Article:
Wu, S., Li, H., Deng, L. et al. (5 more authors) (2025) FoggyFuse: Infrared and visible image fusion method based on saturation line prior in foggy conditions. Optics & Laser Technology, 190. 113075. ISSN: 0030-3992

https://doi.org/10.1016/j.optlastec.2025.113075

FoggyFuse: Infrared and Visible Image Fusion Method Based on Saturation Line Prior in Foggy Conditions

Shengkun Wu[a,b,1], Hao Li[a,b,1], Lei Deng[a,b,c,*], Heng Yu[e], Hanrui Chen[a,b], Zhixiang Chen[d], Mingli Dong[a,b,c,*], Lianqing Zhu[a,b,c]

[a]*Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science & Technology University, Beijing, 100192, China*
[b]*Beijing Laboratory of Optical Fiber Sensing and System, Beijing Information Science & Technology University, Beijing, 100016, China*
[c]*Guangzhou Nansha Intelligent Photonic Sensing Research Institute, Guangzhou, Guangdong Province, 511462, China*
[d]*Department of Computer Science, The University of Sheffield, Sheffield, South Yorkshire, S1 4DP, UK*
[e]*Robotics Institute, Carnegie Mellon University, Pittsburgh, 15213, USA*

Abstract

Infrared and visible image fusion is widely used to enhance image details and information. However, in foggy environments or military smoke bomb scenarios, the scattering and absorption of light significantly degrade the quality of both infrared and visible images, leading to poor fusion performance. Existing fusion methods struggle to effectively restore degraded image details, making them unsuitable for practical applications in such adverse conditions. To address this challenge, we propose a novel fusion architecture based on the saturation line prior (SLP). This method consists of three main modules: the Dehazing Module (DM), the Auxiliary Enhancement Module (AEM), and the Edge Enhancement Module (EEM). The DM optimizes SLP using weighted guided filtering to obtain refined transmission maps for visible images, which are then used to further enhance the infrared image. The AEM and EEM, combined with a non-subsampled shearlet transform (NSST), further process the enhanced visible and infrared images. This approach effectively restores intricate details and achieves natural color reproduc-

*Corresponding author
Email addresses: dally211@163.com (Lei Deng), dongml@bistu.edu.cn (Mingli Dong)
[1]These authors contributed equally to this work

tion in hazy environments, significantly improving the visual quality of fused images. Given the limited research in this area and the absence of relevant datasets, we constructed an infrared and visible image pair dataset, Foggy, specifically designed for foggy conditions. Qualitative and quantitative evaluations demonstrate that the proposed method outperforms state-of-the-art fusion techniques on the Foggy dataset.

Keywords:
Infrared and visible image fusion, Image dehazing, Foggy image fusion, Detail enhancement,

---

1. Introduction

The fusion of infrared and visible images plays a crucial role in image processing due to its wide range of applications. However, in the presence of smoke and fog, some information captured by visible-light cameras is lost. The negative impact on the image will also be different under different smoke and fog conditions. Under uniform smoke or fog conditions, the image will be whitened over a large area, the color information will be weakened, and the tiny details will be difficult to distinguish. Under non-uniform smoke or fog conditions, the distribution of fog is obviously random, and the effective information behind it is looming. Under dense smoke or fog conditions, the effective information of the image is greatly destroyed, and the texture information is seriously lost. As shown in Fig. 1. Although infrared cameras are less affected, certain details may still be obscured by fog. The impact of fog on both infrared and visible images leads to texture loss and color distortion in the final fusion result. Therefore, developing image fusion algorithms tailored for foggy conditions is of great significance.

Current infrared and visible image fusion approaches can be broadly categorized into learning-based methods [1, 2, 3, 4, 5, 6, 7, 8] and traditional fusion methods [9, 10, 11, 12, 13]. Gao *et al.* [14] proposed an infrared and visible image fusion network, BCMFIFuse, which is based on bidirectional cross-modal feature interaction. This method effectively integrates cross-modal information and extracts more comprehensive multi-scale features to enhance fusion performance. Yao *et al.* [15] proposed HG-LPFN, a fusion method combining Laplacian pyramid and hierarchical saliency guidance. Although these commonly used fusion methods produce satisfactory results in most cases, they fail to address the texture loss and color distortion caused by fog, as illustrated in Fig. 2. Yang *et al.* [16] introduced an infrared and visible image fusion framework combining an adap-

(a) uniform smoke or fog (b) non-uniform smoke or fog (c) dense smoke or fog

Figure 1: Demonstration of the effects of varying smoke or fog conditions on images.

(a) NSCT based (b) IFEVIP (c) DATFuse (d) CDDFuse

Figure 2: Performance of common fusion methods under foggy conditions. (a) and (b) both show image distortion. (c) and (d) both show a large amount of texture loss.

tive differential fusion module and a salient target perception module. Zhou *et al.* [17] proposed a night vision context enhancement algorithm that combines adaptive enhancement and perception-based parameter selection to improve visibility in low-light conditions. Yao *et al.* [18] developed a color-aware fusion framework that leverages color injection and knowledge distillation to achieve natural color fusion of infrared and visible images under low-light conditions. Yao *et al.* [19] proposed a semantic-powered image enhancement and fusion framework that leverages uncertainty modeling to improve perceptual quality and detail preservation. Although these methods perform well in low-light scenarios such as nighttime environments, they do not design dedicated mechanisms for the foggy imaging degradation problem, resulting in limitations in foggy conditions.

To mitigate the adverse effects of fog on images, current mainstream methods typically employ defogging algorithms to process hazy images and restore image information obscured by fog. These algorithms can generally be categorized into two types: single-image defogging based on visible images and multi-image defogging assisted by near-infrared images. Single-image dehazing methods can be further divided into physical and prior-based methods [20, 21, 22, 23, 24], as well as learning-based methods [25, 26, 27, 28, 29]. He *et al.* [30] discovered the dark channel prior (DCP), which can accomplish the dehazing task simply and effi-

ciently. Shen *et al.* [31] proposes a iterative dehazing method with polarization, which uses joint depth–chromaticity regularization to suppress noise amplification during dehazing. However, these defogging methods focus solely on processing visible images without considering the characteristics of infrared images, making them challenging to apply directly to infrared image fusion tasks. Multi-image defogging assisted by near-infrared images [32, 33, 34, 35, 36] leverages the properties of near-infrared images to compensate for details lost in visible images due to fog, thereby achieving an assisted defogging effect. However, in infrared and visible image fusion, far-infrared images are typically used to enhance images by incorporating both visible and far-infrared information, with both modalities holding equal significance. Despite their advantages, these multi-image defogging methods are not specifically designed for the fusion of far-infrared and visible images. Additionally, due to its longer wavelength, far-infrared is more sensitive to thermal radiation and provides complementary information to visible images. In contrast, near-infrared has a shorter wavelength and produces images that are more similar to visible images, making it less suitable for the infrared and visible image fusion tasks we aim to address.

Current infrared and visible image fusion methods rarely focus on foggy conditions. To address this gap, we propose a fusion framework specifically designed for infrared and visible image fusion in foggy environments. The framework consists of three main modules: the Dehazing Module (DM), the Auxiliary Enhancement Module (AEM), and the Edge Enhancement Module (EEM). In the DM, we redesigned SLP based on weighted guided filtering and used the resulting transmission map of the visible image to guide the enhancement of the infrared image. The AEM performs an initial fusion of infrared and visible images, preserving essential color information. The EEM enhances the texture details of the fused image. The proposed framework effectively restores image details lost due to fog occlusion, resulting in superior visual quality of the fused images. Furthermore, due to the lack of publicly available foggy infrared and visible image fusion datasets, we constructed the Foggy dataset to evaluate the effectiveness of our proposed method.

Our contribution could be summarized as follows.

- We propose a novel fusion framework specifically designed for infrared and visible image fusion in foggy environments. Unlike existing methods, we redesign the Saturation Line Prior (SLP) for image dehazing and introduce a weighted guided filtering optimization strategy to refine the transmission map. Furthermore, we leverage the transmission map of the visible image

to guide the enhancement of the infrared image, thereby preserving richer textures throughout the fusion process.

- We design two specialized enhancement modules to further optimize the fusion process. The Auxiliary Enhancement Module (AEM) incorporates visual saliency maps and weighted least squares optimization to integrate useful color and detail information from the input images, enhancing image quality before fusion. Meanwhile, the Edge Enhancement Module (EEM) is developed to effectively refine image textures, mitigate texture loss in a simple yet efficient manner, and significantly enhance the overall visual quality of the fused image.

- We construct the Foggy Dataset, the first dataset specifically designed for far-infrared and visible image fusion in foggy conditions. This dataset encompasses diverse smoke and fog scenarios, including uniform smoke or fog, non-uniform smoke or fog, and dense smoke or fog, providing a comprehensive benchmark for future research in this domain.

2. Related Work

This section provides an overview of the current commonly used image fusion and dehazing methods.

2.1. Multi-scale decomposition-based image fusion methods

The multi-scale decomposition approach employs specialized techniques to decompose images into multiple layers, with different fusion strategies applied to the corresponding sub-images at each layer. The final fused image is obtained by performing an inverse transformation on the decomposed layers. Wang *et al.* [37] utilized the Laplacian pyramid for image decomposition. Da *et al.* [38] proposed the Non-subsampled Contourlet Transform (NSCT) for image decomposition. However, NSCT suffers from low computational efficiency, making it unsuitable for real-time applications. To address this limitation, Easley *et al.* [13] introduced the Non-subsampled Shearlet Transform (NSST), which offers higher computational efficiency and better fusion performance compared to NSCT.

The decomposition and reconstruction of NSST can be simply expressed as Eq. (1) and Eq. (2):

$$nsst_{de}(I_{in}) = \{L_k, H_k\}, \tag{1}$$

$$nsst_{re}(L_k, H_k) = I_{re}, \tag{2}$$

where I_{in} and I_{re} represent the input image and the reconstructed image respectively, L_k refers to the low-frequency subband, H_k represents the high-frequency subband.

