DOI:10.11918/202409008

一种 Restormer 结合细节补偿的红外与可见光图像融合方法

杨艳春,李佳龙

(兰州交通大学 电子与信息工程学院,兰州 730070)

摘 要:为提升融合图像的质量和信息完整性,解决红外与可见光图像融合中存在的特征提取能力不足、缺乏纹理细节以及全局上下文信息丢失等问题,提出一种红外与可见光图像的融合与分解网络架构。首先,利用 Restormer 和 Res2Net 的并联结构,通过多个深度卷积头转置注意力机制和多尺度残差连接,协同捕获全局上下文信息和局部细节特征;其次,通过带有仿射耦合结构的可逆神经网络,将红外与可见光图像浅层特征分为两部分,利用交替耦合变换实现特征无损保留;然后,重建模块利用拼接及卷积操作生成高质量融合图像;最后,分解网络通过最小化分解损失函数,将融合图像逆向分解为源图像。实验结果表明:在 RoadScene 数据集上,本文方法的主客观结果均优于多数对比方法,其中标准差、差异相关系数、平均梯度和空间频率较其他对比方法分别平均提升了 8.5%、23.1%、49.0% 和 56.1%;在 MSRS 数据集上,本文方法较 SDCFusion 方法在标准差、视觉信息保真度、平均梯度、差异相关系数和空间频率方面分别提升了 1.4%、0.4%、0.6%、4.3% 和 3.4%。所提方法在提升融合图像质量、保留纹理细节和全局信息方面展现出显著优势。

关键词:图像融合;并联结构;细节补偿;可逆神经网络;分解网络

中图分类号:TN911.73 **文献标志码**:A **文章编号**:0367-6234(2025)09-0149-12

A Restormer-based fusion method with detail compensation for infrared and visible images

YANG Yanchun, LI Jialong

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: To enhance the quality and information integrity of fused images and tackle issues like inadequate feature extraction, insufficient texture details, and loss of global contextual information in infrared and visible images fusion, a fusion and decomposition network architecture for infrared and visible images is proposed. Firstly, a parallel structure of Restormer and Res2Net is utilized. Multiple deep convolutional heads with transposed attention mechanisms and multi-scale residual connections are employed to collaboratively capture the global contextual information and local detail features. Secondly, an invertible neural network with affine coupling structure is adopted to divide the shallow-level features of infrared and visible images into two parts, using alternating coupling transformations to achieve lossless feature preservation. Then, the reconstruction module generates high-quality fused images through concatenation and convolution operations. Finally, the decomposition network reverses the fusion image into source images by minimizing the decomposition loss function. Experimental results show that on the RoadScene dataset, the objective and subjective results of this method surpass most comparative methods. Specifically, compared to other methods, the standard deviation improves by an average of 8.5%, the difference correlation coefficient by 23.1%, the average gradient by 49.0%, and the spatial frequency by 56.1%. On the MSRS dataset, the proposed method outperforms SDCFusion method by 1.4% in standard deviation, 0.4% in visual information fidelity, 0.6% in average gradient, 4.3% in difference correlation coefficient, and 3.4% in spatial frequency. The proposed method shows significant advantages in improving the quality of fused images, preserving texture details, and retaining global information.

Keywords: image fusion; parallel structure; detail compensation; invertible neural network; decomposition network

收稿日期:2024-09-04;**录用日期**:2024-11-14;**网络首发日期**:2025-06-12
网络首发地址:https://link.cnki.net/urlid/23.1235.t.20250611.1732.004
基金项目:国家自然科学基金(62462043,62067006);甘肃省重点研发计划(25YFGA047);甘肃省自然科学基金(23JRRA847,21JR7RA300)
作者简介:杨艳春(1979—),女,副教授,硕士生导师
通信作者:杨艳春,yangyanchun102@sina.com

由于硬件设备的理论和技术限制,从单一模态传感器或在单一拍摄环境下获得的信息无法有效、全面地描述成像场景。为此,图像融合技术应运而生,其目的是将多模式传感器或在不同拍摄环境下获取的互补信息结合起来[1]。在图像融合技术应用领域,红外与可见光图像融合的应用尤为广泛。可见光图像通过捕捉物体反射的光线,能够呈现出丰富的纹理细节,这与人类视觉习惯相契合,然而这种图像容易受到光照变化等环境因素的干扰,可能会影响其对目标的识别能力。与此同时,红外图像则通过检测物体发出的热辐射来成像,其优势在于高对比度,能够清晰地区分背景与目标,即使在光线不足或环境恶劣的情况下,红外图像依然能够保持较高的成像质量,对目标的识别具有较高的稳定性。通过红外与可见光图像融合,能够得到一种更为全面和鲁棒的图像表示,它不仅能够提供丰富的视觉细节,还能够在各种环境条件下保持对目标的准确识别。这种融合技术在军事侦察、医学成像、遥感监测等多个领域都有着重要的应用价值[2-3],为图像处理领域带来了新的视角和可能性。

现有的红外与可见光图像融合方法主要分为两大类:传统方法和基于深度学习的方法。图像融合的传统技术中,通常采用数学变换来实现源图像的转换和融合。这一过程被划分为特征提取、特征融合和特征重构 3 个主要步骤。首先,通过特定的数学变换,从源图像中提取关键特征,这是整个特征提取阶段的核心任务。然后,进入特征融合阶段,利用特定的融合规则将提取的特征进行有效整合。最后,在特征重构阶段,对融合后的特征应用逆变换,以重构出最终的融合图像。根据应用数学变换,传统算法可进一步分为 4 类:基于多尺度分解的方法、基于子空间聚类的方法[4]、基于稀疏表示的方法[5]和基于显著性方法[6]。然而,手工设计的转换过程往往涉及复杂的操作,而融合策略在应对多样化场景时可能会遇到局限。

近年来,随着深度学习的迅速发展,神经网络强大的非线性拟合能力和出色的特征提取能力推动了图像融合领域的发展。目前基于深度学习的融合方法分为基于自动编码器(autoencoder,AE)的方法,基于卷积神经网络(convolutional neural network,CNN)的方法及基于生成对抗网络(generative adversarial network,GAN)的方法。对于 AE 的融合方法,Li 等[7]提出了一种基于密集块的 AE,通过密集连接强化细节特征,生成信息丰富的融合图像。考虑到手工融合规则的局限性,Li 等[8]提出了一种名为 RFN-Nest 的残差融合网络,该网络采用巢状连接和注意力机制多尺度捕捉关键信息,确保融合过程的信息完整性。范焱等[9]考虑到频域中潜在的全局信息,提出了一种基于空间和频率特征解耦的融合方法,该方法不仅关注空间域的特征,还结合了频域的信息,通过交互式融合,提高融合算法的鲁棒性。对于 CNN 的融合方法,李永萍等[10]在变换域中通过多尺度引导滤波器提取基础与细节信息,并利用拉普拉斯能量融合基础层,VGGNet19 网络提取细节层特征,最终通过加权平均策略合成融合图像,实现红外与可见光图像的有效融合。Zhang 等[11]通过梯度路径和强度路径分别提取图像的高频纹理和像素强度信息,并采用特征重用和信息交换优化融合效果,实现了更好的视觉效果。对于 GAN 的融合方法,Ma 等[12]提出了一种名为 FusionGAN 的融合方法,该方法首先将 GAN 引入图像融合领域,可以在没有监督信息的情况下保留重要特征。为避免单鉴别器网络丢失红外图像对比度信息的问题,许光宇等[13]设计了一种双路径双鉴别器生成对抗网络的红外与可见光图像融合方法,该方法通过构建梯度路径和对比度路径,以及多尺度特征提取,提升了融合图像的细节和对比度。利用双鉴别器避免信息丢失,并通过主辅梯度和强度损失函数增强模型的信息提取能力,能够更好地保留和增强图像中的纹理细节。

上述方法虽取得了较好的融合效果,但是目前红外与可见光图像融合多聚焦于从源图像到融合图像的过程,未考虑能否通过分解融合图像来重建源图像,以最小化信息损失;而且融合图像的质量不仅受局部感知区域内像素的影响,还与全局结构的像素强度和纹理细节紧密相关。部分现有方法侧重于局部处理,通过有限的感受野来构建局部深度特征,这种做法虽然能够捕捉到局部细节,但往往忽视了特征之间的长期依赖性,可能导致关键的上下文信息丢失,从而影响融合图像的整体质量和信息的完整性。

