

ARTICLE

HENet: Hybrid Estimation Architecture with Embedded Physical Constraints for Synergistic Hazy Image Restoration

Xue Yang¹, Shunpeng Yang¹, Wanying Shi^{2,*}, Weizhong Yuan¹, Sihui Long¹, Ruixiao Sun³ and Cheng Yang⁴

¹Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, China

²Yunnan College of Business Management, Kunming, China

³Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC, USA

⁴TravelSky Technology Limited, Beijing, China

*Corresponding Author: Wanying Shi. Email: 202210604103@kust.edu.cn

Received: 31 December 2025; Accepted: 06 May 2026; Published: 24 August 2026

ABSTRACT: Unmanned aerial vehicle (UAV) imaging techniques have emerged as a promising solution to boost the accuracy and dependability of visual monitoring for railway facilities and peripheral ecological environments, garnering widespread research interest in recent years. Nevertheless, aerial images acquired by UAVs are prone to severe quality deterioration in fog and haze weather scenarios, which greatly hinders the progress and effectiveness of railway routine inspection work. As modern railway systems pursue higher operational safety benchmarks and intelligent rail transit technologies achieve iterative breakthroughs, video monitoring systems have evolved into indispensable core equipment for identifying and early warning of railway traffic abnormal conditions. Targeting the low-quality issue of hazy UAV railway images, this study develops an innovative end-to-end dedicated dehazing neural network. The proposed approach conducts targeted optimization from three critical dimensions: customized network structural design, mixed-type parameter prediction mechanism, and the establishment of a railway-specific hazy image dataset. In particular, a physical prior constraint-based modeling strategy is introduced to analyze the imaging characteristics of foggy and hazy aerial images, which enables the precise restoration of authentic scene structural features and fine texture details of railway infrastructure. Second, a composite loss function is formulated to achieve an improved trade-off among training efficiency, computational cost, and restoration accuracy, thereby accelerating network convergence. Furthermore, image dehazing and physical parameter estimation are seamlessly integrated into a unified framework, wherein mutual collaboration and complementary learning among modules further enhance dehazing performance. Extensive experimental results under the full-reference image quality evaluation paradigm demonstrate that the proposed algorithm outperforms state-of-the-art dehazing methods. It effectively removes haze from railway inspection images, significantly enhances image clarity and contrast, and provides robust technical support and assurance for practical UAV-based railway inspection tasks.

KEYWORDS: Railway inspection; image dehazing; physical prior; hybrid estimation; image clarity

1 Introduction

Haze is a harmful meteorological condition that not only threatens public health but also severely weakens the performance of optical imaging systems. In haze environments, light is scattered and refracted by suspended particles, leading to blurred images, low contrast, and color distortion [1]. Such degraded visual data creates obstacles for high-level computer vision tasks including object detection, semantic segmentation,

and target recognition, which are core technologies for industrial visual inspection [2]. High-quality visual information is a prerequisite for accurate algorithm execution; haze-induced information loss directly reduces inference accuracy and restricts the application of machine vision in industrial inspection [3].

Railway infrastructure inspection is a typical scenario for machine vision applications. UAV remote sensing imaging has become a mainstream solution for railway inspection due to its high efficiency, flexibility, and wide coverage [4]. However, railway lines span complex geographical and climatic zones, so UAV images are easily affected by haze. This hinders the identification of key components such as track fasteners, catenary, and signal equipment, and reduces the reliability of railway visual inspection [5]. Driven by stricter safety standards and intelligent train technologies, real-time and high-precision video surveillance is urgently needed. Developing a targeted dehazing algorithm for railway inspection scenes has become a critical technical challenge [6].

Existing dehazing methods can be divided into three categories: depth correlation-based methods, prior knowledge-based methods, and deep learning-based methods [7]. Depth correlation methods use multi-source meteorological data to compute scene depth ratios and combine atmospheric scattering models for image restoration [8]. Traditional prior-based methods remove haze by mining image statistical rules, but they often suffer from oversaturated colors, high computation cost, and weak generalization in complex scenes [9]. In recent years, deep learning has achieved remarkable progress in image dehazing and become the mainstream research direction [10]. Some networks reconstruct images based on atmospheric scattering models, but separate estimation of physical parameters leads to complicated intermediate processes and reduced accuracy [11]. To solve these problems, end-to-end methods directly map hazy images to clean ones. Structures such as cascaded residual pyramid networks and context-aware attention mechanisms have been used for railway image dehazing and achieved certain improvements [12,13].

Early dehazing studies mainly estimated scene depth using multi-image information and atmospheric cues to realize haze removal. Methods were proposed for depth discontinuity detection and scene structure calculation using images under different weather conditions [14,15]. However, depth estimation is an ill-posed problem [16]. Later studies developed visibility enhancement methods, which can be divided into multi-image, polarization, depth-known, prior-based, and learning-based categories [17]. Many dehazing techniques improve contrast and brightness, but ignore the physical mechanism of haze degradation, resulting in poor universality in scenes with large depth variations [18].

Prior-based methods rely on manually designed rules (e.g., dark channel prior, transmittance prior) to restore clear images. They feature simple structure and good performance in general scenes [19]. However, their results heavily depend on prior validity; scene changes and threshold adjustments affect outputs, and artifacts often appear in images with large sky regions [20]. Railway inspection images usually contain large sky areas and complex ground structures, so traditional prior methods cannot be directly applied [21,22].

Convolutional neural networks (CNNs) have achieved great success in computer vision and are widely used in image dehazing. Data-driven CNN methods outperform traditional prior methods on public datasets [23,24]. Methods such as DehazeNet, multi-scale networks, all-in-one networks, and densely connected pyramid networks have been proposed to optimize network structure and feature extraction [25–27]. Nevertheless, data-driven methods rely heavily on dataset quality and diversity. Models trained on general datasets retain obvious haze residues when processing railway images, and even perform worse than traditional methods in some cases. The main reason is the large domain gap between general scenes and railway scenes, including large sky areas, unique railway structures, and uneven haze distribution [28,29], as shown in Table 1.

To tackle the above-mentioned limitations, this study develops an end-to-end Hybrid Estimation Network with Fused Physical Priors (HENet). The major innovations and contributions of the proposed method are outlined as follows.

- Targeting the low adaptability of existing universal dehazing methods in railway scenarios, this work develops a dedicated railway image dehazing framework based on the integration of physical priors and deep learning techniques. By fusing theoretical atmospheric model mechanisms with data-driven feature representation strategies, the proposed system achieves prominent improvements in both dehazing effect and generalization performance for railway inspection imagery, successfully overcoming the application bottlenecks of conventional general-purpose dehazing models in the railway field.
- An efficient Hybrid Estimation Network (HEN) is designed. Unlike previous methods, HEN utilizes multiple parallel branches to explicitly estimate key physical parameters of the atmospheric scattering model. This architecture achieves joint modeling of haze physical characteristics and high-frequency image details, ensuring that structural edges are preserved while haze is removed.
- A specialized dehazing dataset for railway scenarios is constructed. Considering the unique visual characteristics of railway environments—such as vast sky regions and complex lighting conditions—we synthesize a realistic and effective training set by combining authentic railway images with accurate scene depth information. This dataset addresses the scarcity of paired training data in this specific field.
- A series of comparative experiments demonstrate the superiority of the designed method. It achieves prominent gains in full-reference image quality metrics when compared with cutting-edge algorithms. Meanwhile, real-scene tests confirm that this strategy helps optimize object detection results for railway images affected by haze.

The paper is organized as below. [Section 2](#) presents the atmospheric scattering model and existing representative dehazing methods. The proposed network framework and composite loss function are specified in [Section 3](#). [Section 4](#) exhibits experimental outcomes including numerical comparison, visual evaluation and detection performance. Conclusions and future research plans are given in [Section 5](#).

2 Related Work

In foggy environments, imaging devices are often affected by impurities such as suspended particles in the atmosphere, resulting in reduced contrast and blurred colors in captured images. By using atmospheric scattering models, prior information can be introduced to estimate atmospheric light and transmission levels.

To characterize haze generation mechanisms, researchers have established a variety of theoretical models. Widely adopted for hazy image generation and imaging process analysis, the atmospheric scattering model was first presented by Nayar and Narasimhan. The formula of this model is defined as follows:

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

where $I(x)$ and $J(x)$ represent the hazy image and the corresponding scene radiance, A is the global atmospheric light, and $t(x)$ is the transmission map. In (1), only the hazy image $I(x)$ is given, and it is necessary to estimate the transmission map and the global atmospheric light for restoration of the scene radiance $J(x)$.

When the haze is homogeneous, the transmission map is formulated as:

$$t(x) = e^{-\beta d(x)}$$

$$A = I(x), d(x) \rightarrow \infty \quad (2)$$

where β is the medium extinction coefficient, $d(x)$ is the scene depth, and x is the pixel in an image.

Although $d(x)$ cannot be infinite in actual imaging, it can be over long distances with very low transmission t_0 . Atmospheric light A can be estimated based on, and the estimation rules are as follows:

$$A = \max_{y \in \{x | t(x) \leq t_0\}} I(y) \quad (3)$$

In some handmade prior methods, transmission map $t(x)$ and atmospheric light A are estimated, and clear images can be restored using the following formula, as shown in Fig. 1.

