

# ASMF: A Self-Supervised Atmospheric Scatter Model-Based Fusion Network for Infrared Image Enhancement

Luyuan Liu , Rui Song, *Member, IEEE*, Jiaojiao Li, *Senior Member, IEEE*, Wenman Wang, and Qingrong Wen

**Abstract**—Infrared camera sensors typically capture thermal radiation data, which can be influenced by random scattering and reflection during transmission. As a result, the infrared images produced often exhibit low contrast and clarity. This problem is especially evident in uncooled long-wave infrared (LWIR) images, which experience a decline in image quality, making it more difficult to enhance contrast. To tackle this issue, we propose a self-supervised network model that utilizes an atmospheric scattering model (ASM) specifically designed for the distinct characteristics of infrared imaging. Moreover, leveraging the analysis of ASM and the characteristics of infrared imaging, we introduce an image enhancement framework that integrates multi-scale dehazing with fusion techniques to facilitate self-supervised learning. Our method includes two self-supervised sub-networks, ASM-Net and F-Net, each equipped with tailored loss functions to address the challenges arising from the limited availability of high-quality datasets and the variability in imaging from different manufacturers. We improve self-supervised learning by integrating degradation models into the network architecture and incorporating prior knowledge into the loss functions. To evaluate our proposed method, we have developed a new LWIR dataset featuring real-world scenes. Extensive experiments conducted on established benchmark datasets, as well as our dataset, demonstrate that our approach significantly surpasses existing methods in both qualitative and quantitative assessments. The source code and dataset are available at <https://github.com/ANDYLLY/Infrared-image-enhancement-method-ASMF-and-LWIR-Dataset-LGC>.

**Index Terms**—Infrared image, image dehazing, self-supervised, infrared dataset

## I. INTRODUCTION

**I**NFRARED imaging is notable for its ability to detect thermal radiation from objects, setting it apart from visible light (VIS) imaging. This feature allows infrared imaging to function without reliance on external light sources, making it more robust against changes in lighting and atmospheric conditions. These qualities provide significant resistance to interference and adaptability to environmental variations, enabling the use of infrared systems in various fields such as

This work was supported in part by the Youth Innovation Team of Shaanxi Universities, the "Scientist + Engineer" team of the Qin Chuang Yuan in Shaanxi Province, the Xi'an Science and Technology Program, the 'Leading the Charge' initiative for the industrialization of core technologies in key industrial chains in Shaanxi Province, the National Nature Science Foundation of China under Grant 62371359, the Key Research and Development Program of Shaanxi Grant No.2024GX-ZDCYL-02-09, the Key Research Program of the Chinese Academy of Sciences , Grant NO. KGFZD-145-2023-15. (*Corresponding author: Rui Song, Jiaojiao Li*)

Luyuan Liu, Rui Song, Jiaojiao Li and Wenman Wang are with the State Key Laboratory of Integrated Service Network, Xidian University, Xi'an 710071, China (e-mail: andy\_lly@163.com; rsong@mail.xidian.edu.cn; jjli@xidian.edu.cn; wenmandeltaw@outlook.com).

Qingrong Wen is with the Optoelectronic System Division of the 11th Research Institute of China Electronics Technology Group Corporation.(qrwen@sina.com).

Manuscript received November xx, 2024; revised March xx, 2025.

military operations, security, autonomous driving, and medical imaging[1]. However, the high cost of infrared detectors poses a major barrier to their widespread use. Among these detectors, uncooled long-wave infrared (LWIR) sensors are more cost-effective, but they face issues like lower signal-to-noise ratios, reduced contrast, and blurred details due to detector non-uniformity and background thermal noise[2]. Thus, improving the contrast of infrared images is essential for enhancing the performance of LWIR systems.

Recent advancements in contrast enhancement algorithms, especially those based on the atmospheric scattering model (ASM), have shown significant promise. These algorithms typically involve two main steps: dehazing and contrast adjustment. However, their effectiveness is often hindered by reliance on global grayscale mapping, which complicates the balance between enhancing contrast and preserving details. In contrast, deep learning methods often depend on datasets created using simple degradation models or perceptually defined high-quality images, limiting their generalizability and enhancement quality. Additionally, while self-supervised methods have gained popularity in contrast enhancement, they mainly focus on incorporating perceptual image quality into loss functions, neglecting the underlying physical imaging model, which can lead to less effective results.

To tackle these issues, we propose a two-stage self-supervised framework for enhancing the contrast of infrared images that integrates the ASM model into its design. We also introduce a specialized infrared dataset aimed at testing, particularly for long-range imaging scenarios crucial for infrared applications. Through evaluations on the FLIR, MF3D, and our LGC dataset, which contains over 7,000 images, demonstrate that our method surpasses existing contrast enhancement techniques in both objective and subjective evaluations. Furthermore, our approach complements non-ASM-based methods, effectively addressing their shortcomings and advancing the field of infrared image enhancement.

The main contributions of this work are as follows:

- 1) We introduce a novel two-stage infrared image enhancement framework that incorporates ASM for stable multi-scale dehazing and fusion to indirectly estimate transmittance by statistical features or perceptual priors of target images, resulting in clearer and more detailed infrared images.

- 2) We create a dehazing network ASM-Net whose design is closely aligned with the ASM, using loss functions based on priors to achieve a consistent dehazing effect suited to the brightness of the target image.

- 3) We develop a fusion network that processes multi-scale dehazed images and employs self-supervised loss functions to maintain details while performing effective dehazing.

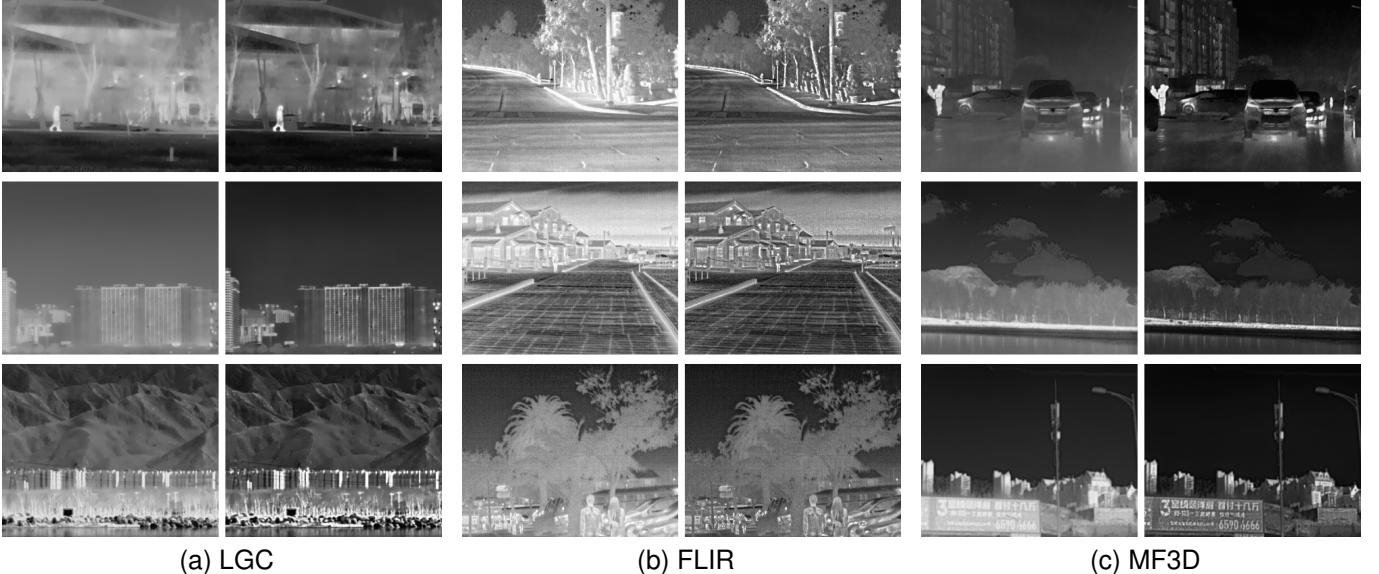


Fig. 1. Image enhancement results for three datasets: LGC, FLIR, and MF3D. The left column displays the original captured images for each group, while the right column presents their enhanced counterparts. Our algorithm significantly improves image quality across various scenarios, including streets, pedestrians, thickets, and buildings. (a) LGC. (b) FLIR. (c) MF3D.

4) We release the LGC dataset, an uncooled LWIR dataset consisting of 668 images taken with a 72mm focal length, covering a range of long-range scenarios to aid research in infrared image enhancement.

## II. RELATED WORK

### A. Conventional Methods

Histogram equalization (HE) is a prevalent technique for contrast enhancement, aiming to achieve a uniform distribution of gray levels. However, HE often results in over-enhancement and loss of detail. Variants such as Contrast Limited Adaptive Histogram Equalization (CLAHE) [3] and Double Plateau Histogram Equalization (DPHE) [4] are improvements upon HE by employing strategies such as histogram segmentation, spatial segmentation, and adaptive plateau value calculation. Further research has utilized metrics such as entropy and fuzziness to suppress noise and enhance contrast [5], [6]. Frequency-domain methods, including wavelet transform and homomorphic filtering, also aim to enhance image details while mitigating high-frequency noise [7], [8].

