中图法分类号:TP18;TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2026)02-0573-16

论文引用格式:Song C C, Hu J W and Jin Q W. 2026. High-quality fusion of infrared and visible images by integrating frequency information with atmospheric scattering models. Journal of Image and Graphics, 31(2):0573-0588(宋程程, 胡辑伟, 靳淇文. 2026. 融合频率信息与大气散射模型的高质量红外与可见光图像融合. 中国图象图形学报, 31(2):0573-0588)[DOI:10.11834/jig.250159]

融合频率信息与大气散射模型的高质量红外与可见光图像融合

宋程程,胡辑伟,靳淇文*

武汉理工大学信息工程学院,武汉　430070

摘　要:**目的**　红外与可见光图像融合(infrared and visible image fusion,IVIF)旨在融合多源图像中的互补信息,生成包含丰富纹理细节和明显热辐射特征的融合图像。然而,现有的融合方法未能充分考虑大气传输过程中的光能损失和散射效应,这对图像质量产生了不利影响,尤其是在光照变化显著的复杂场景中。**方法**　为克服这一问题,提出一种创新的IVIF框架,将大气散射物理模型与频域特征组件相结合,精确预测并估计大气散射模型中的关键物理参数——传输图和大气光。通过这一框架,增强可见光图像,减轻散射和能量衰减的影响。此外,结合傅里叶变换与空间通道注意力机制,选择性地增强频域中的幅度和相位特征,有效提升融合图像的纹理保真度和细节表现。**结果**　在RoadScene、TNO(TNO multiband image data collection)、M^3FD(multi-modal fusion face detection)和VT5000(RGB-T salient object detection dataset VT5000)4个公开数据集以及1个极端场景有雾数据集AWMM-100K(all-weather multi-modality image fusion 100K benchmark)上的实验评估显示,本文方法在复杂场景中的融合效果优于现有方法,融合图像具有清晰的目标轮廓和高视觉质量。在定量评估中,与其他深度学习方法相比,本文方法在信息熵、标准差、空间频率、平均梯度、峰值信噪比和相关系数6项指标上均值分别提升7.44%、44.22%、97.89%、91.01%、59.15%和83.68%,表明该方法在目标检测等高层次视觉任务中具有显著优势。**结论**　本研究将传统大气散射模型引入IVIF领域,并结合傅里叶变换与空间通道注意力机制,提升了图像融合的精度和细节保真度,具有重要的实际应用价值。该方法能够有效减轻大气散射带来的负面影响,提升图像质量,适用于复杂动态环境中的实际应用,如目标识别、自动驾驶、遥感监测和安防监控等任务,在实际应用中展现出强大的潜力。

关键词:图像融合;红外与可见光图像;大气散射增强;融合算法;深度学习

High-quality fusion of infrared and visible images by integrating frequency information with atmospheric scattering models

Song Chengcheng, Hu Jiwei, Jin Qiwen*

School of Information Engineering, Wuhan University of Technology, Wuhan　430070, China

Abstract: **Objective** In the domain of image processing, infrared and visible image fusion (IVIF) holds a pivotal position. Its main aim is to synthesize a single image that can optimally integrate the complementary aspects of infrared and visible light images. Visible light images are rich in texture details and color information, while infrared images offer valuable ther-

收稿日期:2025-04-21;**修回日期**:2025-09-06;**预印本日期**:2025-09-13

*通信作者:靳淇文　qiwenjin@whut.edu.cn

基金项目:国家自然科学基金项目(62401416);湖北省自然科学基金项目(2023AFB153)

Supported by:National Natural Science Foundation of China(62401416);Natural Science Foundation of Hubei Province,China(2023AFB153)

mal radiation characteristics. By fusing them, we can create a comprehensive and useful visual representation that is applicable in numerous scenarios such as surveillance, autonomous driving, and remote sensing. However, real-world atmospheric conditions pose significant challenges. During the transmission of light from the source to the sensor, it undergoes various alterations due to atmospheric scattering and energy attenuation. Such alterations lead to issues including light energy loss, scattering effects that blur images, and inaccurate color contrast. In complex outdoor or dynamic environments where the atmosphere has a strong influence, these problems become pronounced, resulting in a significant degradation of image quality and making it difficult to obtain accurate and useful fusion results. Hence, developing a method that can effectively overcome these atmospheric-induced limitations is crucial to enhance the overall quality and practicality of fused images. **Method** To address these challenges, we propose an innovative IVIF framework. This framework combines the physical model of atmospheric scattering with frequency-domain feature components. First, within the atmospheric scattering model, we focus on accurately estimating and predicting two critical parameters: transmission map and atmospheric light. The transmission map reflects how light propagates through the atmosphere and gets attenuated, while the atmospheric light represents the ambient light conditions in the scene. By precisely determining these parameters, we can deeply understand the impact of the atmosphere on images and use this knowledge to enhance visible light images. This enhancement helps in reducing the negative effects of energy loss and scattering. Additionally, to counteract the artifacts and texture loss that might occur as a result of applying the scattering model, we integrate a Fourier transform and a spatial-channel attention mechanism. The Fourier transform allows us to analyze the images in the frequency domain, where we can selectively amplify the amplitude and phase features. Meanwhile, the spatial-channel attention mechanism helps in focusing on the most relevant regions and features in the spatial and channel dimensions, ensuring that the texture fidelity of the fused images is improved and the fine details are well preserved. **Result** We conducted extensive experiments on four publicly available datasets, namely, RoadScene, TNO, M3FD, and VT5000, along with an extreme foggy scene dataset (AWMM-100K). The outcomes of these experiments were quite compelling and provided solid evidence of the superiority of our proposed method. In the qualitative comparison, our method outperformed the existing state-of-the-art fusion techniques, especially when dealing with complex scenes. The fused images generated by our approach boasted clear object contours, which is of great significance for subsequent object recognition and detection tasks. Moreover, these fused images had a high visual quality, presenting a visually appealing and information-rich appearance. They were able to vividly showcase the details that mattered, making them conducive to in-depth analysis and practical applications. Quantitatively, on the RoadScene and TNO datasets, we carried out a meticulous comparison with other deep learning methods in terms of several key metrics that are crucial for evaluating image quality and fusion performance. Specifically, in the aspects of information entropy, standard deviation, spatial frequency, mean gradient, peak signal-to-noise ratio, and correlation coefficient, our method demonstrated remarkable advantages. Detailed calculations and comparisons revealed that compared with other deep learning methods, our method showed increased mean values in these six important indicators by 7.44%, 44.22%, 97.89%, 91.01%, 59.15% and 83.68%. Such significant improvements in these metrics clearly indicated that our method had a strong ability to enhance image quality from multiple dimensions and was highly capable of supporting high-level visual tasks such as object detection. Notably, in the extreme foggy scenes of the AWMM-100K dataset, which are particularly challenging because of the heavy influence of fog, our method shone brightly. It effectively mitigated the adverse effects of fog, significantly reducing the blurriness that is commonly associated with foggy conditions. By doing so, it enhanced the clarity of the fused images and managed to produce higher-quality fused images that were more suitable for further analysis and application in various fields. Overall, the experimental results comprehensively demonstrated the effectiveness and superiority of our proposed IVIF framework in different scenarios and laid a solid foundation for its potential real-world applications. **Conclusion** Our study has introduced a novel approach to IVIF by integrating traditional atmospheric scattering models with deep learning techniques. Through the combination of Fourier transforms and spatial-channel attention mechanisms, we have achieved enhanced image fusion accuracy and improved preservation of critical image details. This approach offers a promising solution for real-world applications in the field of computer vision. The ability to mitigate the effects of atmospheric scattering, along with advanced frequency-domain processing, enables our frame-

work to deliver superior fusion results. Consequently, it is highly suitable for various high-level tasks such as object recognition, segmentation, autonomous driving, remote sensing, and security surveillance in dynamic environments. The potential of our proposed approach for practical use across diverse fields is truly significant. It sets a new benchmark for future research in this area, inspiring researchers to conduct further investigations and improvements in IVIF. This breakthrough will likely lead to the development of more effective and efficient methods for dealing with complex atmospheric conditions and attaining enhanced visual representations, thus advancing the entire domain.

Key words: image fusion; infrared and visible images; atmospheric scattering enhancement; fusion algorithm; deep learning

0 引　言

由于固有的技术限制和成像过程中环境因素的影响，通过同一设备获取的单幅图像往往无法全面呈现场景的所有信息(Liu等，2025)。为了解决这一局限性，图像融合技术应运而生，作为一种有效的解决方案，它能够从多幅源图像中提取有意义的信息，并将其融合成一幅单一的图像。在众多图像融合技术中，红外与可见光图像融合(infrared and visible image fusion，IVIF)已成为最广泛研究和应用的技术之一(Tian等，2024；Toet等，1989)。具体而言，可见光图像高度依赖于环境光照条件，具有较高的空间分辨率、丰富的色彩信息和细腻的纹理细节，使其在光照良好的情况下非常适合感知。然而，在光线不足或恶劣天气条件下，其图像质量会显著下降。与此相对，红外图像捕捉物体辐射的热能，能够有效突出诸如车辆和行人等关键目标，并且不受天气或光照条件的显著影响。然而，红外图像通常具有较低的空间分辨率，且缺乏可见光图像中的细节纹理信息。通过融合这两种模态，得到的图像将两者的互补优势——可见光图像的空间和纹理丰富性，以及红外图像在目标检测中的鲁棒性——结合起来。这样的协同融合提供了更为全面的场景表现，对于多模态显著性检测(Wang等，2023)、目标跟踪(Zhang等，2022)、目标检测(金迪 等，2025)等高级应用具有重要价值。

红外与可见光图像融合(IVIF)方法可分为两大类：传统方法和基于深度学习的技术。传统方法通常遵循3步框架。第1步是特征提取，通过应用特定的数学变换，从源图像中提取显著信息；第2步是特征融合，在此步骤中，预定义的融合规则或策略用于整合提取的特征；第3步是特征重构，通过对融合后的特征应用相应的逆变换，生成融合图像。根据所使用的数学变换类型，传统方法可以进一步细分为多个子类，包括基于多尺度变换的方法(Chen等，2022)、稀疏表示方法(Wei等，2015)、显著性分析方法(Liu等，2017)、子空间聚类方法(Mou等，2013)和混合优化方法(Ma等，2017)。尽管这些方法在实现图像融合目标方面已显示出一定的有效性，但手工设计的变换过程往往较为复杂，而相关的融合规则也常常难以有效处理复杂场景。