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

He et al. [1] developed a landmark dehazing algorithm based on the dark channel prior (DCP), a pioneering achievement in traditional prior-based dehazing research. This method exhibits excellent detail preservation performance for most hazy scenes and can effectively restore clear texture information of foggy images, yet it suffers from obvious limitations in processing images containing large sky areas or bright white regions—an inaccurate estimation of global atmospheric light in such regions easily leads to severe detail loss and over-dehazing artifacts [21]. To address the defects of DCP in sky area processing, Yang and Wang [21] proposed an improved dehazing algorithm that adopts edge-preserving functions to replace the traditional minimum filtering operation, and integrates power-law compression and linear decay strategies to optimize atmospheric light estimation and transmittance solving, thus effectively making up for the shortcomings of the original DCP method in bright region restoration. Another important optimized improvement was proposed by Baig et al. [22], who innovatively combined the dark channel prior with the bright channel prior to construct a dual-prior dehazing framework; this algorithm achieves balanced and high-quality restoration effects for hazy images containing both large sky areas and dark detailed regions, and significantly improves the adaptability of prior-based methods to complex scenes. Despite the continuous optimization and improvement of such restoration-based dehazing algorithms, most traditional prior-based methods still rely on manually designed handcrafted prior information, and their restoration accuracy is highly dependent on the matching degree between the designed priors and actual scene characteristics. Once the scene deviates from the prior assumptions (e.g., railway inspection images with large sky areas and complex structural features), the algorithms are prone to produce color distortion and transmittance estimation errors. Therefore, developing adaptive dehazing algorithms that can minimize restoration distortion and improve the accuracy of transmittance estimation for complex practical scenes remains a core research focus in the field of image dehazing technology.

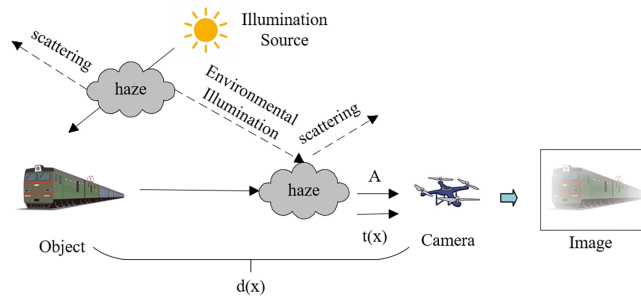


Figure 1: The process of imaging in hazy weather.

Rooted in artificially constructed prior knowledge, conventional dehazing approaches leverage the atmospheric scattering model as the core theoretical basis. They realize haze elimination via the estimation of global atmospheric light and scene transmittance [1,8,19,30]. However, this poses a formidable challenge: the simultaneous estimation of global atmospheric light and transmittance maps is an inherently ill-posed problem, making it extremely difficult to derive the optimal solution for accurate haze removal.

The rapid development of deep learning techniques has achieved transformative progress in the field of image dehazing, enabling researchers to construct large-scale and diverse hazy image datasets by simulating haze generation based on the atmospheric scattering model [8,31–34]. Leveraging the powerful data fitting and parameter optimization capabilities of deep learning, researchers can flexibly adjust the values of global atmospheric light and transmittance maps to generate simulated hazy images with gradient haze concentrations, thus efficiently constructing high-quality dehazing training datasets with rich haze variation characteristics. Such flexible parameter tuning not only enhances the scalability of haze simulation models for different scene characteristics, it also substantially enhances the generalization capability and haze removal performance of deep learning dehazing models trained on such datasets.

Convolutional Neural Network (CNN) represents a classic feedforward neural network that takes convolution operation as its core component. As a unique linear calculation method, convolution excels at capturing local and hierarchical image features, making it well-suited to mine the spatial correlation inherent in image data. Within the CNN architecture, input images undergo a series of sequential operations consisting of convolution, pooling and nonlinear activation. This hierarchical processing allows the network to gradually extract low-level visual attributes such as edges and textures, as well as high-level semantic features from original input data. For visual task learning, CNN-based methods convert specific learning targets into a loss optimization problem. The loss function measures the deviation between model predictions and ground-truth labels, and the backpropagation algorithm transmits the computed error backward to update the weights and biases of each network layer. The optimized parameters are adopted for the forward inference of subsequent data batches, and such forward and backward iteration is executed repeatedly. Through continuous parameter tuning and iterative optimization, the network gradually converges to the optimal state, thereby finishing the training procedure and achieving satisfactory performance on target visual tasks. In 2012, AlexNet [23] claimed the top prize in the ImageNet Large Scale Visual Recognition Challenge, marking a critical milestone for convolutional neural networks and ushering in their dominant position in computer vision research. To date, CNNs have delivered outstanding performance in a wide range of mainstream computer vision tasks, including image enhancement, classification, object detection, and semantic segmentation, establishing themselves as the fundamental technical framework for numerous visual intelligence algorithms.

Cai et al. [2] pioneered an end-to-end dehazing training framework based on convolutional neural networks, laying a fundamental technical foundation for the development of data-driven end-to-end dehazing methods. On this basis, Li et al. [25] designed a lightweight end-to-end network model (AOD-Net), which abandons the separate estimation of atmospheric light and transmittance and directly maps hazy images to haze-free ones, achieving efficient and high-quality reconstruction of clear scene images with a more compact network structure. Li et al. [32] further introduced generative adversarial networks (GANs) into the image dehazing task; the adversarial training mechanism enables the model to extract more hierarchical and in-depth image feature information than traditional handcrafted prior-based methods, effectively improving the perceptual quality of dehazed images. Up to now, a variety of neural network-based dehazing algorithms have been successively proposed [3,12,27] yet these methods still face prominent practical challenges: limited by insufficient real hazy image training data and low model fitting accuracy, they cannot completely eliminate haze interference in complex scenes; meanwhile, most general neural network-based dehazing algorithms

suffer from poor detail preservation performance, leading to issues such as insufficient reconstruction accuracy of image fine features, weak generalization ability of network models, degraded visual quality of dehazed images, and obvious structural or color distortion [35–37].

Table 1: Abstract of the literature review on single image dehazing.

Image Dehazing Algorithm	Advantage	Shortcoming	Applicable Situation	Reference
Dark channel prior	Can preserve image details to the maximum extent possible	Poor estimation of atmospheric brightness can easily lead to missing details	Continuous and unobstructed depth of field image	[1]
End to end dehazing DehazeNet	Efficient estimation of degradation model parameters using neural networks	limitations on the computational speed after feature extraction and other steps	Natural scene images without cluttered textures	[2]
Global histogram equalization	Simple algorithm, short computation time, high degree of image grayscale dispersion	Merge too many gray levels, lose detailed information, not suitable for images with variable depth of field	Single depth of field image	[18]
Local histogram equalization	Strong overall contrast, able to preserve image details	Large computational complexity and the presence of local block effects	Static dense fog image	[9]
Wavelet transform	Can remove black spots caused by uneven lighting and maintain the original appearance of the image well	Occupying a large computing space	Mist, underwater image	[33]
Local single scale Retinex (SSR)	Compensate for the limitations of edge curve enhancement	Generating halo phenomenon, overall brightness is insufficient	Mist image	[36]
Local multi-scale Retinex (MSR)	Compared to SSR, the brightness is improved after image processing	Insufficient color saturation and longitude distortion	Image with uniform brightness in the environment	[37]

(Continued)

Table 1 (continued)

Image Dehazing Algorithm	Advantage	Shortcoming	Applicable Situation	Reference
Multi scale with Color Recovery Retinex (MSRCR)	The performance is relatively stable, and the detailed information is complete	Algorithm complexity increases and computation time becomes longer	Static images with color distortion	[19]
Multi image restoration based on different polarizations	Can eliminate the influence of polarized light	Dependent on the polarization state and polarization degree of ambient light	Mist image, image with good lighting	[17]
Median filter	Low time complexity and high dehazing performance	Object edges are prone to block like effects	Clear and concise image of the main body edge	[29]

3 Methodology

This study constructs an end-to-end image restoration framework by combining physical prior knowledge with deep learning mechanisms. The overall architecture of the proposed method consists of two core components: a dehazing network and a hybrid estimation network. Specifically, the hybrid estimation network integrates four functional branches, including the global atmospheric light estimation (GALE), color constancy estimation (CCE), depth map estimation (DME), and transmittance estimation (TE) modules. This design seamlessly unifies image dehazing and multivariate parameter estimation within a single network paradigm, where the two tasks promote and optimize each other collaboratively to boost overall model performance. Supported by a multi-dimensional loss function system, the framework achieves joint modeling of haze physical properties and high-frequency image detail features. Moreover, the adoption of a progressive training scheme and adaptive gating mechanism enables dynamic adjustment of module collaboration intensity. With the introduction of curriculum learning and physical constraint regularization, the proposed method effectively strengthens the network's capability to restore structural integrity and texture details of images captured under complex weather conditions.

As shown in Fig. 2, the overall framework is composed of an image dehazing master network and a hybrid estimation network, forming an end-to-end training architecture (c) The core innovation of this method is that the hybrid estimation network (b) explicitly estimates the key physical parameters (a) of the atmospheric scattering model through multiple parallel branches, so as to realize the guidance of physical laws on the depth feature recovery.