2.2. Image dehazing

The introduction of image priors provides additional constraints when processing hazy images, compensating for the lack of information. At the same time, saturation information has been widely utilized in prior-based dehazing methods [24, 39, 40, 41]. Ling *et al.* [24] observed that in fog-free images, pixels with identical surface reflectance tend to appear in local patches, revealing a linear relationship between saturation and the inverse of the brightness component. This relationship enables reliable and accurate transmission estimation for foggy images, leading to color restoration with rich details.

Another approach involves leveraging the fusion of near-infrared and visible images for assisted defogging [32, 33, 34]. This method enhances image details by utilizing near-infrared characteristics to restore fog-obscured information. However, these methods focus solely on fog removal and do not extend to far-infrared image fusion. While far-infrared imaging has the ability to penetrate fog, dense smoke or fog conditions still degrade image details, resulting in color distortion and artifacts when fused with visible images. Moreover, the scarcity of publicly available foggy infrared-visible image fusion datasets hinders experimental validation. To address this, we collected and curated the Foggy dataset, which we incorporated into our experiments. In this work, we propose a novel fusion framework that integrates the saturation line prior with the NSST architecture. In addition, we introduce two specialized modules: the Auxiliary Enhancement Module (AEM) and the Edge Enhancement Module (EEM). The saturation line prior is employed to pre-dehaze visible images while simultaneously guiding the enhancement of infrared images. AEM and EEM further refine the fusion process by preserving essential image details, mitigating texture loss, and improving overall visual quality. Our proposed method has been extensively evaluated on the Foggy dataset, demonstrating superior visualization results and enhanced performance metrics compared to state-of-the-art fusion algorithms.

3. Proposed Method

In this section, we propose a method for the fusion of infrared and visible images under foggy conditions. The framework of the proposed method is shown

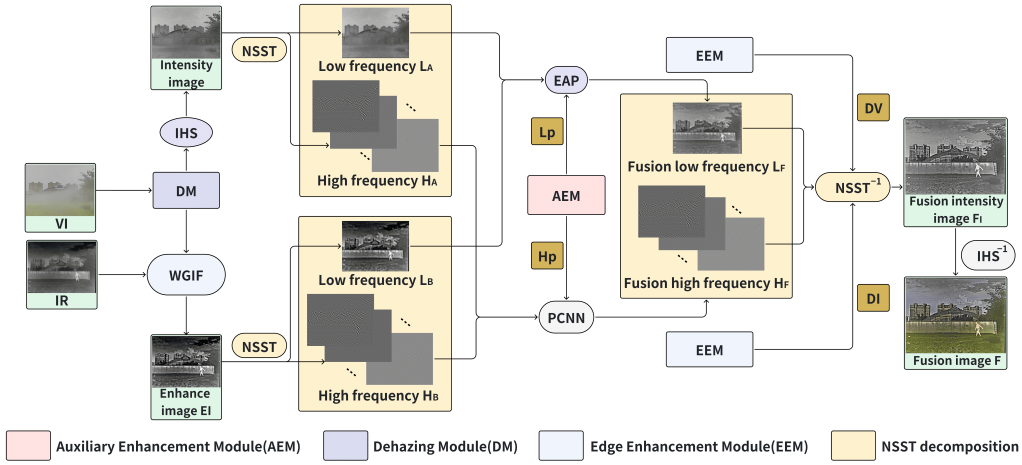


Figure 3: The proposed method framework. The Dehazing Module generates a dehazed visible image and an enhanced infrared image, which are further combined with the pre-fused image obtained by the Auxiliary Enhancement Module after NSST decomposition, and finally the Edge Enhancement Module is used to improve the texture clarity.

in Fig. 3. It primarily consists of three components: the Dehazing Module (DM), the Auxiliary Enhancement Module (AEM), and the Edge Enhancement Module (EEM). These modules are explained in detail in Sections 3.1, 3.2, and 3.3. The first step of the method is to obtain a transmission map of the visible image using DM. We use the acquired transmission map as a guide to enhance the infrared image. The enhanced image is then subjected to intensity-hue-saturation (IHS) transformation to generate an intensity image. Finally, the image undergoes NSST decomposition, processing, and reconstruction in combination with AEM and EEM, followed by an inverse IHS transformation to obtain the fused image.

3.1. Dehazing Module

The process of the DM module is shown in Fig. 4. It mainly obtains the transmission map through the Saturation line model, uses the transmission map optimized by weighted guided filtering as a guide to enhance the infrared image, and reconstructs a clear visible image at the same time.

Haze imaging model According to the currently recognized atmospheric scattering model [42], the formation model of foggy blurred images can be constructed using Eq.(3):

$$I(x,y) = J(x,y) \cdot t(x,y) + A \cdot (1 - t(x,y)), \tag{3}$$

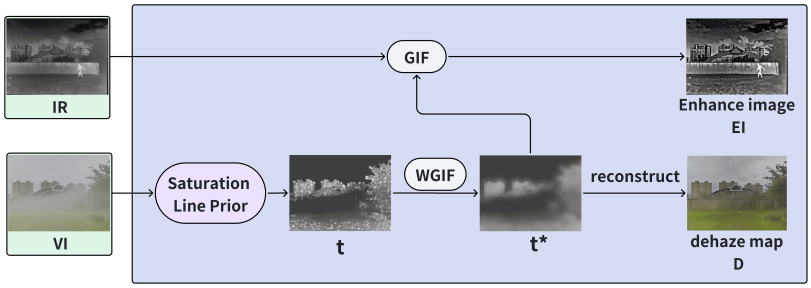


Figure 4: Dehazing Module (DM). IR and VI represent infrared and visible images. t and t* represent the rough transmission map and the optimized transmission map. EI and D represent the enhanced infrared image and dehazed image.

where I represents the haze image, J represents the haze-free image, (x, y) represents the pixel coordinate, t represents the medium transmission, and A represents the global atmospheric light. As can be seen from Eq. (3), the pixels in the haze image I that correspond to these regions are primarily composed of A. Therefore, we can estimate the atmospheric light A by analyzing the sky region in the haze image.

Saturation line model Based on the currently recognized saturation model, the saturation component of an image $S_P(x, y)$ can be expressed as:

$$S_P(x, y) = 1 - \frac{\min\limits_{c\in\{r,g,b\}} P^c(x, y)}{\max\limits_{c\in\{r,g,b\}} P^c(x, y)}, \tag{4}$$

where P represents the input image, (x, y) represents the pixel coordinate, $c \in \{r, g, b\}$ represents the color channel.

Combining Eq. (3) and Eq. (4), and denoting the maximum and minimum color components of J as J_N^{max} and J_N^{min}, the saturation component of the normalized haze image I_N can be expressed as:

$$\begin{aligned} S_{I_N}(x, y) = 1 &- \frac{J_N^{min}(x, y) \cdot t(x, y) + (1 - S_{J_N}(x, y)) \cdot (1 - t(x, y))}{J_N^{max}(x, y) \cdot t(x, y) + (1 - t(x, y))} \\ &- \frac{S_{J_N}(x, y) \cdot (1 - t(x, y))}{J_N^{max}(x, y) \cdot t(x, y) + (1 - t(x, y))}. \end{aligned} \tag{5}$$

where S_{J_N} is the saturation component of the normalized haze-free image J_N.

According to the definition of HSV color space, assuming that the brightness component V_{I_N} of I_N is the largest color component, that is:

$$V_{I_N}(x,y) = I_N^{max}(x,y) \tag{6}$$

where I_N^{max} is the maximum color component of I_N. So S_{I_N} can be expressed as:

$$S_{I_N}(x,y) = S_{J_N}(x,y) \cdot (1 - \frac{1 - t(x,y)}{V_{I_N}(x,y)}). \tag{7}$$

Ling et al. [24] modeled natural images as trajectories of projected surface tiles [43, 44], and by following the local monochromatic surface assumption, the clear image $J(x,y)$ is represented as:

$$J(x,y) = k(x,y) \cdot R(x,y), \tag{8}$$

$$R(x,y) = R^*, \quad (x,y) \in \Omega, \tag{9}$$

where R is a three-dimensional vector representing the surface reflectance, R^* is a constant three-dimensional vector in a given local patch, and $k(x,y)$ is a one-dimensional shadow factor describing the radiation intensity. The local patch can be further untangled by assuming R to be piecewise constant. By combining the Eq. (8) and Eq. (9), it can be proved that the saturation component S_{J_N} is a local constant under the assumption of a local monochromatic surface:

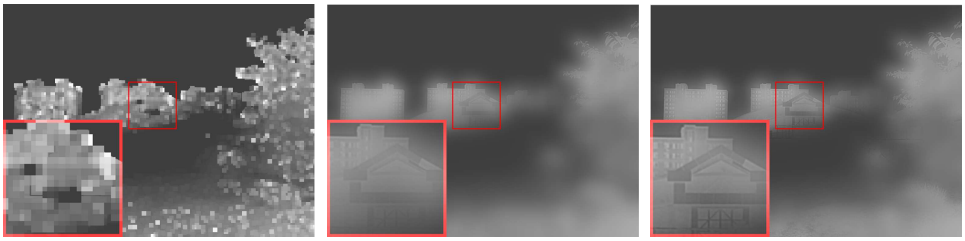
$$S_{J_N}(x,y) = 1 - \frac{\min\limits_{c\in\{r,g,b\}} \frac{R^c(x,y)}{A^c} \cdot k(x,y)}{\max\limits_{c\in\{r,g,b\}} \frac{R^c(x,y)}{A^c} \cdot k(x,y)} = 1 - \frac{\min\limits_{c\in\{r,g,b\}} \frac{\overline{R}^c}{A^c}}{\max\limits_{c\in\{r,g,b\}} \frac{\overline{R}^c}{A^c}}. \tag{10}$$

Then S_{J_N} is a constant in the local patch, expressed as $S_{J_N}^*$. In addition, assuming that the propagation t is locally constant, which we denote as t^*, a given local patch can be expressed by Eq. (7):

$$S_{I_N}(x,y) = S_{J_N}^* - \frac{S_{J_N}^* \cdot (1 - t^*)}{V_{I_N}(x,y)}. \tag{11}$$

Eq. (7) can be further rewritten as:

$$S_{I_N}(x,y) = k \cdot V_{I_N}^{-1}(x,y) + b, \tag{12}$$



(a) initial transmission map (b) guided filter refinement (c) weighted guided filter refinement

Figure 5: Comparison of transmission map refinement effects using different methods. The transmission map obtained by our improved method has richer texture.

where $V_{I_N}^{-1}(x, y)$ is the reciprocal of $V_{I_N}(x, y)$, and the slope k and intercept b can be expressed as:

$$k = t^* \cdot S_{J_N}^* - S_{J_N}^*, \tag{13}$$

$$b = S_{J_N}^*. \tag{14}$$

From this we can directly construct the transmission t^* in the corresponding local patch:

$$t^* = k/b + 1. \tag{15}$$

In Eq. (12), S_{I_N} and $V_{I_N}^{-1}$ have a linear relationship, which is called the saturation line prior. We can use this relationship to accurately estimate the value of transmittance t by constructing the saturation line.