随着深度学习的发展，红外与可见光图像融合取得了显著的进展，并且在性能上持续超越了传统方法(Zhang等，2021)，大致可分为基于卷积神经网络(convolutional neural network，CNN)(Ma 等，2021a)、基于自编码器(autoencoder，AE)(Long等，2021)、基于生成对抗网络(generative adversarial network，GAN)(王丽丹 等，2025)和基于变换器(Transformer)(杨天宇 等，2024)的方法。基于CNN的方法采用精确设计的架构和损失函数来促进特征提取、融合和重构，确保高质量的融合输出(王彦舜 等，2023)。Xu等人(2020)提出U2Fusion，一种无监督融合框架，能够自适应地估计源图像的重要性，并保留对各种融合任务至关重要的信息，避免了需要真实数据和特定度量的限制。基于AE的方法利用自编码器进行精确的特征提取和图像重构。Shi等人(2024)通过DAE-Nest(depth-aware autoencoder with nest connection)进一步发展了这些方法，采用结构化的编码器—增强—融合—解码器架构，有效地从红外和可见光图像中提取并融合多尺度特征。基于GAN的方法利用其强大的建模能力，在无监督环境下表现出显著的图像融合效果。Ma等人(2021b)提出GANMcC(generative adversarial network with multi-classification constraints)，该方法通过生成对抗网络与多分类约束相结合，并采用定制内容损失函数，平

衡了可见光图像和红外图像的分布，从而增强了对比度并丰富了纹理表现。基于Transformer的框架近年来在各种视觉任务中取得了显著成功。Rao等人（2023）提出TGFuse（Transformer-based generative adversarial network for infrared and visible image fusion），该方法结合了Transformer模块与对抗学习，捕捉全局交互并精炼红外与可见光图像融合的特征表示。

尽管现有的基于深度学习的方法已展示出有效整合可见光和红外图像中关键互补信息的能力，但仍存在一些未解决的挑战。这些方法大多针对理想的光照条件和大气环境进行定制，因此常常忽视了可见光图像在白天和夜间条件下的光照衰减问题，以及大气悬浮粒子（如雾霾和水蒸气）引起的光散射和吸收造成的质量退化。为缓解可见光图像中的光照衰减，先前的融合方法主要依赖红外信息弥补场景缺失。然而，这种方法往往无法保留和传递低光照可见光图像中丰富的场景细节，偏离了红外—可见光图像融合的根本目标。此外，大气中的悬浮粒子引起的散射效应会导致颜色偏差和失真，表现为低对比度和不准确的色彩表现（Ju等，2017；Guo等，2023）。而早期的融合方法很少考虑大气粒子对图像质量的负面影响，从而影响了融合任务的效果。此外，在低照明条件下进行图像融合时，通常难以保持纹理细节，纹理信息常因过度增强光照而丧失。

为了解决这些问题，本文提出一种基于大气散射模型（atmospheric scattering model，ASM）和傅里叶变换的红外可见光图像融合网络框架，具体包含大气散射增强模块和频率域集成的空间通道注意力模块。大气散射增强模块专门用于恢复可见光图像中的光照和准确的色彩表现。与之相辅的是，频率域集成的空间通道注意力模块旨在有效增强不同粒度下的纹理差异，确保更强的特征保留和更优的融合质量。

本文主要贡献如下：1）提出一种新颖的框架，首先增强全局特征，然后在融合过程中精细化纹理细节，同时强调显著目标。大气散射增强模块旨在基于预测的传输图和大气光，增强可见光图像的光照和色彩对比，有效解决传输过程中光能损失和多重散射的问题。2）设计了频率域集成的空间通道注意力模块，通过空间和通道注意力自适应增强频域中的幅度和相位特征，减少边缘模糊和纹理丧失，同时提高纹理保真度和鲁棒性，实现互补信息的有效融合。3）大量实验表明，本文所提出的融合算法在增强光照、对比度和纹理保真度方面超越了几种最先进的方法。此外，其在下游任务如显著性目标检测（salient object detection，SOD）中的有效性也得到了验证。

1 融合方法

1.1 网络结构

本文提出一种基于大气散射模型和频率域空间通道注意力的红外与可见光图像融合算法，算法结构如图1所示，整体网络结构包含两个基本模块：大气散射增强模块和频率域集成的空间通道注意力模块。

首先，本文将可见光图像输入到大气散射增强模块中，采用编码器—解码器框架预测大气光$\boldsymbol{V}_p^{\mathrm{AL}}$和传输图$\boldsymbol{V}_p^{\mathrm{TM}}$，其中编码器由6个卷积块组成，通过在解码器中的两个平行分支以预测传输图和大气光。传输图表示原始场景辐射与尚未被大气散射的入射光之比。大气光是由大气散布并到达相机的全球照明。然后通过ASM这种数学算法，模拟光在与大气相互作用时的行为方式，考虑散射、吸收和反射因子的影响，以获得保留重要特征并增强照明和可见性的可见光图像$\boldsymbol{V}^E(x)$。

随后，将增强后的可见光图像$\boldsymbol{V}^E(x)$和红外图像$\boldsymbol{I}$分别输入到频率域集成的空间通道注意力模块中，通过两个卷积层对特征进行粗提取，从而提取到浅层特征$\boldsymbol{F}_c^v(x)$和$\boldsymbol{F}_c^i(x)$，接着对$\boldsymbol{F}_c^v(x)$和$\boldsymbol{F}_c^i(x)$进行傅里叶变换，生成对应于幅度和相位分支。将幅度特征和相位特征输入到空间和通道补偿模块（spatial and channel attention，SCA）中以提取深层多尺度特征。空间和通道补偿模块（SCA）主要通过空间域和通道中的注意力补偿机制来改善幅度分量和相位分量中的纹理细节并突出显示显著对象，以此保持显著性和纹理保真度。然后，将经过SCA模块后的幅度和相位特征进行傅里叶反变换，从而促进可见光图像和红外图像在空间域中的重建，得到$\tilde{\boldsymbol{F}}_v$和$\tilde{\boldsymbol{F}}_i$。最后，对$\tilde{\boldsymbol{F}}_v$和$\tilde{\boldsymbol{F}}_i$进行逐元素求和，采用串行残留块将融合图像重

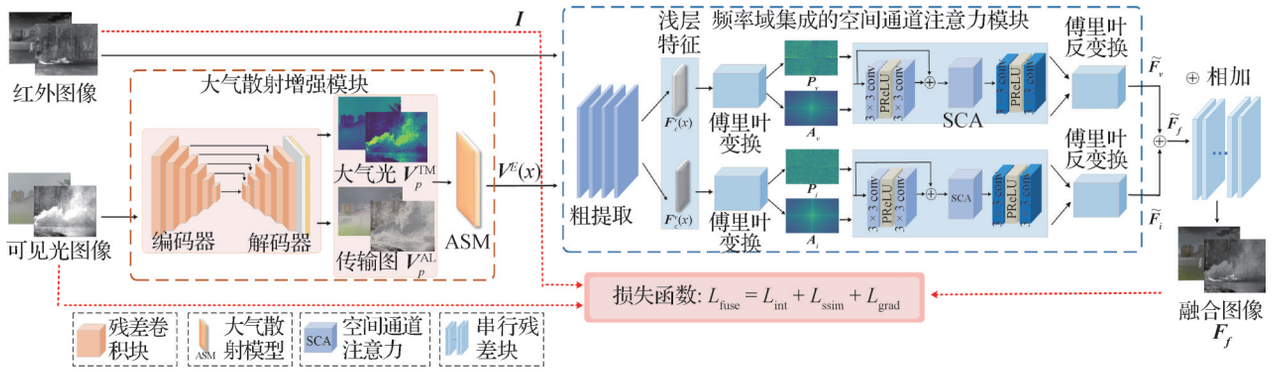

图 1　本文网络框架

Fig. 1　Structure of proposed network

建为 F_f。

1.2　大气散射增强模块

将输入的配对红外图像和可见光图像分别表示为 $\boldsymbol{I} \in \mathbf{R}^{H \times W}$ 和 $\boldsymbol{V} \in \mathbf{R}^{H \times W \times 3}$。采用编码器—解码器网络，其中编码器用 $\mathcal{S}(\cdot)$ 表示，由 6 个层次递进的残差卷积块构成，如图 2 所示。每个残差块均包含 3 × 3 卷积层、归一化层和 GELU（Gaussian error linear unit）激活函数，并结合挤压与激励机制，以提升特征通道的重要性，并自适应调整通道间的特征表示，实现有效的多尺度特征提取。