Despite their advantages, conventional methods often rely on handcrafted assumptions, which limits their adaptability to varying imaging conditions. The atmospheric scattering model (ASM) has emerged as a promising approach for infrared image enhancement. Inspired by the dark channel prior (DCP) in visible light image processing, DCP has been adapted for infrared images [9], [10]. Li et al. [11] applied ASM for infrared image enhancement, focusing on dehazing with gauss filter and contrast adjustment with DPHE. However, these ASM-based methods are typically shallow models that depend on single-scale global grayscale mapping, making it challenging to balance contrast enhancement with detail preservation. In contrast, our method employs a deep neural network to generate dehazed images and incorporates priors

into the loss function of the network, rather than relying on fixed handcrafted priors. This approach enhances adaptability and leads to significant improvements in image quality.

### B. Learning-Based Methods

Deep learning has achieved notable success in various low-level vision tasks, including low-light image enhancement and multi-exposure fusion, leading to its application in infrared image enhancement. Kuang et al. [12] introduced a GAN-based architecture for contrast enhancement in infrared images, utilizing a contrast degradation model to generate training datasets [13]. Wang et al. [14] explored nonlinear degradation models and achieved distinct visual results. However, these methods are heavily reliant on degradation models and perceptually defined high-quality ground truth images, which limits their generalizability. Marnissi et al. [15] proposed a Vision Transformer-based GAN trained on high and low-contrast images selected from the KAIST dataset, incorporating a super-resolution component for enhanced results. Despite their success, these methods are constrained by the limited availability of high-quality infrared datasets.

Recently, self-supervised approaches [16], [17], [18], inspired by low-level tasks such as ZeroDCE [19], YOLY [20], and DeepFuse [21], have emerged. For instance, MMFF-Net [22] employs perceptual priors to adjust brightness in infrared images. Xiao et al. [23] utilized a CNN-based network with perceptual metric-driven loss functions for contrast enhancement. While these methods leverage perceptual priors, they do not address the physical degradation models inherent in infrared imaging. In contrast, our method directly incorporates ASM into the network, bridging the gap left by non-ASM-based approaches and significantly enhancing contrast enhancement results.

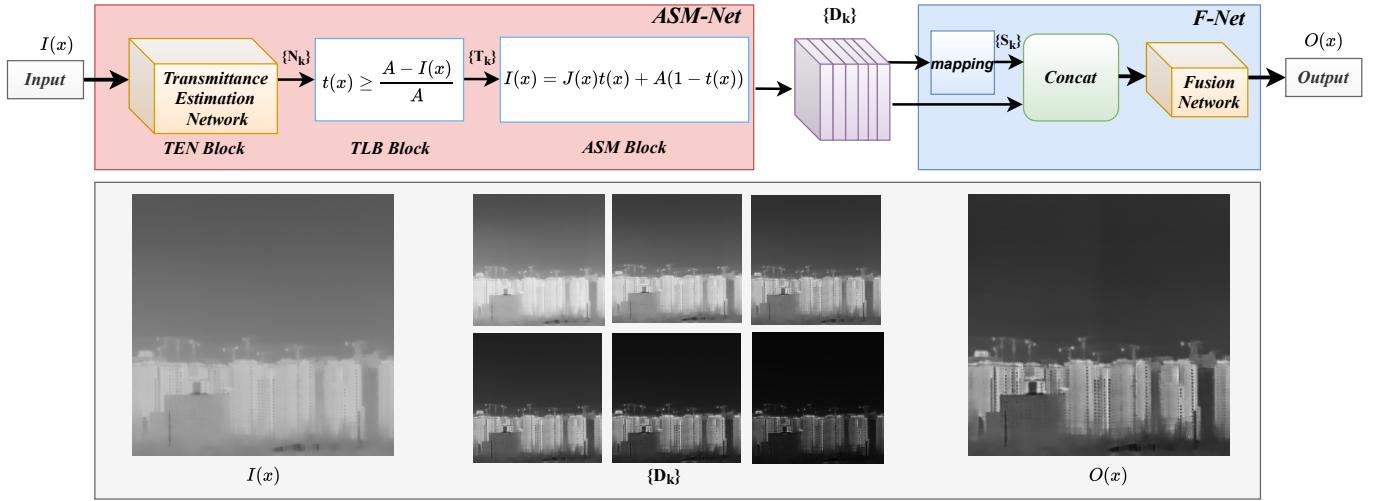


Fig. 2. The workflow of our method. The upper row illustrates the structure and data flow of the approach. The hazy images  $I(x)$  are first fed into ASM-Net to obtain multi-scale dehazed images  $D_k$ . Then, F-Net fuses the  $D_k$  images to generate the final enhanced image  $O(x)$ . The lower row demonstrates the image transformation from the input image  $I(x)$  to the multi-scale dehazed images  $D_k$  and the final output image  $O(x)$ .

### III. PROPOSED METHODS

The primary objective of our research is to develop a network architecture that incorporates the fundamental physical model and constraints pertinent to the task at hand. We integrate prior knowledge and quality assessment into loss functions, which serve to direct the network towards effective learning outcomes. Our methodology consists of two distinct sub-networks: a multi-scale dehazing network grounded in a physical model (ASM-Net) and a fusion network (F-Net). To facilitate a comprehensive understanding, we will commence with an analysis of the task and a thorough exposition of the theoretical framework that supports our pipeline, elucidating the rationale for adopting a two-stage approach. Subsequently, we will detail the network architecture, highlighting the manner in which these task-specific constraints are integrated into our design. Finally, we will outline the self-supervised loss functions that guide and enhance the training process.

#### A. Motivation and Theory

**ASM Model:** The Atmospheric Scattering Model (ASM) is frequently utilized for the dehazing of visible images and serves as a fundamental degradation model for enhancing the contrast of infrared images, thereby facilitating the reduction of blurriness. This is attributable to the fact that thermal radiation is similarly influenced as it traverses the atmosphere. In the context of machine vision, the mathematical formulation of the atmospheric scattering model is expressed in Equation 1:

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

In this equation,  $x$  represents the coordinates of the image pixels,  $I(x)$  denotes the observed hazy image,  $J(x)$  corresponds to the clean image,  $t(x)$  signifies the transmittance, and  $A$  indicates the atmospheric light value. The first term,  $J(x)t(x)$ , is referred to as direct attenuation, which represents an exponential decay of the intrinsic scene radiance. The

second term,  $A(1 - t(x))$ , introduces a white atmospheric veil, commonly known as airlight. We define airlight as  $L(x)$ :

$$L(x) = A(1 - t(x)) \quad (2)$$

Assuming that  $t(x)$  and  $A$  have been determined,  $J(x)$  can be computed using Equation 3:

$$J(x) = \frac{I(x) - L(x)}{t(x)} \quad (3)$$

The atmospheric light value  $A$  and the transmittance  $t(x)$  are two critical parameters that require estimation. The primary workflow of our approach, based on the analysis of the ASM, is delineated below and comprises two stages: multi-scale dehazing through transmittance estimation and the fusion of dehazed images, as illustrated in Figure 2.

**Transmittance Estimation:** The prior knowledge utilized in this study is predicated on the brightness of the image. By subtracting  $J(x)$  from  $I(x)$ , we can derive Equation 4, which facilitates the estimation of atmospheric light  $A$ :

$$I(x) - J(x) = \frac{1 - t(x)}{t(x)}(A - I(x)) \quad (4)$$

In Equation 4, it is important to note that  $t(x)$  is constrained within the interval  $[0, 1]$ . This indicates that the value of  $A$  influences whether  $J(x)$  appears brighter or darker than  $I(x)$ . Specifically, if  $A$  exceeds  $I(x)$ , then  $J(x)$  will be darker than  $I(x)$ ; conversely, if  $I(x)$  is greater than  $A$ ,  $J(x)$  will be brighter than  $I(x)$ .

An analysis of infrared datasets, such as FLIR and MF3D [38], reveals that these images display hazy characteristics and are predominantly bright, with numerous details appearing saturated and indistinct. Consequently, we anticipate that  $J(x)$  will be darker than  $I(x)$ , which serves as a critical prior for the estimation of  $A$ . Therefore, we assign  $A$  a value of 1, suggesting that  $J(x)$  should indeed be darker than

$I(x)$ . Furthermore, maintaining  $A$  as a constant facilitates the mapping from  $t(x)$  to  $J(x)$ .

Given that  $A$  is constant and  $I(x)$  serves as the input, we define the Darkness  $D$  as expressed in Equation 5:

$$D = \frac{I(x) - J(x)}{A - I(x)} \quad (5)$$

From this definition, we can derive Equation 6, which elucidates the relationship between transmittance and the dehazed image  $J(x)$ . This relationship indicates that the control of the output image's brightness directly influences the transmittance; however, it only reflects a global uniform scale of dehazing. Additionally, it poses challenges in directly manipulating  $J(x) - I(x)$ .

$$t(x) = \frac{1}{D + 1} \quad (6)$$

**Fusion for Multi-Scale Dehazing:** Although stable dehazing can be achieved, identifying the optimal transmittance remains a significant challenge. To tackle this issue, we initially employ the Atmospheric Scattering Model (ASM) to elucidate observable characteristics of hazy images, particularly their low contrast. Specifically, we utilize Equation 7 to quantify contrast by calculating the variance within a local image patch:

$$F(x) = |x - \mathbb{E}[x]| \quad (7)$$

In this context,  $x$  denotes the coordinates of a pixel within a local patch, while  $\mathbb{E}[x]$  represents the mean value of that local patch. Assuming that the transmittance  $t(x)$  is smooth and approximately constant within a local patch, we can infer from Equation 8 that dehazed images exhibit enhanced contrast:

$$F(I) = t|J - \mathbb{E}[J]| = tF(J) < F(J) \quad (8)$$

This analysis indicates that the contrast of a dehazed image, denoted as  $F(I)$ , is typically greater than that of the hazy image, represented as  $F(J)$ . Furthermore, given that brightness can be systematically manipulated to regulate the degree of dehazing, we produce a sequence of dehazed images with varying levels of haze removal, referred to as multi-scale dehazed images  $D_k(x)$ . These images progressively exhibit darker tones, akin to a multi-exposure fusion mask. To construct the final dehazed image, we employ a fusion methodology based on the following formulation.