Refine the transmission map with weighted guided filter For the transmission map t generated by the saturation line prior, Ling *et al.* [24] used the traditional guided filtering to optimize the transmission map t. However, this method does not achieve good edge preservation. For this reason, we improved it to weighted guided filtering [45], using its stronger edge preservation characteristics, combined with reasonable weight assignment, to better refine the transmission map, obtain more image details, and achieve better defogging effect. The acquisition of the refined transmission map t' can be simply expressed by Eq. (16). and the comparison of the transmission map effect before and after the weighted guided filtering optimization is shown in Fig. 5. It can be seen that the optimized transmission map has a clearer texture and better visualization effect.

$$t'(x, y) = WGIF(t(x, y), I(x, y)). \tag{16}$$

Figure 6: Defogging effect display. The left side of each image set is before DM processing, and the right side is after DM processing. Our method obtains dehazed images with realistic colors and clear details.

Figure 7: Infrared image enhancement display. The left side of each image set is before IR enhancement, and the right side is after IR enhancement. The enhanced IR has richer texture and better visualization effect.

Reconstruct image Currently, common dehazing methods usually use the pixels in the farthest area of the foggy image as a reference to estimate the atmospheric light value A. Therefore, based on the atmospheric scattering model in Eq. (3), we can obtain the restored scene radiance J, which is what we need dehaze map D:

$$D = J(x, y) = A + \frac{I(x, y) - A}{t'(x, y)}. \tag{17}$$

The effect of the DM is shown in Fig. 6, the smoke part is removed as much as possible, and the real image color is restored well.

Infrared image enhancement We propose an infrared image dehazing and enhancement method based on weighted guided filtering (WGIF). This method uses the transmission map as a guide to enhance the infrared image through weighted guided filtering. The transmission map can reflect the distribution of fog in the image, so it can effectively guide the enhancement process of the infrared image, thereby retaining more detailed information while removing haze. As shown in Fig. 7, the enhanced infrared image has richer texture and is less affected by fog.

First, a weight map is generated based on the transmission map, with the following equation:

$$W(x, y) = 1 - \frac{T(x, y)}{255}, \tag{18}$$

where $T(x, y)$ represents the pixel value of the transmission map.

Afterward, we use the obtained transmission weight map to weight the infrared image, and the formula is:

$$I_{weighted}(x, y) = I_{ir}(x, y) \cdot W(x, y), \tag{19}$$

where $I_{ir}(x, y)$ is the pixel value of the infrared image.

Then, the weighted infrared image is filtered using a fast guided filter (FGIF) [46], with the filter process equation as follows:

$$I_{wgif}(x, y) = FGIF(I_{weighted}, I_{ir}, r, \epsilon, s), \tag{20}$$

where r is the radius of the guided filter, ϵ is the regularization parameter, and s is the subsampling ratio.

Finally, a residual enhancement technique is applied to further improve the details and contrast of the image, with the equation:

$$I_{enhanced}(x, y) = (I_{ir}(x, y) - I_{wgif}(x, y)) \cdot \alpha + I_{wgif}(x, y), \tag{21}$$

where α is the enhancement parameter, empirically set to 5 in this study.

3.2. Auxiliary Enhancement Module design

Since the original image has been preprocessed for dehazing, the dehazing itself will actually lose some details of the original image, so we designed the Auxiliary Enhancement Module, as shown in Fig. 8. To meet our final experimental purpose, we apply a pre-fusion technique based on visual saliency maps and weighted least squares optimization. The results of the pre-fusion are shown in Fig. 9. This method effectively improves scale separation in multi-scale decomposition while preserving edges. It also enhances the transfer of useful visual details to the fused image, while reducing noise in the infrared image [47].

The pre-fused image will be processed by guided filtering to decompose the base layer L_P and detail layer H_P of the image. The dehazed image D and IR image will be decomposed by NSST [13] to obtain the base layer L_A, L_B and detail layer $H_A^{l,k}$, $H_B^{l,k}$, as shown in Eq. (1). The base layer L_P will enter the EAP fusion together with L_A and L_B to obtain fusion Base Layer L_F, as shown in Eq. (22) This can avoid the loss of some effective details of the original visible image. At the same time, the detail layer H_P is fused with $H_A^{l,k}$ and $H_B^{l,k}$ using Pulse-Coupled Neural Network (PCNN) [48] to obtain H_F, it can better capture the distinctive features of high-frequency images and make full use of its adaptability and parallelism to improve the fusion effect.

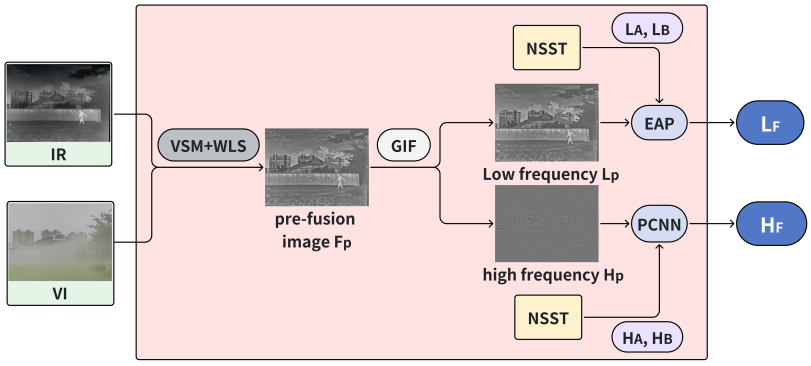


Figure 8: Auxiliary Enhancement Module (AEM). Fp represents the pre-fusion image. Lp and Hp represent the low-frequency image and high-frequency image obtained by decomposing Fp through guided image filtering.

(a) VI (b) IR (c) pre-fusion results

Figure 9: VI, IR, and pre-fusion results. The pre-fusion results successfully retain the thermal radiation data from the infrared image and the texture details from the visible image.

$$L_F = EAP(L_A, L_B, L_P), \tag{22}$$

where the EAP module will be introduced later.

3.3. Edge Enhancement Module design

The Edge Enhancement Module as shown in Fig. 10, we use NSST decomposition to obtain the low-frequency subband L_B of the infrared enhanced image EI and the low-frequency subband L_A of the visible enhanced image D. Then we subtract them from the IR and VI images respectively to obtain the high-frequency edges DI and DV, as shown in Eq. (23) and Eq. (24). Then fuse them with the low-frequency subband L_F and the high-frequency subbands $H_A^{l,k}$ to obtain the final edge-enhanced fused image. This operation can effectively improve the dehazing effect and obtain more textures.

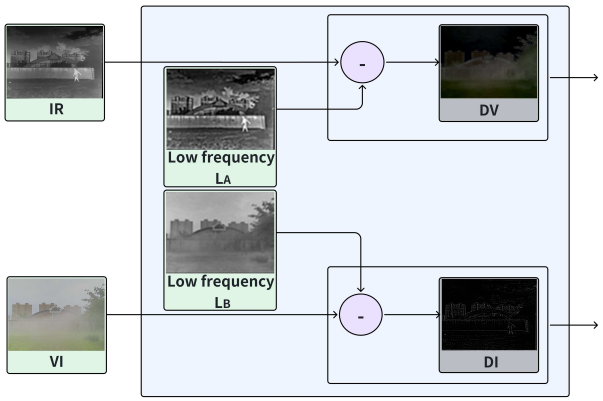


Figure 10: Edge Enhancement Module (EEM). LA and LB represent the intensity image of D and the low-frequency image of EI after NSST decomposition. DV and DI represent the extracted detail layers.

$$DV = VI - L_A, \tag{23}$$
$$DI = IR - L_B. \tag{24}$$

3.4. Image reconstruction

NSST decomposition is one of the common frequency domain decomposition methods. We use NSST decomposition to divide the image into low-frequency subbands and multiple high-frequency subbands, use EAP weighted fusion to fuse the low-frequency subbands of VI and IR, and use PCNN [48] Multiple high-frequency subbands are fused, and finally the NSST inverse transform is performed on the initially fused low-frequency subband and multiple initially fused high-frequency subbands to restore the final fused image.

Low frequency sub-band fusion using EAP The low-frequency subband contains most of the background information of the source image. In low-frequency fusion, Tan *et al.* [49] used an energy attribute fusion strategy EA, but often EA has two inputs. Since our method has multiple inputs, we modified the fusion strategy of EA, which is called EAP, the processing steps are as follows:

$$IPV_A = \mu_A + MI_A, \tag{25}$$
$$IPV_B = \mu_B + MI_B, \tag{26}$$
$$IPV_P = \mu_P + MI_P, \tag{27}$$

where IPV represents the fundamental characteristic values of the low-frequency sub-band, μ represents the mean value, and MI represents the median value of I. The mean of the three base layers I_A, I_B, and I_P is calculated as follows:

$$I_A(x,y) = e^{\alpha|L_A(x,y)-IPV_A|}, \tag{28}$$
$$I_B(x,y) = e^{\alpha|L_B(x,y)-IPV_B|}, \tag{29}$$
$$I_P(x,y) = e^{\alpha|L_P(x,y)-IPV_P|}, \tag{30}$$

where α represents the modulation parameter.