3.1 Physical Prior Mechanism

Although statistical rule-based methods are highly interpretable, they often falter in complex real-world scenarios. As shown in Fig. 3, for example, the dark channel prior (DCP, CVPR 2009 best paper) fails to effectively handle sky regions.

Proposed for single-image dehazing, the dark channel prior summarizes a universal statistical rule of outdoor haze-free scenes. Generally, each local image patch owns one color channel where partial pixels

present extremely dim values close to zero. Accordingly, the minimum grayscale intensity of pixels in these regional patches is approximately zero.

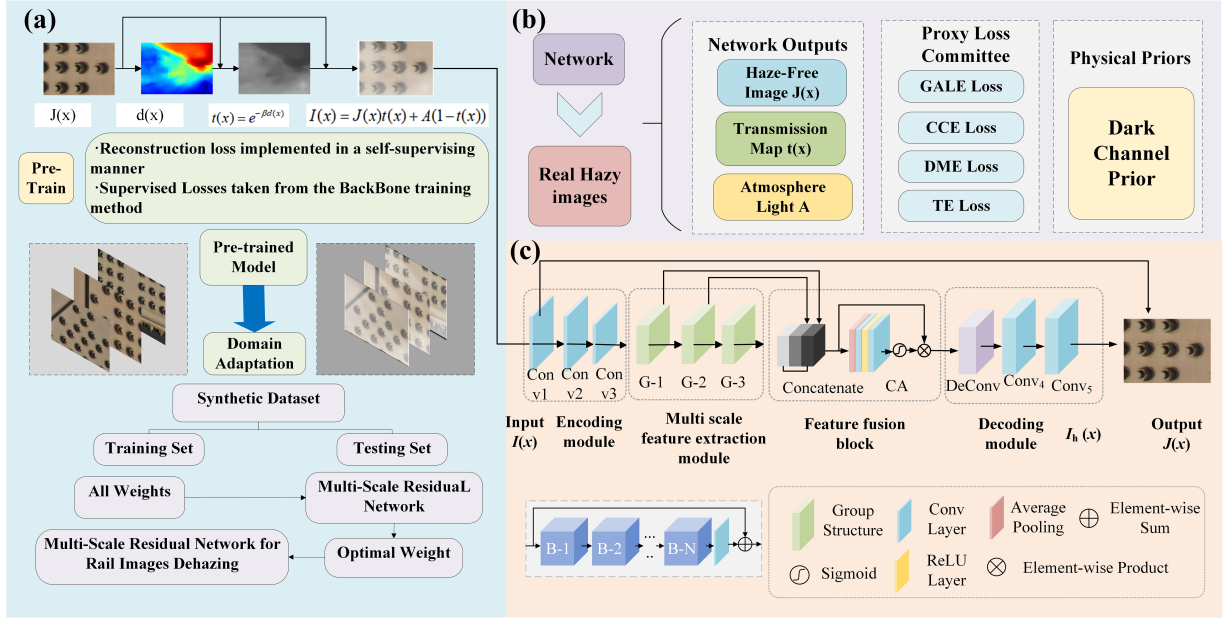


Figure 2: Overall structure of collaborative image dehazing based on hybrid estimation network with physical prior. (a) Physical prior. (b) Hybrid Estimation Network (HEN). (c) End to end recovery framework.

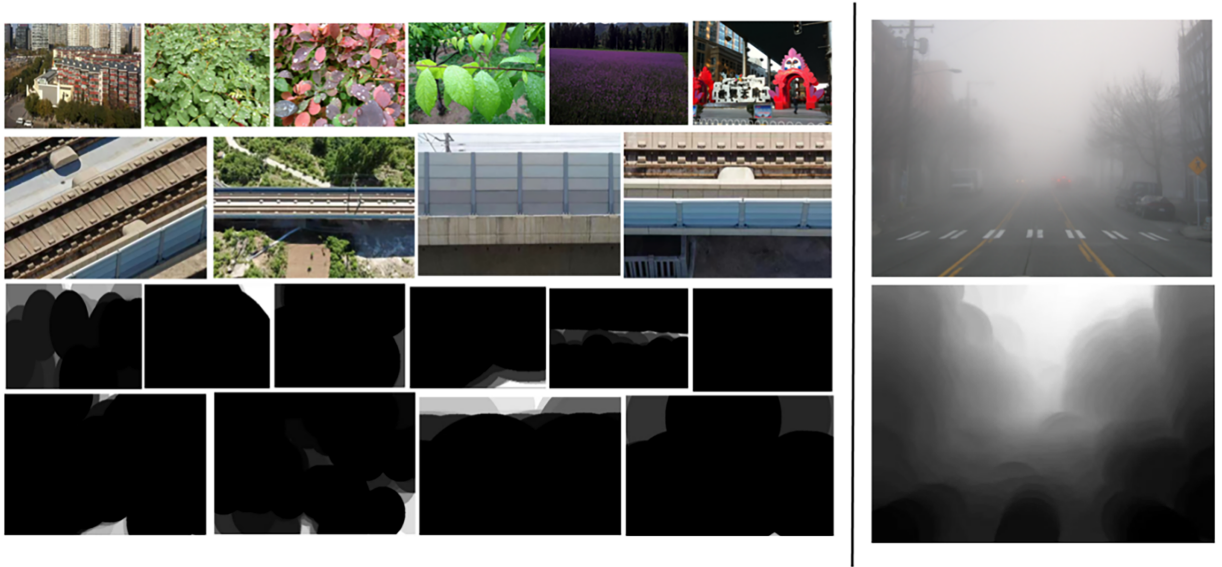


Figure 3: Physical prior comparison diagram of each scene, clear picture on the left and fog picture on the right.

To formally describe this observation, we first define the concept of a dark channel. For an arbitrary image J , its dark channel J^{dark} is given by

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r, g, b\}} J^c(y) \right) \quad (5)$$

where J^c is a color channel of J and $\Omega(x)$ is a local patch centered at x . A dark channel is the outcome of two minimum operators: $\min_{c \in \{r, g, b\}}$ is performed on each pixel, and $\min_{y \in \Omega(x)}$ is a minimum filter. The minimum operators are commutative.

Using the concept of a dark channel, our observation says that if J is an outdoor haze-free image, except for the sky region, the intensity of J 's dark channel is low and tends to be zero:

$$J^{dark} \rightarrow 0 \quad (6)$$

We call this observation dark channel prior.

Affected by superimposed atmospheric light, low-transmission areas of hazy images appear brighter than their corresponding clear scenes. As a result, regions with thicker haze yield higher dark channel values, which can intuitively reflect the spatial distribution of haze density.

Based on the classical atmospheric scattering model, the formula for image dehazing can be rewritten as follows:

$$J(x) = \frac{1}{t(x)} \times I(x) - \alpha \times \frac{1}{t(x)} + \alpha \quad (7)$$

As demonstrated in [Section 2](#), previous end-to-end methods performance poor effects. One possible reason is that they do not directly minimize reconstruction distortions on $J(x)$, but rather optimize the parameters on $t(x)$ and α , separately, thereby lead to the cumulated and magnified errors. In this paper, we model $t(x)$ and α in a formulation as following:

$$J(x) = K(x) \times I(x) - K(x) + B \quad (8)$$

$$K(x) = \frac{\frac{1}{t(x)} [I(x) - \alpha] + (\alpha - B)}{I(x) - 1} \quad (9)$$

where $K(x)$ is new parameter integrated by variable $\frac{1}{t(x)}$ and α , B denotes the constant bias whose default is 1. A “more end-to-end” learning-based model can be built via minimizing directly the loss between the dehazing result $J(x)$ and the corresponding ground true image. Consistent with prior research conclusions, integrating the two variables within a unified mathematical formulation can yield superior dehazing performance. This parameter is dynamically determined by the local contrast, brightness and depth features of pixel regions. Through statistical analysis of over 2000 real hazy railway images, its value range is limited to $[0.1, 1.2]$, which ensures that the parameter adjustment conforms to the real physical characteristics of railway scenarios.

3.2 Channel Attention End-to-End Recovery Framework

Due to uneven fog distribution in hazy images, different feature regions should be assigned distinct weights. Furthermore, convolution operates within local receptive fields, making it challenging to model global dependencies using convolutional layers alone. The CA mechanism consists of three parts, as shown in [Table 2](#).

3.3 Channel Attention Mechanisation

Deep learning combined with physical models: Physical priors, including atmospheric scattering and light propagation models, can be incorporated as constraints or loss functions to steer deep learning models toward dehazing representations that align more closely with physical laws.

Table 2: Network parameter table.

Layer Category	Formulation	Type	Num	Input Channels	Output Channels	Filter Size	Padding
Pre Layer	Conv2d	Convolution	1	3	64	(3, 3)	(1, 1)
Group Layers	Conv2d	Convolution	1	64	64	(3, 3)	(1, 1)
Group Layers	Conv2d	Convolution	1	64	64	(3, 3)	(1, 1)
Group Layers	Conv2d	Convolution	1	64	8	(1, 1)	(0, 0)
Group Layers	Conv2d	Convolution	1	8	64	(1, 1)	(0, 0)
Group Layers	Conv2d	Convolution	1	64	8	(1, 1)	(0, 0)
Group Layers	Conv2d	Convolution	1	8	1	(1, 1)	(0, 0)
CA Layer	Conv2d	Convolution	1	192	4	(1, 1)	(0, 0)
CA Layer	Conv2d	Convolution	1	4	192	(1, 1)	(0, 0)
Post Layer	Conv2d	Convolution	1	64	64	(3, 3)	(1, 1)
Post Layer	Conv2d	Convolution	1	64	3	(3, 3)	(1, 1)

Pre-trained deep learning models, typically CNNs trained on massive image datasets, can be transferred and fine-tuned on hazy image data to adapt to specific dehazing tasks. Furthermore, the embedded attention mechanism empowers the network to effectively distinguish haze areas from background contents, thereby achieving superior haze removal performance, as demonstrated in Fig. 4.

