Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Cgh-Dehazenet: A Lightweight Channel Grouping Half- Convolutional Network With Cross-Attention Refinement for Single Image Dehazing

**Xukang Zheng**

zhengxukang@mails.gdut.edu.cn

Guangxi Radio and Television Technology Center, Liuzhou Sub-cente

**Cheng Tang**

Zhuhai College of Science and Technology

**Shoujing Huang**

Guangxi Radio and Television Technology Center, Liuzhou Sub-cente

**Saibiao Jiang**

Zhuhai College of Science and Technology

**Research Article**

**Keywords:** Single image dehazing, lightweight network, channel grouping half-convolution, cross-attention fusion, image restoration

**Posted Date:** October 12th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11170403/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# CGH-DEHAZENET: A LIGHTWEIGHT CHANNEL GROUPING HALF-CONVOLUTIONAL NETWORK WITH CROSS-ATTENTION REFINEMENT FOR SINGLE IMAGE DEHAZING

**Abstract**.Image dehazing remains a challenging task due to the complex degradation caused by atmospheric scattering and the high computational cost of existing deep learning-based methods. While convolutional neural networks have achieved remarkable performance, their heavy architectures often hinder deployment in real-time and resource-constrained scenarios. To address these issues, we propose a novel lightweight end-to-end network, termed CGH-DehazeNet, for efficient and high-quality image dehazing. Specifically, we employ a Channel Grouping Half-Convolution (CGHalfConv) block as the core feature extractor. By partitioning channels and applying convolutions to only half of the grouped features while preserving the remaining ones via identity mapping, CGHalfConv substantially reduces redundant computations and model parameters without compromising representational capacity. Furthermore, to mitigate the information loss and artifact propagation induced by conventional skip connections, we introduce a Cross-Attention Refinement Fusion (CARF) module. CARF dynamically recalibrates and fuses encoder details with decoder semantics through a channel-wise cross-attention mechanism, effectively suppressing haze-induced noise. Coupled with Detail Enhancement Blocks (DEB) in the decoding path to recover high-frequency textures, the proposed architecture achieves a superior trade-off between efficiency and restoration quality. Extensive experimental results demonstrate the effectiveness of CGH-DehazeNet, achieving a PSNR exceeding 40 dB with merely 1.932M parameters, delivering performance on par with existing state-of-the-art methods.

***Keywords**:* Single image dehazing, lightweight network, channel grouping half-convolution, cross-attention fusion, image restoration.

## 1. INTRODUCTION

Image dehazing is a fundamental yet challenging task in low-level computer vision, aiming to recover clear, haze-free scene content from degraded observations[1]. Haze, caused by the scattering and absorption of light by suspended atmospheric particles, severely degrades image contrast and induces color shifts. Such degradation leads to significant performance deterioration in downstream high-level vision tasks, such as object detection and semantic segmentation[2]. Consequently, developing efficient and effective dehazing algorithms has garnered extensive attention from both academia and industry over the past decades. As a foundational low-level restoration task, image dehazing serves as a critical pre-processing step for subsequent high-level vision applications[3]. In this paper, we aim to develop an effective algorithm to remove haze and recover fine details from hazy inputs.

Early dehazing methods primarily relied on hand-crafted priors and physical models. Representative approaches, including the Atmospheric Scattering Model (ASM)[4], Dark Channel Prior (DCP) [5], and Color Attenuation Prior (CAP) [6], achieved notable success under specific assumptions. However, when real-world scenes violate these assumptions, such methods often suffer from color distortion, halo artifacts, and over-saturation. Early CNN-based methods[7-9] typically estimated the transmission map and atmospheric light separately before reconstructing haze-free images via the ASM . These transmission maps were usually supervised by ground truth generated from synthetic datasets. Nevertheless, inaccurate estimation of either the transmission map or atmospheric light significantly compromises restoration quality. With the advent of deep learning, Convolutional Neural Networks (CNNs) and Vision Transformers have revolutionized the field by learning direct mappings from hazy to clear images in a data-driven manner[8]. To capture long-range dependencies and rich contextual information, mainstream architectures frequently stack heavy convolutions, multi-scale processing modules, or self-attention mechanisms with quadratic complexity[10,11]. This "bigger-is-better" design paradigm renders them impractical for real-time deployment on resource-constrained edge devices, such as UAVs, mobile terminals, and embedded surveillance systems. Although lightweight dehazing networks have been proposed to alleviate efficiency bottlenecks[12-14], they typically achieve computational

reduction at the cost of significant performance degradation; moreover, their computational overhead and GPU memory consumption remain non-negligible. Two core issues persist: first, standard lightweight convolutions (e.g., depth-wise separable convolutions) lack sufficient representational capacity to model the complex non-linear mapping of haze degradation; second, naive skip connections in encoder-decoder architectures—relying on simple concatenation or addition—inevitably propagate haze-corrupted noise and spatial misalignments from the encoder to the decoder, limiting final restoration fidelity.

To address these challenges, we posit that an ideal lightweight dehazing network must simultaneously achieve efficient feature extraction and discriminative feature fusion. Inspired by recent advances in lightweight network design, we propose CGH-DehazeNet, a novel lightweight end-to-end network designed to achieve state-of-the-art dehazing performance with minimal computational overhead. Specifically, departing from traditional heavy convolutions and oversimplified lightweight operators, we design the Channel Grouping Half-Convolution (CGHalfConv) module as the core feature extractor. CGHalfConv partitions input channels into groups, applying standard convolution to only half of the channels while preserving the other half via identity mapping. This asymmetric design substantially reduces FLOPs and parameter count by eliminating redundant computation on low-information channels, while maintaining robust representation of haze-relevant features through the active convolution branch. Furthermore, to overcome the inherent limitations of conventional skip connections, we introduce the Cross-Attention Refinement Fusion (CARF) module at each encoder-decoder junction. Rather than blindly fusing multi-scale features, CARF employs a channel cross-attention mechanism to dynamically recalibrate encoder detail features using decoder semantic features as queries. This selectively suppresses haze-corrupted activations and prevents artifact propagation during upsampling. Additionally, we integrate a Detail Enhancement Block (DEB) into the decoding path to explicitly restore high-frequency textures and edge information smoothed by haze and downsampling operations. CGH-DehazeNet achieves superior or comparable restoration quality while requiring only a fraction of the computational resources of existing state-of-the-art methods.