The  $k$ -th multi-scale dehazed image is defined as follows:

$$D_k(x) = \frac{I - A + At_k(x)}{t_k(x)} \quad (9)$$

where  $t_k(x)$  signifies the estimated transmittance for the  $k$ -th multi-scale dehazed image. To derive the final dehazed image  $O(x)$ , we assign pixel-wise weights  $W_k(x)$  to each multi-scale dehazed image, as articulated in Equation 10:

$$O(x) = \sum_k W_k(x) \cdot D_k(x) \quad (10)$$

In this context,  $x$  represents the pixel coordinates,  $D_k(x)$  is the  $k$ -th multi-scale dehazed image, and  $W_k(x)$  denotes the

corresponding pixel-wise weights. By substituting Equation 9 into Equation 10, we obtain:

$$O(x) = \frac{(I - A) \cdot \sum_{k=1}^m (W_k \cdot \prod_{n=1, n \neq k}^m t_n) + \prod_{k=1}^m t_k \cdot A}{\prod_{k=1}^m t_k} \quad (11)$$

Next, we consider an ideal transmittance  $T(x)$  that corresponds to an ideal dehazed image  $D(x)$ , expressed as:

$$D(x) = \frac{I - A + AT(x)}{T(x)} \quad (12)$$

It is apparent that the fused image also conforms to the ASM, with its effective transmittance  $T_f(x)$  defined by:

$$T_f(x) = \frac{\prod_{k=1}^m t_k}{\sum_{k=1}^m W_k \cdot \prod_{n=1, n \neq k}^m t_n} \quad (13)$$

The image enhancement framework of the proposed algorithm, which is based on the previously discussed theory, is depicted in Figure 3. In this figure, the estimation of transmittance is derived from the difference between the input image  $I(x)$  and the dehazed image  $D_k(x)$ , under the assumption that the atmospheric light value  $A$  remains constant, as per the prior assumption  $I(x) > O(x)$ . To further improve the accuracy of transmittance estimation, a fusion of multi-scale dehazed images is conducted, which effectively estimates the ideal transmittance by utilizing the statistical characteristics or perceptual priors of the enhanced image  $O(x)$ . Additionally, the perceptual prior of dehazed images, as articulated in Equation 8, provides essential guidance for the synthesis of the final image  $O(x)$ , thereby ensuring the retention of intricate details and enhancing the overall efficacy of the dehazing process.

## B. Network Structure

As illustrated in Fig 2, our proposed network comprises two subnetworks: ASM-Net and F-Net. ASM-Net is designed for multi-scale dehazing, while F-Net is responsible for fusing the multi-scale dehazed images. This section provides a comprehensive description of both components.

**ASM-Net:** ASM-Net is structured around three primary components: the Transmittance Estimation Network (TEN), the Transmittance Low Boundary (TLB) block, and the Atmospheric Scattering Model (ASM) block. The fundamental design principle of ASM-Net is to ensure that the network's inference adheres strictly to the Atmospheric Scattering Model, drawing inspiration from recent advancements in low-light image enhancement. Previous studies frequently employed the Retinex model as a loss function, incorporating priors related to environmental and reflectance components [24], [25]. In contrast, more contemporary approaches have integrated the Retinex model directly into the network architecture [26], [27], thereby ensuring that the output images align with the Retinex model without interference from other priors.

The TEN is a straightforward convolutional neural network that utilizes adaptive normalization [28] to estimate the transmittance for a total of  $G$  images, where  $G$  denotes the number of scales employed in the model for multi-scale dehazing. Subsequently, the ASM block guarantees that the network's

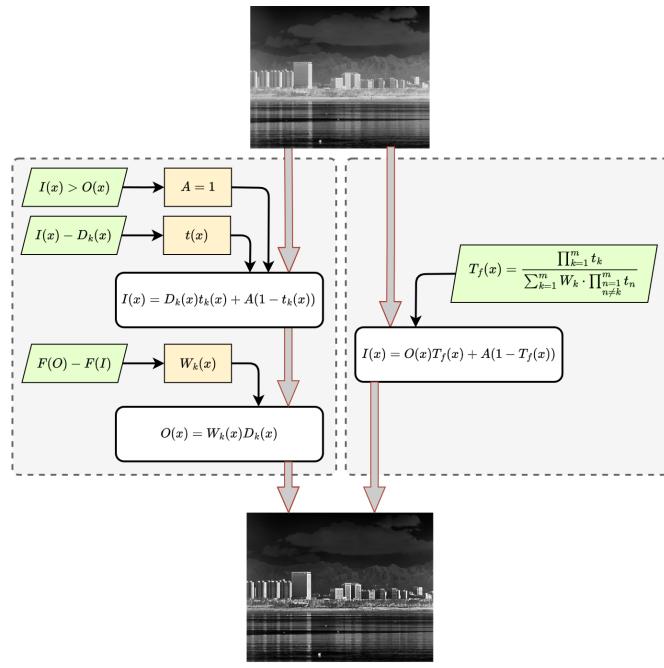


Fig. 3. The image enhancement framework has been constructed upon the theoretical principles outlined in this section. The green boxes illustrate the primary theoretical conclusions, while the yellow boxes signify essential parameters that have been derived from these theoretical insights. The rounded white boxes represent the fundamental components of the image enhancement framework. The pathway on the left pertains to the practical enhancement framework, whereas the pathway on the right encapsulates its theoretical foundation. By utilizing multi-scale dehazing and image fusion techniques, the estimation of the optimal transmittance can be effectively accomplished through the statistical characteristics of the enhanced image  $O(x)$ .

inference complies with Eq.3, while the TLB block ensures that the model adheres to the constraints of the ASM. Notably, the transmittance range is not confined to  $[0, 1]$ . A critical constraint is that the range of  $J(x)$  must remain within  $[0, 1]$ , leading to Eq.14, which illustrates that transmittance possesses a lower boundary:

$$\begin{cases} J(x) \geq 0 & \Rightarrow t(x) \geq \frac{A - I(x)}{A} \\ J(x) \leq 1 & \Rightarrow t(x) \geq \frac{I(x) - A}{1 - A} \end{cases} \quad (14)$$

We establish  $A$  as 1 based on theoretical analysis, resulting in the condition  $t(x) \geq 1 - I(x)$ . Consequently, we posit that the output of TEN, denoted as  $N_k(x)$ , resides within the range of  $[0, 1]$ . Therefore, we can estimate  $t(x)$  that satisfies the constraints using Eq.15:

$$t(x) = 1 - I(x) + N_k(x)I(x) \quad (15)$$

Finally, by substituting Eq.15 into Eq.3 and Eq.2, we derive Eq.16, which encapsulates the structure of the TLB block and ASM block.

$$\begin{cases} T_k(x) &= 1 - I(x) + N_k(x)I(x) \\ L_k(x) &= 1 - T_k(x) \\ D_k(x) &= \frac{I(x) - L_k(x)}{T_k(x)} \end{cases} \quad (16)$$

**F-Net:** The imposition of a constant value for parameter  $A$  results in a dehazed image  $D_k(x)$  that exhibits diminished pixel intensity across all pixels, leading to an underutilization

of the brightest pixels within the grayscale spectrum. Additionally, the darkest pixels are frequently associated with sky regions, where fine textures often manifest as noise. Consequently, our objective is to compress the grayscale range for these extreme pixel values, thereby reallocating more of the grayscale spectrum to mid-tone intensities, which ultimately enhances overall contrast. Previous research has explored gray transformation methods applicable to both visible and infrared images. Given the unique imaging characteristics of infrared images, the sigmoid function has garnered our interest [29]. We employ a four-segment curve to represent the sigmoid function [30] to ensure improved stability.

In the F-Net architecture, a gray transformation block is applied to each dehazed image across various scales, resulting in transformed images denoted as  $S_k(x)$ , as depicted in Figure 2. Thanks to many brilliant research in image fusion task such as HDR[31], [32], [33], [34], [35], we adapt MEF-Net [36] to compute the weight  $W_k(x)$  to facilitate the fusion of  $S_k(x)$  and  $D_k(x)$  according to the following equation:

$$O(x) = \sum_{k=1}^G W_k(x) \cdot D_k(x) + \sum_{k=1}^G W_{G+k}(x) \cdot S_k(x) \quad (17)$$

In this equation,  $O(x)$  represents the output image generated by our methodology.

### C. The Loss Function

In the context of self-supervised learning, we have developed three distinct loss functions for ASM-Net and two for F-Net. The training process is structured into two phases: first, ASM-Net is trained to generate multi-scale dehazed images, which are then utilized as input for F-Net, which is responsible for learning the fusion process to produce the final output image.

**ASM-Net:** The self-supervised loss functions for ASM-Net are formulated based on the theoretical framework presented in Section III-A, incorporating prior knowledge regarding transmittance. Mathematically, ASM-Net is optimized using the following loss functions:

$$L_{ASM} = w_{smooth}L_{smooth} + w_{darker}L_{darker} + L_{av} \quad (18)$$

In this equation,  $L_{smooth}$  denotes the smoothness loss associated with transmittance, with  $w_{smooth}$  representing its corresponding weight. Additionally,  $L_{av}$  enforces uniformity in the dehazing process, while  $L_{darker}$  controls the degree of dehazing.