为了解决上述问题,本文提出了一种基于 Restormer 的融合网络和一种基于 UNet 的分解网络(decomposition network,DN)。融合网络中的特征提取模块采用 Restormer 与 Res2Net 并联构成,共同捕捉全局和局部特征,实现信息的有效整合。由于融合图像的纹理细节被认为多数由可见光图像提供,然而一些情况下,红外图像中也可能有少量的细节信息。为了充分挖掘源图像中的细节信息,本文利用可逆神经网络无损信息保存的特性,构建细节补偿模块(detail compensation,DC),使融合图像保留更多边缘和纹理信息。基于分解结果的质量直接取决于融合的结果这一观点,本文基于 UNet 架构,设

计了一个轻量型的分解网络，最大程度地减少信息损失并保留关键的视觉特征。旨在通过本文方法可以优化融合图像中的纹理细节，提升融合图像的视觉效果。

1　相关理论

1.1　Restormer

Parmar 等[14]提出了 Transformer 架构，主要是为了解决机器翻译问题，并在自然语言处理领域取得了显著的成就。近年来，Transformer 的思想也被引入到计算机视觉领域。2020 年，Dosovitskiy 等[15]提出了视觉领域 Transformer（vision transformer，ViT），这是一种将 Transformer 架构应用于图像处理的创新方法，在目标检测、图像分类等视觉任务方面都取得了令人满意的结果。

在处理图像数据时，Transformer 通常将图像切割成小块并拉伸成向量处理，这会导致空间信息丢失，影响模型捕捉局部和全局依赖的能力。相比之下，CNN 能自然学习局部特征并构建全局特征。此外，Transformer 计算量大，结构复杂，包含更多参数和隐藏层，训练和推理成本较高。

为了克服上述局限性，Liu 等[16]开发了 Swin Transformer，通过移位操作将图像划分为局部和交叉窗口，并在相应窗口内进行注意力计算。然而，该方法将上下文信息的聚合限制在局部邻域内，影响了对长距离依赖关系的捕捉效果。Zamir 等[17]提出了 Restormer 模型，该模型旨在学习长距离依赖关系的同时，保持计算效率。其核心组件包括多个深度卷积头转置注意力（multi-dconv head transposed attention，MDTA）模块和新的门控深度卷积前馈网络（gated dconv feed-forward network，GDFN），如图 1、2 所示。MDTA 模块有效聚合局部与非局部像素交互，保留细节并捕捉全局信息，使 Restormer 在处理高分辨率图像时既高效又准确。GDFN 模块通过受控特征变换，抑制低信息性特征，仅传递有用信息，增强模型性能，提高泛化能力和鲁棒性。

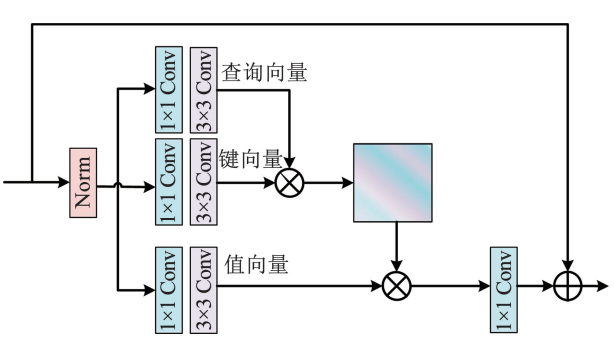


图 1　多个深度卷积头转置注意力

Fig. 1　Multi-dconv head transposed attention

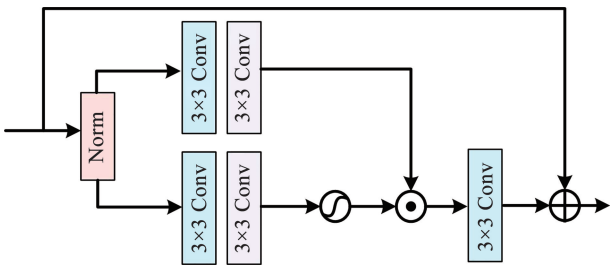


图 2　门控深度卷积前馈网络

Fig. 2　Gated dconv feed-forward network

1.2　可逆神经网络

可逆神经网络（invertible neural network，INN）作为一种特殊的神经网络架构，独具显著特色，即能够实现输入与输出之间的双向、无歧义映射。这种独特的可逆性不仅确保了信息的完整传递，还在反向传播和梯度计算中展现了极高的效率和可靠性。INN 的概念最初由 Dinh 等[18]提出，旨在实现一种特殊的网络结构。其中，输入与输出之间存在明确的可逆关系。给定一个变量 y 和通过正向计算 f_θ 得到的某种表示或输出，INN 能够通过逆向计算过程 f_θ^{-1} 直接且准确地恢复原始的 y 值。逆向计算的关键在于其反函数的设计，而反函数被构建为与正向计算过程共享参数 θ，从而保证了计算的可逆性和效率。

INN 的可逆特性为数据重构和恢复提供了强大支持。同时，在图像隐藏、上色、压缩和视频超分辨率等推理任务中也展现了卓越的性能，有效推动了计算机视觉领域的发展。

INN 的基本网络架构是采用仿射耦合层来构建的。其核心思想是将输入数据 x 分为两个部分：u_1 和 u_2。随后，通过两个学习函数 s_i 和 t_i 对这两部分数据进行转换，并通过交替耦合的方式进行处理，得到两个结果：v_1 和 v_2，经过一系列变换后得到结果 y。函数 s_i 和 t_i 可以是任意复杂度的函数，并且其本身不需要满足可逆性的要求。这种设计为网络提供了更大的灵活性，允许模型学习更复杂的数据转换，同时保持了正向和逆向传播的高效性。正向传播过程的如图 3 所示，相关的计算式为：

$$v_1 = u_1 \odot \exp(s_2(u_2)) \oplus t_2(u_2) \tag{1}$$

$$v_2 = u_2 \odot \exp(s_1(v_1)) \oplus t_1(v_1) \tag{2}$$

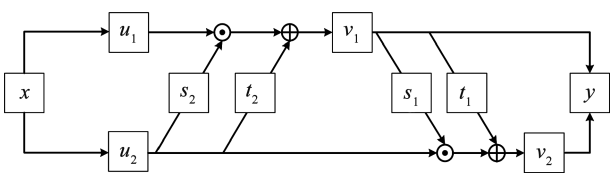


图 3　正向传播过程

Fig. 3　Forward propagation process

INN 的逆向过程如图 4 所示，相关的计算式为：

$$u_2 = (v_2 - t_1(v_1)) \odot \exp(-s_1(v_1)) \quad (3)$$

$$u_1 = (v_1 - t_2(u_2)) \odot \exp(-s_2(u_2)) \quad (4)$$

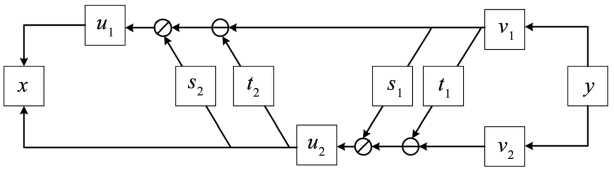


图 4 逆向传播过程

Fig. 4 Backward propagation process

2 本文方法

2.1 网络总体结构

本文提出的融合与分解网络由融合网络和分解网络构成，如图 5 所示。其中，融合网络由特征提取模块、细节补偿模块和重建模块组成。融合网络作为图像融合的目标网络，专注于将多源图像融合成一幅单一的、信息丰富的图像。分解网络由双路径 UNet 组成，致力于从融合结果中分解近似源图像的结果，从而形成一种自我约束的机制，确保融合图像在多个层面上均保持原始图像的信息和细节。整体而言，融合与分解网络架构可以使融合结果包含更丰富的场景细节，从而具有更好的融合效果。

2.2 特征提取模块

特征提取模块由浅层特征提取和深层特征提取两部分组成。浅层特征提取由两个 3 × 3 的卷积层组成，而深层特征提取采用 Restormer 和 Res2Net[19] 的并联结构，如图 6 所示。Restormer 凭借其卓越的全局上下文信息提取能力，能有效把握图像整体结构；Res2Net 则通过在残差块内构建分层的类残差连接，实现多尺度特征的细粒度表示，丰富特征层次并扩大感受野范围，提升图像处理任务表现，尤其在细节纹理信息提取方面表现出色。两者并联协同作用，使深层特征提取模块既能快速捕捉全局上下文信息，又能更好地保留纹理信息，即让模块在关注图像整体特征的同时，深入局部细节，提取出对后续任务至关重要的关键信息。