The main contributions of this work are summarized as follows:

We propose CGH-DehazeNet, a novel lightweight end-to-end single image dehazing network that achieves an exceptional balance between restoration quality and computational efficiency. By combining a symmetric encoder-decoder architecture with specialized lightweight modules, the network effectively recovers high-quality haze-free images while significantly reducing computational complexity and parameter count, making it highly suitable for real-time and resource-constrained applications.

We design the Channel Grouping Half-Convolution (CGHalfConv) module to replace traditional heavy convolutions. By partitioning feature channels and applying convolution to only half while retaining the other half via identity mapping, CGHalfConv drastically reduces redundant computation and memory footprint without sacrificing the network's representational capacity for modeling haze degradation.

We introduce the Cross-Attention Refinement Fusion (CARF) module at the skip connections of the encoder-decoder framework. Unlike conventional simple concatenation or addition, CARF leverages channel cross-attention to dynamically recalibrate and fuse low-level spatial details from the encoder with high-level semantic cues from the decoder. This effectively suppresses haze-induced noise and prevents artifact propagation during upsampling.

Extensive experiments on multiple benchmark datasets validate the superiority of CGH-DehazeNet. The proposed method achieves excellent quantitative metrics and visual quality while significantly reducing FLOPs and improving inference speed compared to state-of-the-art approaches.

## 2. APPROACH

As illustrated in Figure 1, our CGH-DehazeNet consists of three components: an encoder, a feature transformation module, and a decoder. Serving as the core of CGH-DehazeNet, the feature transformation module adopts stacked Detail Enhancement Blocks (DEB) and Channel Grouping Half-Convolution (CGHalfConv) blocks to learn haze-free feature representations. The architecture comprises three stages, within which we employ identical modules—namely, CGHalfConv and DEB—to extract corresponding

features, while CARF is introduced in Stage 2 and Stage 3 for feature fusion. Given a hazy input image Input$\in\mathbb{R}^{3\times H\times W}$, CGH-DehazeNet aims to recover the corresponding haze-free output image Output$\in\mathbb{R}^{3\times H\times W}$.

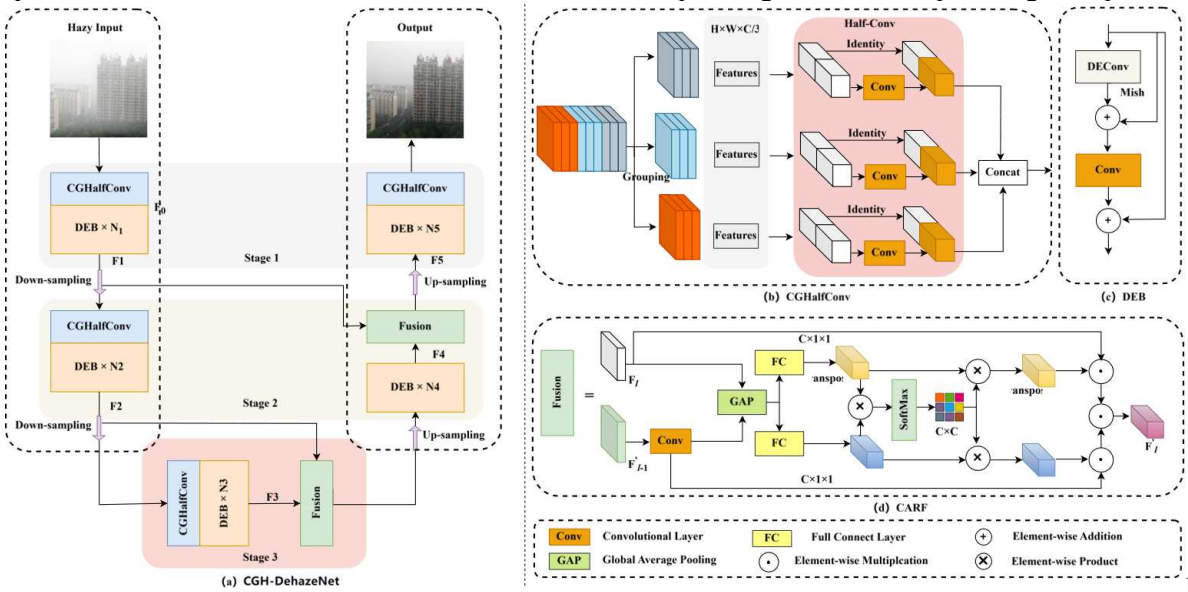



Fig. 1 – Overall architecture of the proposed CGH-DehazeNet.