The loss  $L_{smooth}$  is specifically designed to promote the smoothness of transmittance, as transmittance is intricately linked to scene depth, and ASM can be likened to a soft-matting problem. This is achieved by minimizing the following expression:

$$L_{smooth} = w_{tv}L_{TV} + w_{median}L_{median} \quad (19)$$

Here,  $L_{TV}$  and its weight  $w_{tv}$  are derived from a gradient-based approach, while  $L_{median}$  and its weight  $w_{median}$  are formulated from a spatial domain perspective. The inputs for these two

loss functions are both the transmittance  $t(x)$  estimated by ASM-Net.

$$L_{\text{median}} = \frac{1}{2m} \sum_{i=1}^m \left( t(x_i) - \frac{1}{|N(x_i)|} \sum_{y_i \in N(x_i)} t(y_i) \right)^2 \quad (20)$$

In this equation,  $N(x_i)$  represents the second-order neighborhood centered at  $x_i$ , and  $m$  denotes the total number of pixels in the image.

$$L_{\text{TV}} = \frac{1}{N} \sum_{i=1}^N (|\nabla_x t(x)| + |\nabla_y t(x)|)^2 \quad (21)$$

In this equation,  $\nabla_x t(x)$  and  $\nabla_y t(x)$  signify the gradients in the horizontal and vertical directions, respectively, while  $N$  represents the total number of pixels in the image. The darkening loss,  $L_{\text{darker}}$ , serves to limit the degree to which the output image is permitted to be darker than the input image. According to the derivation presented in Equation 6, regulating the difference  $I(x) - J(x)$  is tantamount to regulating  $t(x)$ , given that  $A$  is a constant and  $I(x)$  denotes the input image. For the  $G$  output images, the darkening loss is formulated as follows:

$$L_{\text{darker}} = \sum_{i=1}^G L_{\text{dark}_i} \quad (22)$$

where

$$L_{\text{dark}_i} = \frac{1}{N} \sum_{j=1}^N \left| \frac{\mathbb{E}[X]}{G+1} \cdot i + Y_j - X_j \right| \quad (23)$$

In these equations,  $G$  indicates the number of scales,  $N$  denotes the number of pixels,  $X$  represents the input image processed by ASM-Net,  $Y$  is the output image generated by ASM-Net, and  $\mathbb{E}[X]$  signifies the mean value of  $X$ . Furthermore, the uniformity loss,  $L_{\text{av}}$ , is designed to ensure consistency in dehazing intensity across various regions of the image. It is defined as follows:

$$L_{\text{av}} = \frac{1}{N} \sum_{i=1}^N \sigma(Y_i - X_i) \quad (24)$$

In this context,  $\sigma$  represents the variance.

**F-Net:** In the context of F-Net, we have modified the MEF-SSIM-based[37] loss function  $L_{\text{mef}}$  specifically for infrared imagery and have introduced a fuzziness-based loss function  $L_{\text{fuzzy}}$ . The comprehensive loss function for F-Net is articulated as follows:

$$L_{\text{fusion}} = w_{\text{fuzzy}} \cdot L_{\text{fuzzy}} + L_{\text{mef}} \quad (25)$$

In this equation,  $w_{\text{fuzzy}}$  denotes the weight assigned to  $L_{\text{fuzzy}}$ . The term  $L_{\text{mef}}$  is designed to enhance the output contrast, leveraging the prior established in Equation (8). The Structural Similarity Index Measure (SSIM) value, denoted as  $S(x, k)$ , for a candidate image  $y_k$  is computed by  $S(x, k) = \frac{2\sigma_{y_k y_f} + C}{\sigma_{y_k}^2 + \sigma_{y_f}^2 + C'}$ . In this context,  $x$  represents the coordinates of the image,  $y_k$  signifies the  $k$ -th candidate image,  $y_f$  indicates

the output from the network, while  $\sigma_y^2$  and  $\sigma_{y_k y_f}$  represent the variance and covariance of  $y_k$  and  $y_f$ , respectively, calculated over a specified patch.

To effectively guide the network, we select the patch exhibiting the highest contrast, which is quantified as follows:

$$c_k = |\mathbf{y}_k - \mu_{\mathbf{y}_k}|$$

Here,  $\mathbf{y}_k$  refers to the input image,  $\mu_{\mathbf{y}_k}$  is its mean, and  $c_k$  denotes the contrast. The ultimate score,  $S_c(x)$ , is defined as the SSIM score corresponding to the pixel whose patch demonstrates the highest contrast:

$$S_c(x) = \max_{k \in \{1, 2, \dots, 2G\}} S(x, k)$$

Subsequently,  $L_{\text{mef}}$  is expressed as:

$$L_{\text{mef}} = 1 - \mathbb{E}[S_c(x)] \quad (26)$$

The fuzziness loss  $L_{\text{fuzzy}}$  is formulated based on the linear index of fuzziness [6], which aims to mitigate the clustering of pixel intensities around the mid-range, thus averting a hazy visual effect. For the output image  $O$  generated by F-Net,  $L_{\text{fuzzy}}$  is computed as follows:

$$L_{\text{fuzzy}} = \frac{2}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W \min(Q(O_{i,j}), 1 - Q(O_{i,j})) \quad (27)$$

$$Q(O_{i,j}) = \sin \left( \frac{\pi}{2} \cdot \left( 1 - \frac{O_{i,j}}{M} \right) \right)$$

In these equations,  $O_{i,j}$  denotes the pixel value at the coordinates  $(i, j)$  of the image  $O$ , and  $M = \max O(x)$  represents the maximum pixel value within the image  $O$ .

#### IV. LGC DATASET

Numerous publicly accessible infrared datasets exist, including FLIR, MF3D [38], and KAIST [39]. Nevertheless, in comparison to the visible spectrum domain, the quantity and variety of datasets pertaining to infrared imagery remain inadequate. Additionally, long-distance observation constitutes a primary application of long-wave infrared (LWIR) technology. To mitigate these deficiencies, we present a novel dataset, referred to as *LGC*, which was captured utilizing the IRay Tech LGC6122 uncooled infrared core, a device also employed in practical applications.

A significant advantage of our dataset lies in the 72 mm focal length of the LGC6122, which facilitates the detection of human subjects at distances of up to 1.3 kilometers and vehicles at distances of 1.7 kilometers. In contrast, the FLIR dataset features a focal length of 13 mm, while the KAIST dataset has a focal length of 7.5 mm. The LGC6122 operates at a standard resolution of  $640 \times 512$  pixels and functions within the  $8-12 \mu\text{m}$  wavelength range, characteristic of typical LWIR detectors.

The infrared images included in our dataset were captured at Kunming Pool Qixi Park in Xi'an, Shaanxi, China. Data collection occurred during nighttime and afternoon hours in both spring and winter seasons. The scenes represented in

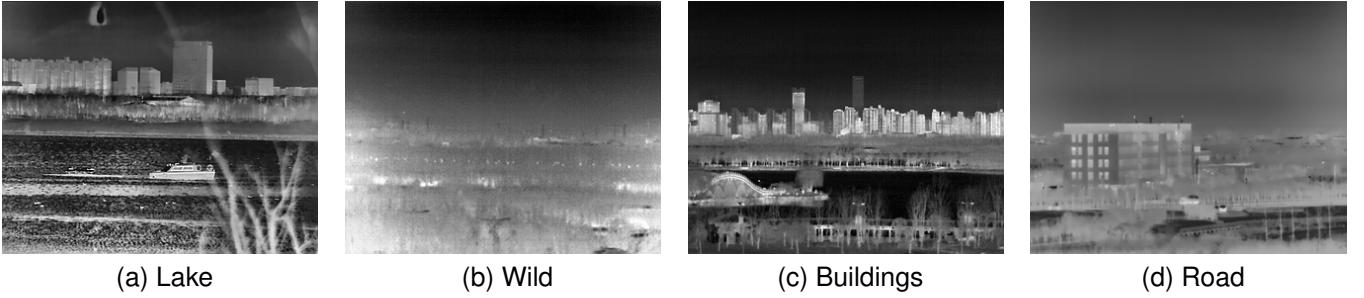


Fig. 4. Typical images of the LGC dataset. The dataset covers extensive scenarios with various environments, illumination, season, and weather. (a) Lake. (b) Wild. (c) Buildings. (d) Road.

TABLE I  
COMPARISON WITH SOTA. BOLD REPRESENTS THE BEST, UNDERLINED REPRESENTS THE SECOND BEST.

<b>Metric</b>	<b>BRISQUE</b> ( $\downarrow$ )			<b>PIQE</b> ( $\downarrow$ )			<b>FADE</b> ( $\downarrow$ )			$\eta$ ( $\downarrow$ )		
	<b>Dataset</b>	<b>LGC</b>	<b>FLIR</b>	<b>MF3D</b>	<b>LGC</b>	<b>FLIR</b>	<b>MF3D</b>	<b>LGC</b>	<b>FLIR</b>	<b>MF3D</b>	<b>LGC</b>	<b>FLIR</b>
LiY et al.[11]	<u>25.434</u>	22.996	<u>24.183</u>	33.194	29.681	<u>35.28</u>	0.636	0.736	1.134	<u>0.221</u>	0.467	0.239
MMFF[22]	30.805	21.621	24.928	35.836	21.964	39.832	1.100	1.223	2.020	0.566	0.642	0.597
C2PNet[44]	26.710	21.426	36.319	<u>32.663</u>	<u>21.741</u>	51.581	0.655	0.975	1.201	0.226	0.484	0.262
DeHamer[45]	30.227	20.490	39.916	37.844	26.152	53.224	0.560	0.660	1.102	0.235	<u>0.445</u>	<u>0.234</u>
MITNet[46]	29.973	<u>32.677</u>	42.112	41.113	31.182	53.445	<b>0.473</b>	1.108	<b>0.514</b>	0.238	0.449	0.236
ZeroDCE[19]	31.772	22.475	36.188	35.689	22.885	47.411	1.330	1.222	2.628	0.704	0.740	0.729
Ours	<b>25.424</b>	<b>20.019</b>	<b>22.752</b>	<b>32.486</b>	<b>21.455</b>	<b>31.367</b>	0.541	<b>0.627</b>	1.060	<b>0.143</b>	<b>0.194</b>	<b>0.193</b>

our dataset encompass buildings, lakes, and trees observed from considerable distances, thereby offering a diverse array of challenging scenarios for infrared image enhancement. Given the labor-intensive process of identifying appropriate locations for long-distance target detection, our dataset comprises a total of 668 images. Representative examples of these images are illustrated in Figure 4.