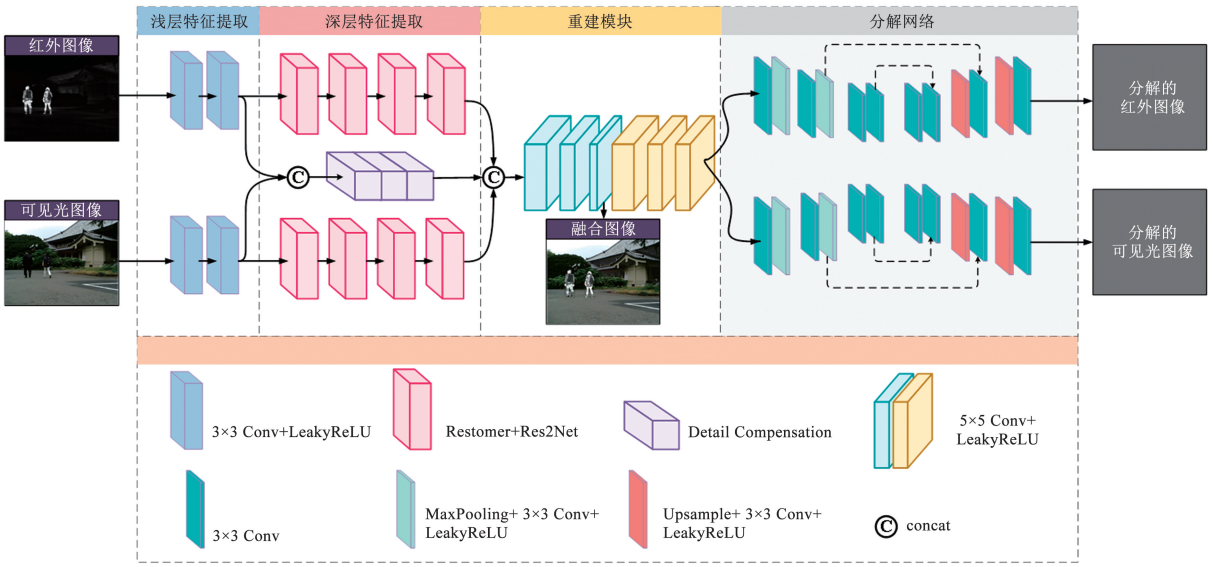


图 5 网络总体架构

Fig. 5 Overall network architecture

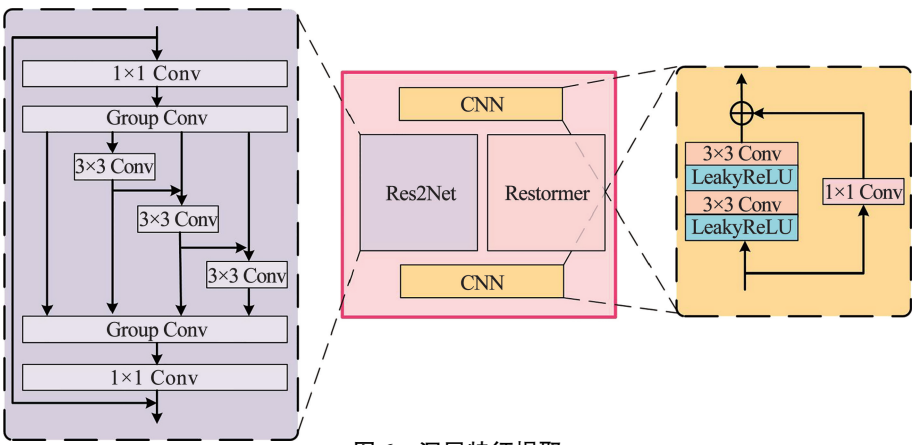


图 6 深层特征提取

Fig. 6 Deep feature extraction

浅层与深层特征提取过程表示为：

$$\boldsymbol{\phi}_{\mathrm{ir}}^{\mathrm{s}}=E_{\mathrm{s}}(\boldsymbol{I}_{\mathrm{ir}}) \tag{5}$$

$$\boldsymbol{\phi}_{\mathrm{vis}}^{\mathrm{s}}=E_{\mathrm{s}}(\boldsymbol{I}_{\mathrm{vis}}) \tag{6}$$

$$\boldsymbol{\phi}_{\mathrm{ir}}^{\mathrm{d}}=E_{\mathrm{d}}(\boldsymbol{\phi}_{\mathrm{ir}}^{\mathrm{s}}) \tag{7}$$

$$\boldsymbol{\phi}_{\mathrm{vis}}^{\mathrm{d}}=E_{\mathrm{d}}(\boldsymbol{\phi}_{\mathrm{vis}}^{\mathrm{s}}) \tag{8}$$

式中：$\boldsymbol{I}_{\mathrm{ir}}$和$\boldsymbol{I}_{\mathrm{vis}}$分别为一对严格配准的红外和可见光图像，$\boldsymbol{\phi}_{\mathrm{ir}}^{\mathrm{s}}$和$\boldsymbol{\phi}_{\mathrm{vis}}^{\mathrm{s}}$分别为经过浅层特征提取后的红外和可见光特征，$\boldsymbol{\phi}_{\mathrm{ir}}^{\mathrm{d}}$和$\boldsymbol{\phi}_{\mathrm{vis}}^{\mathrm{d}}$分别为经过深层特征提取后的红外和可见光特征，$E_{\mathrm{s}}(\cdot)$和$E_{\mathrm{d}}(\cdot)$分别为浅层特征提取操作和深层特征提取操作。

2.3　细节补偿模块

考虑到纹理细节不仅存在于可见光图像中，红外图像也可能包含少量细节，因此将浅层特征提取后的红外与可见光特征拼接在一起，输入至 INN 中，利用可逆网络的信息无损的这一特点，实现细节信息的有效保留和补偿。

由 INN 构成的细节补偿模块（DC）如图 7 所示，通过输入和输出特征相互生成机制，INN 模块能够在特征提取过程中最大限度地保留原始输入信息，从而确保细节信息的无损传递。为了进一步增强 INN 模块的性能，本文设计了带有仿射耦合层的 INN 块，采用并行映射结构（parallel mapping structure，PMS）（图 8）将输入特征分为两部分，分别提取不同方面的特征，通过 1×1 卷积重新加权和组合特征通道，实现变换操作$\varphi(\cdot)$、$\rho(\cdot)$、$\eta(\cdot)$。上述计算过程为：

$$\boldsymbol{\phi}_{\mathrm{cat}}^{\mathrm{I}}=F_{\mathrm{concat}}(\boldsymbol{\phi}_{\mathrm{ir}}^{\mathrm{s}},\boldsymbol{\phi}_{\mathrm{vis}}^{\mathrm{s}}) \tag{9}$$

$$\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}[\lambda+1{:}C]=\boldsymbol{\phi}_{\mathrm{cat},k}^{\mathrm{I}}[\lambda+1{:}C]+\varphi(\boldsymbol{\phi}_{\mathrm{cat},k}^{\mathrm{I}}[1{:}\lambda]) \tag{10}$$

$$\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}[1{:}\lambda]=\boldsymbol{\phi}_{\mathrm{cat},k}^{\mathrm{I}}[1{:}\lambda]\odot\exp(\rho(\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}[\lambda+1{:}C]))+\eta(\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}[\lambda+1{:}C]) \tag{11}$$

$$\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}=F_{\mathrm{concat}}(\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}[1{:}\lambda]+\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}[\lambda+1{:}C]) \tag{12}$$

式中：$\boldsymbol{\phi}_{\mathrm{cat}}^{\mathrm{I}}$为红外与可见光浅层特征的拼接结果，$F_{\mathrm{concat}}(\cdot)$为拼接操作的功能函数，$k\in\{1,2,3\}$为细节补偿模块个数，$\lambda$为输入特征划分的通道数，$C$为输入特征的通道总数，$\varphi(\cdot)$、$\rho(\cdot)$、$\eta(\cdot)$为 INN 的变换函数，$\boldsymbol{\phi}_{\mathrm{cat},k}^{\mathrm{I}}[1{:}\lambda]$为第 k 个可逆神经网络中从 1 到 λ 的特征，$\boldsymbol{\phi}_{\mathrm{cat},k+1}^{\mathrm{I}}[\lambda+1{:}C]$为第 $k+1$ 个可逆神经网络中从 $\lambda+1$ 到 C 的特征。