## 2.1. Channel Grouping Half-Convolution

As shown in Figure 1(b), to ensure the diversity and integrity of the extracted features, CGHalfConv uniformly partitions the input feature map $X_{input}\in\mathbb{R}^{h\times w}$ into three independent groups along the channel dimension, a process termed Channel Grouping. For instance, given c total channels, each group contains c/3 channels, with any remainder channels assigned to the last group.This process is formulated as follows：

$$X_{input} = Concat\{X^1_{input}, X^2_{input}, X^3_{input}\} \tag{1}$$

Here,$X^i_{input}\in\mathbb{R}^{h\times w\times c}$ with a = c/3,Convolution operations are performed independently within each group, enabling targeted extraction of the aforementioned three types of features while substantially reducing model parameters and computational overhead. Let X denote the input feature map and Y the output. The FLOPs (floating-point operations) of Half-Convolution amount to approximately $h\times w\times k^2\times c/2$.Compared to standard convolution, the memory access is also reduced by half, to approximately $h\times w\times 2\times c/2$.this enables CGHalfConv to achieve significant lightweight efficiency while maintaining robust feature extraction capability.This process is formulated as follows：

$$Y = Concat\{(Conv(X[h,w,0:c/2]), X[h,w,c/2+1:c])\} \tag{2}$$

## 2.2. Detail Enhancement Convolution Module

In the field of image dehazing, prior methods [5,6] commonly adopt standard vanilla convolution (VC) layers for feature extraction and learning. However, ordinary convolutional layers operate without any constraints when searching the vast solution space (often starting from random initialization), which inherently limits their representational and modeling capacity. Notably, high-frequency information, such as edges and contours, plays a crucial role in restoring images degraded by hazy scenes. Motivated by this observation, several researchers[8]have incorporated edge priors into dehazing models to facilitate the recovery of sharper contours. Inspired by these works[8], we design a Detail Enhancement Convolution (DEConv) layer, as illustrated in Figure 1(c), which enables the integration of carefully designed priors into an ordinary convolutional layer.

Before elaborating on the proposed DEConv, we first revisit differential convolution (DC). Prior works [15] typically characterize DC as convolution over pixel differences—where pixel differences are first

computed and subsequently convolved with kernel weights to produce feature maps—thereby enhancing the representational capacity and generalization ability of vanilla convolution. Central difference convolution (CDC) and angular difference convolution (ADC) are two representative DC variants that reduce computational cost and memory consumption by rearranging the learned kernel weights [15]. They have proven effective in edge detection and face anti-spoofing tasks .

We introduce DC to address the single-image dehazing problem. As depicted in Figure 2(b), our DEConv comprises five convolution layers deployed in parallel: vanilla convolution (VC), central difference convolution (CDC), angular difference convolution (ADC), horizontal difference convolution (HDC), and vertical difference convolution (VDC). Within DCs, prior information can be explicitly encoded into CNNs by designing specific strategies for computing pixel-pair differences. For our DEConv, in addition to CDC and ADC, we further derive HDC and VDC to integrate traditional local descriptors—such as Sobel[16] , Prewitt[17] , or Scharr [18]—into the convolutional layers. The horizontal gradient is computed by calculating the differences of selected pixel pairs. After training, we equivalently rearrange the learned kernel weights and directly apply convolution to the unprocessed input features. It is worth noting that the equivalent kernel shares a similar format with traditional local descriptors, where the sum of horizontal weights equals zero. The horizontal kernels of Sobel , Prewitt , and Scharr ] can be regarded as special cases of this equivalent kernel. VDC follows an analogous derivation by replacing the horizontal gradient with its vertical counterpart. Both HDC and VDC explicitly encode gradient priors into convolutional layers, thereby enhancing representational capacity and generalization ability through learning beneficial gradient information.

Given an input feature map $F_{in}$, DEConv produces the output $F_{out}$ with the same computational cost and inference time as a standard convolution layer, enabled by re-parameterization techniques. This process is formulated as follows (bias terms omitted for simplicity):

$$F_{out} = DEConv(F_{in}) = \sum_{i=1}^{5} F_{in} * K_i = F_{in} * \left( \sum_{i=1}^{5} K_i \right) = F_{in} * K_{kvt} \tag{3}$$

where $DEConv(\cdot)$ denotes the DEConv operation, $K_i\{i=1:5\}$ represent the convolution kernels of VC, CDC, ADC, HDC, and VDC, respectively, ast denotes the convolution operation, and $K_{cvt}$ represents the converted kernel that fuses the parallel convolutions together[15]. Figure 2(b) intuitively illustrates the processing pipeline of the re-parameterization technique. During the backpropagation phase, the kernel weights of the five parallel convolutions are updated individually using the chain rule of gradient propagation. In the forward propagation phase, the kernel weights of the parallel convolutions are frozen, and the converted kernel weights are computed by summing the weights at corresponding positions. It is worth noting that the re-parameterization technique accelerates both training and inference, as both stages involve the forward propagation phase[15]. In our design, standard convolution is employed to capture intensity-level information, while differential convolution serves to enhance gradient-level information. We simply sum the learned features to obtain the output of DEConv. We believe that more sophisticated pixel difference computation strategies could further advance image restoration tasks; however, this is beyond the primary scope of this work. Compared with vanilla convolutional layers, the proposed DEConv is capable of extracting richer features while maintaining the same parameter count, and introduces no additional computational cost or memory overhead during inference.

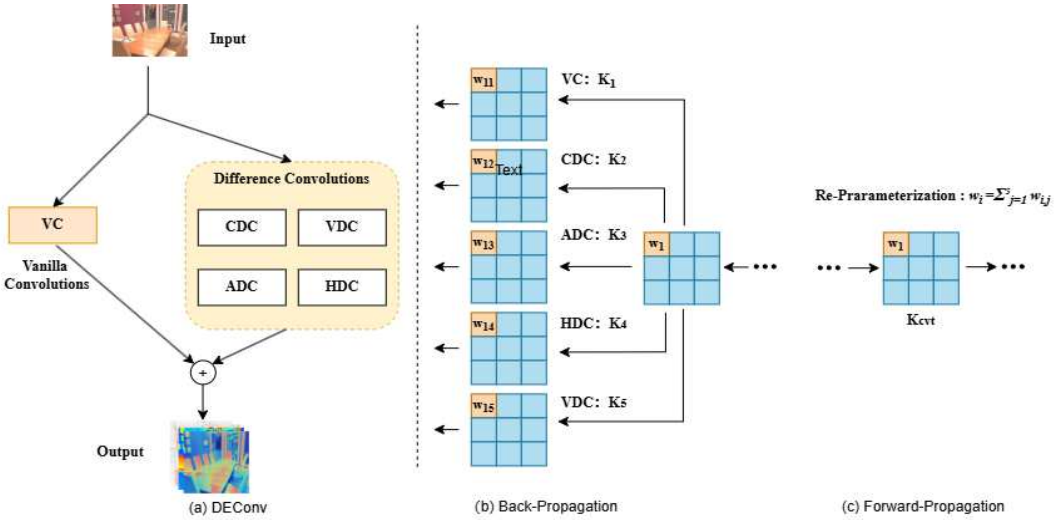



Fig. 2 – Structure of DEConv and the re-parameterization procedure.