Then we can get the low-frequency subband of each branch image L^*:

$$L_A^* = I_A(x,y) \times L_A(x,y), \tag{31}$$
$$L_B^* = I_B(x,y) \times L_B(x,y), \tag{32}$$
$$L_C^* = I_C(x,y) \times L_C(x,y), \tag{33}$$

Ultimately, the fused low-frequency subband L_F is generated through a weighted averaging technique.

$$L_F(x,y) = \frac{L_A^* + L_B^* + L_C^*}{I_A(x,y) + I_B(x,y) + I_C(x,y)}. \tag{34}$$

NSST reconstruction

The fused high-frequency edge is expressed as:

$$EP = DV + DI, \tag{35}$$
$$H_F^{*\,l,k} = nsst_{re}(H_F^{l,k}, EP), \tag{36}$$

where DV and DI represent the two high-frequency edges obtained by the EEM module. $H_F^{l,k}$ represents the high-frequency edge from AEM fused by PCNN [48].

Then we can get the fused brightness channel image F_I

$$F_I = nsst_{re}(L_F, H_F^{*\,l,k}), \tag{37}$$

where $nsst_{re}$ represents the inverse NSST, as shown in Eq. (2).

Finally, the lightness channel image F_I is restored to the final fused image F through the inverse IHS transform, as shown in Eq. (38).

$$F = IHS^{-1}(F_I). \tag{38}$$

where IHS^{-1} represents the inverse IHS transform.

(a) uniform smoke or fog (b) non-uniform smoke or fog (c) dense smoke or fog

Figure 11: An example of the foggy dataset we created. It includes three common types of infrared fusion image pairs: (a) uniform smoke or fog, (b) non-uniform smoke or fog, and (c) dense smoke or fog.

4. Experiments

4.1. Dataset

Since there is no open-source infrared and visible image dataset specifically for foggy scenes, in order to meet the needs of the experiment, we decided to collect and organize from multiple open-source datasets to build our experimental dataset Foggy. The screening conditions are color, hazy, smoky, or foggy images. At the same time, we also considered the glare caused by vehicles or street lights at night, because we believe that glare can be similar to the occlusion of fog to a certain extent. Initially, we collected a total of 1,747 pairs of infrared and visible images to form our dataset. Later, after a lot of research and experiments, we found that high-quality data is more effective than low-quality data. Therefore, we decided to manually select high-quality data, re-screen and remove a large number of duplicate and possibly unsuitable image pairs, and manually classify them according to the classification conditions of uniform smoke or fog, non-uniform smoke or fog, and dense smoke or fog. We obtained the Foggy dataset, which contains 190 pairs of uniform smoke or fog images, 230 pairs of non-uniform smoke or fog images, and 240 pairs of dense smoke or fog images, a total of 660 image pairs, and used them in our experiments. The 660 pairs of images in the Foggy dataset come from the M3FD dataset (610 pairs)[50], the MSRS dataset (28 pairs)[51], and the VTUAV dataset (22 pairs)[52]. Some examples of Foggy dataset are shown in Fig. 11.

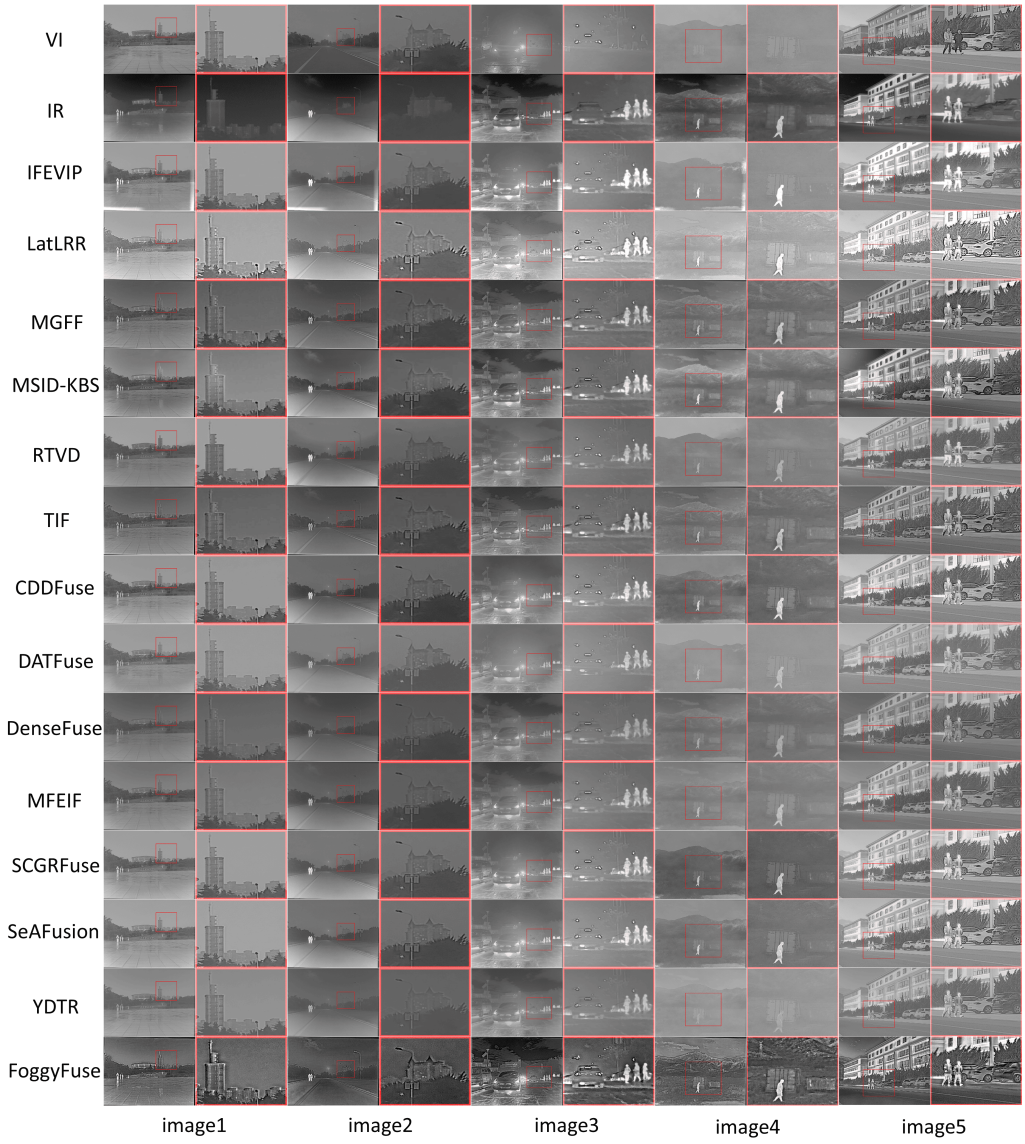


Figure 12: Comparison of visualization effects of different methods on uniform smoke or fog classification of Foggy dataset. Our proposed method can better restore more effective image details under foggy conditions, with clearer edges and more prominent infrared targets to be identified.

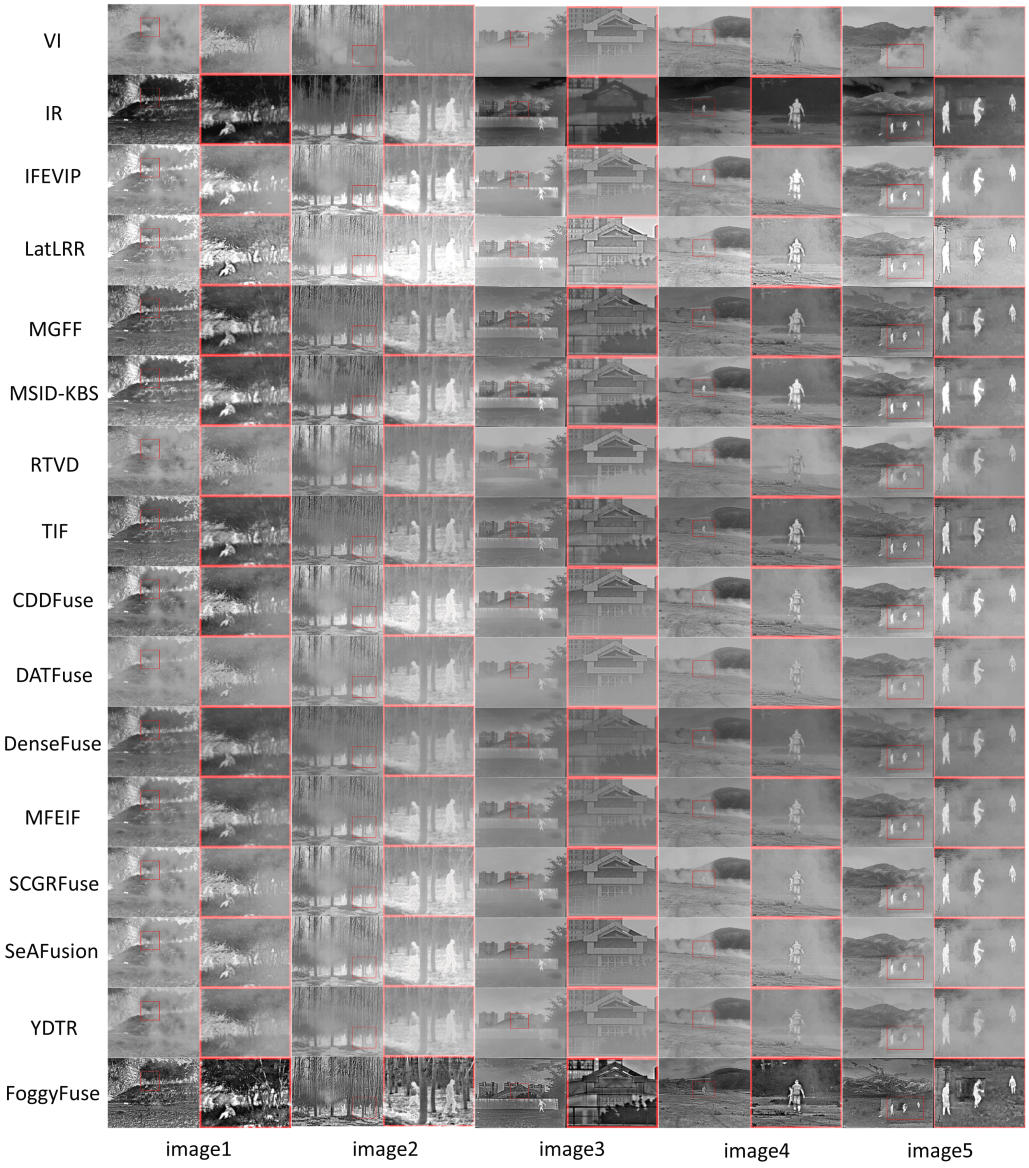


Figure 13: Comparison of visualization effects of different methods on non-uniform smoke or fog classification of Foggy dataset. Our proposed method can capture as much detail information as possible in foggy scenes with large variations.