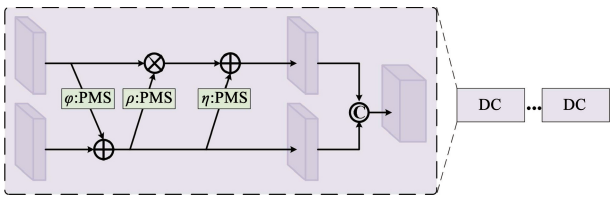


图 7　细节补偿模块

Fig. 7　Detail compensation module

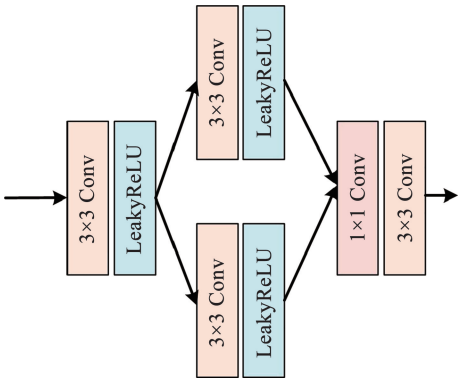


图 8　并行映射结构

Fig. 8　Parallel mapping structure

2.4　重建模块

重建模块由多个相同的子模块依次堆叠而成，每个子模块包含一个大小为 3×3 的卷积层和 LeakyReLU 激活函数。将经过深层特征提取的红外与可见光特征以及补偿信息通过拼接操作一起输入至重建模块中，得到融合图像。该过程表达式为：

$$\boldsymbol{\phi}_{\mathrm{F}}=F_{\mathrm{concat}}(\boldsymbol{\phi}_{\mathrm{ir}}^{\mathrm{d}},\boldsymbol{\phi}_{\mathrm{vis}}^{\mathrm{d}},\boldsymbol{\phi}_{\mathrm{detail}}) \tag{13}$$

$$\boldsymbol{I}_{\mathrm{fu}}=F_{\mathrm{Rconv}}(\boldsymbol{\phi}_{\mathrm{F}}) \tag{14}$$

式中：$\boldsymbol{\phi}_{\mathrm{detail}}$为经过细节补偿模块后的特征，$\boldsymbol{\phi}_{\mathrm{F}}$为红外与可见光图像的深层特征及补偿特征的拼接结果，$\boldsymbol{I}_{\mathrm{fu}}$为融合结果，$F_{\mathrm{Rconv}}(\cdot)$为特征重建操作。

2.5　分解网络

分解网络是对融合后的图像进行分解，以恢复出与原始源图像在细节和结构上高度一致的独立图像。图像分解与融合为互补设计，以确保融合结果的完整性，进而提高图像的质量。本文基于 UNet 架构设计了轻量型的分解网络，如图 5 所示。每个分解网络分为下采样层和上采样（Upsample）层两部分。下采样层由 3 对 3×3 卷积构成，前两对采用最大池化操作（MaxPooling）作为下采样方法。上采样层也由 3 对 3×3 卷积构成，前两对采用双线性插值法作为上采样方法。分解过程表达式为：

$$\boldsymbol{I}_{\mathrm{ir}}^{\mathrm{D}}=N_{\mathrm{D}}(F_{\mathrm{Conv3}}(\boldsymbol{I}_{\mathrm{fu}})) \tag{15}$$

$$\boldsymbol{I}_{\mathrm{vis}}^{\mathrm{D}}=N_{\mathrm{D}}(F_{\mathrm{Conv3}}(\boldsymbol{I}_{\mathrm{fu}})) \tag{16}$$

式中：$\boldsymbol{I}_{\mathrm{ir}}^{\mathrm{D}}$和$\boldsymbol{I}_{\mathrm{vis}}^{\mathrm{D}}$分别为经过分解网络分解出的红外与可见光图像，$F_{\mathrm{Conv3}}(\cdot)$为对融合图像进行的 3 × 3 卷积操作，$N_{\mathrm{D}}(\cdot)$为分解网络对应的操作。

2.6 损失函数

本文网络结构主要分为融合网络和分解网络，相应的损失函数也分为融合损失 L_{fu}和分解损失 L_{de}两部分。总损失 L_{total}为

$$L_{\mathrm{total}} = L_{\mathrm{fu}} + L_{\mathrm{de}} \tag{17}$$

2.6.1 融合损失

融合损失从强度和梯度两方面约束融合结果，表达式为

$$L_{\mathrm{fu}} = L_{\mathrm{int}} + L_{\mathrm{grad}} \tag{18}$$

式中：L_{int}为强度损失，L_{grad}为梯度损失。

L_{int}为了突出红外与可见光图像中的显著目标，将结果的强度值最大化，以保证目标的显著性。其表达式为

$$L_{\mathrm{int}} = \frac{1}{HW} \| \boldsymbol{I}_{\mathrm{fu}} - \max(\boldsymbol{I}_{\mathrm{ir}}, \boldsymbol{I}_{\mathrm{vis}}) \|_1 \tag{19}$$

式中：H、W 分别为图像的高度、宽度，$\| \cdot \|_1$ 为 l_1 范数。

L_{grad}保留了两个源图像的最大边缘，以获得更清晰的纹理表示。其表达式为

$$L_{\mathrm{grad}} = \frac{1}{HW} \| \nabla \boldsymbol{I}_{\mathrm{fu}} - \max(\nabla \boldsymbol{I}_{\mathrm{ir}} - \nabla \boldsymbol{I}_{\mathrm{vis}}) \|_1 \tag{20}$$

式中 ∇ 为梯度算子，用于测量图像的细粒度纹理信息。

2.6.2 分解损失

分解损失要求融合图像的分解结果尽可能与源图像相似。其表达式为

$$L_{\mathrm{de}} = \frac{1}{HW} \sum_i \sum_j (\boldsymbol{I}_{\mathrm{ir}}^{\mathrm{D}} - \boldsymbol{I}_{\mathrm{ir}})^2 + (\boldsymbol{I}_{\mathrm{vis}}^{\mathrm{D}} - \boldsymbol{I}_{\mathrm{vis}})^2 \tag{21}$$

式中 i、j 分别为 H、W 的像素索引。

融合图像的质量对于其分解结果与源图像的相似程度具有决定性的影响。因此，分解损失确保融合过程不仅保留了源图像的重要特征，还尽可能地包含了更多的场景内容，进而提升整体的融合性能。

3 实验结果与分析

3.1 数据集和实验设置

本文在 MSRS[20] 和 RoadScene[21] 两个公开数据集上进行对比实验。本文方法在 MSRS 训练集上进行训练，并在其测试集上进行测试以评估其性能。在 RoadScene 上进行泛化性实验。训练样本随机裁剪成 128 × 128 像素的图像块，训练 epoch 设置为 200，训练批次设置为 8。

采用 Adam 优化器进行参数更新，学习率设置为 1×10^{-4}，所有实验均在 NVIDIA GeForce RTX 4060ti 和 2.5 GHz Intel Core i5-13400 CPU 上采用 PyTorch 框架进行。

3.2 评价指标和对比方法

为了更好地比较本文方法与其他融合算法的融合性能，选用 6 种客观评价指标进行评价：标准差 D_{S}，从统计学角度衡量融合图像的分布特性及对比度；视觉信息保真度 F_{VI}，评价融合图像在人类视觉系统中的信息保真度；平均梯度 G_{A}，反映了融合图像中纹理信息的丰富程度；差异相关系数 D_{SC}，衡量源图像与融合图像之间的差异，并主要估计从源图像到融合图像传递的信息量；质量指标 $Q^{\mathrm{AB/F}}$，主要计算从源图像传输到融合图像的边缘信息量；空间频率 F_{S}，测量融合图像中存在的空间频率信息。以上 6 种指标均为值越大，融合效果越好。

将本文方法与目前主流的 DenseFuse[7]、RFN-Nest[8]、PMGI[11]、FusionGAN[12]、U2Fusion[21]、IFCNN[22]、SDNet[23]、SeAFusion[24] 以及 SDCFusion[25] 9 种深度学习的方法进行比较，以验证其有效性。