## 2.3. Cross-Attention Refinement Fusion

As illustrated in Figure 1(d), the core of the Channel-Aware Cross-Resolution Fusion (CAFR) module is a lightweight channel-wise cross-attention mechanism[19]. Compared with conventional cross-resolution feature fusion modules, CAFR introduces bidirectional channel interaction in cross-resolution fusion. Specifically, cross-attention weights are generated via matrix multiplication of global descriptors from low- and high-level features, enabling high-level features to guide the modulation of low-level features while simultaneously allowing low-level features to refine high-level feature representations. This design achieves bidirectional complementary enhancement, rather than unidirectional guidance. By extracting channel-wise global statistics through global average pooling, CAFR mitigates the interference of local feature noise in attention weight learning, allowing channel modulation to better align with global feature distributions and thereby improving modulation accuracy[19]. As shown in Figure 3 shows four randomly selected channels of SIMs, by learning channel-specific similarity matrices (SIM), CAFR is able to recalibrate features more effectively, thus attending to haze distribution discrepancies across channels.

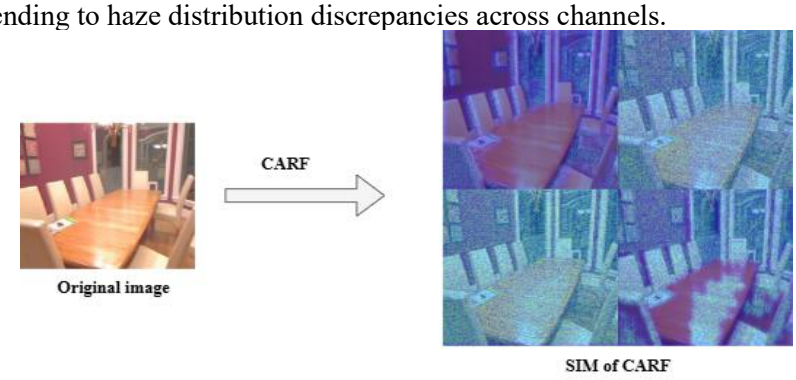



Fig. 3 – Learned similarity matrices of CAFR and the associated restoration results.

First, the module employs a 3×3 downsampling convolution for resolution alignment. Compared with bilinear interpolation, this approach preserves local feature structures during downsampling while incurring only marginally higher computational cost than interpolation, thus striking a favorable balance between effectiveness and efficiency. The module consists solely of a downsampling convolution, global pooling, 1×1 projection, cross-attention, and a fusion convolution, with no complex architectural components, resulting in extremely low parameter and computational overhead that renders it well-suited for lightweight models. Second, the module leverages cross-attention to precisely modulate the dual-branch channel features and integrates them via element-wise addition followed by a refinement convolution, thereby mitigating feature conflicts and redundancy after fusion and enhancing feature consistency[19].Specifically, to adaptively fuse the aggregated spectral features with the enhanced spatial details, we first align the spatial dimensions of $F_l$and

$F'_{l-1}$using a 3×3 convolution, thereby strengthening foreground–background contrast. Subsequently, global average pooling (GAP) is applied to each to compute global confidence scores. These scores are then concatenated and passed through a 1×1 convolution followed by an activation function to generate adaptive fusion weights.

$$W = f_s\left(\phi_{1\times1}\left(\left[AVG(F_l), AVG\left(Conv(F'_{l-1})\right)\right]\right)\right) \tag{4}$$

These are followed by a fully connected layer that produces global spectral vectors $V, V_{l-1} \in R^{C\times1\times1}$, . The resulting vectors are employed in cross-attention with a softmax function to derive fusion weights $W_l, W_{l-1}$ , alleviating spectral bias between layers.

$$V_l, V_{l-1} = FC(W) \tag{5}$$

$$A = softmax\left(V_l^T V_{l-1}\right) \tag{6}$$

$$W'_l, W'_{l-1} = A^T V_l, AV_{l-1} \tag{7}$$

Finally, we embed $F_l$and $F'_{l-1}$ into the low-level features via adaptive weighted summation, thereby enhancing foreground–background contrast.

$$F'_l = F \odot W'_l + F'_{l-1} \odot W'_{l-1} \tag{8}$$

### 2.4. Activation Function

In deep networks, image data first pass through convolutional layers for feature extraction, followed by Batch Normalization (BN) layers that normalize the extracted features, and subsequently through nonlinear activation functions that remap the data distribution to increase the model's nonlinear capacity, thereby enhancing its learning and representational power. Currently, most segmentation networks employ the ReLU [20]activation function, illustrated in Fig. 4(a). However, ReLU suffers from a notable limitation: when the input is negative, the output is zero and the gradient during backpropagation is also zero, preventing gradient updates—a phenomenon known as the vanishing gradient problem—which hampers effective model training. For image data containing negative values, ReLU renders training updates difficult, leading to degraded performance.