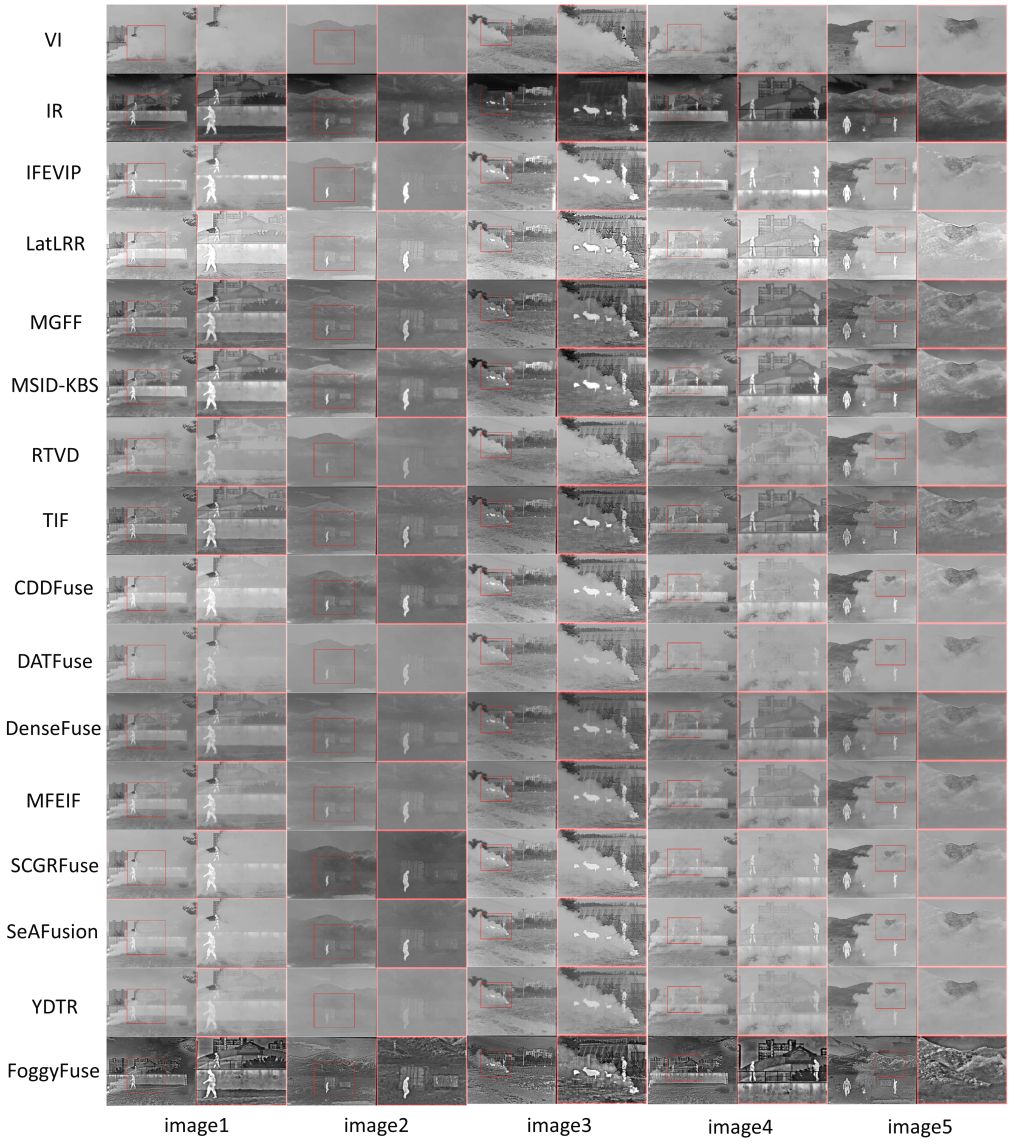


Figure 14: Comparison of visualization effects of different methods on dense smoke or fog classification of Foggy dataset. Our proposed method can make full use of potential detail textures in dense smoke or fog scenes and enhance the final fusion result.

Table 1: *Comparison of FoggyFuse with Fusion Methods on Foggy Dataset. Red and Blue are used to indicate the 1^{st} and 2^{nd} ranks, respectively.*

Method	AG	CE	EI	EN	SCD	SD	SF	VIF
IFEVIP	3.3162	1.5247	35.7963	6.5388	1.3423	32.6775	10.4868	0.583
LatLRR	6.9466	2.6295	72.1500	6.8431	1.6662	36.4967	21.4438	0.6793
MGFF	3.9775	1.8031	41.4592	6.5962	1.6421	26.9782	11.5388	0.6533
MSID-KBS	3.8219	1.5164	39.9922	6.9182	1.6242	35.5908	11.6308	0.7391
RTVD	2.7391	1.6066	28.0010	6.4252	0.9062	28.4466	8.5273	0.673
TIF	4.1332	1.677	43.7676	6.5678	1.6512	27.586	12.4994	0.6293
CDDFuse	4.5782	1.8941	46.8972	6.7187	1.6556	34.1565	13.7882	0.8492
DATFuse	3.2394	1.6972	32.0528	6.3271	1.2926	22.9864	9.6551	0.6879
DenseFuse	2.1637	2.1428	22.4318	6.2415	1.5352	21.5559	6.3530	0.6301
MFEIF	3.0281	1.8092	29.5538	6.3931	1.6485	25.8859	8.178	0.7148
SCGRFuse	4.3228	1.7966	44.9652	6.6908	1.6071	31.9872	13.056	0.8095
SeAFusion	3.9942	1.9297	41.2998	6.5496	1.5812	30.1173	11.7510	0.775
YDTR	2.6802	1.8678	27.4902	6.1835	1.5834	22.5535	8.2828	0.6879
FoggyFuse	10.4736	1.3456	106.7249	7.3055	1.3746	46.4692	26.2423	0.8179

4.2. Experimental setup

All experiments were conducted on a computer equipped with an Intel(R) Core(TM) i5-10200H CPU @ 2.40GHz (8 cores) and an NVIDIA GeForce GTX 1650 Ti GPU. The algorithms were implemented using MATLAB R2023b.

4.3. Comparison methods

We compared the classic and recent mainstream methods in the field of infrared and visible image fusion, including 6 traditional methods and 7 learning-based methods: IFEVIP [53], LatLRR [54], MGFF [55], MSID-KBS [56], RTVD [57], TIF [58], CDDFuse [59], DATFuse [60], DenseFuse [61], MFEIF [62], SCGRFuse [63], SeAFusion [64], YDTR[65].

4.4. Parameter settings

In our proposed method, the NSST decomposition level N is set to 4; the window radius r of the guided filter is set to 30; the regularization parameter λ of the guided filter is set to 1/64; the standard deviation $sigma$ of the Gaussian weight function of the weighted guided filter is set to 5.

Figure 15: RGB fusion image. The RGB images obtained by our proposed method have clear texture and realistic colors.

Table 2: *Average Computation Time of Different Methods on the Foggy Dataset.*

Method	**Average Time (s)**
IFEVIP	0.433
LatLRR	1166.579
MGFF	4.007
MSID-KBS	0.522
RTVD	2.754
TIF	0.401
CDDFuse	2.916
DATFuse	0.117
DenseFuse	0.162
MFEIF	0.355
SCGRFuse	1.531
SeAFusion	0.215
YDTR	1.087
FoggyFuse	165.496

Table 3: *Ablation Experiment Results. Red and Blue are used to indicate the* 1^{st} *and* 2^{nd} *ranks, respectively.*

Method	AG	CE	EI	EN	SD	SF	VIF
No DM	5.0567	1.6726	52.4073	6.6707	29.72	15.3465	0.6556
No EEM	9.0725	1.3584	94.6946	7.2647	45.3501	22.062	0.8031
No AEM	9.7773	1.6146	100.3314	7.1595	44.8855	24.2832	0.8416
FoggyFuse	10.4736	1.3456	106.7249	7.3055	46.4692	26.2423	0.8179

4.5. Evaluation metrics

To evaluate the performance of the proposed algorithm quantitatively, we use 8 indicators: EN [66], CE [67], EI [68], SF [69], SD [70], VIF [71], AG [72], SCD [73] to evaluate different fusion methods.

4.6. Experimental results

Comparing the performance of our proposed method on 8 indicators with the other 8 mainstream traditional and learning-based methods, it is evident that our method outperforms other approaches in 6 of them, and achieves better visualization effect and restoration of details is also far better than other algorithms, which verifies the feasibility of our proposed method.