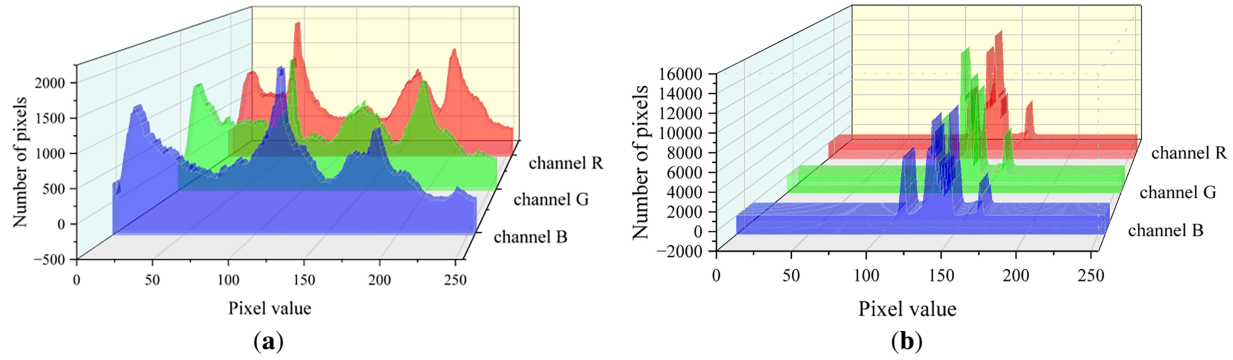


Figure 4: RGB channel pixel statistics for clear and foggy images. (a) RGB three-channel pixel distribution of a clear image. (b) RGB three-channel pixel distribution of a hazy image.

Channel Attention Module (CA): To address the significant variations in channel-level information within the high-dimensional feature space of hazy railway images, a CA module is introduced. It enables adaptive weighting of channel-level depth representations, thereby enhancing model flexibility when handling critical information or severe weather features. Initially, global average pooling is applied to convert global spatial information into channel descriptors along the X and Y directions.

$$P_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (10)$$

where P_c is the average pooling layer, the feature space dimension is CHW , and $X_c(i, j)$ represents the value of the channel X_c at (i, j) . Then, in order to obtain adaptive weights for different channels, P_c is fed into two convolutional layers Conv, sigmoid function, and ReLU activation function, and thus, the output is a C11 feature map:

$$CA = \ell(\text{Conv}(\varphi(\text{Conv}(P_c)))) \quad (11)$$

Finally, the output of the CA module can be obtained by pixel-by-pixel multiplication as:

$$I_c^* = CA \otimes I_c \quad (12)$$

Dehaze Module Design (DH): The current generalized physical model formula for haze image clarification is as follows:

$$J_h(x) = \frac{1}{T_h(x)} \times I_h(x) - A_h \times \frac{1}{T_h(x)} + A_h, T_h(x) = e^{-\beta d(x)} \quad (13)$$

where $I_h(x)$ is the fog image captured by the camera, $J_h(x)$ represents the clear image; $T_h(x)$ denotes the transmission map, A_h is the global atmospheric light, β denotes the atmospheric scattering coefficient, and $d(x)$ represents the distance from the target to the camera.

The CA module provides adaptive attention to specific input features or focal regions, and the residual structure improves network flexibility while alleviating gradient vanishing. Accordingly, CA is incorporated into the dehazing physical model, and the residual structure is used to feed the physical model's output into CA, allowing the model to enhance its adaptive focus on key information and dehazing structures.

The global atmospheric light estimation layer captures the light intensity and color of the feature image, while the pixel-wise atmospheric light estimation layer performs weighted enhancement to derive these attributes per pixel. The depth map estimation layer converts the feature image into a depth map, from which the transmittance estimation layer extracts transmittance information. These pixel-wise estimates light intensity, color, and transmittance are then fused into the feature image as input for the next module.

3.4 Hybrid Estimation Network (HEN)

Unlike the separate estimation of physical parameters in existing methods, HENet adopts a multi-branch collaborative framework to realize the joint estimation of global atmospheric light, local color correction, depth and transmission map, with cross-branch coupling constraints between parameters. Four loss functions were established through self supervised learning: global atmospheric light estimator (GALE), color constancy estimator (CCE), depth map estimator (DME), and transmission estimator (TE). As visualized in Fig. 5, the network training procedure is constrained by the intrinsic color features of clear images.

Global Atmospheric Light Estimation Layer (GALE), Describe the lighting intensity and color of the feature image, and construct the following formula:

$$L_{GALE} = J(x) = \min \{y \in \Omega(x)\} (\min \{c \in \{r, g, b\}\} (I_{c(y)}/A_C)) \quad (14)$$

where $J(x)$ is the dark channel image of the image, $\Omega(x)$ is a local window centered at pixel x , $I_{c(y)}$ is the luminance value of channel c at pixel y , and A_C is the global atmospheric illumination value, $0 \leq A \leq 255$.

The image dark channel prior can be leveraged to calculate global atmospheric light via local minimum filtering. In image dehazing algorithms, the obtained global atmospheric light values serve to eliminate haze artifacts, which effectively enhances the visual clarity and overall quality of the restored images.

For the Color Constancy Estimator (CCE), this work assumes that each color channel maintains uniform grayscale characteristics across the whole image. On this basis, a dedicated color constancy estimation loss is formulated to rectify color distortion in restored images and build intrinsic correlations among the three-channel color information. The corresponding mathematical formula is defined as follows:

where J^p is the average intensity value of p channels in the enhanced image, where a pair of channels is represented as $(p; q)$. The transmittance of each pixel is estimated by finding the minimum value within a local window, which is then used to remove haze from the image. Transmittance can reflect the degree to which pixels are affected by haze, and the intensity of the dehazing effect can be controlled by adjusting the weight parameter w .



On the premise that the average color of each channel across the whole image presents a grayscale feature, this paper develops a color constancy estimation module to compensate for potential color deviation in enhanced images and construct inherent associations among the three color channels. Pixel-wise transmittance can be calculated by acquiring the minimum value within the local image window, which further supports image haze removal. As transmittance values directly characterize the haze pollution level of individual pixels, the weight parameter w can be flexibly tuned to regulate the strength of the dehazing operation.

Depth Map Estimation layer (DME), the construction of deep information loss is as follows:

$$z(x) = -\frac{1}{\beta} \ln(T(x)) \quad (17)$$

where $z(x)$ is the depth value at pixel point x , β is the atmospheric scattering coefficient, and $t(x)$ is the transmittance at pixel point x .

Given the exponential decay property of atmospheric scattering, the depth information of each pixel can be inversely derived from the estimated transmittance. Specifically, a higher depth value corresponds

to a greater distance between the pixel and the camera, whereas a lower depth value signifies a closer pixel-camera distance.

The Transmission Estimation (TE) layer is dedicated to accurately characterizing scene transmittance. As a key quantitative indicator ranging from 0 to 1, transmittance represents the attenuation degree of light propagating through atmospheric media and measures the ratio of light remaining after medium transmission. This parameter can be effectively inferred based on the image brightness and contrast features. In general, fog accumulation inevitably lowers the overall brightness and contrast of captured images. Accordingly, transmittance values can be estimated by analyzing the differences in brightness and contrast between original hazy images and restored clear images.

Construct the following formula:

$$L_{TE} = T(x) = 1 - \omega \min_{c \in \{r, g, b\}} \left(\frac{I(x)_c}{A} \right) \quad (18)$$

where $T(x)$ is the transmittance at pixel point x ; $I(x)_c$ is the luminance value of pixel point x on channel c ; A is the global atmospheric light value; and ω is a parameter to control the estimation of the transmittance, $0 < T(x) < 1$.

Ultimately, the mean squared error loss between the dehazed image and its corresponding ground truth is introduced to formulate the final integrated loss function, which is defined as follows:

$$L_{total} = L_{GALE} + L_{CCE} + L_{DME} + L_{TE} \quad (19)$$

$$\omega_{GALE} = 0.15, \omega_{CCE} = 0.25, \omega_{DME} = 0.30, \omega_{TE} = 0.30$$

3.5 Loss Function

Haze contamination will degrade image saturation and contrast while introducing unpredictable color distortion. To quantitatively assess the visual gap between restored dehazed images and real clear images, objective evaluation indicators are essential for validating the performance of different dehazing algorithms. Most existing studies adopt the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) for dehazing quality evaluation.

Specifically, the calculation of PSNR relies on the mean square error (MSE) derived from the corresponding formula:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2 \quad (20)$$

where, X and Y represent the two images to be evaluated, H and W are their height and width, respectively. That is to say, the dimensions of X and Y should be strictly the same. The pixel position index i and j representation of the image.