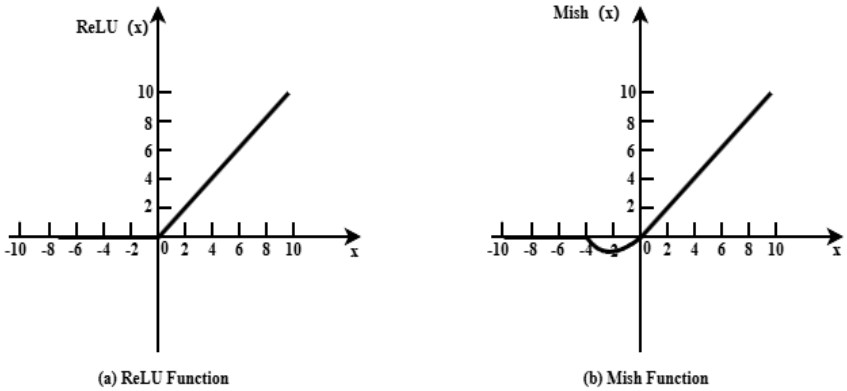



Fig. 4 – ReLU and Mish activation functions

To address these issues, this paper adopts the Mish activation function proposed by Misra et al., as shown in Fig. 4(b). Mish possesses the following desirable properties: unbounded above, bounded below, smooth, and non-monotonic[21]. Its unbounded positive range avoids saturation caused by capping, while it amplifies small negative inputs and attenuates large negative ones. Theoretically, the slight allowance of negative values facilitates better gradient flow, and the smoothness of the curve enables more feature information to propagate

into the model, effectively mitigating the vanishing gradient problem and improving both model accuracy and training stability. The formulations of the two activation functions are given as follows:

$$ReLU(x) = max(0,1) \tag{9}$$

$$Mish(x) = x \times tanh\big(\ln\big(1 + exp(x)\big)\big) \tag{10}$$

## 3. RESULTS AND ANALYSIS

### 3.1. Image dataset AND Implementation Details

We adopt ITS and OTS for training, and SOTS for testing. It is worth noting that SOTS is partitioned into two subsets, namely SOTS-indoor and SOTS-outdoor, for separately evaluating models trained on ITS and OTS, respectively. ITS contains 1,399 indoor clean images, and for each clean image, 10 synthetic hazy images are generated based on the physical scattering model. For OTS, we select approximately 3,000 images for the training process. SOTS-indoor and SOTS-outdoor comprise 500 indoor and 500 outdoor test images, respectively. In addition, the Haze4K dataset [22] contains 3,000 synthetic training images.

We implement the proposed model on the PyTorch deep learning platform using a single NVIDIA RTX 4070 Ti Super GPU. The numbers of DEB blocks deployed at different stages, denoted as $[N_1, N_2, N_3, N_4, N_5]$, are set to [4, 4, 8, 4, 4]. The Adam optimizer[23] is adopted with $\beta 1$、$\beta 2$ and $\varepsilon$ set to their default values of 0.9, 0.999, and 1e-8, respectively. The initial learning rate and batch size are set to 1e-4 and 16, respectively. A cosine annealing strategy[24] is employed to decay the learning rate from the initial value to 1e-4. For training, we randomly crop patches of size 256 × 256 from the original images and apply two data augmentation techniques: rotation by 90°,180° or 270° ,and vertical or horizontal flipping. The model is trained for 1.5 million iterations throughout the entire training process.

### 3.2. Evaluation Metrics

For quantitative evaluation, we adopt two widely used image quality assessment metrics in computer vision, namely Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [40], to evaluate the dehazing performance. For fair comparison, both metrics are computed on RGB color images without pixel cropping. SSIM measures image similarity from three aspects: luminance, contrast, and structure, and is formulated as:

$$SSIM(I, I_0) = l(I, I_0) \cdot c(I, I_0) \cdot s(I, I_0) \tag{11}$$

In this context, $I$ and $I_O$ denote the original image and the restored image, respectively. The terms $l(I, I_O)$,$c(I, I_O)$ and $s(I,I_O)$, and represent the luminance, contrast, and structural similarity components, respectively, which are defined by Eqs. (12), (13), and (14).

$$l(I, I_O) = \frac{2\mu_I\mu_{I_O} + C_1}{\mu_I^2 + \mu_{I_O}^2 + C_1} \tag{12}$$

$$c(I, I_O) = \frac{2\sigma_I\sigma_{I_O} + C_2}{\sigma_I^2 + \sigma_{I_O}^2 + C_2} \tag{13}$$

$$s(I, I_O) = \frac{\sigma_{II_O} + C_3}{\sigma_I\sigma_{I_O} + C_3} \tag{14}$$

In this context, $\mu_I$and $\mu_{I_O}$ denote the mean intensity of images $I$ and $I_O$, respectively. The standard deviations of the intensity of $I$ and $I_O$ are denoted by$\sigma_I$and $\sigma_{I_O}$, and their covariance is denoted by $\sigma_{II_O}$. Three constants $C_1$,$C_2$ and $C_3$, each much smaller than 1, are introduced to prevent the denominators from vanishing.

PSNR, which primarily measures the pixel-wise error between the restored image and the original image, is defined in Eq. (15).

$$PSNR[dB] = 10\lg\left(\frac{MAX_I^2}{MSE}\right) \tag{15}$$

Here, $MAX_I$ is the maximum possible pixel value (255 for 8-bit images), and *MSE* refers to the mean squared error.

### 3.3. Experimental Comparisons

We compare CGH-DehazeNet against a wide range of state-of-the-art dehazing methods, including the traditional prior-based approach DCP[12]and recent deep learning-based methods:

DehazeNet[2],AOD-Net[3],GFN[29],FFA-Net[5],MSBDN[10],DMT-Net[22],AECR-Net[6],SGID-PFF[27],UDN[28],PMDNet[11],Dehamer[17],and DEA-Net[15].