## V. EXPERIMENTS

### A. Experimental Setting

**Implementation Details:** The present study was conducted utilizing the PyTorch framework on a single NVIDIA RTX 4090 GPU. The scale parameter  $G$  for the network was established at 6. The loss function parameters for ASM-Net were configured as follows:  $w_{\text{smooth}} = 100$ ,  $w_{\text{darker}} = 20$ ,  $w_{\text{tv}} = 1$ , and  $w_{\text{median}} = 5 \times 10^{-5}$ . For F-Net, the parameter  $w_{\text{fuzzy}}$  was set to 2. The upsampling window size for the guided filter was determined to be 3[40], while the calculation window size for  $L_{\text{mef}}$  was set to 17. The optimization process employed the ADAM optimizer, with learning rates fixed at  $1 \times 10^{-6}$  for ASM-Net and  $1 \times 10^{-3}$  for F-Net. A batch size of 4 was utilized throughout the training process. ASM-Net underwent training for a total of 200 epochs, followed by an additional 50 epochs dedicated to F-Net training across all datasets.

**Dataset:** The evaluation of our methodology was conducted using the LGC, FLIR, and MF3D datasets. These datasets were captured using infrared cameras produced by FLIR, IRay, and Hikvision, which are recognized as leading manufacturers in the field of infrared imaging. Due to the necessity for camera-specific tuning of image processing parameters, we established distinct training and testing datasets for each infrared dataset included in this study. Specifically, the training and testing

datasets for LGC comprised 200 and 468 images, respectively; for FLIR, they included 200 and 3517 images; and for MF3D, the datasets contained 600 and 3245 images.

**Evaluation Metrics:** A significant challenge in the enhancement of infrared images is the limited availability of high-quality datasets, which serves as a primary motivation for this research. To address this challenge, we employed four no-reference image quality assessment (IQA) metrics: BRISQUE [41], PIQE [42], FADE [43], and the linear index of fuzziness ( $\eta$ ). These metrics provide complementary insights into haze removal. Both BRISQUE and PIQE are grounded in natural scene statistics (NSS) and assume a reference distribution as the ground truth; however, they are primarily tailored for visible light images, which may result in domain discrepancies when applied to infrared images. FADE, which is also based on NSS, emphasizes fog density and has been adapted for infrared images as demonstrated by Fabian et al. The fuzziness-based metric  $\eta$  is specifically designed for infrared images, as elaborated in [6].

**Baseline:** In order to conduct a thorough comparative analysis, we assess our proposed methodology ASMF (Atmospheric Scatter Model-Based Fusion Network) against six distinct techniques, which are organized into two categories. The first category encompasses two infrared image enhancement methods: an enhanced version of the approach by Li Y et al. [11] and MMFF-Net [22]. The second category consists of four visible image dehazing techniques: C2PNet [44], DeHamer [45], MITNet [46], and ZeroDCE [19]. Notably, ZeroDCE and MMFF-Net are classified as self-supervised methods, whereas the remaining techniques employ supervised learning paradigms. The method proposed by Li Y et al. is representative of traditional approaches. For the

self-supervised techniques, we conduct training on the same dataset utilized in our study, while for the supervised methods, we leverage their officially released pretrained weights. All learning-based methodologies are implemented using publicly accessible code, with the exception of the approach by Li Y et al., which does not have available code. Consequently, we have developed an improved version of this method, enhancing the dehazing algorithm and substituting DPHE with CLAHE, thereby exemplifying a classic representation of conventional algorithms.

### B. Comparison with State-of-the-Art Methods

This section presents a comparative analysis of the proposed method against several leading approaches in the field. The evaluation is conducted from both quantitative and qualitative perspectives.

**Quantitative Comparison.** Due to the absence of ground truth images, the quantitative assessment is carried out using non-reference metrics across three distinct datasets. The results of this quantitative analysis are summarized in Table I. The proposed method ASMF demonstrates superior performance in BRISQUE, PIQE, and  $\eta$  metrics across all three datasets, indicating a clearer spatial structure and improvements in both local and global fuzziness. In terms of FADE, ASMF surpasses the competition on the FLIR dataset, while it ranks second to MITNet on both the MF3D and LGC datasets. However, as illustrated in Figures 5, 6, and 7, MITNet tends to generate overexposed images, resulting in the loss of numerous details. Overall, our method achieves the highest scores on the quantitative metrics, and further experiments will substantiate its superiority.

**Qualitative Comparison.** The qualitative assessment across the LGC, FLIR, and MF3D datasets is depicted in Figures 5, 6, and 7, respectively. It is evident that the algorithm proposed by Li et al. struggles with adaptive dehazing, likely due to its dependence on a near-fixed transmittance estimation pattern. MMFF-Net and ZeroDCE, which emphasize brightness, tend to introduce greater fuzziness. MITNet and DeHamer enhance the brightness of target regions for dehazing but neglect finer details. C2PNet is particularly sensitive to haziness in the sky and generally exhibits a more subdued dehazing effect. In contrast, our method demonstrates enhanced clarity, particularly in the details of vehicles, buildings, and streets. However, the images processed by our method in the MF3D dataset appear relatively dark, a consequence of the absence of corresponding contrast enhancement algorithms within the image signal processing (ISP) pipeline.

### C. Post-Processing

Infrared image enhancement techniques incorporated within cameras typically consist of three primary stages: non-uniform correction (NUC), digital detail enhancement (DDE), and contrast enhancement. The latter encompasses a range of methodologies, with dehazing often employed as an initial step for techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE). This approach is consistent with previous contrast enhancement methods based on the Atmospheric

Scattering Model (ASM), which integrate histogram equalization (HE)-based algorithms as essential components.

To evaluate the effectiveness of our methodology, we analyze the impact of incorporating HE-based algorithms as a post-processing step to improve performance. We have selected CLAHE as the post-processing technique following our method. The visual results are illustrated in Figure 8. The findings demonstrate that our method achieves enhanced performance when CLAHE is utilized as a post-processing step, particularly on the MF3D dataset. This improvement can be attributed to the effective enhancement of image detail contrast facilitated by our approach.

Furthermore, Table II presents the effects of post-processing based on quantitative metrics. It is evident that ASMF+CLAHE outperforms other methods in terms of BRISQUE and FADE scores, indicating a clearer spatial structure and a significant reduction in fog density. However, ASMF+CLAHE exhibits a slightly lower score on PIQE and  $\eta$ . This is attributable to the inherent noise present in infrared images, which is exacerbated by CLAHE, resulting in a decrease in PIQE. Specifically, the FLIR and LGC datasets demonstrate higher noise levels, leading to a decline in PIQE, whereas the MF3D dataset, characterized by relatively lower noise, shows an increase in PIQE. Furthermore,  $\eta$  discourages the clustering of pixel intensities around the mid-range, and the brightness enhancement induced by CLAHE slightly diminishes  $\eta$ .

In summary, there exists a minor trade-off that results in a darker image; however, this is effectively addressed through the application of appropriate post-processing techniques such as CLAHE, yielding visually satisfactory outcomes.

### D. Ablation Study

In order to evaluate the effectiveness of each critical component, we conducted a series of ablation experiments using the LGC dataset. Specifically, we investigated the impact of self-supervised losses, the gray transformation block (GTB) incorporated in F-Net, the TLB Block within ASM-Net, and the multi-scale dehazing process. Our findings are presented through both qualitative and quantitative comparisons in the ablation study.

Self-supervised loss functions integrate prior knowledge into the estimation process of the target network. This study specifically assesses the roles of  $L_{\text{smooth}}$  in ASM-Net and  $L_{\text{fuzzy}}$  in F-Net. A comparative analysis of Figures 9a, 9b, and 9f indicates that each loss function plays a significant role in haze removal and the enhancement of image clarity. In particular,  $L_{\text{smooth}}$  improves the efficacy of multi-scale dehazing, whereas  $L_{\text{fuzzy}}$  effectively directs F-Net in the fusion of multi-scale dehazed images.

Furthermore, the GTB module in F-Net offers an alternative methodology for modeling image degradation. A comparison of Figures 9c and 9f illustrates that the GTB module successfully reduces haze by emphasizing target regions while simultaneously compressing the grayscale range in both the darkest and brightest areas.