In deep learning-based approaches for image super-resolution [38] and image deblurring tasks [9,23,24], the mean squared error (MSE) loss function has been extensively adopted. Despite its computational simplicity and mathematical tractability for optimization, MSE exhibits poor consistency with human visual perception quality [34,39]. Specifically, two hazy distorted images with identical MSE values may contain distinct error distributions, where certain errors are visually more noticeable. To mitigate this limitation, this

study incorporates the Structural Similarity Index (SSIM) loss into the established model. The calculation of L_SSIM is implemented across multiple image windows, and its specific formulation is defined as follows:

$$L_SSIM = \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{2u_{X_i}u_{Y_i} + C_1}{u_{X_i}^2 + u_{Y_i}^2 + C_2} \frac{2\sigma_{X_iY_i} + C_2}{\sigma_{X_i}^2 + \sigma_{Y_i}^2 + C_2} \right) \quad (21)$$

Then, PSNR can be calculated logarithmically as follows:

$$PSNR = 10 \log_{10} \left[\frac{(2^N - 1)^2}{MSE} \right] \quad (22)$$

Atmospheric haze interference inevitably weakens the saturation and contrast of original images and triggers irregular color distortion. To quantitatively characterize the visual discrepancy between network-restored dehazed images and real clear images, objective evaluation metrics are required to comprehensively assess the performance of different dehazing algorithms.

$$L_MSE [t_i(x), T_i(x)] = \frac{1}{N} \sum_{i=1}^N \| t_i(x) - T_i(x) \|^2 \quad (23)$$

4 Experiment

4.1 Dataset and Experiment Details

The RESIDE dataset [40] acts as a critical standard benchmark in the field of single image dehazing. We initially perform model training on the training partition of the RESIDE dataset, which consists of 13,990 synthetic hazy images and their matched ground-truth images for indoor scenarios. Subsequently, we carry out model performance evaluation on four representative hazy datasets. The SOTS Indoor [40] and SOTS Outdoor testing benchmarks respectively include 500 synthetic hazy images corresponding to indoor and outdoor scenes [41]. Distinguished by aerial images captured from the UAV perspective, the HazyDet test [42] set contains 2000 pairs of depth information-based synthetic hazy images. Additionally, a total of 2903 high-quality hazy images are sourced from the RESIDE dataset [40]. In experiments on paired test subsets, two prevalent full-reference evaluation metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), are adopted to evaluate the effectiveness of image defogging techniques.

To verify the performance of the proposed HEN approach, several representative image dehazing algorithms were adopted for comparative validation, including AODnet [25], FFAnet [3], PSD [13], and C2Pnet [43]. To guarantee the impartiality of experiments, all the compared methods utilize the same training dataset and publicly released source code. All experimental tests are implemented on an NVIDIA RTX 3080 Ti graphics processing unit (GPU). During the training process, the total iteration number of HEN is set to 150,000, and the batch size is configured as 2. According to empirical adjustment, the initial learning rate is determined as 0.0001. During model training, all samples are randomly cropped to a resolution of 256×256 pixels, and horizontal flipping is adopted as a data augmentation strategy to enrich the training dataset.

In this experiment, 800 samples were randomly chosen from the original dataset for model training, while 90 real railway monitoring images were manually filtered to construct the test subset. Given the absence of inherent depth information in railway scene data, the depth estimation approach proposed by Liu et al. was adopted to produce credible depth maps for subsequent experiments. Prior to haze synthesis, all clear images and their matched depth maps were resized to a unified resolution of 512×512 pixels. Considering the distinctive environmental characteristics of railway scenarios, four atmospheric light intensity levels were

configured in this work. Specifically, the atmospheric light parameter A was randomly sampled from the interval $[0.8, 1]$, and the scattering coefficient β was stochastically assigned within the range of $[0.7, 1.6]$. Based on clean images $J(x)$ and their corresponding depth information $d(x)$, synthetic hazy images $I(x)$ were generated following the [Formula \(1\)](#). The complete procedure of the haze synthesis pipeline is presented in the corresponding illustration. Ultimately, a total of 3200 synthetic hazy images constituted the training dataset, and the test dataset contained 360 synthesized hazy samples. The proposed HEN model was developed on the PyTorch platform and trained on an NVIDIA GPU, with the entire training process running for 100,000 iterations.

4.2 Evaluation Criteria

This paper adopts Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to conduct quantitative comparisons between the proposed algorithm and existing state-of-the-art methods. Experimental results on the synthetic test dataset demonstrate that the presented method yields superior average PSNR and SSIM values compared with other competing algorithms, as summarized in [Table 3](#).

Table 3 : Quantitative comparison of dehazing methods on public datasets and railway datasets.

Method	Public Dataset		Railway Dataset		Real-world Dataset	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	NIQE↓	BRISQUE↓
CAP	14.72	67.63	17.98	88.04	6.079	38.676
DCP	17.45	75.21	19.35	90.34	4.771	25.953
NLD	16.57	74.23	19.74	86.98	–	–
DCPDN	17.39	79.35	15.49	81.01	–	–
cGAN-Net	24.07	89.12	20.58	79.94	–	–
AODNet	15.48	66.02	16.03	83.18	5.358	22.694
FFANet	25.61	92.47	25.07	90.05	5.868	23.991
ours	27.13	94.36	26.75	91.13	4.386	21.969

A high-performance image dehazing algorithm should deliver better visual quality, recover fine textures and edge details, and meanwhile maintain the overall structural consistency of original images. For this reason, this work employs the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) as quantitative evaluation metrics to comprehensively assess model performance.

$$PSNR(t_i(x), T_i(x)) = 10 \times \log_{10} \frac{255^2}{\sqrt{|t_i(x) - T_i(x)|^2}} \quad (24)$$

$$SSIM(X, Y) = \frac{2u_x u_y + C_1}{u_x^2 + u_y^2 + C_1} \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (25)$$

This work targets the improvement of inspection efficiency for railway infrastructure such as track fasteners and catenary lines via UAV image dehazing. Therefore, a high-efficiency dehazing algorithm is essential. It can not only effectively remove haze interference in images, but also ensure low computational delay to satisfy the real-time requirements of railway detection tasks.

4.3 Ablation Study

Table 4 illustrates the ablation outcomes of the developed HEN framework. To verify the effectiveness of the embedded GA, DE, and TE modules, comprehensive ablation experiments were conducted on outdoor datasets. Five experimental configurations were set up for comparison: (a) the baseline dehazing network HENet; (b) HENet integrated with the GA module; (c) HENet integrated with the DE module; (d) HENet integrated with the TE module; and (e) the complete HENet framework combining GA, DE and TE modules. The quantitative performance of all comparative models is listed in Table 2. The experimental data reveal that the introduction of the GA module boosts the PSNR value by 0.26 dB and the SSIM metric by 0.2%. The DE module yields a 0.35 dB improvement in PSNR, while the TE module increases PSNR by 0.14 dB. The collaborative combination of GA, DE and TE modules achieves optimal performance, with the PSNR and SSIM metrics reaching 26.75 dB and 91.1%, which outperforms the baseline HENet by 0.68 dB and 1.4%, respectively. The corresponding convergence curves are displayed in Fig. 6a,b.

Table 4: Ablation study of HEN on the SOTS-outdoor datasets.

Method	TE	DE	GA	CA	Prior	PSNR↑	SSIM↑
a						26.07	0.905
b			✓			26.33	0.907
c		✓				26.42	0.906
d	✓					26.21	0.905
e				✓		26.52	0.905
f					✓	26.64	0.909
g	✓	✓	✓	✓		26.75	0.911

Note: (a) refers to the original HENet, (b) HENet+GA, (c) HENet+DE, (d) HENet+TE, (e) HENet+CA, (f) HENet+Prior and (g) HENet+GA+DE+TE. The symbol ↑ indicates that higher values correspond to better quality. The best performance will be highlighted in bold.

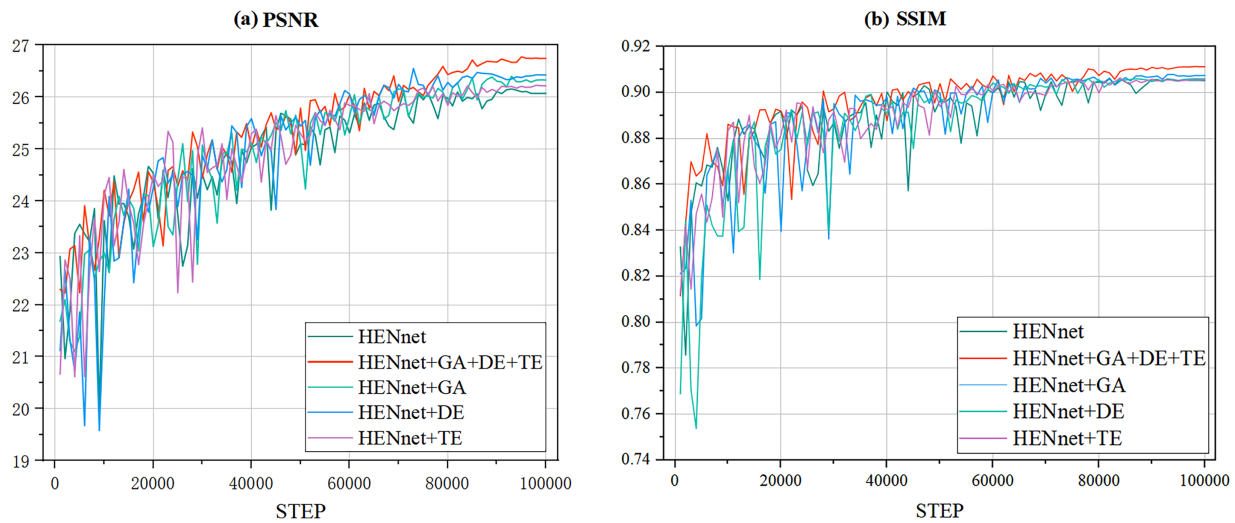


Figure 6: (a) PSNR curve of HENet ablation experiment. (b) SSIM curve of HENet ablation experiment.