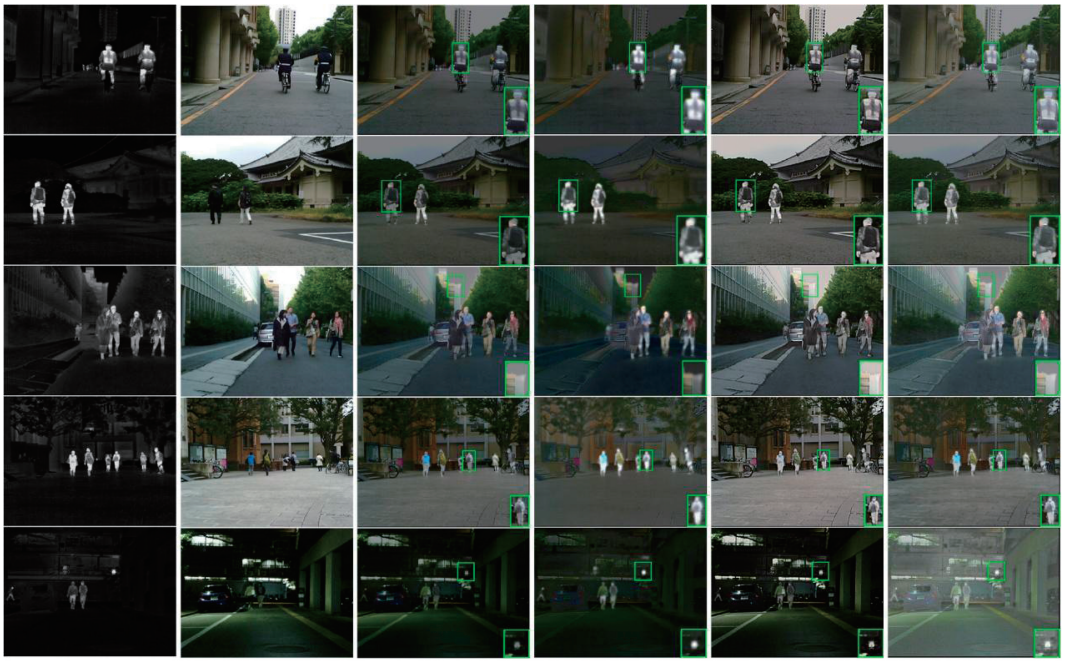
3.3 MSRS 数据集上的对比实验

3.3.1 主观结果分析

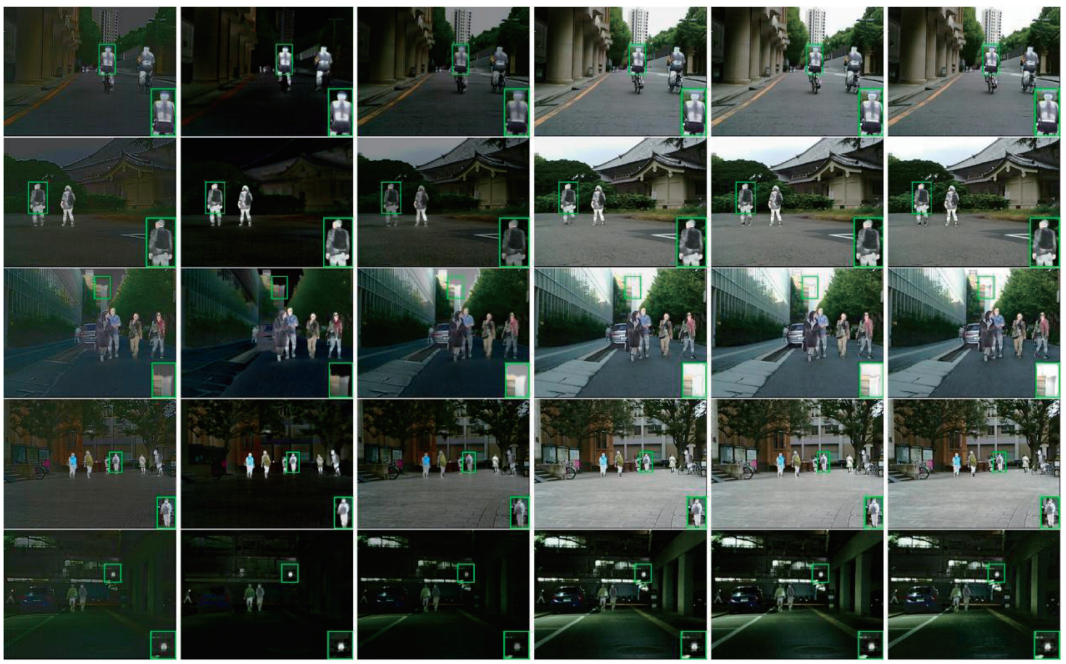
在 MSRS 测试集中选取 5 组图像作为主观分析，结果如图 9 所示。由图 9 可以看出，DenseFuse 在第 4 组图像的融合结果中，地面纹理不清晰，缺乏细节；第 5 组图像的融合结果则表现为灯源模糊且亮度偏暗。FusionGAN、SDNet 和 RFN-Nest 的整体融合结果偏暗，导致细节严重丢失，影响了图像的可视性和信息保留。PMGI 生成的融合结果具有较高的对比度，但在某些区域出现了局部失真，尤其是在第 5 组融合结果中，出现了过曝现象。U2Fusion 在白天场景中具有理想结果，但在低光场景中表现不佳，红外目标被淹没在黑暗场景中。SeAFusion 在第 1 组和第 2 组融合结果中，红外目标轮廓清晰，但具体细节模糊。IFCNN 和 SDCFusion 在第 3 组融合结果中，近景目标清晰，但远景建筑缺少细节，未能有效融合不同距离的目标信息。相比之下，本文提出的方法在 5 组实验中，无论是在白天还是低光场景，均展现出了卓越的性能，融合结果不仅目标清晰，而且纹理细节丰富，有效地克服了其他方法中存在的一些问题。

3.3.2　客观结果分析

表 1 为 MSRS 数据集上 20 组测试结果的平均值。由表 1 可知，本文方法在 D_S、F_{VI}、G_A、D_{SC}、F_S 中均获得了最优结果。在 D_S、F_{VI}、G_A、D_{SC}、F_S 指标方面，本文方法较最新的 SDCFusion 方法分别提升了 1.4%、0.4%、0.6%、4.3%、3.4%。其中，D_S 和 F_{VI} 说明本文方法融合的图像具有较高的对比度和满意的视觉效果；D_{SC} 说明本文方法融合的图像中伪信息较少，源图像与融合图像之间的相关性最大，这归功于本文设计的分解网络；G_A 和 F_S 指标排名第一，说明本文方法融合的结果包含丰富的纹理细节，这得益于本文设计的细节补偿模块；$Q^{AB/F}$ 仅次于最高的 SeAFusion，说明本文方法对源图像边缘信息的融合也有不错的表现。



(a) Infrared　(b) Visible　(c) DenseFuse　(d) FusionGAN　(e) IFCNN　(f) PMGI



(g) SDNet　(h) RFN-Nest　(i) U2Fusion　(j) SeAFusion　(k) SDCFusion　(l) Ours

图 9　MSRS 数据集主观实验结果

Fig. 9　Subjective experimental results on the MSRS dataset

表 1 MSRS 数据集 20 组融合结果平均值

Tab. 1 Mean of 20 sets of fusion results on the MSRS dataset

Methods	D_S	F_{VI}	G_A	D_{SC}	$Q^{AB/F}$	F_S
DenseFuse	8.024 4	0.683 8	2.428 9	1.269 9	0.369 8	0.027 5
FusionGAN	6.164 9	0.459 8	1.630 8	0.997 0	0.122 6	0.018 6
IFCNN	8.555 9	0.822 9	4.531 6	1.584 6	0.626 7	0.053 1
PMGI	8.101 7	0.635 6	3.337 3	1.351 6	0.414 3	0.036 1
SDNet	5.988 6	0.396 5	2.742 5	0.956 2	0.321 4	0.034 4
RFN-Nest	6.456 3	0.881 3	1.589 7	0.457 5	0.242 1	0.017 5
U2Fusion	7.352 6	0.597 8	2.492 3	1.362 1	0.332 8	0.029 1
SeAFusion	8.978 9	0.904 8	4.447 3	1.674 5	**0.691 7**	0.051 5
SDCFusion	8.866 5	0.993 7	4.662 4	1.635 4	0.657 3	0.053 3
Ours	**8.988 3**	**0.998 1**	**4.690 1**	**1.705 1**	0.684 5	**0.054 1**

为了清晰地呈现各个指标的表现差异，对 MSRS 数据集中 20 组融合结果的评价指标进行可视化分析，如图 10 所示。由图 10 可以看出，在 D_S、F_{VI}、G_A、D_{SC}、F_S 指标中，本文方法在个别指标上虽有轻微波动，但总体趋势优于其他比较方法，表明了本文方法的有效性和可靠性。