A significant characteristic of the Atmospheric Scatter Model-Based Fusion Network (ASMF) is its capacity to

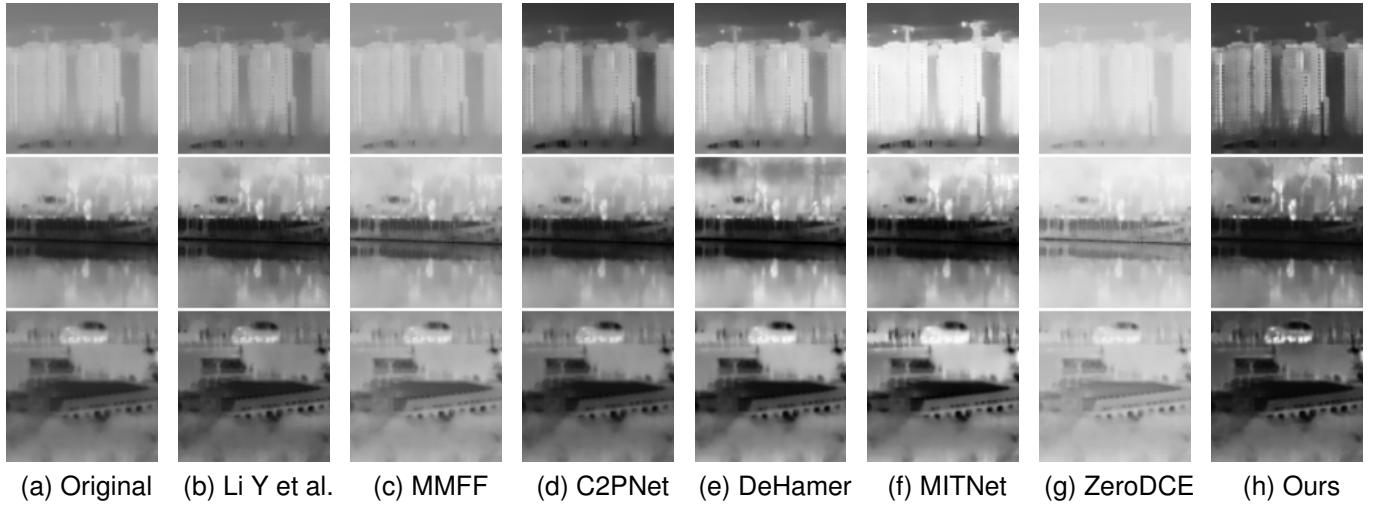


Fig. 5. Visual comparisons on infrared image patches sampled from the LGC dataset. (a) Original. (b) Li Y et al. (c) MMFF-Net. (d) C2PNet. (e) DeHamer. (f) MITNet. (g) ZeroDCE. (h) Ours.

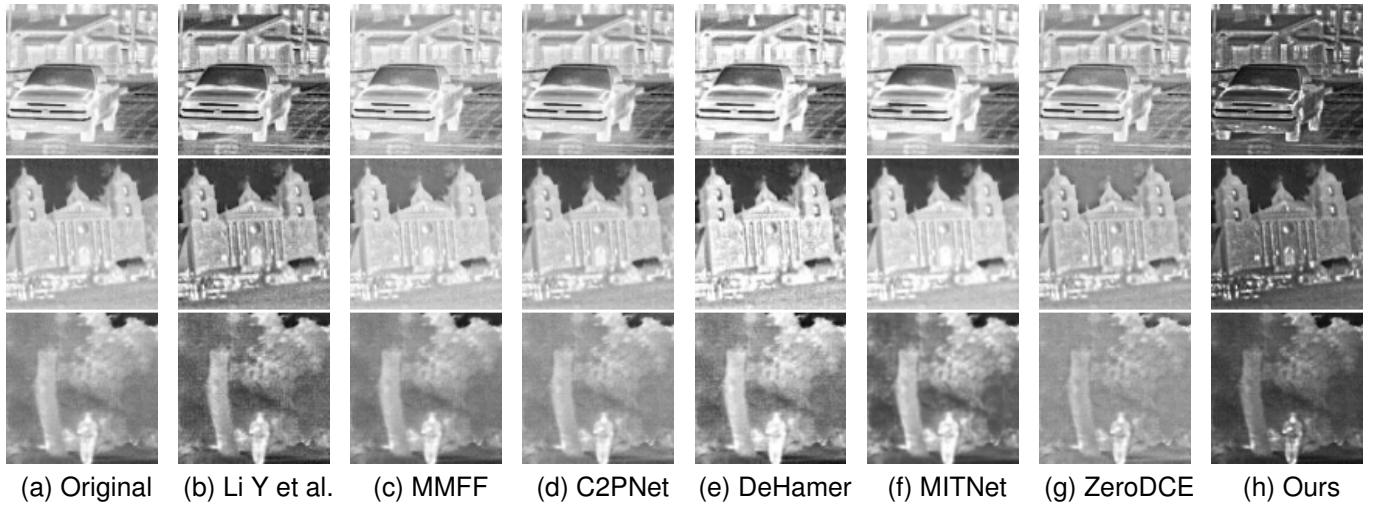


Fig. 6. Visual comparisons on infrared image patches sampled from the FLIR dataset. (a) Original. (b) Li Y et al. (c) MMFF-Net. (d) C2PNet. (e) DeHamer. (f) MITNet. (g) ZeroDCE. (h) Ours.

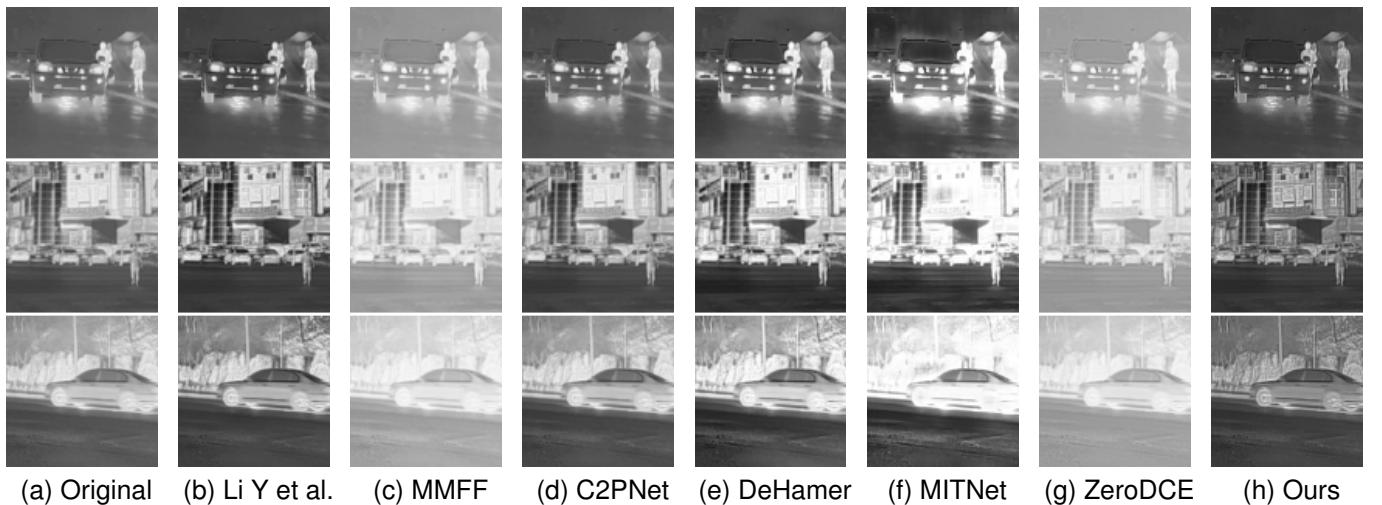


Fig. 7. Visual comparisons on infrared image patches sampled from the MF3D dataset. (a) Original. (b) Li Y et al. (c) MMFF-Net. (d) C2PNet. (e) DeHamer. (f) MITNet. (g) ZeroDCE. (h) Ours.

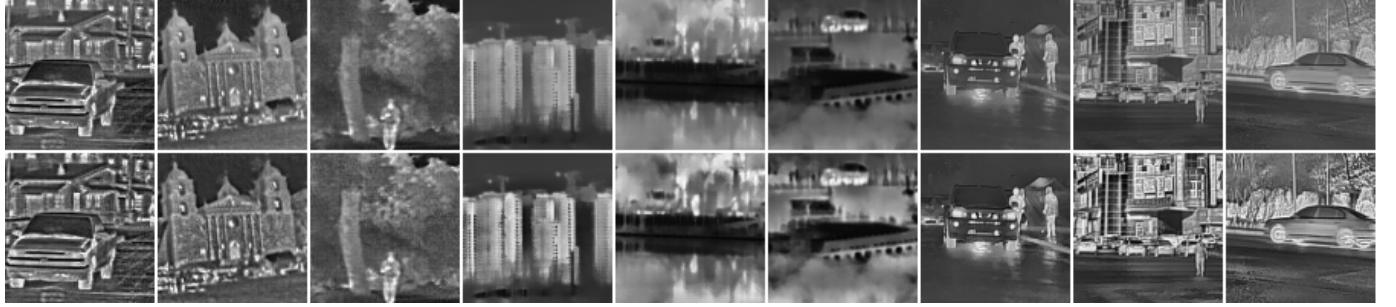


Fig. 8. This figure illustrates visual comparisons between our method and our method enhanced with CLAHE across three distinct datasets. The upper row presents the outcomes of our method, whereas the lower row showcases the results obtained with the incorporation of CLAHE.

TABLE II  
IMPACT OF POST-PROCESSING ON QUANTITATIVE METRICS. THE BOLDED ENTRIES INDICATE THE OPTIMAL RESULTS.