**Quantitative Comparison**. Table 1 presents the quantitative results on the SOTS-Indoor, SOTS-Outdoor, and Haze4K datasets. As shown, our method achieves the best performance across all metrics on all three datasets. Specifically, on SOTS-Indoor, CGH-DehazeNet attains 40.83 dB PSNR and 0.9945 SSIM, outperforming the second-best method (DEA-Net) by 0.63 dB in PSNR. On SOTS-Outdoor, our method reaches 36.77 dB PSNR, surpassing DEA-Net by 0.74 dB. On the more challenging Haze4K dataset, CGH-DehazeNet achieves 33.61 dB PSNR and 0.99 SSIM, demonstrating remarkable robustness across varying haze densities.

More importantly, CGH-DehazeNet attains such state-of-the-art performance with substantially lower computational overhead. As shown in the "Overhead" column of Table 1, our model requires only 1.932 M parameters and 18.75 G FLOPs. In contrast, heavy models such as Dehamer and FFA-Net demand 132.4 M / 48.93 G and 4.456 M / 287.5 G, respectively. Even compared with the highly competitive DEA-Net (3.673 M / 32.23 G), our method achieves superior PSNR/SSIM scores while reducing parameters by 47.4% and FLOPs by 41.8%. These compelling results validate that CGH-DehazeNet establishes a new trade-off between restoration quality and computational efficiency.

Table 1 Comparison with different methods:

| Method | SOTS-indoor | | SOTS-outdoor | | Haze4K | | Overhead | |
|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | #Param.(M) | #FLOPs(G) |
| DCP[12] | 16.61 | 0.8546 | 19.14 | 0.8605 | 14.01 | 0.76 | - | - |
| DehazeNet[2] | 19.82 | 0.8209 | 27,75 | 0.9269 | 19,12 | 0.84 | 0.008 | 0.5409 |
| AOD-Net[3] | 20.51 | 0.8162 | 24.14 | 0.9198 | 17.15 | 0.83 | 0.0018 | 0.1146 |
| GFN[29] | 22.30 | 0.8800 | 21.55 | 0.8444 | - | - | 0.4990 | 14.94 |
| FFA-Net[5] | 36.39 | 0.9886 | 33.57 | 0.9840 | 26.97 | 0.95 | 4.456 | 287.5 |
| MSBDN[10] | 32.77 | 0.9812 | 34.81 | 0.9857 | 22.99 | 0.85 | 31.35 | 24.44 |
| DMT-Net[22] | - | - | - | - | 28.53 | 0.96 | 51.79 | 75.56 |
| AECR-Net[6] | 37.17 | 0.9901 | - | - | - | - | 2.611 | 52.20 |
| SGID-PFF[27] | 38.52 | 0.9913 | 30.20 | 0.9754 | - | - | 13.87 | 152.8 |
| UDN[28] | 38.62 | 0.9909 | 34.92 | 0.9871 | - | - | 4.250 | - |
| PMDNet[11] | 38.41 | 0.9900 | 34.74 | 0.9850 | 33.49 | 0.98 | 18.90 | - |
| Dehamer[17] | 36.63 | 0.9881 | 35.18 | 0.9860 | - | - | 132.4 | 48.93 |
| DEA-Net[15] | 40.20 | 0.9934 | 36.03 | 0.9891 | 33.19 | 0.99 | 3.673 | 32.23 |
| Ours | 40.83 | 0.9945 | 36.77 | 0.9889 | 33.61 | 0.99 | 1.932 | 18.75 |

**Qualitative Comparison**. Figure 5 and Figure 6 presents a qualitative visual comparison on representative samples from the SOTS-Outdoor and SOTS-Indoor datasets. The traditional DCP method suffers from severe color distortion and incomplete haze removal. While deep learning-based methods such as GDN, FFA, and Dehamer improve overall visibility, they tend to introduce residual haze, over-saturation, or fail to recover fine textures (highlighted by the red zoomed-in boxes). For instance, in the first row, existing methods leave visible haze along building edges, whereas our method restores sharp structural details. In the

second row, our method accurately recovers the train textures and background vegetation without introducing artifacts. Similarly, in the indoor scenes (third and fourth rows), our method effectively removes dense haze while preserving natural color fidelity and high-frequency details, yielding results that are visually closest to the ground truth (GT). These visual results strongly corroborate the quantitative superiority of the proposed architecture.

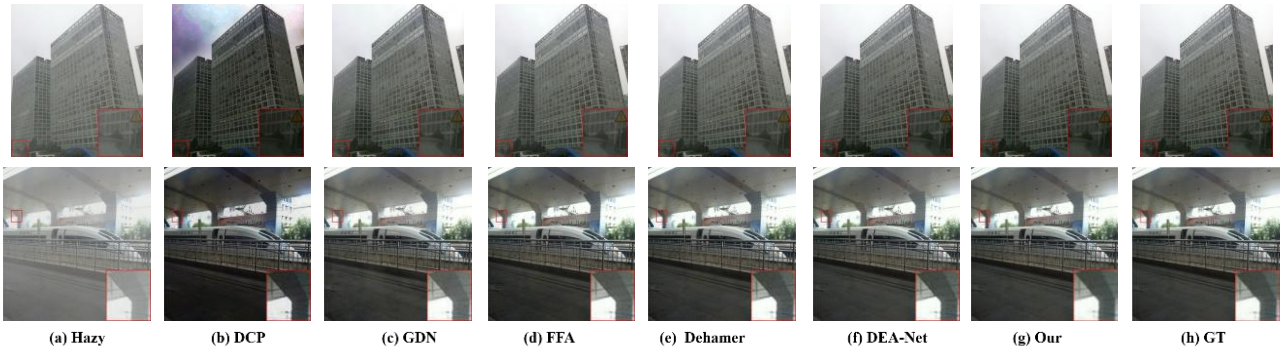



Fig. 5 – Visual comparison of different methods on the SOTS-Outdoor dataset.