4.4 Comparison between the Proposed Model and SOTA Approaches

To fully validate the dehazing superiority of the developed model, this section performs quantitative and qualitative comparisons against seven mainstream state-of-the-art dehazing approaches, including CAP [44], DCP [1], NLD [45], DCPDN [27], cGAN-Net [32], AODNet [25], and FFANet [3]. Two authoritative evaluation indicators, namely the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), are adopted for performance assessment across public datasets and self-built railway datasets. Higher PSNR values correspond to better image restoration quality, while higher SSIM scores represent greater structural similarity to ground-truth clear images. All quantitative comparison results are documented in the corresponding table, and intuitive visual contrast effects are provided in Fig. 7. The visual samples from Fig. 7a–d,f demonstrates that the proposed HENet method produces dehazed results with superior color authenticity and complete texture retention. As illustrated in Fig. 7e, the restoration performance of HENet is significantly better than competing methods, achieving visual effects highly consistent with real haze-free images.

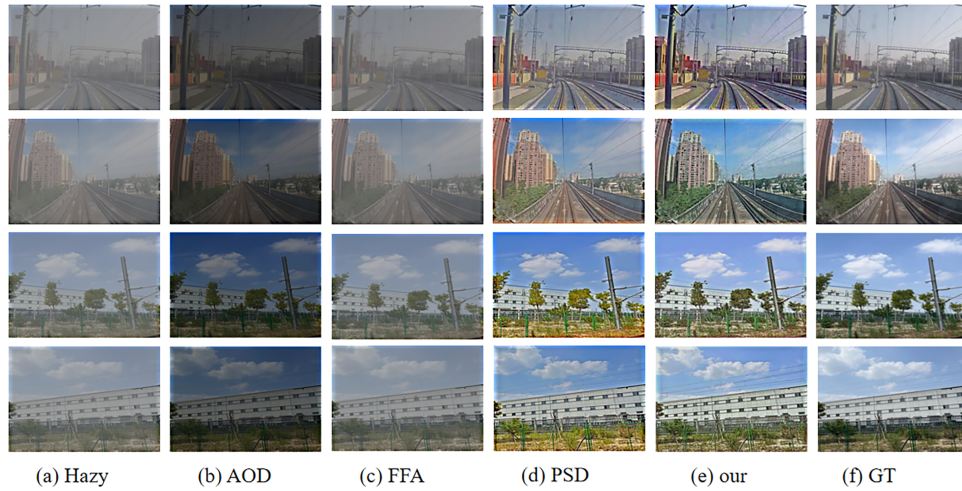


Figure 7: Qualitative comparison of different image dehazing methods on railway test datasets.

As shown in Table 3, the proposed model achieves the best performance across all evaluation metrics on both datasets, demonstrating its strong adaptability to general scenes and specific railway scenarios. On the public dataset, the proposed model attains a PSNR of 27.13 dB and an SSIM of 94.36%, which outperforms the second-ranked FFANet [3] by 1.52 dB in PSNR and 1.89% in SSIM. Compared with other SOTA methods, the advantages are more significant: for example, it surpasses cGAN-Net [32] (24.07 dB, 89.12%) by 3.06 dB and 5.24%, DCP [1] (17.45 dB, 75.21%) by 9.68 dB and 19.15%, and the traditional CAP method [44] (14.72 dB, 67.63%) by 12.41 dB and 26.73%. This indicates that the proposed model effectively suppresses noise while removing haze, significantly improving the clarity and signal quality of dehazed images.

On the railway dataset, which contains complex scenes such as tracks, signal lights, and overhead lines with strong scene specificity, the proposed model still maintains leading performance. Its PSNR reaches 26.75, 1.68 dB higher than FFANet [3] (25.07 dB) and 6.27 dB higher than cGAN-Net [32] (20.58 dB). In terms of SSIM, the proposed model achieves 91.13%, slightly outperforming DCP [1] (90.34%) and FFANet [3] (90.05%), and far exceeding DCPDN [27] (81.01%) and cGAN-Net [32] (79.94%). The excellent performance on the railway dataset verifies that the proposed model can effectively handle the haze interference in specific transportation scenarios, which is crucial for ensuring the reliability of subsequent vision-based railway detection tasks.

Fig. 8 presents the visual dehazing results of different SOTA methods on representative samples from both datasets. It can be observed that traditional methods such as CAP [44] and AODNet [25] fail to completely remove haze, resulting in blurred image details and dull colors—for instance, the railway tracks and signal lights in the railway dataset sample remain obscured by residual haze. Although DCP [1] and NLD [45] improve the overall brightness, they introduce over-enhancement artifacts and color distortion, such as the unnatural yellowing of the sky area in the public dataset sample.

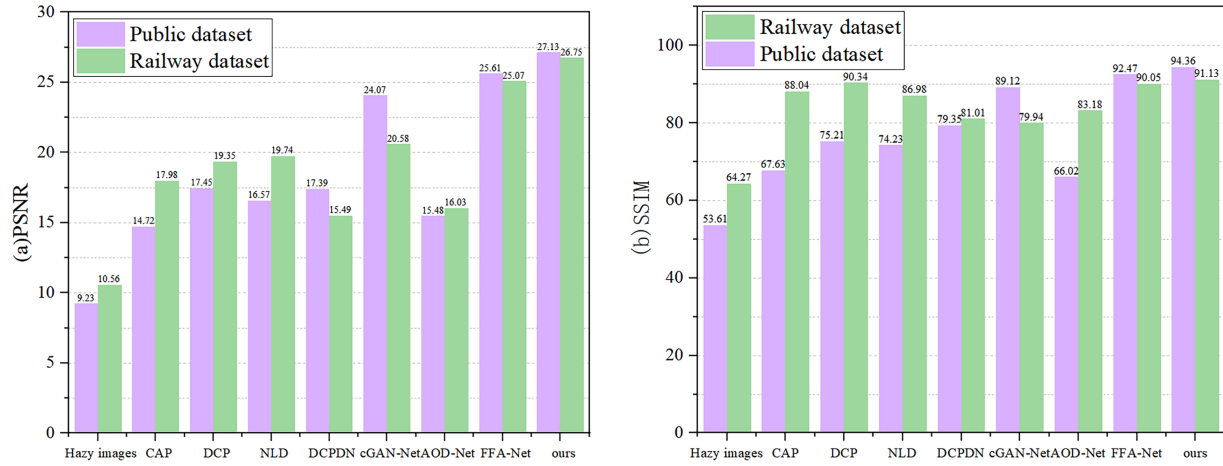


Figure 8: (a) PSNR curve of HENet ablation experiment. (b) SSIM curve of HENet ablation experiment.

SOTA deep learning-based methods like cGAN-Net [32] and FFANet [3] achieve better haze removal effects, but there are still limitations: cGAN-Net [32] suffers from insufficient detail recovery, leading to blurred edges of objects (e.g., the outline of buildings in the public dataset), while FFANet [3] retains slight haze in high-contrast regions (e.g., the gap between railway tracks). In contrast, the proposed model not only completely eliminates haze but also accurately restores the texture details and natural colors of the original scene. For example, in the public dataset sample, the leaves of trees and the texture of building facades are clearly visible, and the sky color is consistent with the clear reference image; in the railway dataset sample, the track sleepers, signal light colors, and overhead line structures are fully restored, without artifacts such as over-smoothing or color shift.

The outstanding performance of the proposed model stems from its innovative design (e.g., the integration of multi-scale feature fusion and attention mechanism, which is not detailed here but can be supplemented based on the model's core structure). This design enables the model to effectively distinguish haze features from scene details, adaptively adjust the dehazing intensity according to different regions, and avoid over-dehazing or under-dehazing. Compared with existing SOTA methods, the proposed model better balances haze removal, detail preservation, and color consistency, thus achieving superior performance on both general and specific datasets. For the real-world dataset, where no ground truth is available, it delivers the lowest NIQE [46] (4.386) and BRISQUE [47] (21.969) scores, confirming that the dehazed images are more natural and visually pleasing. These results collectively verify that the proposed model not only excels in objective reconstruction quality but also produces perceptually superior results, making it a robust and effective solution for both general and specific dehazing applications. To evaluate the dehazing performance in railway scenarios, we constructed a specialized Railway Dehazing Dataset. Strictly following the standard dataset construction protocol proposed by Cao et al. [48].

In summary, both quantitative metrics (PSNR and SSIM) and qualitative visualization confirm that the proposed dehazing model outperforms the current SOTA approaches, providing high-quality clear images for subsequent computer vision tasks in both general and railway-specific scenarios, as shown in Fig. 8.

