文章编号　1004-924X(2025)07-1152-17

基于多尺度空间注意力互补的红外与可见光图像融合

张永兴[1,2,3]，连博文[1,2,3]，顾乃庭[4*]，李方召[4]，李　杨[1,2*]
(1. 自适应光学全国重点实验室，四川 成都 610209；
2. 中国科学院 光电技术研究所，四川 成都 610209；
3. 中国科学院大学，北京 100049；
4. 国防科技大学 前沿交叉学科学院，湖南 长沙 410073)

摘要：针对当前红外与可见光图像融合方法过度引入红外冗余信息导致复杂场景下无法平衡复杂场景信息，融合效果不佳的现状，提出基于多尺度空间注意力互补的红外和可见光图像融合方法，采用双分支卷积网络分别提取红外和可见光图像特征信息并进行差异互补，利用多尺度空间注意力互补处理后回归叠加至图像特征中，实现互补特征中途回归叠加的图像融合，有效平衡复杂场景信息。实验结果表明：相比于Densefuse，PIAFusion等主流融合方法，该方法在通用性较强的互信息(MI)方面分别提升了4.1%和4.3%，在视觉信息保真度(VIF)方面分别提升了5.0%和2.3%，有效保留了复杂场景下的目标特征信息并实现对冗余特征的有效抑制，具有良好的特征平衡能力，在复杂场景下目标检测和识别中具有潜在应用价值。

关　键　词：图像融合；红外和可见光图像；双分支卷积网络；差异互补；多尺度空间注意力；回归叠加
中图分类号：TP391.4　　**文献标识码**：A
doi：10.37188/OPE.20253307.1152　　**CSTR**：32169.14.OPE.20253307.1152

Infrared and visible image fusion based on multi-scale spatial attention complementary

ZHANG Yongxing[1,2,3], LIAN Bowen[1,2,3], GU Naiting[4*], LI Fangzhao[4], LI Yang[1,2*]

(1. *National Key Laboratory of Adaptive Optics*, *Chengdu* 610209, *China*;
2. *Institute on Optics and Electronics Technology*, *Chinese Academy of Sciences*, *Chengdu* 610209, *China*;
3. *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*;
4. *College of Frontier Interdisciplinary Sciences*, *National University of Defense Technology*, *Changsha* 410073, *China*)
* *Corresponding author*, *E-mail*: *gnt*7328@163.*com*;*liyang@ioe.ac.cn*

Abstract: Current infrared and visible image fusion methods tend to introduce excessive redundant infrared

收稿日期：2024-12-31；**修订日期**：2025-01-15.
基金项目：国家重点研发计划资助项目(No. 2021YFC2202200)；国家自然科学基金资助项目(No. 12293031, No. 12403088)；四川省自然科学基金资助项目(No. 2024NSFSC1391)

information, impairing the ability to balance complex scene details and resulting in suboptimal fusion outcomes. To address these limitations, a novel fusion approach based on multi-scale spatial attention complementarity is proposed. This method employs a dual-branch convolutional network to separately extract features from infrared and visible images, followed by difference-based complementary processing. Multi-scale spatial attention mechanisms are then applied to the feature maps, culminating in regression-based superposition to achieve balanced fusion of complementary features. Experimental evaluations demonstrate that, compared to mainstream methods such as Densefuse and PIAFusion, the proposed approach achieves improvements of 4.1% and 4.3% in mutual information (MI), and 5.0% and 2.3% in visual information fidelity (VIF), respectively. These results indicate enhanced retention of target features and effective suppression of redundant information within complex scenes. The method exhibits strong feature balancing capabilities and holds significant potential for applications in target detection and recognition under challenging environmental conditions.