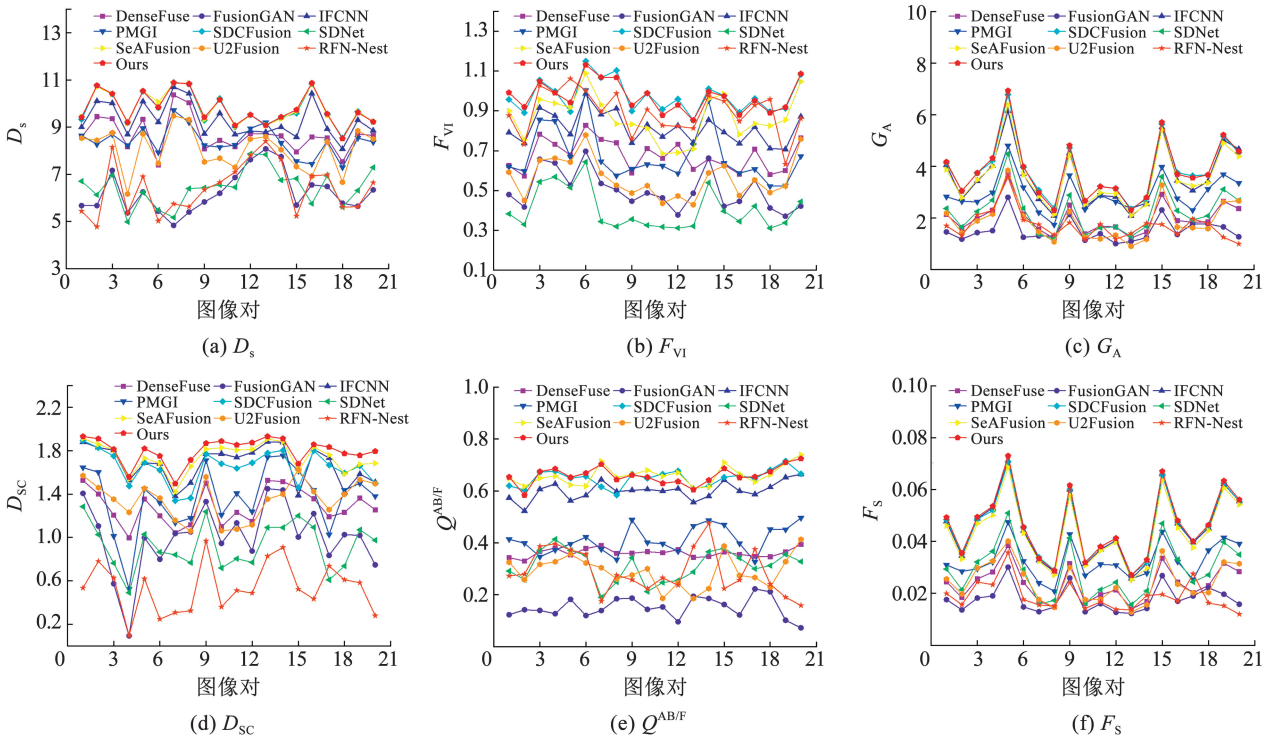


图 10 MSRS 数据集上 20 组图像的 6 种指标分析

Fig. 10 Plot of the six metrics analyzed for the 20 images on the MSRS dataset

3.4 RoadScene 数据集上的对比实验

3.4.1 主观结果分析

在 RoadScene 数据集中选取 5 组图像进行主观分析，如图 11 所示。由图 11 可以看出：DenseFuse 第 4 组融合结果对于曝光中的物体缺乏细节纹理；FusionGAN、SDNet 和 RFN-Nest 整体融合结果偏暗，物体周围产生伪影；IFCNN 第 2 组融合结果中，地面标识被隐藏在阴影中，模糊不清；PMGI 和 SDNet 的融合结果虽有清晰的红外目标，但缺乏细节，更偏向于红外图像；U2Fusion 第 3 组融合结果中，红外目标的对比度较弱，结果更偏向于可见光图像；SeAFusion 第 5 组融合结果中，门框被完全掩盖于灯光中；SDCFusion 第 1 组和第 2 组融合结果中，建筑物的边缘出现伪影，树干纹理不清晰。以上对比方法在 5 组结果中均有不同程度的缺陷，而本文方法在 RoadScene 数据集中泛化性最优，融合结果保留了源图像的互补信息。

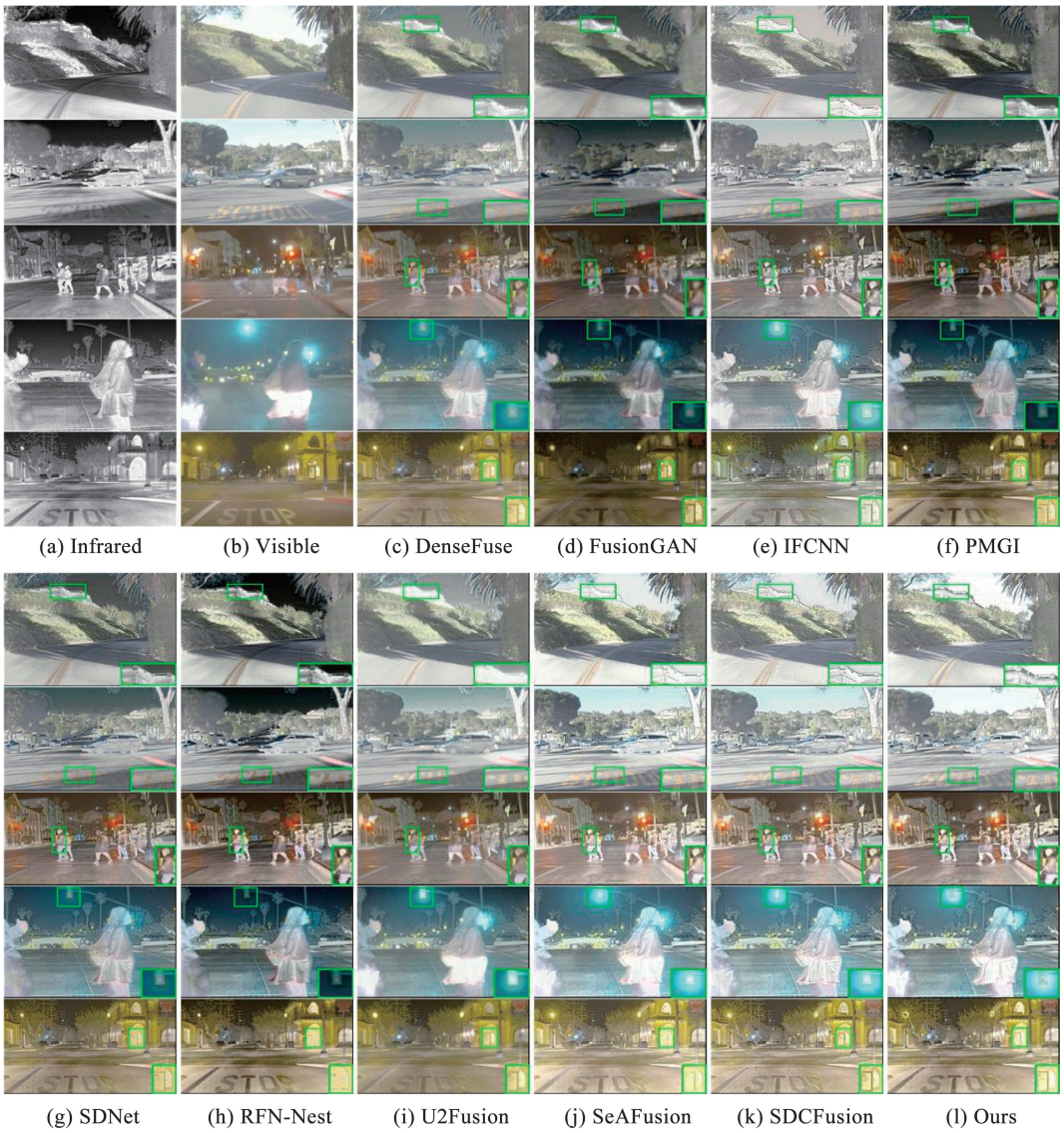


图 11　RoadScene 数据集主观实验结果

Fig. 11　Subjective experimental results on the RoadScene dataset

3.4.2　客观结果分析

表 2 为 RoadScene 数据集上 20 组测试结果的平均值。由表 2 可以看出，本文方法在 D_S、D_{SC}、G_A、F_S 中均获得了最优结果。在 D_S、D_{SC}、G_A 和 F_S 指标方面，本文方法较其他对比方法分别平均提升了 8.5%、23.1%、49.0% 和 56.1%。说明本文方法的融合图像具有较高的对比度，亮度和对比度分布与源图像保持一致，且在纹理细节和结构相似性方面也表现出色。原因是本文设计的分解网络和细节补偿模块：分解网络能够有效地从融合图像中恢复出源图像，从而约束融合图像质量；细节补偿模块则有效地增强了图像的纹理特征。在 F_{VI} 指标方面，本文方法仅次于最优的 SDCFusion，说明本文方法融合的图像提供了令人满意的视觉效果，在视觉上与人的感受相匹配。在 $Q^{AB/F}$ 指标方面，本文方法仅次于最高的 SeAFusion，说明本文方法能够有效地捕捉并融合源图像中的关键边缘特征。