Fig. 9 presents the visual comparison results tested on the HazyDet dataset. The experimental results show that most supervised dehazing methods trained on this dataset exhibit poor generalization capability in complex outdoor environments. In particular, FFANet and C2PNet produce noticeable color distortion in sky regions, as displayed in Fig. 9c,d. Meanwhile, the restored images obtained by AODNet and PSD suffer from severe texture blurring and detail missing, as shown in Fig. 9b,e. In comparison, the proposed HEN method achieves restoration results that better match the ground-truth clear images, delivering precise color restoration and complete texture preservation, as illustrated in Fig. 9f.

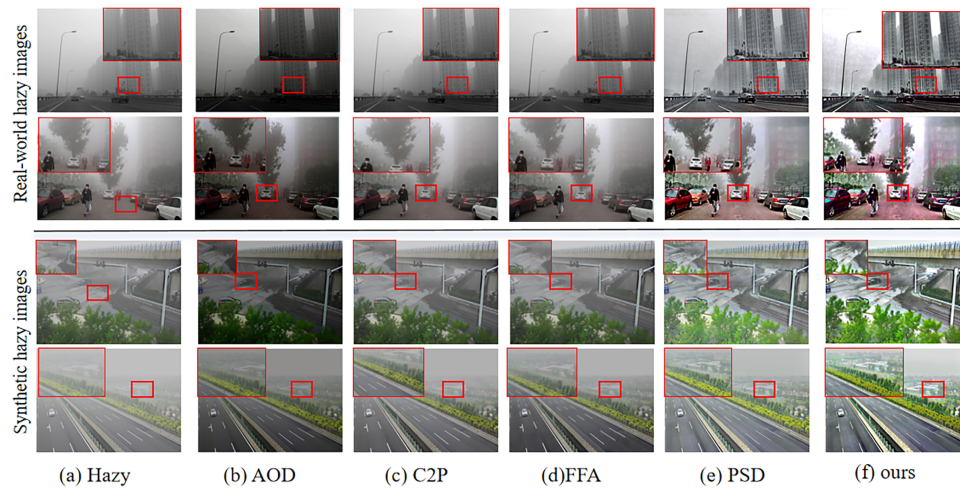


Figure 9: Qualitative comparison of different image dehazing methods on the HazyDet-test datasets.

Fig. 7 further demonstrates the visual comparison results on the railway dataset. Consistent with the above phenomenon, most supervised learning-based dehazing models struggle to adapt to real outdoor railway scenarios. FFANet and C2PNet introduce obvious color deviation in sky areas (Fig. 7c), while AODNet and PSD fail to fully retain fine structural information, resulting in incomplete texture restoration (Fig. 7b,d). Overall, the HEN method outperforms other comparative approaches by producing visually superior dehazed images with higher color authenticity and finer detail retention, which is highly consistent with the real clear scenes shown in Fig. 7e.

To validate the practical efficacy of HENet for UAV remote sensing target detection, this study employs the YOLO11 model to compare the detection performance before and after HENet-based dehazing. The numerical values marked on each bounding box represent the confidence scores of the detected targets. As observed from Fig. 10, direct detection on hazy images easily causes false detection and target omission. In contrast, images restored by HENet can significantly boost detection confidence and further improve detection accuracy in complex scenarios. As summarized in Table 5, the detection metrics including precision (P), recall (R), and mean average precision (mAP) are all substantially higher after image dehazing, compared with the detection results on original hazy inputs.

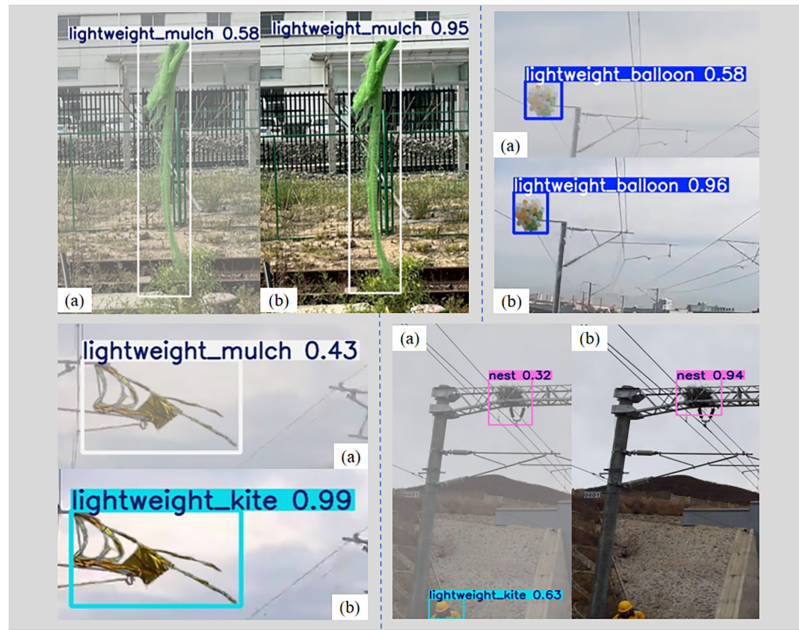


Figure 10: Performance comparison of traffic infrastructure target detection before and after dehazing using YOLOv11. (a) Fog detection result (b) Mist removal image detection result. Testing include slightweight_balloon, lightweight_kite, lightweight_mulch, nest.

Table 5: Quantitative comparison of detection results before and after dehazing.

Class	Before (%)			After (%)		
	P	R	mAP	P	R	mAP
All	86.0	81.4	88.8	90.6	87.6	93.8
Nest	99.0	85.9	96.5	98.8	97.9	99.3
lightweight_mulch	70.3	75.3	78.9	85.8	79.9	89.2
lightweight_kite	90.5	81.0	89.6	92.6	86.3	93.9
lightweight_balloon	84.1	83.5	90.3	85.1	86.2	92.6

Note: P, R, and mAP represent Precision, Recall, and mean Average Precision, respectively

5 Conclusion

Based on the aforementioned experiments and related research surveys, our analytical conclusions are summarized as follows: Prior-based dehazing approaches can effectively accomplish haze elimination without the requirement for cumbersome parameter tuning or large-scale training procedures. Nevertheless, such methods frequently generate images with sharp edge effects and over-contrasted colors; some of them also demand substantial computational resources, taking the NLD algorithm as a typical example, the DCP method can calculate the transmission matrix more accurately based on dark channel prior theory. Nevertheless, its performance degrades significantly when the atmospheric light shares close similarity with foreground scene targets. This inherent limitation leads to the observation that the DCP algorithm delivers superior processing outcomes for outdoor images rather than indoor scenes.

In contrast, learning-based methods like DCPDN which generate intermediate outputs by separately estimating transmission maps and atmospheric light are capable of delivering superior dehazing results while featuring simpler computational processes. However, these approaches rely heavily on massive volumes

of synthetic datasets for model training. It is noteworthy that both prior-based and the aforementioned learning-based methods obtain dehazed images by estimating transmission maps and atmospheric light. If either the transmission map or atmospheric light is not accurately inferred, the restored images often suffer from color artifacts or distortion issues.

In contrast, learning-based approaches that skip intermediate parameter generation, such as AOD Net, cGAN Net, and the newly proposed DPRNet, employ a joint estimation scheme for transmission maps and atmospheric light, thereby achieving outstanding dehazing performance. Although AOD Net and DCPDN have made considerable progress in dehazing quality and computational speed, their restored images still suffer from residual blur and visual artifacts. Such defects are mainly attributed to the cascaded structural design adopted by their network architectures.

According to our experimental findings, accumulating evidence indicates that hybrid estimation mechanisms play a pivotal role in enhancing model performance; numerous visual processing tasks benefit substantially from ultra-deep network architectures. Consequently, inadequate extraction of depth information caused by limited network depth can exert adverse impacts on the final results. This serves as the key rationale for adopting hybrid depth estimation as the core architectural design of the proposed method.

Acknowledgement: Not applicable.

Funding Statement: This work was supported in part by the Yunnan Xing Dian Talents Plan Young of China under Grant KKRD202202112.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Xue Yang, Shunpeng Yang; methodology, Xue Yang; software, Xue Yang; formal analysis, Xue Yang, Weizhong Yuan; investigation, Xue Yang; resources, Xue Yang, Ruixiao Sun; data curation, Xue Yang; writing—original draft preparation, Xue Yang; writing—review and editing, Xue Yang, Shunpeng Yang, Wanying Shi, Sihui Long; visualization, Xue Yang; supervision, Wanying Shi, Cheng Yang, Sihui Long; project administration, Wanying Shi; funding acquisition, Wanying Shi. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the author [Xue Yang].