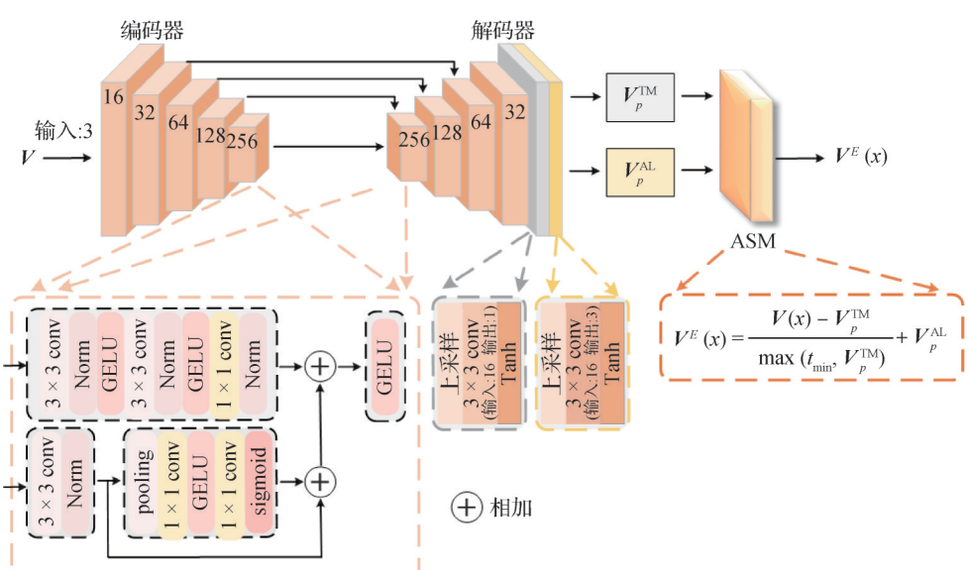

图 2　局部—大气散射增强模块

Fig. 2　Local-atmospheric scattering enhancement module

在编码阶段，将可见光图像 $\boldsymbol{V}$ 输入到编码器 $\mathcal{S}(\cdot)$，通过层级递进的特征提取，使其逐步编码为多尺度表征 $\boldsymbol{F}_V^S$。每一层的通道数依次增加，以提取更加丰富的上下文信息，并通过池化操作逐步压缩空间维度。在较深层的残差块中，采用 1 × 1 卷积进行通道映射，以降低计算复杂度，同时确保不同尺度的特征能够有效融合。具体为

$$\boldsymbol{F}_V^S = \mathcal{S}(\boldsymbol{V}) \tag{1}$$

在解码阶段，采用对称的逐步上采样策略，并设计两个并行的预测分支，分别用于估计传输图和大气光。其中，传输图预测分支 $\mathcal{E}_1(\cdot)$ 采用 1 × 1 卷积映射通道数至 1，并通过 sigmoid 函数进行归一化，使其值范围限定在（0，1），以表征不同区域的光学传输率；大气光预测分支 $\mathcal{E}_3(\cdot)$ 采用 3 × 3 卷积，并利用 Tanh 激活函数，以确保大气光估计的稳定性，输出通道数为 3。传输图 $\boldsymbol{V}_p^{TM}$ 和大气光 $\boldsymbol{V}_p^{AL}$ 表示为

$$\boldsymbol{V}_p^{\mathrm{TM}} = \mathcal{E}_1(\boldsymbol{F}_V^S) \tag{2}$$

$$\boldsymbol{V}_p^{\mathrm{AL}} = \mathcal{E}_3(\boldsymbol{F}_V^S) \tag{3}$$

式中，$\boldsymbol{V}_p^{\mathrm{TM}}$表示原始场景辐射与未被大气散射的入射光之间的比例，用于刻画光在大气传播过程中保持的透射特性。而$\boldsymbol{V}_p^{\mathrm{AL}}$则表示到达相机传感器的全局散射光照，该成分在整幅图像中被视为恒定值。大气散射现象是由光与悬浮在大气中的颗粒相互作用所引起的。这两个成分对于准确估计场景辐射至关重要，从而能够生成具有更优亮度和色彩平衡的增强图像。

随后，$\boldsymbol{V}_p^{\mathrm{TM}}$和$\boldsymbol{V}_p^{\mathrm{AL}}$被输入ASM模型用于模拟光与大气颗粒相互作用的行为。在机器学习的背景下，ASM通常表述为

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

式中，$\boldsymbol{I}(x)$表示含雾图像，$\boldsymbol{J}(x)$为无雾图像，$t(x)$为透射率，A代表大气光照值。获取的$\boldsymbol{V}_p^{\mathrm{TM}}$和$\boldsymbol{V}_p^{\mathrm{AL}}$被输入上述ASM模型。可见光图像$\boldsymbol{V}(x)$对应于含雾图像$\boldsymbol{I}(x)$，而生成的增强图像$\boldsymbol{V}^E(x)$对应于无雾图像$\boldsymbol{J}(x)$，从而得到最终的增强图像。具体为

$$\boldsymbol{V}^E(x) = \frac{\boldsymbol{V}(x) - \boldsymbol{V}_p^{\mathrm{AL}}}{\boldsymbol{V}_p^{\mathrm{TM}}} + \boldsymbol{V}_p^{\mathrm{AL}} \tag{5}$$

然而，由于$\boldsymbol{V}_p^{\mathrm{TM}}$的固有特性，其值受限于范围$\boldsymbol{V}_p^{\mathrm{TM}} \in [0, 1]$。因此，本文为$\boldsymbol{V}_p^{\mathrm{TM}}$设定一个最小阈值$t_{\min}$，即当$\boldsymbol{V}_p^{\mathrm{TM}} < t_{\min}$时，将其限制为$\boldsymbol{V}_p^{\mathrm{TM}} = t_{\min}$。经过处理后的最终增强图像可表示为

$$\boldsymbol{V}^E(x) = \frac{\boldsymbol{V}(x) - \boldsymbol{V}_p^{\mathrm{TM}}}{\max(t_{\min}, \boldsymbol{V}_p^{\mathrm{TM}})} + \boldsymbol{V}_p^{\mathrm{AL}} \tag{6}$$

1.3 频率域集成的空间通道注意力模块

傅里叶变换是分析图像频率特性的关键技术。对于一个具有多个颜色通道的图像$\boldsymbol{x} \in \mathbf{R}^{H \times W \times C}$，每个通道都会分别进行傅里叶变换。该变换将图像转换为复数域表示，其数学表达式为

$$\mathcal{F}(x)(u, v) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(h, w) \mathrm{e}^{-2\pi \mathrm{j}\left(\frac{h}{H}u + \frac{w}{W}v\right)} \tag{7}$$

式中，u和v为傅里叶空间中的坐标。逆傅里叶变换表示为$\mathcal{F}^{-1}$，H和W表示图像的高和宽。傅里叶变换及其逆变换均可通过快速傅里叶变换（fast Fourier transform，FFT）和逆快速傅里叶变换（inverse fast Fourier transform，IFFT）算法高效实现。其对应的幅度和相位分量定义为

$$\boldsymbol{A}(x)(u, v) = \sqrt{R^2(x)(u, v) + I^2(x)(u, v)} \tag{8}$$

$$\boldsymbol{P}(x)(u, v) = \arctan\left(\frac{I(x)(u, v)}{R(x)(u, v)}\right) \tag{9}$$

式中，$R(x)$和$I(x)$分别对应傅里叶变换$\mathcal{F}(x)$的实部和虚部，$\boldsymbol{A}(x)$和$\boldsymbol{P}(x)$分别对应傅里叶逆变换得到的幅度和相位特征。根据傅里叶理论，幅度分量主要编码图像的风格信息，而相位分量则捕捉语义信息（Xu等，2021）。首先，将增强后的可见光图像$\boldsymbol{V}^E(x)$和红外图像$\boldsymbol{I}(x)$输入两个卷积层进行初步特征提取，得到浅层特征$\boldsymbol{F}_c^v(x)$和$\boldsymbol{F}_c^i(x)$。随后，本文对$\boldsymbol{F}_c^v(x)$和$\boldsymbol{F}_c^i(x)$进行傅里叶变换，获取其对应的幅度和相位分量，分别表示为$[\boldsymbol{A}_v, \boldsymbol{P}_v]$和$[\boldsymbol{A}_i, \boldsymbol{P}_i]$，其定义为

$$[\boldsymbol{A}_v, \boldsymbol{P}_v] = \mathcal{F}(F_c^v(x)) \tag{10}$$

$$[\boldsymbol{A}_i, \boldsymbol{P}_i] = \mathcal{F}(F_c^i(x)) \tag{11}$$

然后，$[\boldsymbol{A}_v, \boldsymbol{P}_v]$和$[\boldsymbol{A}_i, \boldsymbol{P}_i]$通过空间和通道补偿模块（SCA）进行处理，以提取深层次的多尺度特征。因此，这些深层次的多尺度特征可以表示为

$$\begin{cases} \boldsymbol{F}_v^A = Conv_{1\times1}([SA(\boldsymbol{A}_v), CA(\boldsymbol{A}_v)]) \\ \boldsymbol{F}_v^P = Conv_{1\times1}([SA(\boldsymbol{P}_v), CA(\boldsymbol{P}_v)]) \end{cases} \tag{12}$$

$$\begin{cases} \boldsymbol{F}_i^A = Conv_{1\times1}([SA(\boldsymbol{A}_i), CA(\boldsymbol{A}_i)]) \\ \boldsymbol{F}_i^P = Conv_{1\times1}([SA(\boldsymbol{P}_i), CA(\boldsymbol{P}_i)]) \end{cases} \tag{13}$$

式中，SA表示空间注意力模块，CA表示通道注意力模块，$\boldsymbol{F}_v^A$、$\boldsymbol{F}_v^P$表示可见光图像对应的幅度和相位多尺度特征，$\boldsymbol{F}_i^A$、$\boldsymbol{F}_i^P$表示红外图像对应的幅度和相位多尺度特征。空间和通道补偿模块专门设计用于通过在空间和通道域中利用基于注意力的补偿机制来增强幅度和相位分量中的纹理细节。有效地突出显著目标，同时保持显著性和纹理保真度。

局部—空间通道补偿模块如图3所示。其中GMaP-C和GMeP-C分别表示沿通道维度的全局最大池化和全局平均池化，而GAP-C表示沿空间维度的全局平均池化。

随后，与可见光和红外图像对应的幅度和相位特征经过逆傅里叶变换（$\mathcal{F}^{-1}$），从而实现其在空间域中的重建。具体为

$$\begin{cases} \tilde{\boldsymbol{F}}_v = \mathcal{F}^{-1}(\boldsymbol{F}_v^A, \boldsymbol{F}_v^P) \\ \tilde{\boldsymbol{F}}_i = \mathcal{F}^{-1}(\boldsymbol{F}_i^A, \boldsymbol{F}_i^P) \end{cases} \tag{14}$$

式中，$\tilde{\boldsymbol{F}}_v$和$\tilde{\boldsymbol{F}}_i$表示重建后的红外与可见光图像。然后，通过以下方式融合来自源图像的保留特征，具体为

$$\tilde{\boldsymbol{F}}_f = \tilde{\boldsymbol{F}}_v \oplus \tilde{\boldsymbol{F}}_i \tag{15}$$