对 RoadScene 数据集中 20 组融合结果的评价指标进行可视化分析，如图 12 所示。由图 12 可以看出，本文方法在多个指标中处于优势地位，F_{VI} 和 $Q^{AB/F}$ 的折线虽然略低于 SDCFusion 和 SeAFusion，但仍然保持在较高水平。这些结果证实了本文方法在图像融合领域的有效性和优越性。

表 2 RoadScene 数据集 20 组融合结果平均值

Tab. 2 Mean of 20 sets of fusion results on the RoadScene dataset

Methods	D_S	F_{VI}	G_A	D_{SC}	$Q^{AB/F}$	F_S
DenseFuse	9.333 5	0.642 9	3.266 3	1.561 9	0.375 1	0.033 5
FusionGAN	10.287 9	0.578 1	3.507 6	1.377 9	0.269 1	0.035 8
IFCNN	10.340 3	0.713 5	5.654 7	1.340 7	0.478 3	0.061 3
PMGI	9.853 9	0.774 9	4.625 8	1.314 9	0.400 2	0.044 8
SDNet	9.841 6	0.698 9	5.102 3	1.457 4	0.464 5	0.050 7
RFN-Nest	10.370 4	0.776 5	4.563 1	1.051 2	0.502 9	0.042 9
U2Fusion	9.872 5	0.704 1	4.069 8	1.655 2	0.372 3	0.034 1
SeAFusion	10.629 1	0.714 4	6.966 3	1.612 1	**0.568 4**	0.061 6
SDCFusion	10.823 2	**0.811 7**	6.754 5	1.573 8	0.468 7	0.068 3
Ours	**10.995 3**	0.808 6	**6.968 2**	**1.740 4**	0.532 4	**0.070 6**

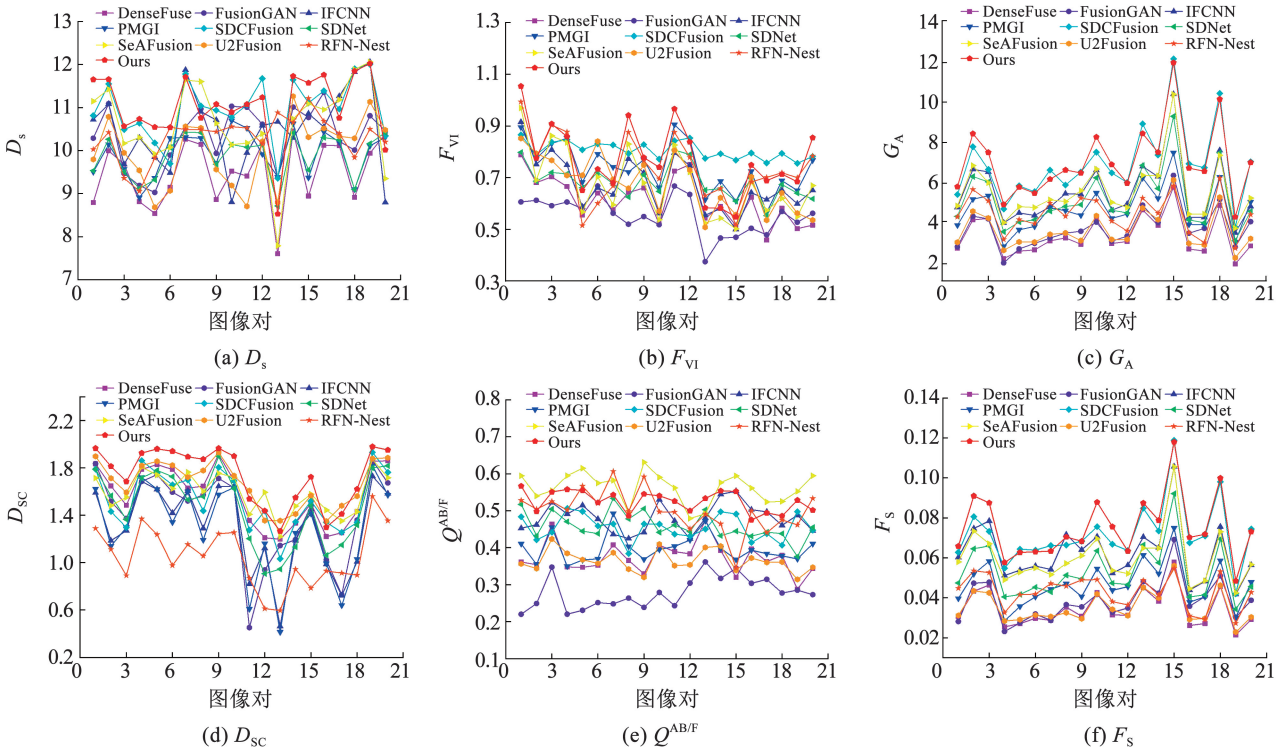


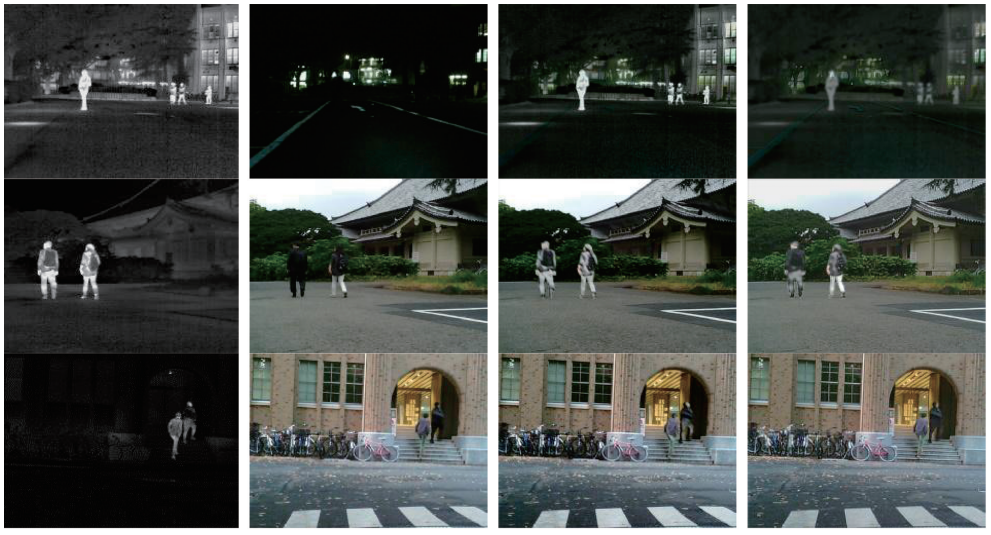
图 12 RoadScene 数据集上 20 组图像的 6 种指标分析

Fig. 12 Plot of the six metrics analyzed for the 20 images on the RoadScene dataset

3.5 消融实验

为了验证本文提出的不同模块的有效性，在 MSRS 数据集进行消融实验。对所提出的并联式深层特征提取模块采用单一模型实验：1）只采用 Restormer 作为特征提取；2）只采用 Res2Net 作为特征提取；3）Restormer 与 Res2Net 结合；4）网络仅去除细节补偿模块（DC）；5）网络仅去除分解网络（DN）；6）本文方法。结果如图 13 所示。

由图 13 可知，只采用 Restormer 作为特征提取时，在捕获全局上下文信息方面具有优势，但在局部细节特征的提取方面存在不足，导致地面标识模糊不清。只采用 Res2Net 作为特征提取时，局部特征和纹理细节的提取表现很好，但在全局信息的整合方面不如 Restormer。Restormer 与 Res2Net 结合的方式提供了最佳的全局和局部特征提取能力，人物细节和地面纹理得到了很好地保留。去除细节补偿模块会减少对输入数据细节信息的增强和保留，导致融合图像在纹理细节和边缘清晰度上的损失。没有分解网络，融合图像无法进行有效的逆向分解，导致目标细节丢失，且存在轻微过曝。本文方法结合了 Restormer 和 Res2Net 的特征提取能力，细节补偿模块及分解网络的完整网络架构，在图像融合质量、纹理细节的保留及信息的完整性方面提供最佳的表现。



(a) Infrared　(b) Visible　(c) Restormer　(d) Res2Net

(e) Restormer+Res2Net　(f) Restormer+Res2Net+DN　(g) Restormer+Res2Net+DC　(h) Ours

图 13　消融实验结果

Fig. 13　Results of ablation experiments

表 3 为消融实验定量分析结果。由表 3 可以看出，在本文方法中，信息熵 E_N 和结构相似性 S_S 均取得最优，说明整合各个模块显著改善了融合图像的质量。

表 3　消融实验定量分析

Tab. 3　Quantitative analysis of ablation experiments

Different Modules	E_N	S_S
Restormer	5.853 5	0.698 5
Res2Net	5.452 8	0.672 3
Restormer + Res2Net	6.011 7	0.711 6
Restormer + Res2Net + DN	6.485 1	0.742 7
Restormer + Res2Net + DC	6.528 5	0.767 5
Ours	**6.729 8**	**0.798 4**