Metric	BRISQUE ( $\downarrow$ )			PIQE ( $\downarrow$ )			FADE ( $\downarrow$ )			$\eta$ ( $\downarrow$ )		
Dataset	LGC	FLIR	MF3D	LGC	FLIR	MF3D	LGC	FLIR	MF3D	LGC	FLIR	MF3D
ASMF	25.424	20.019	22.752	<b>32.486</b>	<b>21.455</b>	31.367	0.541	0.627	1.060	<b>0.143</b>	<b>0.194</b>	<b>0.193</b>
ASMF+CLAHE	<b>24.916</b>	<b>19.73</b>	<b>20.755</b>	32.849	22.088	<b>30.678</b>	<b>0.456</b>	<b>0.520</b>	<b>0.799</b>	0.172	0.234	0.299

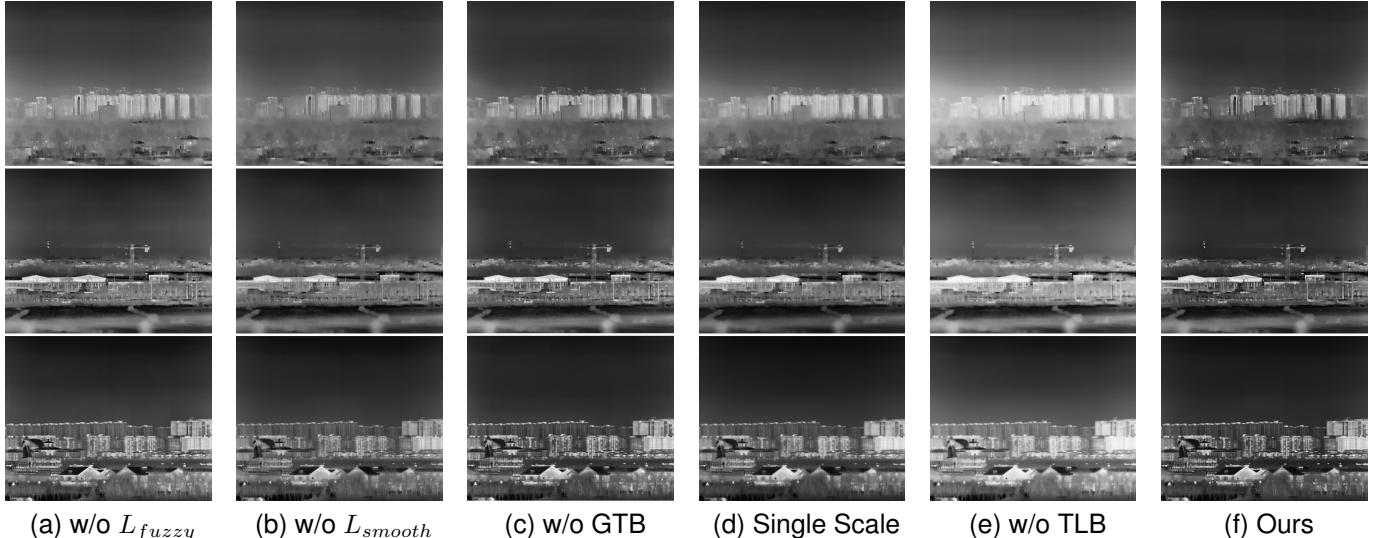


Fig. 9. Ablation Results. (a) w/o  $L_{fuzzy}$ . (b) w/o  $L_{smooth}$ . (c) w/o GTB. (d) Single Scale. (e) w/o TLB. (f) Ours.

employ image fusion techniques to estimate the optimal transmittance based on the statistical attributes of the target image. In Figures 9d and 9f, we present a comparison between single-scale dehazing without the fusion variant and ASMF, which illustrates that a fixed-scale dehazing approach lacks the necessary adaptability, particularly in localized regions. Conversely, ASMF exhibits a dynamic adjustment in the dehazing process, resulting in improved contrast in local details.

The TLB Block incorporates atmospheric scatter model (ASM) constraints, which are essential for achieving stable dehazing outcomes. An analysis of the intermediate results depicted in Figure 10 reveals that ASM-Net is unable to achieve stable dehazing in the absence of the TLB Block. Specifically, the multi-scale dehazed images that include the TLB Block (Figure 10a) exhibit a progressive darkening effect due to the influence of their inherent darken loss  $L_{darker}$ . In contrast, the multi-scale dehazed images that do not utilize the

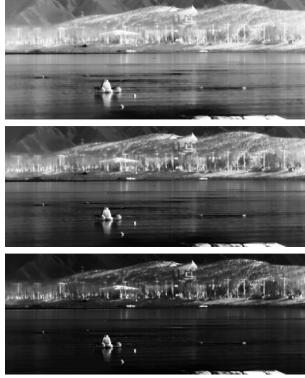
TLB Block (Figure 10b) maintain consistent brightness levels despite variations in darken losses. Consequently, without the TLB Block, ASMF-Net encounters difficulties in effectively propagating gradients, which results in a decline in visual quality, even performing inferiorly compared to single-scale dehazing. As previously discussed, all components have demonstrated their efficacy in qualitative assessments.

The scale parameter  $G$  is also a crucial factor. We present results for  $G$  values of 3, 6, and 8 in Figure 11. The results indicate that increasing  $G$  enhances the contrast of distant buildings; however, the rate of improvement diminishes at higher values. Considering the trade-off with training time, we set  $G = 6$  for this study.

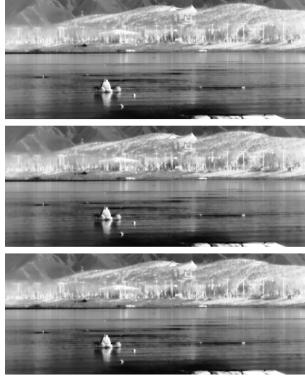
**Quantitative Comparison.** The quantitative findings from the ablation study are detailed in Table III, which aligns with the qualitative analysis. The ASMF method demonstrates superior performance in terms of BRISQUE, FADE, and  $\eta$ ,

TABLE III  
ABLATION STUDY. BOLD REPRESENTS THE BEST, UNDERLINED REPRESENTS THE SECOND BEST.

Metrics	w/o $L_{fuzzy}$	w/o $L_{smooth}$	w/o GTB	Single Scale	w/o TLB	G=3	Ours	G=8
BRISQUE ( $\downarrow$ )	25.627	26.281	25.441	28.879	27.741	25.730	<u>25.424</u>	<b>25.383</b>
PIQE ( $\downarrow$ )	32.658	<b>31.057</b>	<u>31.115</u>	34.723	32.754	32.419	32.486	32.040
FADE ( $\downarrow$ )	0.579	0.645	0.642	0.727	0.807	0.604	<u>0.541</u>	<b>0.508</b>
$\eta$ ( $\downarrow$ )	0.17104	0.17112	0.16922	0.14620	0.23928	0.16751	<u>0.14276</u>	<b>0.14123</b>



(a) The multi-scale dehazed images of ASMF



(b) The multi-scale dehazed images of w/o TLB

Fig. 10. Comparison of multi-scale dehazed images with and without the TLB Block in ASMF. (a) The multi-scale dehazed images with TLB. (b) The multi-scale dehazed images w/o TLB.



Fig. 11. Influence of scale parameter G. (a) Scale G = 3. (b) Scale G = 6(Ours). (c) Scale G = 8.

although it ranks marginally lower than the configurations without  $L_{smooth}$  and without GTB in PIQE. Each objective image quality assessment (IQA) metric evaluates image quality from a distinct perspective, and the overall results suggest that ASMF provides the most effective performance. Specifically, when compared to the configurations without  $L_{smooth}$  and without GTB, ASMF shows a reduction of 4.6% and 4.4% in PIQE, respectively. Conversely, it achieves significant enhancements of 16.1% and 15.7% in FADE, as well as 16.6% and 15.6% in  $\eta$ , indicating that all components contribute positively to the performance regarding fog density and clarity, albeit with a minor trade-off in increased noise amplification. Additionally, Table III reveals that as the value of  $G$  increases, the performance of ASMF also improves. In summary, the quantitative analysis further substantiates the subjective evaluations illustrated in Figures 9 and 11.

### E. Time complexity analysis

To assess the complexity of the proposed algorithm, we conducted a comparative analysis of the running times of various deep learning methods utilizing images from the LGC dataset. The experiments were performed on a workstation featuring an Intel i7-12700H CPU and a GTX 3060 GPU, with the average running times for each method presented in Table IV. Given that both MEF-Net [36] and TEN [28] are classified as lightweight networks, ASMF exhibited superior performance, underscoring its potential for real-time image processing applications. In terms of hardware implementation, a lightweight network can effectively approximate the Image Signal Processing (ISP) pipeline, functioning as an ISP network. This strategy negates the necessity for circuit redesign, as ISP processing can be accomplished by merely reloading the network weights. Our approach not only showcases low computational complexity but can also be further optimized through the ISP network, rendering it exceptionally suitable for real-time infrared image enhancement in camera devices.

## VI. CONCLUSION

This study introduces an innovative methodology for the enhancement of infrared images, which consists of two primary components: multi-scale dehazing and fusion. To support this approach, we have developed two self-supervised networks along with corresponding loss functions, which are elaborated upon in detail. Comprehensive experimental evaluations demonstrate the efficacy and superiority of our proposed method.

The current approach primarily focuses on the ASM for dehazing, thereby neglecting the exploration of alternative degradation models. Investigating these additional models may yield improved outcomes. Moreover, the utilization of 14-bit raw infrared images presents significant potential for enhancing results. Incorporating related techniques, such as Digital Detail Enhancement(DDE), represents another promising avenue for future research endeavors.