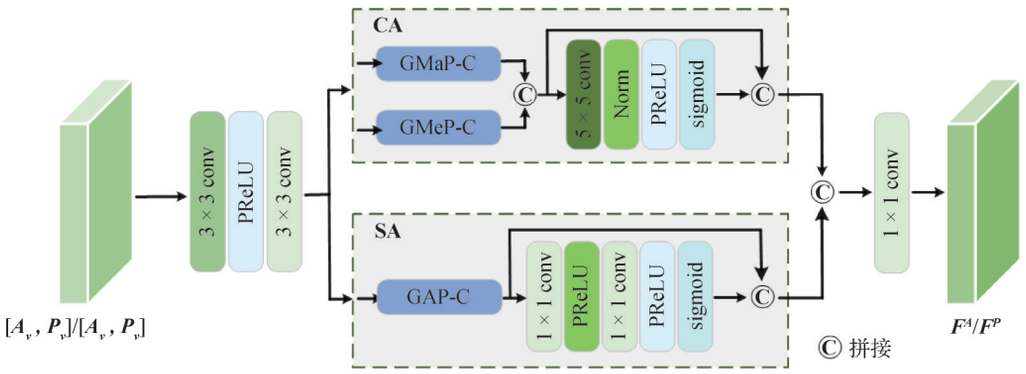

图 3　局部—空间通道补偿模块

Fig. 3　Local-space channel compensation module

式中，⊕ 表示逐元素求和操作。本文采用串行残差块来重建融合图像，记为 $\boldsymbol{F}_f$。

1.4　损失函数

为了保留可见光和红外图像中的显著目标，同时最小化增强图像中的像素级差异，本文设计了一个显著性强度损失函数，定义为

$$\mathcal{L}_{\text{int}} = \left\| (\boldsymbol{\omega}_i \otimes \boldsymbol{I} + \omega_v \otimes \boldsymbol{V}), \boldsymbol{F}_f \right\|_1 \tag{16}$$

此外，受 Liu 等人（2021）的启发，本文设计了一个结构相似性损失函数，以确保增强图像与原始图像保持结构相似性。该损失函数的定义为

$$\mathcal{L}_{\text{ssim}} = 1 - SSIM((\boldsymbol{\omega}_i \otimes \boldsymbol{I} + \omega_v \otimes \boldsymbol{V}), \boldsymbol{F}_f) \tag{17}$$

在式（16）和式（17）中，$\boldsymbol{\omega}_i$ 和 $\boldsymbol{\omega}_v$ 分别是通过 $\boldsymbol{\omega}_i = \boldsymbol{S}_{f_i} / \left(\boldsymbol{S}_{f_i} - \boldsymbol{S}_{f_v} \right)$ 和 $\boldsymbol{\omega}_v = 1 - \boldsymbol{\omega}_i$ 获得的可见光图像和红外图像的加权图。$\boldsymbol{S}_{f_i}$ 和 $\boldsymbol{S}_{f_v}$ 分别表示通过计算得到的可见光和红外图像的显著性矩阵（Ghosh 等，2019）。

在创建融合图像的过程中，本文旨在实现源图像的显著信息与细节的平衡，从而最大化集成源图像与融合图像之间共享的信息。为了实现这一目标，本文设计了一个基于梯度方向的细粒度纹理损失函数，定义为

$$\mathcal{L}_{\text{grad}} = \left\| \nabla \boldsymbol{F}_f, \max\left(\nabla \boldsymbol{I}, \nabla \boldsymbol{V} \right) \right\|_1 \tag{18}$$

式中，∇ 表示拉普拉斯梯度算子，max (·) 表示从两个源图像中聚合细粒度纹理的最大值。

网络的整体损失可以表示为

$$\mathcal{L}_{\text{fuse}} = \alpha_1 \mathcal{L}_{\text{int}} + \alpha_2 \mathcal{L}_{\text{ssim}} + \alpha_3 \mathcal{L}_{\text{grad}} \tag{19}$$

式中，α_1、α_2 和 α_3 表示目标函数的超参数，分别控制显著性强度损失、SSIM 损失和梯度损失在总损失中的贡献程度。

2　实　验

2.1　数据集

为了评估本文算法的有效性，在 3 个广泛认可的基准数据集 RoadScene（Xu 等，2020）、TNO（TNO multiband image data collection）（Toet 和 Hogervorst，2012）、M^{3}FD（multi-model fusion face detection）（Liu 等，2022）以及有雾数据集 AWMM-100K（all-weather multi-modality image fusion look benchmark）（Li 和 Wu，2024）上进行实验。为了进一步展示模型在显著性目标检测（SOD）任务中的泛化能力，本文使用 VT5000（RGB-T salient object detection dataset VT5000）和 VT1000（Zhang 等，2020）数据集，通过将生成的融合图像作为第三输入，直接预测最终的显著性图。同时，为了验证将通过本文算法得到的融合图像作为第三输入在 SOD 任务中的有效性，本文还分析了在 VT5000 数据集上的融合结果。VT5000 数据集包含 5 000 对红外热成像和可见光图像，以及专门为 SOD 设计的相应二值标签，其中，2 500 对用于训练，其余 2 500 对用于测试。

2.2　实验配置

本文提出的融合方法使用 PyTorch 框架实现，并在一台单 GPU 的 NVIDIA GTX 3090 上执行。在训练阶段，本文模型使用来自 VT5000 数据集训练集中的 2 500 对图像进行训练，每幅图像的分辨率为 480 × 640 像素。此外，除了在 VT5000 测试集上评估本文方法的有效性外，还严格评估了在 M^{3}FD、RoadScene、TNO 和 AWMM-100K 数据集上的泛化性能。

整体参数使用 Adam 进行优化，配置为 $\beta_1 = 0.9$ 和 $\beta_2 = 0.99$，经过 18 K 次迭代，每次迭代的批量大

小为4。初始学习率设置为10^{-6}，并在每0.625 K次迭代时按0.97的因子逐步衰减。式(6)中的超参数$t_{\min}$设置为0.15。根据实验中的超参分析(如表1所示)，式(19)中的超参数α_1、α_2和α_3分别设置为20、5和5。

表1 α_1、α_2和α_3的超参数分析

Table 1 Hyperparametric analysis on α_1, α_2 and α_3

超参数	EN ↑	SD ↑	SF ↑
α_1 = 5	6.960	37.797	14.241
α_1 = 10	6.890	35.291	14.266
α_1 = 15	6.933	36.375	14.624
α_1 = 20	**7.036**	**46.948**	**14.751**
α_1 = 25	6.837	34.008	14.376
α_2 = 1	6.882	36.166	9.736
α_2 = 3	6.845	35.364	10.183
α_2 = 5	**7.036**	**46.948**	**13.259**
α_2 = 7	6.800	33.461	8.902
α_2 = 10	6.793	31.907	8.798
α_3 = 1	6.930	37.029	9.809
α_3 = 3	6.904	35.130	10.543
α_3 = 5	**7.036**	**46.948**	**14.751**
α_3 = 7	6.931	36.212	14.402
α_3 = 10	6.830	34.007	14.181

注：加粗字体分别表示α_1、α_2、和α_3中的最优结果。“↑”表示值越高越好。

2.3 评价指标

本文选择6个统计评估指标进行量化评估，包括熵(entropy, EN)、标准差(standard deviation, SD)、空间频率(spatial frequency, SF)、平均梯度(average gradient, AG)、峰值信噪比(peak signal to noise ratio, PSNR)和相关系数(correlation coefficient, CC)。EN常用于量化图像的复杂度。较高的图像熵表示信息更加复杂、纹理更丰富，而较低的熵则表示图像结构较为简单。SD是像素强度值的离散度量，反映了图像中亮度或颜色的变化。SF指的是图像中强度或颜色在空间上的变化率，代表了图像中的细节或纹理水平。AG是用于测量图像清晰度或锐度的指标，通过计算相邻像素之间的平均强度变化率来实现，通常用于评估图像边缘的对焦和对比度。PSNR是一个广泛使用的图像质量评估指标，通过将重建图像或压缩图像与原始图像进行比较评估图像质量。较高的PSNR值通常表示图像质量更好，失真或噪声较少。CC通过测量对应像素值之间的线性关系量化两幅图像的相似度，这一指标常用于图像配准和比较任务。

2.4 对比实验

为了全面评估本文方法的性能，进行了定量和定性实验，与7种最先进的方法进行性能对比，包括CrossFusion（Li和Wu, 2024）、EMMA(equivariant multi-modality image fusion)(Zhao等, 2024)、IRFS(interactively reinforced paradigm for joint fusion and saliency detection)(Wang等, 2023)、PAIF(perception aware infrared visible image fusion)(Wang等, 2023)、LRRNet(learning representation with regularization network)(Li等, 2023)、SwinFusion(Ma等, 2022)和SuperFusion(Tang等, 2022)。

2.4.1 定性实验

在M^3FD、TNO和VT5000数据集上的视觉结果如图4—图7所示。可以看出，PAIF、SwinFusion和CrossFusion的融合效果都有严重的物体模糊问题，EMMA在保持可见光图像纹理保真度方面存在困难，相较于IRFS和LRRNet，本文方法在保持红外图像中的纹理细节特征的同时，也保留了明显的边缘。此外，从整体图像中也可以看出，本文方法有效地增强了可见光图像的色彩对比度。

考虑到本文将使用从本文方法获得的输出作为显著性目标检测(SOD)任务的第三方输入，以展示模型的多功能性，因此在VT5000测试集上进行了视觉比较，结果如图8所示。

为了进一步展示本文方法的鲁棒性，在AWMM-100K数据集上进行定量分析，如图9所示，该数据集包含了在雾霾条件下捕获的图像。在实验中，本文相比其他融合方法(如PAIF、SwinFusion和CrossFusion)表现出优越的性能，尤其在提高对比度和保留雾霾环境中的细节方面表现突出。我们观察到在由本文算法生成的融合图像中，物体与背景之间的对比度显著增强，过度曝光得到了有效抑制。这一增强使得与其他当前多任务框架相比，显著性图的预测更加准确，如图10和图11所示。