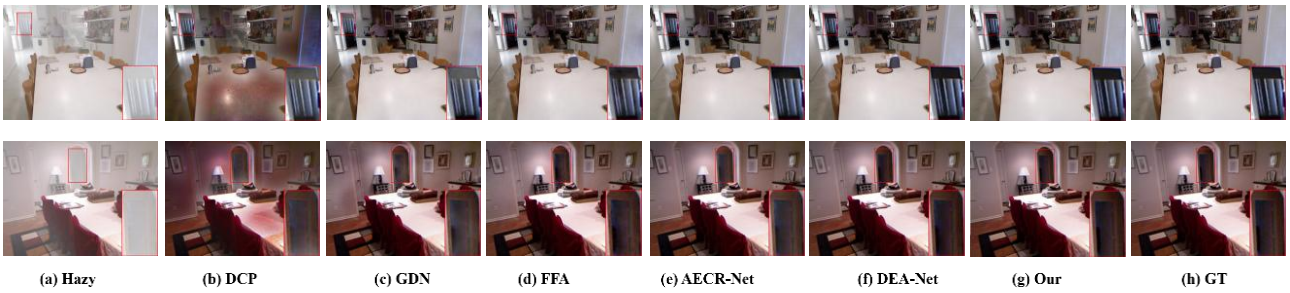



Fig. 6 – Visual comparison of different methods on the SOTS-Indoor dataset.

### 3.4. Ablation Studies

To validate the effectiveness of each proposed component, we conduct comprehensive ablation experiments on the SOTS and Haze4K datasets.

**Effectiveness of Core Modules.**Table 2 investigates the individual and combined contributions of Channel Grouped Half-Convolution (CGHalfConv), the Detail Enhancement Block (DEB), and the Cross-Attention Refinement Fusion (CARF) module. Starting from a baseline using standard convolution with CARF (Conv+CARF), replacing standard convolution with CGHalfConv (CGHalfConv+CARF) maintains competitive performance while substantially reducing computational cost (as evidenced by the remarkably low overhead of the final model). More notably, integrating the DEB module yields significant performance gains. On SOTS-Indoor, comparing "Conv+CARF" (38.63 dB) with "Conv+DEB+CARF" (39.71 dB), the inclusion of DEB brings a PSNR improvement of up to 1.08 dB, confirming its critical role in restoring high-frequency textures. Ultimately, the full configuration (CGHalfConv+DEB+CARF) achieves the optimal performance of 40.83 dB, demonstrating that these modules are highly complementary: CGHalfConv ensures efficient feature extraction, DEB enhances local details, and CARF facilitates discriminative multi-scale feature fusion.

Table 2 Ablation study on core modules:

| Method | SOTS-indoor | | SOTS-outdoor | | Haze4K | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Conv+CARF | 38.63 | 0.9874 | 34.88 | 0.9795 | 31.87 | 0.97 |
| CGHalfConv+CARF | 38.16 | 0.9844 | 35.07 | 0.9823 | 32.31 | 0.98 |
| Conv+DEB+CARF | 39.71 | 0.9895 | 35.78 | 0.9851 | 33.26 | 0.98 |
| CGHalfConv+DEB+CARF | 40.83 | 0.9945 | 36.77 | 0.9889 | 33.61 | 0.99 |

**Effect of Activation Functions.**Table 3 compares the performance of the ReLU and Mish activation functions in our network. The results show that Mish consistently outperforms ReLU across all datasets. On SOTS-Indoor, Mish achieves 40.83 dB, compared to 39.31 dB with ReLU, yielding a substantial improvement of 1.52 dB. This gain is attributed to the smooth, non-monotonic nature of Mish, which facilitates better gradient flow and richer feature representations during the optimization of lightweight networks.

Table 3 Comparison of different activation functions:

| Method | SOTS-indoor | | SOTS-outdoor | | Haze4K | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| ReLU | 39.31 | 0.9928 | 35.78 | 0.9827 | 33.56 | 0.98 |
| Mish | 40.83 | 0.9945 | 36.77 | 0.9889 | 33.61 | 0.99 |

**DEB Block Allocation.** Table 4 investigates the impact of the DEB block allocation across the five decoder stages, denoted as $[N_1, N_2, N_3, N_4, N_5]$,. We compare a lighter configuration [3,3,6,3,3] against our chosen configuration [4,4,8,4,4] . Increasing the number of DEB blocks from the lighter configuration to [4,4,8,4,4] raises the PSNR on SOTS-Indoor from 38.91 dB to 40.83 dB (+1.92 dB). Although the parameter count increases from 1.123 M to 1.932 M and FLOPs from 11.42 G to 18.75 G, the resulting model remains remarkably lightweight compared to existing state-of-the-art methods (e.g., 32.23 G FLOPs for DEA-Net), while delivering a substantial leap in restoration quality. This confirms that allocating sufficient detail enhancement capacity—particularly in the middle stage (N=8 )—is essential for achieving high-fidelity dehazing without compromising the lightweight nature of the network.