Key words: image fusion; infrared and visible image; double branch convolutional network; complementary difference; multi-scale spatial attention; regression overlay

1　引　言

随着光电探测水平的提升，对目标检测与识别的需求不断增强。由于探测维度不够、噪声干扰、分辨率低等因素影响，传统单一探测图像往往无法提供足够的目标特征或场景信息，导致目标信息获取不足、关键信息丢失以及冗余信息过多等一系列问题[1]，从而制约复杂场景下目标检测与识别能力。图像融合技术通过提取多源图像中的目标特征信息并进行整合处理，生成一幅包含多维特征的融合图像，从而解决单一数据源信息不足的问题，显著增强目标探测与识别应用中的信息表达能力，提升图像的质量，改善视觉效果，并有力支持目标有关的复杂场景分析、资源优化和智能决策，显著提升目标探测与识别系统的性能。相比较而言，红外与可见光探测是常用的目标探测方式，因此成为目前应用价值最高、应用最广泛的图像融合数据来源[2-3]。可见光图像与红外图像从不同光谱角度描述目标成像场景，反映不同条件下的场景和目标特征[4-5]。其中，可见光图像主要反映目标反射特性，场景及目标细节丰富、对比度高、更易于探测和识别，但易受光照条件影响，在遮挡、烟雾或夜间等场景下成像效果差；红外图像则主要反映目标自身的辐射特性，能在低光照和复杂环境下突出目标，适合夜间成像、遮挡成像等复杂场景，但对比度较低、分辨率低、缺乏细节，目标识别能力差[6]。因此，将红外图像和可见光图像中目标特征进行叠加并形成兼具二者显著特征的融合图像[7-8]，能够同时获取目标反射和辐射信息，有效抑制冗余场景信息，便于目标检测和识别，在军事监视[9]、目标检测[10-11]和车辆导航[12]等方面具有良好的应用前景。

鉴于红外与可见光图像融合的重要性，多种图像融合方法相继被提出，主要分为传统方法和深度学习方法。传统方法主要包括多尺度变换图像融合[13-16]、数据稀疏表示图像融合[17-18]、子空间描述图像融合[19-21]、特征显著性描述图像融合[22]以及多种方法兼用的混合方法[23]等。这些方法主要依赖事先设定的规则对目标特征进行提取和多源图像融合，对特定应用场景的应用效果较好，但对特征规律不确定的复杂场景效果有限。基于深度学习的图像融合[24-25]是一种端到端的数据驱动型图像融合方法，它利用神经网络[26-27]架构实现对目标特征的准确提取和多维度融合，复杂场景适应性强，具有良好的多源图像融合能力。根据神经网络架构的差异，基于深度学习的图像融合方法归纳为三类：自编码器(Autoencoder, AE)[28]、卷积神经网络(Convolutional Neural Network, CNN)[29]，以及生成对抗网络(Generative Adversarial Network, GAN)[30]。其中，AE方法主要基于图像数据集训练编码器和解码器来实现目标特征的提取和重建，并采用预

先设计好的融合策略在融合层实现图像特征融合，如Densefuse[31]方法；但此类方法中包含的密集块会提取较多无关信息并输出至最终的图像融合结果中，导致融合图像效果不佳。CNN方法一定程度上可以规避预设图像融合规则的局限性，通过设计卷积神经网络架构并辅以代价损失函数来不断修正和优化目标特征提取和融合参数，确保多源图像的融合性能，典型方法如IFCNN[32]，LRRNet[33]，SDNet[34]，PIAFusion[35]等。其中，IFCNN将融合层嵌入到训练过程中，摆脱了单一融合策略的限制，但简化的神经网络架构限制了目标特征的提取能力和融合表达能力，在复杂场景中容易丢失目标的关键特征信息；LRRNet通过可学习的低秩和稀疏分解模块完成特征分解与融合，减少了网络层数和参数量，显著提升了计算性能，但低秩分解会将红外图像中显著的目标信息误归为高频噪声从而显著降低融合图像中的红外目标特征；SDNet方法使用压缩分解网络来提高融合图像的质量，并利用梯度和强度的比例维持设计损失函数，但网络约束不足以及梯度和强度分配权重不合适，高频特征信息的提取会将红外图像中背景热辐射也带入融合图像中，导致多源信息失衡，显著降低多源图像融合的信噪比；PIAFusion方法从多源图像强度分布角度进行平衡，尤其对相同场景下光照度不同的可见光和红外图像，虽能够平衡二者强度并获取亮度适中的图像融合结果，但却存在无法有效识别并去除冗余信息、保证融合质量的问题。上述基于CNN的图像融合方法具有结构简单、计算高效、实时性强，以及红外信息保留度高等优势，不过也存在多源信息失衡、特征信息丢失、复杂场景信息冗余或特征尺度单一[36]等问题。GAN方法则基于生成对抗机制，如FusionGAN[37]，利用鉴别器约束生成器生成融合结果及其分布，在没有监督的情况下尽可能保持融合后特征与源图像特征的一致性，尤其对复杂场景具有良好的多源数据融合效果，其代价是生成器和鉴别器之间的对抗博弈会打破多源图像之间数据分布的平衡，使对抗网络训练不收敛，导致多源信息失衡。综上所述，无论是传统方法还是基于深度学习的方法，在复杂场景下红外和可见光图像融合质量和效率仍有明显的提升空间。