2.4.2 定量实验

对于提供明显训练集和测试集划分的数据集，

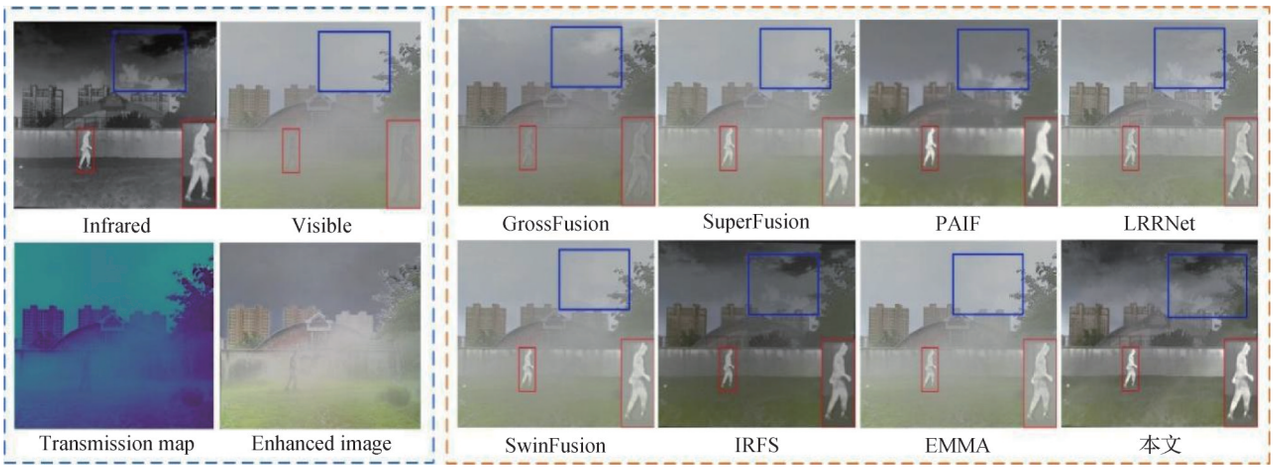

图 4　M^3FD 数据集中的“00349. png”视觉对比

Fig. 4　Visual comparison for “00349. png” in M^3FD dataset

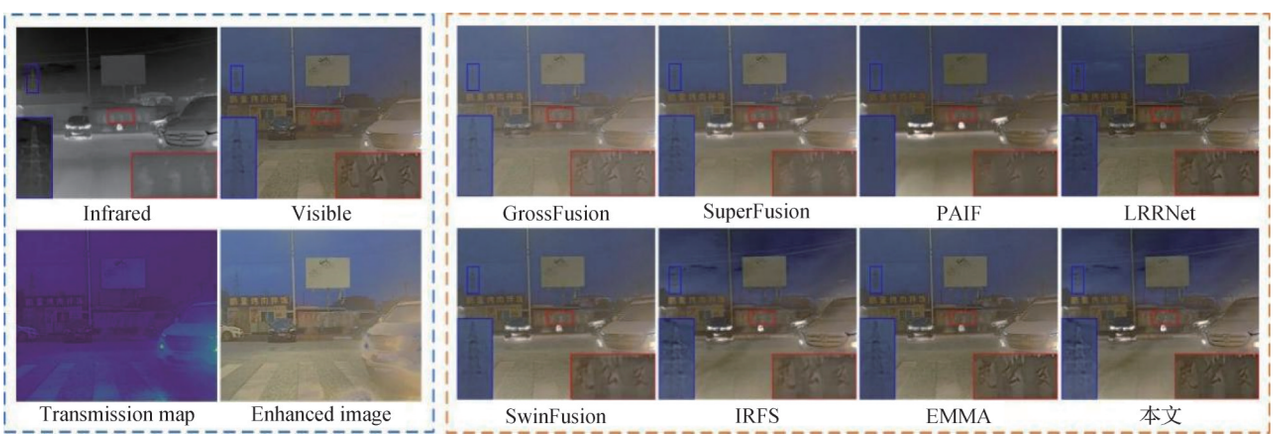

图 5　M^3FD 数据集中的“00671. png”视觉对比

Fig. 5　Visual comparison for “00671. png” in M^3FD dataset

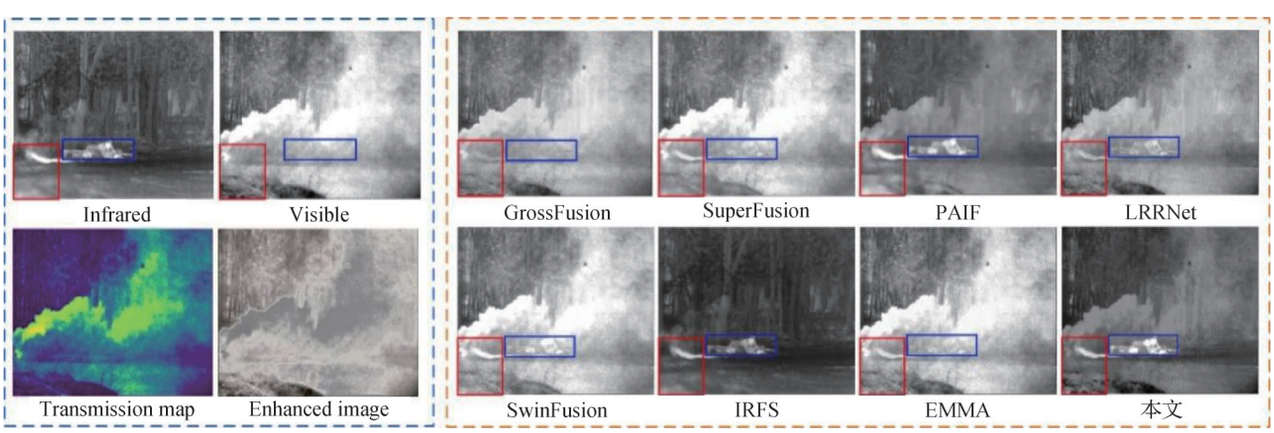

图 6　TNO 数据集中的“008. png”视觉对比

Fig. 6　Visual comparison for “008. png” in TNO dataset

如 M^3FD、RoadScene 和 VT5000 数据集，本文使用指定测试集对不同算法进行评估。相比之下，对于缺乏预定义分割的 TNO 数据集，从中随机选择 37 对图像进行定量比较，定量分析结果如表 2—表 5 所示，指标中 ↑/↓ 表示较大/较小值更为优越。如表 2—表 4 所示，本文方法在 M^3FD 和 TNO 测试集上的各项指标始终超越最先进的其他方法，突出了本文融合结果的丰富性和保真度，更高的 EN 和 SD 得分反映