3.6　运行效率对比分析

表 4 为不同对比方法在 MSRS 和 RoadScene 数据集上的平均运行时间（t_{MSRS}、$t_{RoadScene}$）。由表 4 可以看出，本文方法的平均运行时间位居第四。DenseFuse、IFCNN 和 RFN-Nest 的平均运行时间均比本文方法低，原因是这 3 种方法均采用了包含多个卷积层的传统网络结构，并设计了一种加法融合策略。

表 4　不同方法的平均运行时间

Tab. 4　Average running time of different methods　s

Methods	t_{MSRS}	$t_{RoadScene}$
DenseFuse	0.033 8	0.020 8
FusionGAN	1.788 2	1.093 9
IFCNN	0.140 5	0.023 4
PMGI	0.518 6	0.231 4
SDNet	0.181 2	0.093 7
RFN-Nest	0.137 9	0.029 6
U2Fusion	0.475 6	0.059 6
SeAFusion	0.168 8	0.046 2
SDCFusion	0.191 6	0.050 5
Ours	0.163 2	0.041 3

4 结 论

本文提出了一种融合与分解网络架构,并在公开的 MSRS 和 RoadScene 数据集上进行实验验证,得出以下结论:

1)提出的深层特征提取模块通过结合 Restormer 和 Res2Net 的优势,有效地捕捉了全局和局部特征,增强了网络的表征能力。

2)利用可逆神经网络的特性,本文设计的细节补偿模块显著提升了融合图像的纹理细节和视觉效果,确保了融合结果在保留边缘信息和纹理细节方面的优越性。

3)设计的分解网络能够将融合图像有效分解回原始的红外和可见光图像,确保了融合图像在多个层面上均保持原始图像的信息和细节。

3)RoadScene 数据集上的泛化实验得出,在 D_S、D_{SC}、G_A 和 F_S 指标方面,本文方法较其他对比方法分别平均提升了 8.5%、23.1%、49.0% 和 56.1%。MSRS 数据集上的测试实验得出,在 D_S、F_{VI}、G_A、D_{SC} 和 F_S 指标方面,本文方法较最新的 SDCFusion 方法分别提升了 1.4%、0.4%、0.6%、4.3% 和 3.4%。

4)本文方法不仅在融合结果上表现出色,而且在运行效率上也具有一定的优势。

参考文献

[1] KARIM S, TONG Geng, LI Jinyang, et al. Current advances and future perspectives of image fusion: a comprehensive review[J]. Information Fusion, 2023, 90: 185. DOI:10.1016/j.inffus.2022.09.019

[2] 唐霖峰, 张浩, 徐涵, 等. 基于深度学习的图像融合方法综述[J]. 中国图象图形学报, 2023, 28(1): 3
TANG Linfeng, ZHANG Hao, XU Han, et al. Deep learning-based image fusion: a survey[J]. Journal of Image and Graphics, 2023, 28(1): 3. DOI: 10.11834/jig.220422

[3] 张洲宇, 曹云峰, 丁萌. 采用多层卷积稀疏表示的红外与可见光图像融合[J]. 哈尔滨工业大学学报, 2021, 53(12): 51
ZHANG Zhouyu, CAO Yunfeng, DING Meng. Infrared and visible image fusion via multi-layer convolutional sparse representation[J]. Journal of Harbin Institute of Technology, 2021, 53(12): 51. DOI: 10.11918/202005038

[4] WANG Haozhe, SHU Chang, LI Xiaofeng, et al. Two-stream edge-aware network for infrared and visible image fusion with multi-level wavelet decomposition[J]. IEEE Access, 2024, 12: 22190. DOI: 10.1109/ACCESS.2024.3364050

[5] WU Peicong. Infrared and visible image fusion based on potential low-rank decomposition and anisotropic guided filtering[C]//6th International Conference on Electronics Technology (ICET). Chengdu: IEEE, 2023: 36. DOI: 10.1109/ICET58434.2023.10211821

[6] WANG Yang, CAO Xiaoqian, LIU Weifeng, et al. Improved sparse representation fusion rules based infrared and visible image fusion algorithm[C]//12th IEEE International Conference on Control, Automation and Information Sciences (ICCAIS). Hanoi: IEEE, 2023: 236. DOI: 10.1109/ICCAIS59597.2023.10382352

[7] LI Hui, WU Xiaojun. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614. DOI: 10.1109/TIP.2018.2887342

[8] LI Hui, WU Xiaojun, KITTLER J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images[J]. Information Fusion, 2021, 73: 72. DOI: 10.1016/j.inffus.2021.02.023

[9] 范焱, 刘乔, 袁笛, 等. 空域和频域特征解耦的红外与可见光图像融合[J]. 红外与激光工程, 2024, 53(8): 222
FAN Yan, LIU Qiao, YUAN Di, et al. Spatial and frequency domain feature decoupling for infrared and visible image fusion[J]. Infrared and Laser Engineering, 2024, 53(8): 222. DOI: 10.3788/IRLA20240198

[10] 李永萍, 杨艳春, 党建武, 等. 基于变换域 VGGNet19 的红外与可见光图像融合[J]. 红外技术, 2022, 44(12): 1293
LI Yongping, YANG Yanchun, DANG Jianwu, et al. Infrared and visible image fusion based on transform domain VGGNet19[J]. Infrared Technology, 2022, 44(12): 1293

[11] ZHANG Hao, XU Han, XIAO Yang, et al. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity[C]//34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 12797. DOI: 10.1609/AAAI.V34i07.6975

[12] MA Jiayi, YU Wei, LIANG Pengwei, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48: 11. DOI:10.1016/j.inffus.2018.09.004

[13] 许光宇, 陈浩宇, 张杰. 双路径双鉴别器生成对抗网络的红外与可见光图像融合[J]. 计算机辅助设计与图形学学报, 2024, 36(12): 1946
XU Guangyu, CHEN Haoyu, ZHANG Jie. Infrared and visible image fusion based on dual-path and dual-discriminator generation adversarial network[J]. Journal of Computer-Aided Design & Computer Graphics, 2024, 36(12): 1946. DOI: 10.3724/SP.J.1089.2024.20170

[14] PARMAR N, VASWANI A, USZKOREIT J, et al. Image transformer[C]//35th International Conference on Machine Learning. Stockholm: IMLS, 2018: 6453

[15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words: transformers for image recognition at scale[C]//9th International Conference on Learning Representations. Vienna: ICLR, 2021

[16] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 10012. DOI:10.1109/ICCV48922.2021.00986

[17] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orlean: IEEE, 2022: 5728. DOI:10.1109/CVPR52688.2022.00564

[18] DINH L, SOHL-DICKSTEIN J, BENGIO S. Density estimation using real NVP[C]//5th International Conference on Learning Representations. Toulon: ICLR, 2017

[19] GAO Shanghua, CHENG Mingming, ZHAO Kai, et al. Res2Net: a new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 652. DOI: 10.1109/TPAMI.2019.2938758

[20] TANG Linfeng, YUAN Jiteng, ZHANG Hao, et al. PIAFusion: a progressive infrared and visible image fusion network based on illumination aware[J]. Information Fusion, 2022, 83: 79. DOI: 10.1016/j.inffus.2022.03.007

[21] XU Han, MA Jiayi, JIANG Junjun, et al. U2Fusion: a unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 502. DOI: 10.1109/TPAMI.2020.3012548

[22] ZHANG Yu, LIU Yu, SUN Peng, et al. IFCNN:a general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99. DOI: 10.1016/j.inffus.2019.07.011

[23] ZHANG Hao, MA Jiayi. SDNet: a versatile squeeze and decomposition network for real-time image fusion[J]. International Journal of Computer Vision, 2021, 129(10): 2761. DOI: 10.1007/s11263-021-01501-8

[24] TANG Linfeng, YUAN Jiteng, MA Jiayi. Image fusion in the loop of high-level vision tasks: a semantic-aware real-time infrared and visible image fusion network[J]. Information Fusion, 2022, 82: 28. DOI: 10.1016/j.inffus.2021.12.004

[25] LIU Xiaowen, HUO Hongtao, LI Jing, et al. A semantic-driven coupled network for infrared and visible image fusion[J]. Information Fusion, 2024, 108: 102352. DOI: 10.1016/j.inffus.2024.102352