Qualitative assessment We experimentally compare our proposed method with the current mainstream method on the three categories of uniform smoke or fog, non-uniform smoke or fog, and dense smoke or fog in our Foggy dataset, as shown in Fig. 12,Fig. 13, and Fig. 14. The comparison results show that the FoggyFuse proposed in this paper can effectively perform the fusion task in different foggy scenes and restore texture information obscured by fog, resulting in a fusion with richer texture and better visual quality. Moreover, our method restores the image colors more realistically. The restored image preserves the infrared features while incorporating more color information from the visible image and is enhanced for foggy scenes. The fusion results significantly outperform current mainstream fusion algorithms, as shown in Fig. 15.

Quantitative assessment We compared the indicators of 8 common traditional and learning-based algorithms on the Foggy dataset. The results of the quantitative assessment are shown in Table 1. The results showed that 6 of our indicators ranked first and one indicator ranked second, which is better than the current mainstream algorithms.

From the AG (Average Gradient) metric, FoggyFuse achieves 10.4736, significantly surpassing other methods such as LatLRR (6.9466) and MGFF (3.9775).

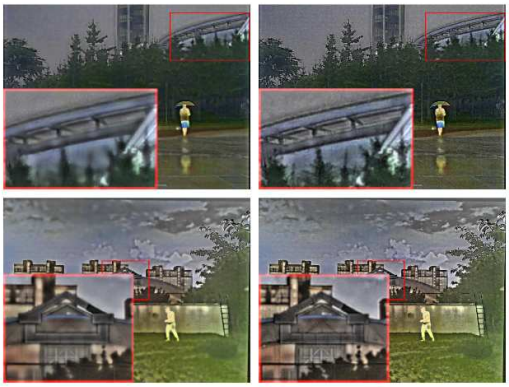
This demonstrates its superior ability to preserve edge details and texture information, thereby enhancing image sharpness. Regarding the EI (Edge Intensity) and EN (Energy) metrics, FoggyFuse also achieves the highest values, 106.7249 and 7.3055, respectively. These results indicate that the proposed method effectively enhances image contrast and retains abundant scene information in foggy conditions. Furthermore, in terms of VIF (Visual Information Fidelity), FoggyFuse achieves 0.8179, which is higher than most existing methods. This result highlights the superior capability of our method in preserving perceptual information. In contrast, traditional methods and most mainstream learning-based methods do not adopt dedicated defogging strategies, leading to lower entropy and edge intensity values. As a result, the fused images tend to suffer from blurring and information loss when processing foggy images. These qualitative indicators also further verify that the fusion method we proposed is very effective in scenes with smoke and fog.

Execution time comparison To evaluate the computational efficiency of different fusion methods, we compared their execution times, as shown in Table 2. Traditional methods, due to their simple computational processes, typically based on mathematical transformations or local filtering, have low computational complexity and faster execution speeds. Learning-based methods, by using pre-trained models and parallel computation, have a clear advantage in processing speed. In comparison with other methods, the proposed approach using Non-subsampled Shearlet Transform (NSST) has a longer execution time. The increased computational cost is primarily attributed to the non-subsampled nature of NSST, which enhances feature extraction by preserving multi-scale and multi-directional structural information. Unlike traditional multi-scale transforms, NSST is unaffected by downsampling, thus avoiding the loss of important high-frequency details. This is the reason why we chose to use it in foggy scene fusion, where preserving more texture information is essential, but this advantage comes at the cost of increased computational complexity.

Despite the longer execution time, the proposed method achieves excellent fusion quality, particularly in preserving texture details and enhancing contrast, which is crucial for infrared-visible image fusion and similar applications. The trade-off between computation time and fusion performance is common in transform-based fusion methods. In the future, improving the algorithm's runtime speed as much as possible will also be a major direction for optimization.

(a) No Dehazing Module

(b) No Edge Enhancement Module



(c) No Auxiliary Enhancement Module

Figure 16: Visualization results of the ablation experiment. The left side of each image set shows the visualization effect without the module, while the right side shows the visualization effect with the module.

4.7. Ablation experiment

Module-level ablation experiment We carried out ablation testing on the three modules we presented: Dehazing Module, Auxiliary Enhancement Module, and Edge Enhancement Module. The indicators we use for ablation experiments are the seven best indicators in Section 4.5. The qualitative results of the ablation experiment are presented in Fig. 16, while the quantitative results are provided in Table 3.

Parameter-level ablation experiment To evaluate the impact of key parameters in the weighted guided filter on the proposed method, we conducted an ablation study by varying the window radius r and the regularization parameter λ.

As shown in Table 4, different parameter settings affect the balance between edge preservation and noise suppression. The results demonstrate that our chosen parameter setting (r = 30, λ = 1/64) achieves the best balance and obtains the best results in multiple indicators of quantitative experiments. It can effectively preserve image details while reducing noise, which is crucial for high-quality fusion.

From the qualitative and quantitative results of the ablation experiment, it can be seen that each module we proposed has greatly improved the visualization effect and indicators of the fused image. The Dehazing Module can effectively restore the color of the image and eliminate the adverse effects of fog on the effective information of the image as much as possible. The Auxiliary Enhancement Module can enhance the texture information of the image, retain part of the original image information, and avoid the loss of infrared information that may be caused by the defogging operation. The Edge Enhancement Module effectively enhances the edge of the image. The ablation experiment shows that the three modules we proposed have a positive impact on the fusion results. Compared with other mainstream algorithms, we have obvious advantages in extreme scenes such as foggy, with impressive visualization effects and fusion results that align better with human visual perception.

Table 4: *Ablation Study on Weighted Guided Filter Parameters. Red and Blue are used to indicate the 1^{st} and 2^{nd} ranks, respectively.*

Parameter Setting	AG	CE	EI	EN	SCD	SD	SF	VIF
$r = 10, \lambda = 1/128$	9.4246	1.3456	96.8189	7.2719	1.3603	45.7465	24.4376	0.8298
$r = 50, \lambda = 1/128$	9.0310	1.3602	93.2131	7.2488	1.3891	45.1877	23.9138	0.8421
$r = 30, \lambda = 1/256$	8.1947	1.3737	85.1416	7.2332	1.3878	44.9761	21.4973	0.8962
$r = 30, \lambda = 1/64$	10.4736	1.3456	106.7249	7.3055	1.3746	46.4692	26.2423	0.8179

4.8. Downstream IVF applications

In advanced computer vision tasks, fused images play a crucial role in visual observation and object detection. To further study the fusion performance of our algorithm and its optimization effect for foggy scenes, we apply the infrared, visible, and fusion images of the past three years' methods to object detection for comparison and analyze the advantages of our algorithm.

Infrared-visible object detection We selected 60 different foggy scene images from the Foggy dataset as the test set, covering various urban environments. We employed YOLOv5 [74] as the object detection model for our experiments. Fig. 17, Fig. 18, and Fig. 19 illustrate the comparison between our fusion results

and those of algorithms developed in the past three years. It is evident that our optimization for foggy scenes significantly benefits object detection. For instance, in Scene 1, due to the presence of fog, the background information in the image becomes blurred. The fusion methods from the past three years fail to address this issue effectively, leading to misclassification in object detection. Specifically, the object detection models misidentify containers in the background as trucks, whereas our method avoids such errors. In Scene 2, the presence of dense smoke or fog has a considerable negative impact on the object detection performance of previous fusion methods. For example, RTVD, SCGRFuse, SeAFusion, and YDTR all mistakenly classify smoke bombs as persons. Additionally, most methods either fail to recognize the sheep obscured by fog or misidentify them as horses. In contrast, our method accurately detects all targets. In Scene 3, the loss of image information caused by fog leads multiple algorithms to fail in identifying a person standing further down the road. Among the algorithms that successfully detect the person, our approach achieves the highest accuracy.

The results of the above object detection experiments demonstrate the effectiveness of our design for foggy scenes. Our method effectively preserves both the semantic information and texture details of the image, leading to a significant improvement in object detection performance.

5. Conclusions

This paper proposes a new method for the fusion of infrared and visible images under foggy conditions based on the saturation line prior. This method optimizes the use of saturation line prior to obtaining a more refined visible transmission map. The obtained transmission map is used to further guide the enhancement optimization of infrared images. The method is simple and efficient. At the same time, we propose an Auxiliary Enhancement Module and an Edge Enhancement Module based on the NSST architecture to make the final fusion result more realistic in color and obtain more detailed textures. Numerous experiments have demonstrated the effectiveness of the proposed method, confirming its superiority over current mainstream algorithms in fusion performance and detail preservation.

6. Declaration of competing interest

All authors disclosed no relevant relationships.

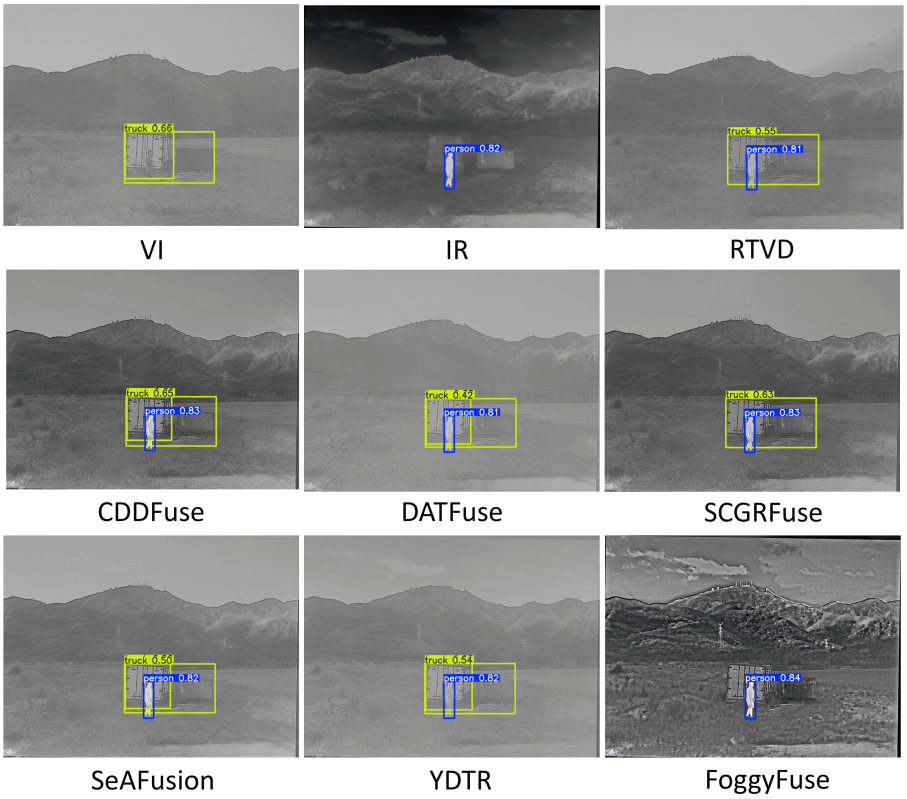


Figure 17: Object detection results on the scene of ”Scene 1” from the Foggy dataset. The blue box represents the detection of a person, and the yellow box represents the detection of a truck.