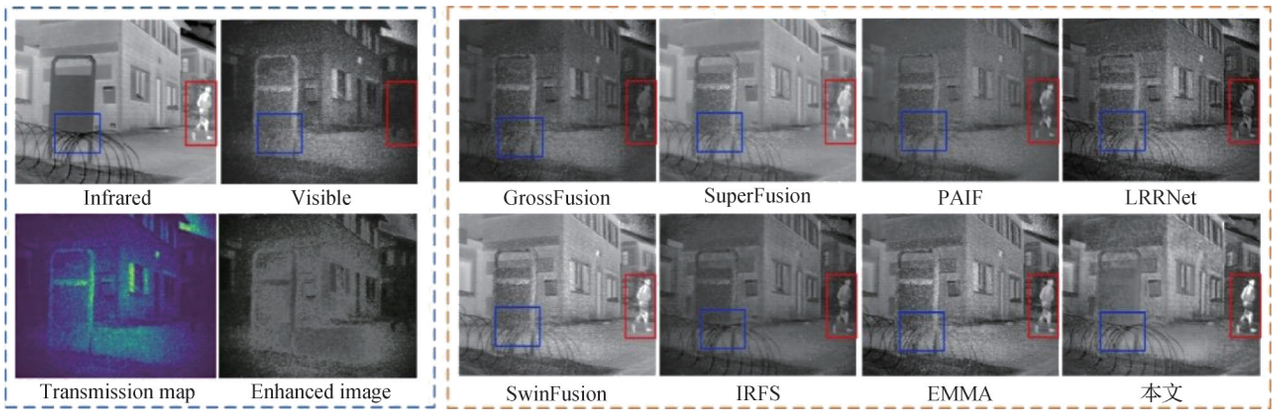

图7　TNO数据集中的“015. png”视觉对比

Fig. 7　Visual comparison for “015. png” in TNO dataset

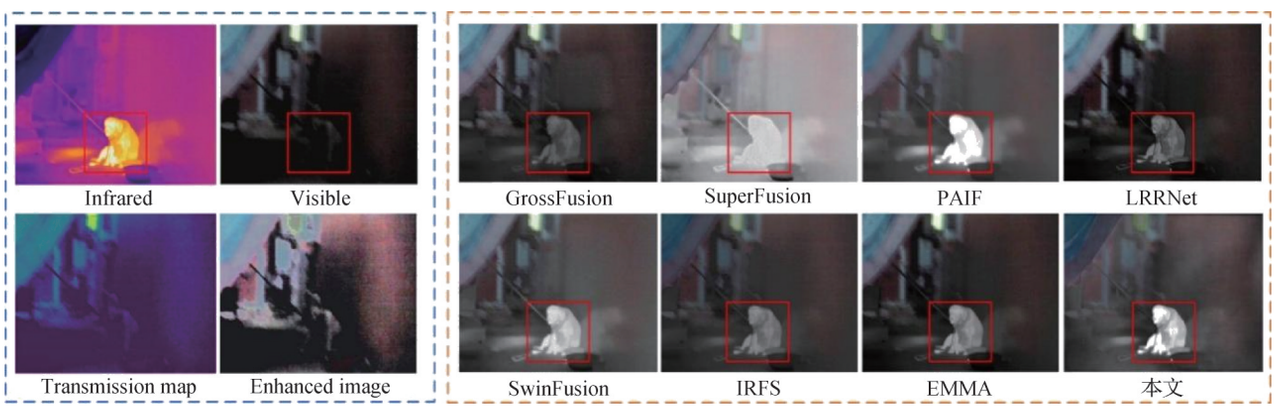

图8　VT5000数据集中的“794. png”视觉对比

Fig. 8　Visual comparison for “794. png” in VT5000 dataset

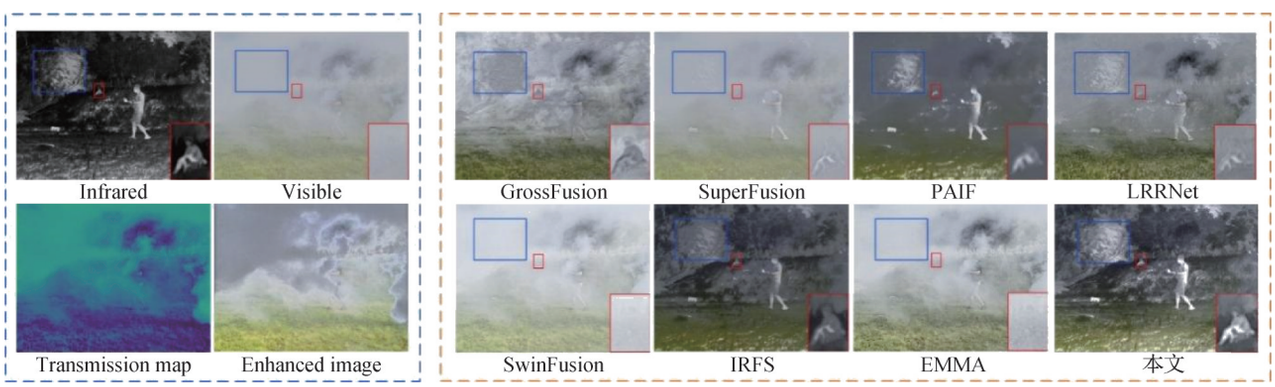

图9　AWMM-100K数据集中的视觉对比

Fig. 9　Visual comparison in AWMM-100K dataset

了本文方法能够有效地将源图像中的细致纹理细节传递到融合输出中，从而提高融合表示的信息熵和结构清晰度。此外，在SF和AG指标上具有明显优势，表明了其在纹理保留方面的全面性，这与本文的定性评估一致。

表6展示了本文方法在有雾场景下AWMM-100K数据集上的定量结果，突出了本文方法在复杂环境中的表现。如表6所示，本文方法在信息熵（SF）、标准差（AG）和峰值信噪比（PSNR）等指标上优于其他融合算法，证明了其在有雾环境中提升图像质量的能力。此外，在VT5000数据集（如表5所示）上，本文方法取得了最高的EN和SD得分，突出了在SOD任务中引入第三方输入后所带来的优越纹理复杂度和细节保真度，从而为显著性目标检测

表 2　不同融合方法在 RoadScene 数据集上的定量比较

Table 2　Quantitative comparison of different fusion methods on the RoadScene dataset

方法	EN ↑	SD ↑	SF ↑	AG ↑	PSNR ↑ /dB	CC ↑
SwinFusion	7.002	44.032	11.744	4.430	13.174	0.623
SuperFusion	6.990	41.296	11.652	4.367	13.978	0.591
IRFS	7.253	44.364	12.470	4.689	11.122	0.478
PAIF	6.892	39.183	7.740	2.795	14.678	0.600
LRRNet	7.139	42.608	11.954	4.587	11.758	0.621
CrossFusion	7.180	44.709	11.459	4.374	12.038	0.539
EMMA	7.401	49.686	13.895	5.489	13.913	0.644
本文	**7.405**	**56.509**	**15.053**	**5.820**	**17.701**	**0.878**

注：加粗、下划线字体表示各列最优、次优结果。

表 3　不同融合方法在 TNO 数据集上的定量比较

Table 3　Quantitative comparison of different fusion methods on the TNO dataset

方法	EN ↑	SD ↑	SF ↑	AG ↑	PSNR ↑ /dB	CC ↑
SwinFusion	6.920	40.929	10.951	4.241	13.394	0.527
SuperFusion	6.788	38.076	9.342	3.558	13.830	0.508
IRFS	6.809	38.720	10.266	3.751	14.652	0.505
PAIF	6.331	34.076	7.311	2.230	14.230	0.516
LRRNet	7.024	42.749	9.595	3.828	13.740	0.501
CrossFusion	6.992	40.531	9.788	3.721	13.704	0.451
EMMA	7.018	46.900	12.378	4.850	13.934	0.519
本文	**7.029**	**46.948**	**12.978**	**4.973**	**14.751**	**0.585**

注：加粗、下划线字体表示各列最优、次优结果。

表 4　不同融合方法在 M^3FD 数据集上的定量比较

Table 4　Quantitative comparison of different fusion methods on the M^3FD dataset

方法	EN ↑	SD ↑	SF ↑	AG ↑	PSNR ↑ /dB	CC ↑
SwinFusion	6.799	35.656	13.631	4.588	13.220	0.519
SuperFusion	6.692	32.382	11.296	3.772	13.295	0.482
IRFS	6.537	28.460	8.392	2.844	13.062	**0.549**
PAIF	6.520	37.627	8.209	2.449	14.304	0.479
LRRNet	6.440	27.185	10.704	3.606	14.377	0.541
CrossFusion	6.588	30.897	11.596	3.860	13.435	0.421
EMMA	5.857	36.388	**14.923**	**5.130**	13.675	0.476
本文	**7.029**	**41.109**	13.325	4.678	**14.395**	0.483

注：加粗、下划线字体表示各列最优、次优结果。

表 5　不同融合方法在 VT5000 数据集上的定量比较

Table 5　Quantitative comparison of different fusion methods on the VT5000 dataset

方法	EN ↑	SD ↑	SF ↑	AG ↑	PSNR ↑ /dB	CC ↑
SwinFusion	6.771	40.609	8.333	3.368	13.046	0.461
SuperFusion	6.245	31.014	7.455	2.960	10.118	0.606
IRFS	6.454	29.205	4.932	1.986	10.836	**0.687**
PAIF	6.556	33.646	5.456	1.945	15.006	0.641
LRRNet	6.902	42.206	9.042	3.443	14.832	0.659
CrossFusion	6.961	41.909	8.753	3.589	14.010	0.669
EMMA	**7.068**	42.805	9.042	**3.832**	14.760	0.654
本文	6.997	**42.905**	**9.118**	3.141	**15.018**	0.623

注：加粗、下划线字体表示各列最优、次优结果。

表 6　不同融合方法在 AWMM-100K 数据集上的定量比较

Table 6　Quantitative comparison of different fusion methods on the AWMM-100K dataset

方法	EN ↑	SD ↑	SF ↑	AG ↑	PSNR ↑ /dB	CC ↑
SwinFusion	6.973	39.124	4.978	1.720	7.383	0.669
SuperFusion	6.968	39.644	4.966	1.700	7.312	0.662
IRFS	6.449	27.266	4.310	1.802	8.976	0.705
PAIF	6.146	21.661	4.466	1.110	9.203	0.713
LRRNet	6.935	37.264	4.912	1.799	9.185	0.651
CrossFusion	7.209	44.844	6.485	2.398	9.002	0.561
EMMA	**7.391**	**52.981**	6.863	2.432	8.130	0.662
本文	7.219	41.305	**7.057**	**2.456**	**9.258**	**0.721**

注：加粗、下划线字体表示各列最优、次优结果。

提供了更具信息量的源图像。尽管 AG 指标未达到最佳值，但最佳与次佳得分之间的性能差距微乎其微。

2.4.3　效率及模型参数对比

本文对 7 种先进的红外与可见光图像融合方法的模型大小和计算效率进行全面比较。在效率评估中，排除了数据加载和预处理时间，使用 M^3FD 数据集（包含 300 对原始红外—可见光图像）作为基准，评估了每种方法的推理时间。如表 7 所示，尽管 IRFS 和 LRRNet 展现了最小的模型大小和最短的平均推理时间，但它们的定性和定量表现不如本文模型。此外，结果还表明，尽管本文模型相较于其中 3 种算法模型稍大，但在推理时间上明显低于其他 5 种算法，主要由于实现框架的差异。本文模型大

小的增加主要源于在编码器和解码器中集成了更深的大气散射模型。尽管如此，本文模型依然展现了竞争力的效率，平均每幅图像的推理时间低于0.03 s。

表7　在RoadScene数据集上300对图像的推理时间

Table 7　The inference time on 300 pairs of images in the RoadScene dataset

方法	参数量/MB ↓	时间/s ↓
SwinFusion	0.974	3.634
SuperFusion	1.962	0.417
IRFS	0.242	**0.005**
PAIF	44.865	0.176
LRRNet	**0.197**	**0.005**
CrossFusion	20.801	0.990
EMMA	1.516	0.071
本文	4.930	0.024

注：加粗、下划线字体表示各列最优、次优结果。

2.4.4　显著性目标检测（SOD）任务结果对比

为了全面研究红外与可见光图像融合技术对显著性目标检测（SOD）性能的影响，本文使用了VT5000和VT1000数据集进行实验，这两个数据集均包含为SOD任务专门设计的二值真实标签。本文采用FC^{2}Net作为SOD的基线模型，将本文方法的融合结果以及7种最先进的融合技术的结果作为输入，生成最终的显著性图。图10和图11展示了在VT5000和VT1000数据集上的预测显著性图的定性可视化，而表8和表9则提供了定量评估。

在评估指标方面，本文选择常用的S_α、F_β、E_ξ和M综合评估SOD性能。具体而言，S_α表示结构相似性度量，F_β结合了精度和召回率，量化了预测显著性图与真实标签之间的对齐程度，E_ξ表示增强对齐度量，M代表均方误差，提供了一种平衡的方式评估显著性图在结构、对齐和像素级精度方面的质量。

表8和表9的定量结果表明，在VT5000和VT1000数据集上，采用本文模型生成的融合图像作为第三方输入，始终在4个主要SOD评估指标上取得了顶级的表现。这一结果表明，预测的显著性图与真实标签之间的对齐程度更高，均方误差显著减少，相比其他方法表现更佳。这些发现凸显了本文模型在红外与可见光图像融合中的有效性，进一步证明了其在提高显著性检测精度方面的强大能力，且具有高度的鲁棒性和可靠性。

表8　显著性目标检测（SOD）任务在VT5000数据集上的定量结果

Table 8　Quantitative results of the SOD tasks on the VT5000 dataset

方法	S_α ↑	F_β ↑	E_ξ ↑	M ↓
SwinFusion	0.847	0.780	0.885	0.049
SuperFusion	0.871	0.791	0.896	0.037
IRFS	0.863	0.823	0.902	0.041
PAIF	0.852	0.797	0.885	0.044
LRRNet	0.846	0.779	0.907	0.048
CrossFusion	0.869	0.803	0.887	0.036
EMMA	0.847	0.781	0.916	0.047
本文	**0.881**	**0.839**	**0.925**	**0.033**