## REFERENCES

- [1] Sousa E, Vardasca R, Teixeira S, et al. A review on the application of medical infrared thermal imaging in hands[J]. *Infrared Physics & Technology*, 2017, 85: 315-323.
- [2] Takasawa S. Uncooled LWIR imaging: applications and market analysis[C]//Image Sensing Technologies: Materials, Devices, Systems, and Applications II. SPIE, 2015, 9481: 87-99.
- [3] Gao M, Bai Y, Liao H, et al. Contrast enhancement algorithm for infrared images based on atmospheric scattering model[J]. *Computers and Electrical Engineering*, 2024, 118: 109318.
- [4] Li S, Jin W, Li L, et al. An improved contrast enhancement algorithm for infrared images based on adaptive double plateaus histogram equalization[J]. *Infrared Physics & Technology*, 2018, 90: 164-174.

TABLE IV  
RUNNING TIME OF DIFFERENT METHODS. BOLD REPRESENTS THE BEST.

<b>Method</b>	<b>MMFF</b>	<b>C2PNet</b>	<b>DeHamer</b>	<b>MITNet</b>	<b>ZeroDCE</b>	<b>Ours</b>
Running time(ms)	75.351	1227.27	143.39	1682.28	54.895	<b>41.47</b>

- [5] Paul A, Sutradhar T, Bhattacharya P, et al. Infrared images enhancement using fuzzy dissimilarity histogram equalization[J]. Optik, 2021, 247: 167887.
- [6] Zhang H, Qian W, Wan M, et al. Infrared image enhancement algorithm using local entropy mapping histogram adaptive segmentation[J]. Infrared Physics & Technology, 2022, 120: 104000.
- [7] Qi Y, He R, Lin H. Novel infrared image enhancement technology based on the frequency compensation approach[J]. Infrared Physics & Technology, 2016, 76: 521-529.
- [8] Mello Roman J C, Vazquez Noguera J L, Legal-Ayala H, et al. Entropy and contrast enhancement of infrared thermal images using the multiscale top-hat transform[J]. Entropy, 2019, 21(3): 244.
- [9] Zheng L, Shi H, Gu M. Infrared traffic image enhancement algorithm based on dark channel prior and gamma correction[J]. Modern Physics Letters B, 2017, 31(19-21): 1740044.
- [10] Yan S, Zhu J, Yun K, et al. An infrared image dehazing method based on modified dark channel prior[C]//International Conference on Biometrics, Microelectronic Sensors, and Artificial Intelligence (BMSAI). SPIE, 2022, 12252: 132-138.
- [11] Li Y, Zhang Y, Geng A, et al. Infrared image enhancement based on atmospheric scattering model and histogram equalization[J]. Optics & Laser Technology, 2016, 83: 99-107.
- [12] Kuang X, Sui X, Liu Y, et al. Single infrared image enhancement using a deep convolutional neural network[J]. Neurocomputing, 2019, 332: 119-128.
- [13] Pang Z, Liu G, Li G, et al. An infrared image enhancement method via content and detail Two-Stream deep convolutional neural network[J]. Infrared Physics & Technology, 2023, 132: 104761.
- [14] Wang D, Lai R, Guan J. Target attention deep neural network for infrared image enhancement[J]. Infrared Physics & Technology, 2021, 115: 103690.
- [15] Marnissi M A, Fathallah A. GAN-based vision Transformer for high-quality thermal image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 817-825.
- [16] Tian C, Zheng M, Li B, et al. Perceptive self-supervised learning network for noisy image watermark removal[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(8): 7069-7079.
- [17] Tian C, Zheng M, Jiao T, et al. A self-supervised CNN for image watermark removal[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024.
- [18] Tian C, Xiao J, Zhang B, et al. A self-supervised network for image denoising and watermark removal[J]. Neural Networks, 2024, 174: 106218.
- [19] Li C, Guo C, Loy C C. Learning to enhance low-light image via zero-reference deep curve estimation[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(8): 4225-4238.
- [20] Li B, Gou Y, Gu S, et al. You only look yourself: Unsupervised and untrained single image dehazing neural network[J]. International Journal of Computer Vision, 2021, 129: 1754-1767.
- [21] Ram Prabhakar K, Sai Srikanth V, Venkatesh Babu R. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//Proceedings of the IEEE international conference on computer vision, 2017: 4714-4722.
- [22] Zhu G, Chen Y, Wang X, et al. MMFF-NET: Multi-layer and multi-scale feature fusion network for low-light infrared image enhancement[J]. Signal, Image and Video Processing, 2024, 18(2): 1089-1097.
- [23] Xiao Y, Zhang Z, Li Z. A Light-Weight Self-Supervised Infrared Image Perception Enhancement Method[J]. Electronics, 2024, 13(18): 3695.
- [24] Wei C, Wang W, Yang W, et al. Deep retinex decomposition for low-light enhancement[J]. arXiv preprint arXiv:1808.04560, 2018.
- [25] Fu Z, Yang Y, Tu X, et al. Learning a simple low-light image enhancer from paired low-light instances[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023: 22252-22261.
- [26] Ma L, Ma T, Liu R, et al. Toward fast, flexible, and robust low-light image enhancement[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2022: 5637-5646.
- [27] Liang Z, Li C, Zhou S, et al. Iterative prompt learning for unsupervised backlit image enhancement[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 8094-8103.
- [28] Chen Q, Xu J, Koltun V. Fast image processing with fully-convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2497-2506.
- [29] Zhong R, Fu Y, Song Y, et al. A fusion approach to infrared and visible images with Gabor filter and sigmoid function[J]. Infrared Physics & Technology, 2023, 131: 104696.
- [30] Daway H G, Daway E G, Kareem H H. Colour image enhancement by fuzzy logic based on sigmoid membership function[J]. International Journal of Intelligent Engineering and Systems, 2020, 13(5): 238-246.
- [31] Yan Q, Yang K, Hu T, et al. From dynamic to static: Stepwisely generate HDR image for ghost removal[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024.
- [32] Zhang Y, Zhang H, Nasrabadi N M, et al. Multi-metric learning for multi-sensor fusion based classification[J]. Information Fusion, 2013, 14(4): 431-440.
- [33] Yan Q, Wang H, Ma Y, et al. Uncertainty estimation in HDR imaging with Bayesian neural networks[J]. Pattern Recognition, 2024, 156: 110802.
- [34] Yan Q, Hu T, Sun Y, et al. Toward high-quality HDR deghosting with conditional diffusion models[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 34(5): 4011-4026.
- [35] Yan Q, Zhang L, Liu Y, et al. Deep HDR imaging via a non-local network[J]. IEEE Transactions on Image Processing, 2020, 29: 4308-4322.
- [36] Ma K, Duanmu Z, Zhu H, et al. Deep guided learning for fast multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2019, 29: 2808-2819.
- [37] Ma K, Zeng K, Wang Z. Perceptual quality assessment for multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3345-3356.
- [38] Liu J, Fan X, Huang Z, et al. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022: 5802-5811.
- [39] Hwang S, Park J, Kim N, et al. Multispectral pedestrian detection: Benchmark dataset and baseline[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 1037-1045.
- [40] Jiang T, Wang C, Li X, et al. Meflut: Unsupervised 1d lookup tables for multi-exposure image fusion[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 10542-10551.
- [41] Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer[J]. IEEE Signal processing letters, 2012, 20(3): 209-212.
- [42] Venkatanath N, Praneeth D, Bh M C, et al. Blind image quality evaluation using perception based features[C]//2015 twenty first national conference on communications (NCC). IEEE, 2015: 1-6.
- [43] Choi L K, You J, Bovik A C. Referenceless prediction of perceptual fog density and perceptual image defogging[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3888-3901.
- [44] Zheng Y, Zhan J, He S, et al. Curricular contrastive regularization for physics-aware single image dehazing[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023: 5785-5794.
- [45] Guo C L, Yan Q, Anwar S, et al. Image dehazing transformer with transmission-aware 3d position embedding[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022: 5812-5820.
- [46] Shen H, Zhao Z Q, Zhang Y, et al. Mutual information-driven triple interaction network for efficient image dehazing[C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023: 7-16.



**Luyuan Liu** (Graduate Student Member, IEEE) received the B.S. degree in Communication Engineering from Xidian University, Xi'an, China, in 2022, and the M.S. degree in Electronic and Information Engineering from the same university in 2025. He is about to join Raytron Technology as an Image Algorithm Engineer. His research interests include image enhancement, image generation, and deep learning accelerator design.



**Rui Song** received the B.S. degree in Electronic and Information Engineering from Xi'an University of Science and Technology, Xi'an, China, in 2022, and the M.S. degree in Information and Communication Engineering from Xidian University, Xi'an, China, in 2025. He is about to join Huawei Technologies Co., Ltd. His research interests include image processing, AI systems, and high-performance computing.



**Jiaojiao Li** received the B.S. degree in Electronic and Information Engineering from Xi'an University of Science and Technology, Xi'an, China, in 2022, and the M.S. degree in Information and Communication Engineering from Xidian University, Xi'an, China, in 2025. He is about to join Huawei Technologies Co., Ltd. His research interests include image processing, AI systems, and high-performance computing.



**Wenman Wang** received the B.S. degree in Electronic and Information Engineering from Xi'an University of Science and Technology, Xi'an, China, in 2022, and the M.S. degree in Information and Communication Engineering from Xidian University, Xi'an, China, in 2025. He is about to join Huawei Technologies Co., Ltd. His research interests include image processing, AI systems, and high-performance computing.



**Qingrong Wen** received the B.S. degree in Electronic and Information Engineering from Xi'an University of Science and Technology, Xi'an, China, in 2022, and the M.S. degree in Information and Communication Engineering from Xidian University, Xi'an, China, in 2025. He is about to join Huawei Technologies Co., Ltd. His research interests include image processing, AI systems, and high-performance computing.