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. He K, Sun J, Tang X. Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell.* 2010;33(12):2341–53. doi:10.1109/TPAMI.2010.168.
2. Cai B, Xu X, Jia K, Qing C, Tao D. DehazeNet: an end-to-end system for single image haze removal. *IEEE Trans Image Process.* 2016;25(11):5187–98. doi:10.1109/TIP.2016.2598681.
3. Qin X, Wang Z, Bai Y, Xie X, Jia H. FFA-Net: feature fusion attention network for single image dehazing. In: *Proceedings of the AAAI Conference on Artificial Intelligence*; 2020 Feb 7–12; New York, NY, USA. p. 11908–15.
4. Liu L, Wang Z. Automatic detection of wind turbine blade surface cracks based on UAV-taken images. *IEEE Trans Ind Electron.* 2017;64(9):7293–303. doi:10.1109/tie.2017.2682037.
5. Wu Y, Qin Y, Wang Z, Ma X, Cao Z. Densely pyramidal residual network for UAV-based railway images dehazing. *Neurocomputing.* 2020;371(1):124–36. doi:10.1016/j.neucom.2019.06.076.
6. Zhang J, Dong Q, Song W. GGADN: guided generative adversarial dehazing network. *Soft Comput.* 2023;27(3):1731–41. doi:10.21203/rs.3.rs-386958/v1.
7. Dwivedi P, Chakraborty S. Single image dehazing using extended local dark channel prior. *Image Vis Comput.* 2023;136:104747. doi:10.21203/rs.3.rs-2387969/v1.

8. Narasimhan SG, Nayar SK. Contrast restoration of weather degraded images. *IEEE Trans Pattern Anal Mach Intell.* 2003;25(6):713–24. doi:10.1109/TPAMI.2003.1201821.
9. Tan RT. Visibility in bad weather from a single image. In: *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*; 2008 Jun 23–28; Anchorage, AK, USA.
10. Song Y, He Z, Qian H, Du X. Vision transformers for single image dehazing. *IEEE Trans Image Process.* 2023;32:1927–41. doi:10.1109/TIP.2023.3256763.
11. Li J, Li G, Fan H. Image dehazing using residual-based deep CNN. *IEEE Access.* 2018;6:26831–42. doi:10.1109/access.2018.2833888.
12. Jeong CY, Moon K, Kim M. An end-to-end deep learning approach for real-time single image dehazing. *J Real Time Image Process.* 2023;20(1):12. doi:10.1007/s11554-023-01270-2.
13. Chen Z, Wang Y, Yang Y, Liu D. PSD: principled synthetic-to-real dehazing guided by physical priors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021*; 2021 Jun 20–25; Nashville, TN, USA. p. 7180–90.
14. Narasimhan SG, Nayar SK. Chromatic framework for vision in bad weather. In: *Proceedings of the Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*; 2000 Jun 15; Hilton Head Island, SC, USA. p. 598–605.
15. Nayar SK, Narasimhan SG. Vision in bad weather. In: *Proceedings of the Proceedings of the Seventh IEEE International Conference on Computer Vision*; 1999 Sep 20–25; Corfu, Greece. p. 820–7.
16. Arampatzakis V, Pavlidis G, Mitianoudis N, Papamarkos N. Monocular depth estimation: a thorough review. *IEEE Trans Pattern Anal Mach Intell.* 2024;46(4):2396–414. doi:10.1109/TPAMI.2023.3330944.
17. Ancuti C, Ancuti CO, Timofte R. Ntire 2018 challenge on image dehazing: methods and results. In: *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 891–901.
18. Zhai Y, Zhang Y. Contrast restoration for fog-degraded images. In: *Proceedings of the 2009 International Conference on Computational Intelligence and Security*; 2009 Dec 11–14; Beijing, China. p. 619–23. doi:10.1109/CIS.2009.243.
19. Fattal R. Single image dehazing. *ACM Trans Graph.* 2008;27(3):1–9. doi:10.1145/1360612.1360671.
20. Chen D, He M, Fan Q, Liao J, Zhang L, Hou D, et al. Gated context aggregation network for image dehazing and deraining. In: *Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*; 2019 Jan 7–11; Waikoloa Village, HI, USA.
21. Yang Y, Wang Z. Haze removal: push DCP at the edge. *IEEE Signal Process Lett.* 2020;27:1405–9. doi:10.1109/LSP.2020.3013741.
22. Baig N, Riaz MM, Fatima A, Ali SS, Ghafoor A, Siddiqui AM. Image dehazing using dark and bright channel priors and multi-scale filters. In: *Proceedings of the 2020 14th International Conference on Open Source Systems and Technologies (ICOSST)*; 2020 Dec 16–17; Lahore, Pakistan.
23. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM.* 2017;60(6):84–90. doi:10.1145/3065386.
24. Ersavas T, Smith MA, Mattick JS. Novel applications of Convolutional Neural Networks in the age of Transformers. *Sci Rep.* 2024;14(1):10000. doi:10.21203/rs.3.rs-3868861/v1.
25. Li B, Peng X, Wang Z, Xu J, Feng D. AOD-net: all-in-one dehazing network. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*; 2017 Oct 22–29; Venice, Italy. p. 4780–8. doi:10.1109/iccv.2017.511.
26. Ren W, Ma L, Zhang J, Pan JS, Cao X, Liu W, et al. Gated fusion network for single image dehazing. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. doi:10.1109/CVPR.2018.00343.
27. Zhang H, Patel VM. Densely connected pyramid dehazing network. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 3194–203. doi:10.1109/cvpr.2018.00337.

28. Zeng Y, Wang X, Zhao H, Jin Y, Giannopoulos GA, Li Y. Image fusion methods in high-speed railway scenes: a survey. *High-Speed Railw.* 2023;1(2):87–91. doi:10.1016/j.hspr.2023.01.001.
29. Goncalves LT, De Oliveira Gaya J, Drews P, Da Costa Botelho SS. DeepDive: an end-to-end dehazing method using deep learning. In: *Proceedings of the 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*; 2017 Oct 17–20; Niteroi, Brazil. p. 436–41. doi:10.1109/SIBGRAPI.2017.64.
30. Qiu Y, Zhang K, Wang C, Luo W, Li H, Jin Z. Mb-taylorformer: multi-branch efficient transformer expanded by taylor formula for image dehazing. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision 2023*; 2023 Oct 1–6; Paris, France.
31. Xie Z, Li Q, Zong S, Liu G. Multiscale feature fusion deep network for single image dehazing with continuous memory mechanism. *Optik.* 2023;287(11):171144. doi:10.1016/j.ijleo.2023.171144.
32. Li R, Pan J, Li Z, Tang J. Single image dehazing via conditional generative adversarial network. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 8202–11. doi:10.1109/CVPR.2018.00856.
33. Appiah DJ, Guan D, Kasule AN, Wei M. WTCL-dehaze: rethinking real-world image dehazing via wavelet transform and contrastive learning. *arXiv:2410.04762*. 2024.
34. Wu Y, Chen P, Qin Y, Qian Y, Xu F, Jia L. Automatic railroad track components inspection using hybrid deep learning framework. *IEEE Trans Instrum Meas.* 2023;72:1–15. doi:10.1109/tim.2023.3265636.
35. Redmon J, Farhadi A. YOLOv3: an incremental improvement. *arXiv:1804.02767*. 2018.
36. Jang CY, Lim JH, Kim YH. A fast Multi-scale Retinex algorithm using dominant SSR in weights selection. In: *Proceedings of the 2012 International SoC Design Conference (ISOC)*; 2012 Nov 4–7; Jeju, Republic of Korea. p. 37–40. doi:10.1109/ISOC.2012.6406919.
37. Barnard K, Funt B. Investigations into multi-scale retinex. In: *Proceedings of the Presented at the Colour Imaging in Multimedia Conference*; 1998 Nov 17–20; Scottsdale, Arizona. p. 9–17.
38. Yang J, Huang T. Image super-resolution: historical overview and future challenges. In: *Super-resolution imaging*. Boca Raton, FL, USA: CRC Press; 2017.
39. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* 2004;13(4):600–12. doi:10.1109/TIP.2003.819861.
40. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, et al. Benchmarking single-image dehazing and beyond. *IEEE Trans Image Process.* 2019;28(1):492–505. doi:10.1109/TIP.2018.2867951.
41. Kumar BV, Chahar V. HUDRS: hazy unpaired dataset for road safety. *Vis Comput.* 2022;38(12):3881–9. doi:10.1007/s00371-022-02534-x.
42. Feng C, Chen Z, Li X, Wang C, Yang J, Cheng MM, et al. HazyDet: open-source benchmark for drone-view object detection with depth-cues in hazy scenes. *arXiv:2409.19833*. 2024.
43. Zheng Y, Zhan J, He S, Dong J, Du Y. Curricular contrastive regularization for physics-aware single image dehazing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023*; 2023 Jun 17–24; Vancouver, BC, Canada.
44. Zhu Q, Mai J, Shao L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans Image Process.* 2015;24(11):3522–33. doi:10.1109/tip.2015.2446191.
45. Berman D, Treibitz T, Avidan S. Non-local image dehazing. In: *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition 2016*; 2016 Jun 27–30; Las Vegas, NV, USA.
46. Mittal A, Soundararajan R, Bovik AC. Making a “completely blind” image quality analyzer. *IEEE Signal Process Lett.* 2013;20(3):209–12. doi:10.1109/LSP.2012.2227726.
47. Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. *IEEE Trans Image Process.* 2012;21(12):4695–708. doi:10.1109/tip.2012.2214050.
48. Cao Z, Qin Y, Jia L, Xie Z, Liu Q, Ma X, et al. Haze removal of railway monitoring images using multi-scale residual network. *IEEE Trans Intell Transport Syst.* 2021;22(12):7460–73. doi:10.1109/tits.2020.3003129.