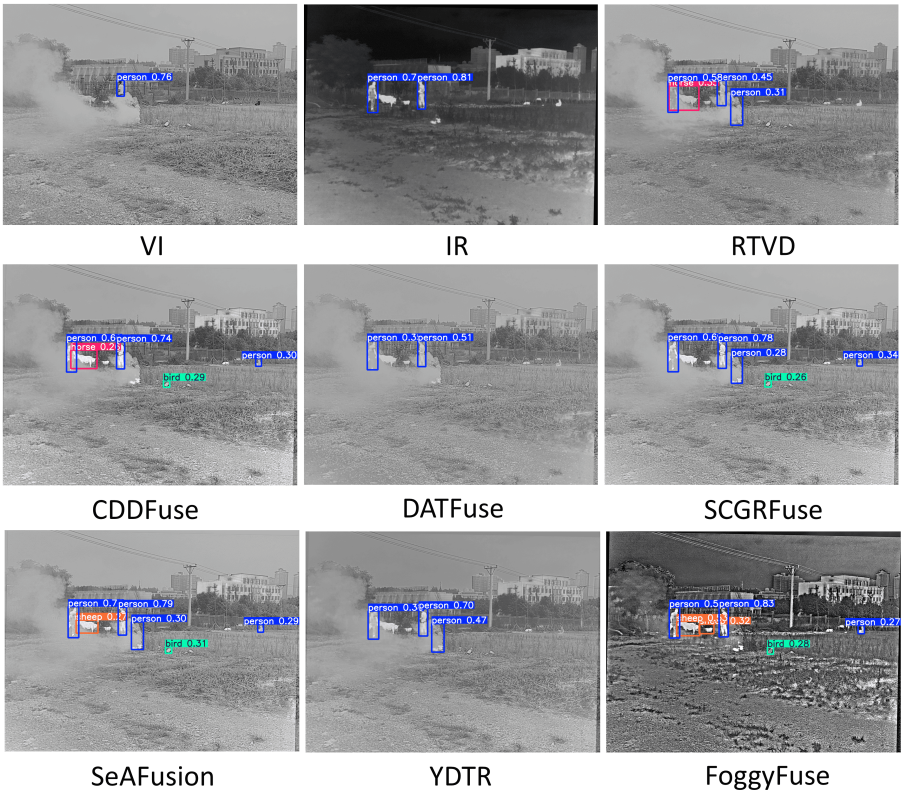


Figure 18: Object detection results on the scene of "Scene 2" from the Foggy dataset. The blue box represents the detection of a person, the pink box represents the detection of a horse, the orange box represents the detection of a sheep, and the green box represents the detection of a bird.

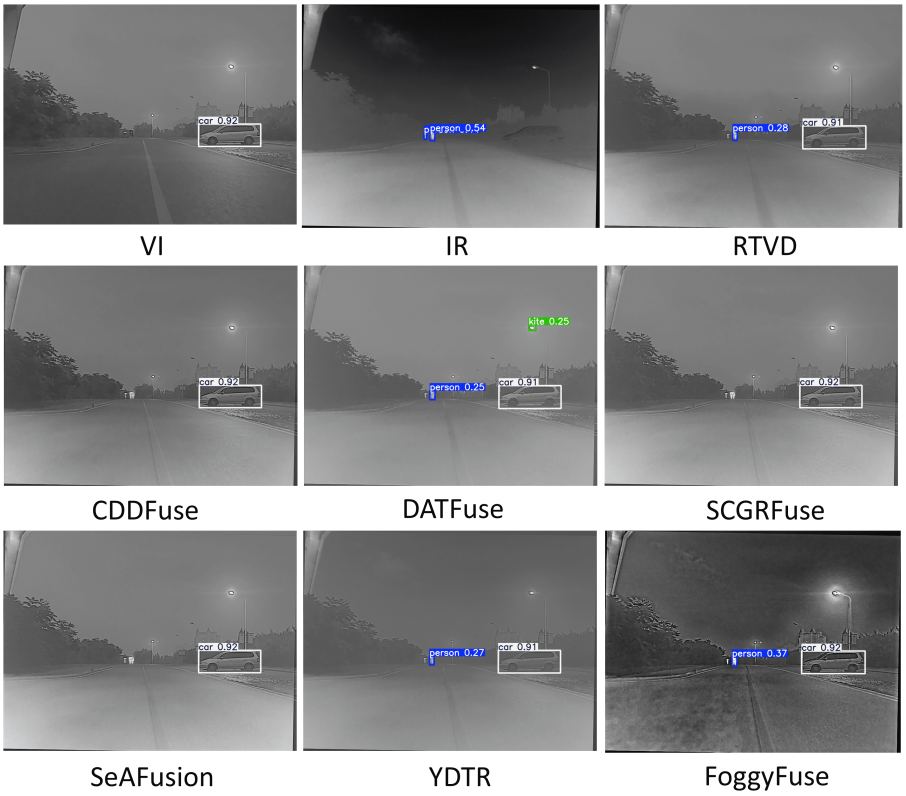


Figure 19: Object detection results on the scene of ”Scene 3” from the Foggy dataset. The blue box represents the detection of a person, the white box represents the detection of a car, and the green box represents the detection of a kite.

7. Date availability

Data will be made available on request.

8. Acknowledgments

This work is supported by the Research Project of Beijing Municipal Natural Science Foundation (No. BJXZ2021-012-00046).

References

[1] Yu Liu, Xun Chen, Hu Peng, and Zengfu Wang. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, 36:191–207, 2017.

[2] Yu Liu, Xun Chen, Rabab K Ward, and Z Jane Wang. Image fusion with convolutional sparse representation. *IEEE signal processing letters*, 23(12):1882–1886, 2016.

[3] Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. Fusiongan: A generative adversarial network for infrared and visible image fusion. *Information fusion*, 48:11–26, 2019.

[4] Yu Liu, Xun Chen, Juan Cheng, Hu Peng, and Zengfu Wang. Infrared and visible image fusion with convolutional neural networks. *International Journal of Wavelets, Multiresolution and Information Processing*, 16(03):1850018, 2018.

[5] Pan Zhu, Yufei Yin, and Xinglin Zhou. Mgrcfusion: An infrared and visible image fusion network based on multi-scale group residual convolution. *Optics Laser Technology*, 180:111576, 2025.

[6] Jinying Zhong, Bin Yang, Yuehua Li, Fei Zhong, and Zhongze Chen. Image fusion and super-resolution with convolutional neural network. In *Pattern Recognition: 7th Chinese Conference, CCPR 2016, Chengdu, China, November 5-7, 2016, Proceedings, Part II 7*, pages 78–88. Springer, 2016.

[7] Hao Li, Shengkun Wu, Lei Deng, Chenhua Liu, Yifan Chen, Hanrui Chen, Heng Yu, Mingli Dong, and Lianqing Zhu. Enhancing infrared and visible image fusion through multiscale gaussian total variation and adaptive local entropy. *The Visual Computer*, 2025.

[8] Jinyuan Liu, Guanyao Wu, Zhu Liu, Di Wang, Zhiying Jiang, Long Ma, Wei Zhong, and Xin Fan. Infrared and visible image fusion: From data compatibility to task adaption. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[9] Peter J Burt and Edward H Adelson. The laplacian pyramid as a compact image code. In *Readings in computer vision*, pages 671–679. Elsevier, 1987.

[10] Haiyan Jin, Licheng Jiao, Fang Liu, and Yutao Qi. Fusion of infrared and visual images based on contrast pyramid directional filter banks using clonal selection optimizing. *Optical Engineering*, 47(2):027002–027002, 2008.

[11] Rekha R Nair and Tripty Singh. Multi-sensor medical image fusion using pyramid-based dwt: a multi-resolution approach. *IET Image Processing*, 13(9):1447–1459, 2019.

[12] Chenhua Liu, Hanrui Chen, Lei Deng, Chentong Guo, Xitian Lu, Heng Yu, Lianqing Zhu, and Mingli Dong. Modality specific infrared and visible image fusion based on multi-scale rich feature representation under low-light environment. *Infrared Physics Technology*, page 105351, 2024.

[13] Glenn Easley, Demetrio Labate, and Wang-Q Lim. Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis*, 25(1):25–46, 2008.

[14] Xueyan Gao and Shiguang Liu. Bcmfifuse: A bilateral cross-modal feature interaction-based network for infrared and visible image fusion. *Remote Sensing*, 16(17):3136, 2024.

[15] Jiaxin Yao, Yongqiang Zhao, Yuanyang Bu, Seong G Kong, and Jonathan Cheung-Wai Chan. Laplacian pyramid fusion network with hierarchical guidance for infrared and visible image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(9):4630–4644, 2023.

[16] Qiao Yang, Yu Zhang, Zijing Zhao, Jian Zhang, and Shunli Zhang. Iaifnet: an illumination-aware infrared and visible image fusion network. *IEEE Signal Processing Letters*, 2024.