Table 4 Ablation study on the number of DEB modules:

| Number of DEB blocks $[N_1, N_2, N_3, N_4, N_5]$ | SOTS-indoor | | SOTS-outdoor | | Haze4K | | Overhead | |
|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | #Param. (M) | #FLOPs (G) |
| [3,3,6,3,3] | 38.91 | 0.9811 | 35.98 | 0.9805 | 33.23 | 0.98 | 1.123 | 11.42 |
| [4,4,8,4,4] | 40.83 | 0.9945 | 36.77 | 0.9889 | 33.61 | 0.99 | 1.932 | 18.75 |

## 4. CONCLUSION

This paper presents CGH-DehazeNet, a novel lightweight end-to-end single-image dehazing network that successfully bridges the trade-off between restoration quality and computational efficiency. To address the redundancy inherent in conventional convolutions, we design the Channel Grouped Half-Convolution (CGHalfConv) module, which eliminates redundant computation through asymmetric channel grouping while preserving strong feature representation capability. Furthermore, to overcome the limitations of naïve skip connections, we introduce the Cross-Attention Refinement Fusion (CARF) module, which dynamically recalibrates and selectively fuses multi-scale features via a channel-wise cross-attention mechanism, effectively suppressing the propagation of haze-induced noise and artifacts. Complemented by the Detail Enhancement Block (DEB) for explicit high-frequency texture recovery and the Mish activation function for improved gradient flow, the proposed architecture achieves superior dehazing performance. Extensive experiments on the SOTS-Indoor, SOTS-Outdoor, and Haze4K datasets demonstrate that CGH-DehazeNet attains performance on par with state-of-the-art methods in both quantitative metrics and visual quality, while requiring only 1.932 M parameters and 18.75 G FLOPs. Future work will explore extending this lightweight paradigm to other low-level vision tasks, such as image deraining and underwater image enhancement, as well as further optimizing the network for deployment on ultra-low-power mobile devices.

## REFERENCES


[1] R. T. Tan, "Visibility in bad weather from a single image," in CVPR.IEEE, 2008, pp. 1–8.

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

[3] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 4770–4778.

[4] P. Li, J. Tian, Y. Tang, G. Wang, and C. Wu, "Deep retinex network for single image dehazing," IEEE Transactions on Image Processing,vol. 30, pp. 1100–1115, 2020.

[5] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, 2020, pp.11 908–11 915.

[6] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10 551–10 560.

[7] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in European conference on computer vision. Springer, 2016, pp. 154–169.

[8] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 3194–3203.

[9] S. Narasimhan and S. Nayar, "Contrast restoration of weather degraded images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pp. 713–724, 2003.

[10] 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 Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2157–2167.

[11] T. Ye, M. Jiang, Y. Zhang, L. Chen, E. Chen, P. Chen, and Z. Lu, "Perceiving and modeling density is all you need for image dehazing," arXiv preprint arXiv:2111.09733, 2021.

[12] K. He, J. Sun, and X. Tang, "Single image haze removal using darkchannel prior," IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 12, pp. 2341–2353, 2010.

[13] D. Berman, S. Avidan et al., "Non-local image dehazing," in Proceedings of the IEEE conference on computer vision and pattern recognition,2016, pp. 1674–1682.

[14] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," IEEE transactions on image processing,vol. 24, no. 11, pp. 3522–3533, 2015.

[15] Chen, Z., He, Z., Lu, Z.-M., & Wu, S. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. IEEE Transactions on Image Processing,vol. 31, pp. 1023–1037. https://doi.org/10.1109/TIP.2024.3354103

[16] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma, "Knowledge Transfer Dehazing Network for NonHomogeneous Dehazing," in CVPR Workshop. IEEE,2020, pp. 1975–1983.

[17] C.-L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3d position embedding," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5812–5820.

[18] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, "Searching central difference convolutional networks for face antispoofing," in CVPR, 2020, pp. 5294–5304.

[19] Han, S., Xu, T., Liu, P., & Li, J, MODA: The first challenging benchmark for multispectral object detection in aerial images. In Proceedings of the AAAI Conference on Artificial Intelligence, vol.40, pp. 4574-4582.

[20] Krizhevsky, Alex, et al. "ImageNet Classification with Deep Convolutional Neural Networks." Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097–1105.

[21] MisraD.Mish:Aselfregularizednon-monotonicneuralactivation-function.https:arxiv.org/vc/arxiv/papers/1908/1908.08681v1.pdf.

[22] Y. Liu, L. Zhu, S. Pei, H. Fu, J. Qin, Q. Zhang, L. Wan, and W. Feng,"From synthetic to real: Image dehazing collaborating with unlabeled real data," in Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 50–58.

[23] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in ICLR, 2015, pp. 1–15.

[24] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in CVPR.IEEE, 6 2019, pp. 558–567.

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.

[26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in CVPR, 2018, pp. 7132–7141.

[27] H. Bai, J. Pan, X. Xiang, and J. Tang, "Self-guided image dehazing using progressive feature fusion," IEEE Transactions on Image Processing, vol. 31, pp. 1217–1229, 2022.

[28] M. Hong, J. Liu, C. Li, and Y. Qu, "Uncertainty-driven dehazing network," Proceedings of the AAAI Conference on Artificial Intelligence,vol. 36, no. 1, pp. 906–913, Jun. 2022.

[29] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang,"Gated fusion network for single image dehazing," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018,pp. 3253–3261.

[30] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attentionbased multi-scale network for image dehazing," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 7314–7323.

[31] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao,"Searching central difference convolutional networks for face antispoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5295–5305.

[32] Z. He, D. Chen, Y. Cao, J. Yang, Y. Cao, X. Li, S. Tang, Y. Zhuang, and Z. ming Lu, "Single image super-resolution based on progressive fusion of orientation-aware features," Pattern Recognition, vol. 133, p.109038, 1 2023.

[33] Tian F ,Zhang R ,Xiaopei L .An efficient structure-aware image dehazing algorithm for underground mining scenes.The Journal of Supercomputing,2026,82(5):344-344.DOI:10.1007/S11227-026-08489-3.
[34] Liu, C., Min, W., Song, J., Yang, Y., Sheng, G., Yao, T., Channel grouping vision transformer for lightweight fruit and vegetable recognition. Expert Systems with Applications, 2025,pp.128636.

*Received September 17, 2026*