注：加粗、下划线字体表示各列最优、次优结果。

表9　显著性目标检测（SOD）任务在VT1000数据集上的定量结果

Table 9　Quantitative results of the SOD tasks on the VT1000 dataset

方法	S_α ↑	F_β ↑	E_ξ ↑	M ↓
SwinFusion	0.912	0.881	0.928	0.026
SuperFusion	0.921	0.887	0.936	0.021
IRFS	0.914	0.870	0.925	0.024
PAIF	0.907	0.880	0.934	0.028
LRRNet	0.911	0.881	0.872	0.781
CrossFusion	0.917	0.872	0.929	0.022
EMMA	0.847	0.781	0.885	0.048
本文	**0.923**	**0.896**	**0.941**	**0.020**

注：加粗、下划线字体表示各列最优、次优结果。

如图10和图11所示，利用本文模型生成的融合图像作为输入生成的显著性图具有更高的准确性，并显著减少了错误检测。图11展示了从VT5000数据集中随机选择的一个示例进行显著性预测比较。与7种其他融合方法生成的图相比，本文模型在路障基部的椭圆轮廓的划分更加精确，并且在路障顶部尖端轮廓的表现更加清晰。如图10所示，本文模型的预测在与真实标签（ground truth，GT）的对齐方

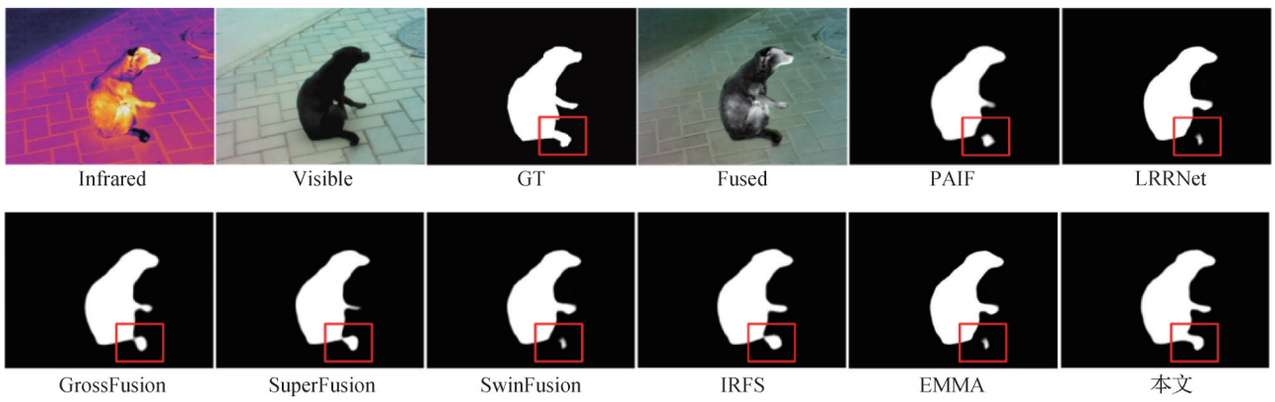

图 10　显著性目标检测在 VT1000 数据集中的视觉对比

Fig. 10　Visual comparison of saliency object detection in the VT1000 dataset

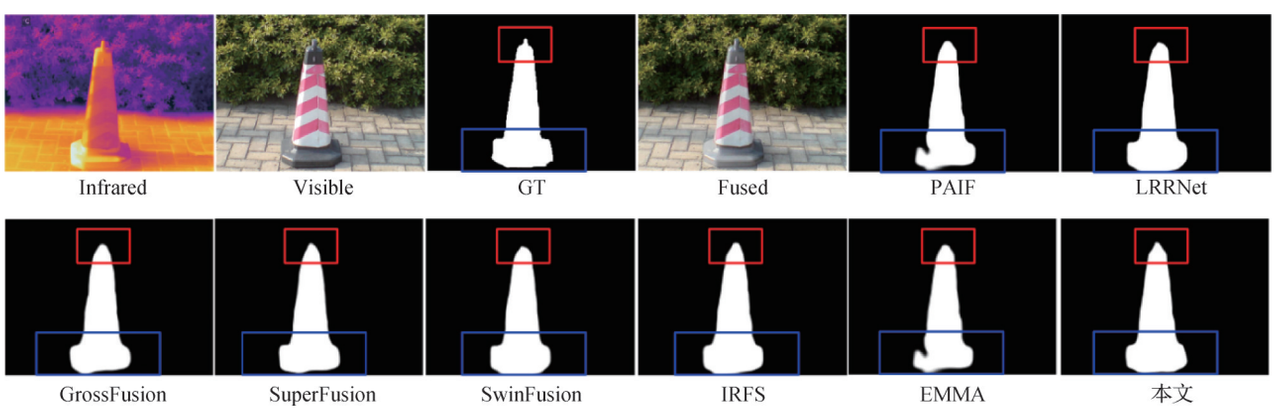

图 11　显著性目标检测在 VT5000 数据集中的视觉对比

Fig. 11　Visual comparison of saliency object detection in the VT5000 dataset

面表现更佳，特别是在捕捉狗爪精确轮廓方面，超越了其他模型在保持结构准确性方面的表现。

2.5　消融实验

为了评估本文模型中关键组件的贡献，专门聚焦于大气散射增强模块（atmospheric scattering enhancement module，ASEM）和频率集成空间通道注意力模块（frequency integrated spatial channel attention module，FSCA），在 VT5000 数据集上进行消融实验。

在消融实验的定性分析中，不同模块组合的消融实验视觉对比如图 12 所示，展示了红外与可见光图像融合的消融实验结果。图 12（a）（b）分别为红外和可见光图像。w/o ASEM 表示不使用 ASEM 模块，w/o FSCA 表示不使用 FSCA 模块。本文方法表示结合了 ASEM 和 FSCA 模块。显然，结合 ASEM 和 FSCA 模块明显改善了电箱侧面和“不锈钢”文字的清晰度，增强了图像的纹理保真度和细节表现，特别是在高亮区域，细节得到了显著恢复。定量结果如表 10 所示。

2.5.1　大气散射增强模块的有效性

大气散射增强模块（ASEM）用于恢复真实光照并增强融合图像的色彩对比度。为了严格评估该组件的影响，在消融实验中，排除了频率集成空间通道注意力模块（FSCA）的影响。

从图 12 可以看出，包含 ASEM 的融合图像相比于没有 ASEM 的图像有显著的增强。具体而言，不使用 ASEM 模块，电箱侧面（蓝色高亮区域）的亮度较低，色彩对比度较弱，导致目标轮廓不够清晰。而电箱侧面（蓝色高亮区域）的亮度在包含 ASEM 的图像中明显更高，并且色彩对比度也显著增加。这一改进不仅恢复了可见光的强度，还清晰地勾画出了物体轮廓，并增强了色彩分辨力。这些发现证明了 ASEM 组件在融合模型中的关键作用，确认了它对提升图像质量和增强视觉感知的贡献。实验结果表

明，ASEM模块有效提升了图像的亮度和色彩对比度。表10的结果进一步验证了ASEM在保持核心特征和突出目标方面的作用，从而实现了更优的融合效果。此次消融分析确认，ASEM不仅在保持色彩完整性方面起到了重要作用，还在提升整体感知质量方面发挥了关键作用，使其成为确保所提模型架构中高保真融合结果的关键组件。

2.5.2　频率集成空间通道注意力模块的有效性

频率集成空间通道注意力模块（FSCA）是本文融合模型中的一个关键组件，在有效地将源图像中的细粒度纹理融入融合结果中起着重要作用，针对FSCA模块进行消融实验，评估了其重要性。

从图12可以看出，在不使用FSCA模块的情况下，红框内的“不锈钢”文字明显模糊，缺乏精细的细节。相比之下，集成了FSCA模块的本文方法生成的融合图像在纹理清晰度上有显著提升。实验结果表明，FSCA模块在纹理清晰度和细节恢复方面发挥了重要作用。从表10可以看出，在不使用FSCA模块的情况下，6个融合指标的得分普遍低于使用本文方法时的得分，尤其在结构保真度（SF）指标上，减少了约40%。表明缺少FSCA模块会导致融合图像更平滑，细节减少。该结果验证了FSCA模块在保持源图像频域视觉保真度方面的关键作用，能够有效利用空间和通道注意力机制改善纹理细节，并保持纹理保真度。表10的定量结果表明，移除FSCA模块后，融合性能显著下降，进一步凸显了FSCA模块的重要性。

(a) 红外图像　(b) 可见光图像　(c) w/o ASEM　(d) w/o FSCA　(e) 本文

图12　在VT5000数据集上不同模块组合的消融实验视觉对比

Fig. 12　Visual comparison of ablation experiments on the VT5000 dataset with various module combination

((a) infrared image; (b) visible image; (c) w/o ASEM; (d) w/o FSCA; (e) ours)

表10　VT5000数据集上不同模块的消融实验定量分析

Table 10　Quantitative analysis of ablation experiments on the VT5000 dataset with various module combinations

方法	EN ↑	SD ↑	SF ↑	AG ↑	PSNR ↑ /dB	CC ↑
w/o ASEM	6.427	28.615	4.559	1.928	10.626	**0.652**
w/o FSCA	6.826	42.807	5.383	2.220	14.930	0.631
本文	**6.997**	**42.905**	**9.118**	**3.141**	**15.018**	0.622

注：加粗字体表示各列最优结果。

3　结　论

本文提出一个用于融合红外与可见光图像的新型网络，包含两个关键组件：大气散射增强模块和频率集成空间通道注意力模块。通过利用大气散射模型，本文预测了可见光图像的传输图和大气光，以解决由散射和传输损失引起的光照衰减和低对比度问题。此外，在估计传输图和大气光时不可避免的误差，可能导致细节纹理的丧失，本文采用傅里叶分析全面评估增强后的可见光和红外图像在频域中的相位和幅度信息。然后，利用空间通道注意力机制自适应地调整和增强这些频率成分，有效恢复细节、丰富纹理，并提高模型的鲁棒性。大量的定性和定量实验表明，本文算法表现优越，特别是在下游显著性检测准确度方面有显著的提升。