[17] Zhiqiang Zhou, Mingjie Dong, Xiaozhu Xie, and Zhifeng Gao. Fusion of infrared and visible images for night-vision context enhancement. *Applied optics*, 55(23):6480–6490, 2016.

[18] Jiaxin Yao, Yongqiang Zhao, Yuanyang Bu, Seong G Kong, and Xun Zhang. Color-aware fusion of nighttime infrared and visible images. *Engineering Applications of Artificial Intelligence*, 139:109521, 2025.

[19] Jiaxin Yao, Yongqiang Zhao, Seong G Kong, and Xun Zhang. Navigating uncertainty: Semantic-powered image enhancement and fusion. *IEEE Signal Processing Letters*, 2024.

[20] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971.

[21] Cosmin Ancuti, Codruta O Ancuti, Christophe De Vleeschouwer, and Alan C Bovik. Night-time dehazing by fusion. In *2016 IEEE international conference on image processing (ICIP)*, pages 2256–2260. IEEE, 2016.

[22] Yongqiang Zhao, Xinbo Qiao, Ning Li, and Quan Pan. "polarzation vision". *Scientia Sinica Informationis*, 54(07):1620–1645, 2024.

[23] Codruta O Ancuti, Cosmin Ancuti, Chris Hermans, and Philippe Bekaert. A fast semi-inverse approach to detect and remove the haze from a single image. In *Asian Conference on Computer Vision*, pages 501–514. Springer, 2010.

[24] Pengyang Ling, Huaian Chen, Xiao Tan, Yi Jin, and Enhong Chen. Single image dehazing using saturation line prior. *IEEE Transactions on Image Processing*, 2023.

[25] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE transactions on image processing*, 25(11):5187–5198, 2016.

[26] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 154–169. Springer, 2016.

[27] Kunal Swami and Saikat Kumar Das. Candy: Conditional adversarial networks based fully end-to-end system for single image haze removal. *arXiv preprint arXiv:1801.02892*, 2018.

[28] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Joint transmission map estimation and dehazing using deep networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(7):1975–1986, 2019.

[29] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proceedings of the european conference on computer vision (ECCV)*, pages 702–717, 2018.

[30] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[31] Linghao Shen, Yongqiang Zhao, Qunnie Peng, Jonathan Cheung-Wai Chan, and Seong G Kong. An iterative image dehazing method with polarization. *IEEE Transactions on Multimedia*, 21(5):1093–1107, 2018.

[32] Ashish V Vanmali and Vikram M Gadre. Visible and nir image fusion using weight-map-guided laplacian–gaussian pyramid for improving scene visibility. *Sādhanā*, 42:1063–1082, 2017.

[33] Vivek Sharma, Jon Yngve Hardeberg, and Sony George. Rgb–nir image enhancement by fusing bilateral and weighted least squares filters. In *Color and Imaging Conference*, volume 25, pages 330–338. Society for Imaging Science and Technology, 2017.

[34] Lex Schaul, Clément Fredembach, and Sabine Süsstrunk. Color image dehazing using the near-infrared. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1629–1632. IEEE, 2009.

[35] Zhang Jingyun, Ding Yifan, Yang Yi, and Sun Jiasong. Real-time defog model based on visible and near-infrared information. In *2016 IEEE international conference on multimedia & expo workshops (ICMEW)*, pages 1–6. IEEE, 2016.

[36] Frederike Dümbgen, Majed El Helou, Natalija Gucevska, and Sabine Süsstrunk. Near-infrared fusion for photorealistic image dehazing. *IS&T EI Proceedings*, 2018.

[37] Wencheng Wang and Faliang Chang. A multi-focus image fusion method based on laplacian pyramid. *J. Comput.*, 6(12):2559–2566, 2011.

[38] Arthur L Da Cunha, Jianping Zhou, and Minh N Do. The nonsubsampled contourlet transform: theory, design, and applications. *IEEE transactions on image processing*, 15(10):3089–3101, 2006.

[39] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing*, 24(11):3522–3533, 2015.

[40] Se Eun Kim, Tae Hee Park, and Il Kyu Eom. Fast single image dehazing using saturation based transmission map estimation. *IEEE Transactions on Image Processing*, 29:1985–1998, 2019.

[41] Lin-Yuan He, Ji-Zhong Zhao, and Du-Yan Bi. Effective haze removal under mixed domain and retract neighborhood. *Neurocomputing*, 293:29–40, 2018.

[42] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International journal of computer vision*, 48:233–254, 2002.

[43] Jinggang Huang and David Mumford. Statistics of natural images and models. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 1, pages 541–547. IEEE, 1999.

[44] Ann B Lee, David Mumford, and Jinggang Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41:35–59, 2001.

[45] Zhengguo Li, Jinghong Zheng, Zijian Zhu, Wei Yao, and Shiqian Wu. Weighted guided image filtering. *IEEE Transactions on Image processing*, 24(1):120–129, 2014.

[46] Kaiming He and Jian Sun. Fast guided filter. *arXiv preprint arXiv:1505.00996*, 2015.

[47] Jinlei Ma, Zhiqiang Zhou, Bo Wang, and Hua Zong. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Physics & Technology*, 82:8–17, 2017.

[48] Zhaobin Wang, Shuai Wang, and Lijie Guo. Novel multi-focus image fusion based on pcnn and random walks. *Neural Computing and Applications*, 29:1101–1114, 2018.

[49] Wei Tan, William Thitøn, Pei Xiang, and Huixin Zhou. Multi-modal brain image fusion based on multi-level edge-preserving filtering. *Biomedical Signal Processing and Control*, 64:102280, 2021.

[50] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5802–5811, 2022.

[51] Linfeng Tang, Jiteng Yuan, Hao Zhang, Xingyu Jiang, and Jiayi Ma. Piafusion: A progressive infrared and visible image fusion network based on illumination aware. *Information Fusion*, 83:79–92, 2022.

[52] Pengyu Zhang, Jie Zhao, Dong Wang, Huchuan Lu, and Xiang Ruan. Visible-thermal uav tracking: A large-scale benchmark and new baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8886–8895, 2022.

[53] Yu Zhang, Lijia Zhang, Xiangzhi Bai, and Li Zhang. Infrared and visual image fusion through infrared feature extraction and visual information preservation. *Infrared Physics & Technology*, 83:227–237, 2017.

[54] Hui Li and Xiao-Jun Wu. Infrared and visible image fusion using latent low-rank representation. *arXiv preprint arXiv:1804.08992*, 2018.

[55] Durga Prasad Bavirisetti, Gang Xiao, Junhao Zhao, Ravindra Dhuli, and Gang Liu. Multi-scale guided image and video fusion: A fast and efficient approach. *Circuits, Systems, and Signal Processing*, 38(12):5576–5605, 2019.

[56] Hui Li, Xianbiao Qi, and Wuyuan Xie. Fast infrared and visible image fusion with structural decomposition. *Knowledge-Based Systems*, 204:106182, 2020.

[57] Jun Chen, Xuejiao Li, and Kangle Wu. Infrared and visible image fusion based on relative total variation decomposition. *Infrared Physics & Technology*, 123:104112, 2022.

[58] Durga Prasad Bavirisetti and Ravindra Dhuli. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Physics & Technology*, 76:52–64, 2016.

[59] Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, and Luc Van Gool. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5906–5916, 2023.

[60] Wei Tang, Fazhi He, Yu Liu, Yansong Duan, and Tongzhen Si. Datfuse: Infrared and visible image fusion via dual attention transformer. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[61] Hui Li and Xiao-Jun Wu. Densefuse: A fusion approach to infrared and visible images. *IEEE Transactions on Image Processing*, 28(5):2614–2623, 2018.

[62] Jinyuan Liu, Xin Fan, Ji Jiang, Risheng Liu, and Zhongxuan Luo. Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(1):105–119, 2021.

[63] Yong Wang, Jianfei Pu, Duoqian Miao, L Zhang, Lulu Zhang, and Xin Du. Scgrfuse: An infrared and visible image fusion network based on spatial/channel attention mechanism and gradient aggregation residual dense blocks. *Engineering Applications of Artificial Intelligence*, 132:107898, 2024.

[64] Linfeng Tang, Jiteng Yuan, and Jiayi Ma. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. *Information Fusion*, 82:28–42, 2022.

[65] Wei Tang, Fazhi He, and Yu Liu. Ydtr: Infrared and visible image fusion via y-shape dynamic transformer. *IEEE Transactions on Multimedia*, 25:5413–5428, 2022.

[66] J Wesley Roberts, Jan A Van Aardt, and Fethi Babikker Ahmed. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing*, 2(1):023522, 2008.

[67] DM Bulanon, TF Burks, and V Alchanatis. Image fusion of visible and thermal images for fruit detection. *Biosystems engineering*, 103(1):12–22, 2009.

[68] B Rajalingam and R Priya. Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis. *International Journal of Engineering Science Invention*, 2(Special issue):52–60, 2018.

[69] Ahmet M Eskicioglu and Paul S Fisher. Image quality measures and their performance. *IEEE Transactions on communications*, 43(12):2959–2965, 1995.

[70] Yun-Jiang Rao. In-fibre bragg grating sensors. *Measurement science and technology*, 8(4):355, 1997.

[71] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444, 2006.

[72] Guangmang Cui, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. *Optics Communications*, 341:199–209, 2015.

[73] V Aslantas and Emre Bendes. A new image quality metric for image fusion: The sum of the correlations of differences. *Aeu-international Journal of electronics and communications*, 69(12):1890–1896, 2015.

[74] Glenn Jocher, Alex Stoken, Jirka Borovec, Liu Changyu, Adam Hogan, Laurentiu Diaconu, Francisco Ingham, Jake Poznanski, Jiacong Fang, Lijun Yu, et al. ultralytics/yolov5: v3. 1-bug fixes and performance improvements. *Zenodo*, 2020.