521.
[11] ZHANG J L, YANG Y. Single image dehazing based on hazy features extraction and enhancement network. Journal of Measurement Science and Instrumentation, 2023, 14(1): 45-54.
[12] JIANG K, WANG Z Y, YI P, et al. Rain-free and residue hand-in-hand: a progressive coupled network for real-time image deraining. IEEE Transactions on Image Processing, 2021, 30: 7404-7418.
[13] GAO G H, GUO Y N, ZHOU L M, et al. Res2Net-based multi-scale and multi-attention model for traffic scene image classification. PLoS One, 2024, 19(5): e0300017.
[14] LIU Y D, WANG X B, ZHU Y R, et al. SDCNet: spatially-adaptive deformable convolution networks for HR NonHomogeneous dehazing//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 17-18, 2024, Seattle, WA, USA. New York: IEEE, 2024: 6682-6691.
[15] HUANG H, HO G T S, GEDA M W, et al. Dehaze-attention: enhancing image dehazing with a multi-scale, attention-based deep learning framework. Scientific Reports, 2025, 15: 44191.
[16] WANG X, ZHANG S F, LIN Y Y, et al. Pixel attention convolutional network for image super-resolution. Neural Computing and Applications, 2023, 35(11): 8589-8599.
[17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. 2014: arXiv: 1409.1556.
[18] LI B Y, REN W Q, FU D P, et al. Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing, 2019, 28(1): 492-505.
[19] ANCUTI C O, ANCUTI C, TIMOFTE R, et al. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 754-762.
[20] ANCUTI C O, ANCUTI C, TIMOFTE R. NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 14-19, 2020. Seattle, WA, USA. New York: IEEE, 2020: 444-445.

雾-细节协同机制驱动的并行特征融合去雾框架

余 璇, 马嘉庆, 于雷易*

重庆人文科技学院 计算机工程学院, 重庆 401524

摘 要: 由于空间变化的雾气与高频背景细节之间的复杂耦合, 图像去雾一直是低级计算机视觉任务中一个极具挑战性的不适定问题。现有的基于深度学习的方法往往难以平衡强力去雾与精细纹理保留之间的矛盾, 常常导致色彩失真、光晕效应或细节丢失。为了解决这些局限性并实现高保真恢复, 本文提出了一种由雾-细节协同机制驱动的并行特征融合框架。该框架采用双分支并行架构来解耦去雾过程。上部分支作为雾层提取网络, 利用基于Res2Net的编码器捕获多尺度语义特征, 并集成了新颖的可变形卷积-残差混合注意力模块。通过动态调整感受野, 该模块能够精确表征非均匀的雾分布并建模长程依赖关系。同时, 下部分支作为细节补偿网络, 利用包含多尺度空洞卷积的上下文细节信息块来聚合上下文线索, 并增强高频纹理细节的表示。随后, 融合网络执行自适应特征集成, 有效地将提取的雾特征与增强的细节信息合并, 以重建无雾图像。为了确保训练的鲁棒性, 本文引入了一种双重监督机制, 结合了用于对齐潜在空间特征分布的特征正则化损失和用于约束像素级内容保真度的重建损失。在合成基准和真实世界数据集上进行的广泛定量和定性实验表明, 所提算法表现出优越的性能, 与最先进的方法相比, 取得了较更高的峰值信噪比和结构相似性得分。视觉对比进一步证实, 该方法在有效去除浓雾的同时, 恢复了生动的色彩和清晰的结构细节, 且未引入伪影。

关键词: 图像去雾; Res2Net; 注意力机制; 可变形卷积; 纹理保持; 特征监督

引用格式: YU Xuan, MA Jiaqing, YU Leiyi. Parallel feature fusion framework for image dehazing driven by haze-detail collaboration. Journal of Measurement Science and Instrumentation, 2026, 17(2): 232-242. DOI: 10.62756/jmsi.1674-8042.2026020