尽管在多种场景中表现强劲，当前框架仍存在两个需要进一步研究的限制。首先，尽管空间通道注意力机制显著提高了融合质量，但其计算复杂度可能会限制实时部署；通过轻量级架构或硬件感知设计优化该模块仍是一个亟待解决的挑战。其次，尽管AFFusion有效应对了常见的退化因素（如雾霾和低光照条件），但其在高度动态环境（如快速光照变化、极端天气）中的泛化能力仍需进一步探索。潜在的解决方案包括自适应的ASM参数估计和动态频域融合策略。解决这些限制将有助于拓宽IVIF系统在实际应用中的适用性和效率。

参考文献(References)

Chen J, Li X J, Luo L B and Ma J Y. 2022. Multi-focus image fusion based on multi-scale gradients and image matting. IEEE Transactions on Multimedia, 24: 655-667 [DOI: 10.1109/TMM.2021.3057493]

Ghosh S, Gavaskar R G and Chaudhury K N. 2019 Saliency guided image detail enhancement//Proceedings of 2019 National Conference on Communications (NCC). Bangalore, India: IEEE: 1-6 [DOI: 10.1109/ncc.2019.8732250]

Guo Y, Gao Y, Liu R W, Lu Y X, Qu J X, He S F, et al. 2023. SCANet: self-paced semi-curricular attention network for non-homogeneous image dehazing//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 1885-1894 [DOI: 10.1109/CVPRW59228.2023.00186]

Jin D, Wu T, Li P, Qiu Z H, He Q and Peng Y. 2025. A target detection method for substation equipment based on feature interaction and guidance of infrared-visible images. Infrared Technology, 1-10 [EB/OL]. [2025-06-09]. (金迪, 吴田, 黎鹏, 邱中华, 何清, 彭勇. 2025. 红外—可见光图像特征交互引导的变电设备目标检测方法. 红外技术, 1-10 [EB/OL]. [2025-06-09]. https://link.cnki.net/urlid/53.1053.TN.20250529.1157.002

Ju M Y, Zhang D Y and Wang X M. 2017. Single image dehazing via an improved atmospheric scattering model. The Visual Computer, 33(12): 1613-1625 [DOI: 10.1007/s00371-016-1305-1]

Li H and Wu X J. 2024. CrossFuse: a novel cross attention mechanism based infrared and visible image fusion approach. Information Fusion, 103: #102147 [DOI: 10.1016/j.inffus.2023.102147]

Li H, Xu T Y, Wu X J, Lu J W and Kittler J. 2023. LRRNet: a novel representation learning guided fusion network for infrared and visible images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(9): 11040-11052 [DOI: 10.1109/TPAMI.2023.3268209]

Liu C H, Qi Y and Ding W R. 2017. Infrared and visible image fusion method based on saliency detection in sparse domain. Infrared Physics and Technology, 83: 94-102 [DOI: 10.1016/j.infrared.2017.04.018]

Liu J Y, Fan X, Huang Z B, Wu G Y, Liu R S, Zhong W, et al. 2022. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5802-5811 [DOI: 10.1109/CVPR52688.2022.00571]

Liu J Y, Wu Y H, Huang Z B, Liu R S and Fan X. 2021. SMoA: searching a modality-oriented architecture for infrared and visible image fusion. IEEE Signal Processing Letters, 28: 1818-1822 [DOI: 10.1109/LSP.2021.3109818]

Liu X W, Huo H T, Yang X and Li J. 2025. A three-dimensional feature-based fusion strategy for infrared and visible image fusion. Pattern Recognition, 157: #110885 [DOI: 10.1016/j.patcog.2024.110885]

Long Y Z, Jia H T, Zhong Y D, Jiang Y D and Jia Y M. 2021. RXDNFuse: a aggregated residual dense network for infrared and visible image fusion. Information Fusion, 69: 128-141 [DOI: 10.1016/j.inffus.2020.11.009]

Ma J L, Zhou Z Q, Wang B and Zong H. 2017. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrared Physics and Technology, 82: 8-17 [DOI: 10.1016/j.infrared.2017.02.005]

Ma J Y, Tang L F, Fan F, Huang J, Mei X G and Ma Y. 2022. SwinFusion: cross-domain long-range learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica, 9(7): 1200-1217 [DOI: 10.1109/JAS.2022.105686]

Ma J Y, Tang L F, Xu M L, Zhang H and Xiao G B. 2021a. STDFusionNet: an infrared and visible image fusion network based on salient target detection. IEEE Transactions on Instrumentation and Measurement, 70: #5009513 [DOI: 10.1109/TIM.2021.3075747]

Ma J Y, Zhang H, Shao Z F, Liang P W and Xu H. 2021b. GANMcC: a generative adversarial network with multiclassification constraints for infrared and visible image fusion. IEEE Transactions on Instrumentation and Measurement, 70: #5005014 [DOI: 10.1109/TIM.2020.3038013]

Mou J, Gao W and Song Z X. 2013. Image fusion based on non-negative matrix factorization and infrared feature extraction//Proceedings of the 6th International Congress on Image and Signal Processing (CISP). Hangzhou, China: IEEE: 1046-1050 [DOI: 10.1109/CISP.2013.6745210]

Rao D Y, Xu T Y and Wu X J. 2023. TGFuse: an infrared and visible image fusion approach based on transformer and generative adversarial network. IEEE Transactions on Image Processing: #3273451 [DOI: 10.1109/TIP.2023.3273451]

Shi P C, Mao F and Zhang R Y. 2024. DAE-Nest: a depth information extraction and enhancement fusion network for infrared and visible images. Optics Communications, 560: #130441 [DOI: 10.1016/j.optcom.2024.130441]

Tang L F, Deng Y X, Ma Y, Huang J and Ma J Y. 2022. SuperFusion: a versatile image registration and fusion network with semantic awareness. IEEE/CAA Journal of Automatica Sinica, 9(12): 2121-2137 [DOI: 10.1109/JAS.2022.106082]

Tian X R, Xianyu X H, Li Z M, Xu T and Jia Y J. 2024. Infrared and visible image fusion based on multi-level detail enhancement and generative adversarial network. Intelligence and Robotics, 4(4): 524-543 [DOI: 10.20517/ir.2024.30]

Toet A and Hogervorst M A. 2012. Progress in color night vision. Optical Engineering, 51(1): #010901 [DOI: 10.1117/1.OE.51.1.010901]

Toet A, Van Ruyven L J and Valeton J M. 1989. Merging thermal and visual images by a contrast pyramid. Optical Engineering, 28(7): #287789 [DOI: 10.1117/12.7977034]

Wang D, Liu J Y, Liu R S and Fan X. 2023. An interactively reinforced paradigm for joint infrared-visible image fusion and saliency object detection. Information Fusion, 98: #101828 [DOI: 10.1016/j.inffus.2023.101828]

Wang L D, Zhao H C, Pan D T, Fang J and Yuan D C. 2025. Infrared and visible image fusion based on twin axial-attention and dual-discriminator generative adversarial network. Computer and Modernization(4): 89-95, 102 (王丽丹, 赵怀慈, 潘多涛, 房建, 袁德成. 2025. 基于孪生轴向注意力与双鉴别器生成对抗网络的红外可见光图像融合. 计算机与现代化, (4): 89-95, 102) [DOI: 10.3969/j.issn.1006-2475.2025.04.014]

Wang Y S, Nie R C, Zhang G C and Yang X F. 2023. Infrared and visible image fusion based on multi-level guided network. Journal of Image and Graphics, 28(1): 207-220 (王彦舜, 聂仁灿, 张谷铖, 杨小飞. 2023. 多级特征引导网络的红外与可见光图像融合. 中国图象图形学报, 28(1): 207-220) [DOI: 10.11834/jig.220638]

Wei Q, Bioucas-Dias J, Dobigeon N and Tourneret J Y. 2015. Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 53(7): 3658-3668 [DOI: 10.1109/TGRS.2014.2381272]

Xu H, Ma J Y, Le Z L, Jiang J J and Guo X J. 2020. FusionDN: a unified densely connected network for image fusion//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI, 34(7): 12484-12491 [DOI: 10.1609/aaai.v34i07.6936]

Xu Q W, Zhang R P, Zhang Y, Wang Y F and Tian Q. 2021. A Fourier-based framework for domain generalization//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14378-14387 [DOI: 10.1109/CVPR46437.2021.01415]

Yang T Y, Huo H T, Liu X W and Zheng B W. 2024. Information interaction linear transformer network for infrared and visible image fusion. Journal of Image and Graphics, 30(7): 2499-2513 (杨天宇, 霍宏涛, 刘晓文, 郑博文. 用于红外与可见光图像融合的信息交互线性Transformer网络. 中国图象图形学报, 30(7): 2499-2513 [DOI: 10.11834/jig.240569]

Zhang H, Xu H, Tian X, Jiang J J and Ma J Y. 2021. Image fusion meets deep learning: a survey and perspective. Information Fusion, 76: 323-336 [DOI: 10.1016/j.inffus.2021.06.008]

Zhang P Y, Zhao J, Wang D, Lu H C and Ruan X. 2022. Visible-thermal UAV tracking: a large-scale benchmark and new baseline//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: 8876-8885 [DOI: 10.1109/CVPR52688.2022.00868]

Zhang Q, Huang N C, Yao L, Zhang D W, Shan C F and Han J G. 2020. RGB-T salient object detection via fusing multi-level CNN features. IEEE Transactions on Image Processing, 29: 3321-3335 [DOI: 10.1109/TIP.2019.2959253]

Zhao Z X, Bai H W, Zhang J S, Zhang Y L, Zhang K and Xu S. 2024. Equivariant multi-modality image fusion//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 25912-25921 [DOI: 10.1109/CVPR52733.2024.02448]

作者简介

宋程程,女,硕士研究生,主要研究方向为图像处理与深度学习。E-mail: 301768@whut.edu.cn

靳淇文,通信作者,男,副教授,主要研究方向为高光谱图像、信息融合以及深度学习。E-mail:qiwenjin@whut.edu.cn

胡辑伟,男,教授,主要研究方向为模式识别、机器学习和智能制造。E-mail:hujiwei@whut.edu.cn