本文提出基于多尺度空间注意力互补的红外和可见光图像融合（Infrared and Visible Image Fusion based on Multi-scale Spatial Attention Complementary，IVF-MSAC）方法，采用双分支卷积网络分别提取红外和可见光图像特征信息并进行差异互补，通过并联增加多尺度空间注意力多级互补模块对图像特征进行多尺度空间注意力互补处理后，回归至特征提取器中与红外和可见光特征叠加融合，有效平衡复杂场景信息并形成融合图像特征。相对于传统方法，该融合方法充分利用信息的跨层连接，并与不同层次特征融合，实现多源图像特征信息平衡提取的同时，兼顾复杂场景中细节特征提取，抑制冗余特征。

2 原 理

2.1 网络总体架构

IVF-MSAC方法的总体架构如图1所示。该方法所涉及的网络架构主要包括特征提取器、图像重建器、多尺度空间注意力多级互补模块、光照判别辅助网络以及损失函数等部分。其中，特征提取器由两个相互独立的卷积分支组成，每个分支包含5个连续卷积层，均用于将红外和可见光图像提取为红外和可见光图像的特征，即特征图，尺寸为$H\times W\times C$，H为高度，W为宽度，C为通道数。通道数指的是特征图的深度或个数，通常可以理解为特征图中每个像素上存储的不同类型的特征的个数。在深度学习中，通道数对应每个特征图所提取的不同信息类别。一个通道就是一个特征图，不同通道可以表示不同的特征类型，如边缘、纹理等特征。多尺度空间注意力多级互补模块（Multi-scale Spatial Attention Multi-level Complementary Module，MSAMCM）主要由多尺度全局平均池化注意力单元（Multi-scale Global Average Pooling Attention Unit，MGAPAU）和多尺度空间注意力单元（Multi-Scale Spatial Attention Unit，MSAU）串联组成。随后进行特征融合时，网络不仅关注红外和可见光图像的特定信息，还关注其互补信息。为捕获两种图像之间的互补信息，将红外和可见光图像进行特征相减得到互补特征。这种相减是每个

特征图的像素相减，如式(1)所示：

$$\begin{aligned}\Delta F_{\mathrm{ir}} &= F_{\mathrm{ir}}^{i} - F_{\mathrm{vis}}^{i}\\ \Delta F_{\mathrm{vis}} &= F_{\mathrm{vis}}^{i} - F_{\mathrm{ir}}^{i},\end{aligned} \tag{1}$$

其中：ΔF_{ir}和ΔF_{vis}分别表示原红外互补特征和原可见光互补特征，F_{ir}^{i}和F_{vis}^{i}分别表示经特征提取器提取的红外图像特征和可见光图像特征。

如图1所示，将3个多尺度空间注意力多级互补模块并联在特征提取器第2～4层卷积层，对红外和可见光互补特征进行注意力机制增强，经注意力增强的互补特征回归叠加融合输入给下一层卷积层，与特征提取器提取的特征逐像素相加融合。特征提取器最后一层卷积层将融合后的红外和可见光图像源特征信息和经处理后的互补特征信息以通道拼接方式整合在一起，形成目标总特征信息，最终输入图像重建器进行图像重建。将特征提取器与图像重建器级联，表示从特征提取器双分支输出的特征实现通道拼接输入给图像重建器，即图像重建器的输入通道数为特征提取器双分支的输出通道数之和。图像重建器由5个连续的卷积层组成，用于重建融合后的特征，生成重建图像。光照判别辅助网络作为辅助部分为损失函数提供光照权重，它由4个卷积层，1个全局平均池化层和2个完全连接层组成，输入为可见光图像，输出为光照权重。在网络实现图像特征提取、融合和图像重建的训练过程中，损失函数对融合图像计算损失，通过不断优化提高融合图像与源图像的相似性，使训练完成的网络在实验验证时获得最优的融合效果。

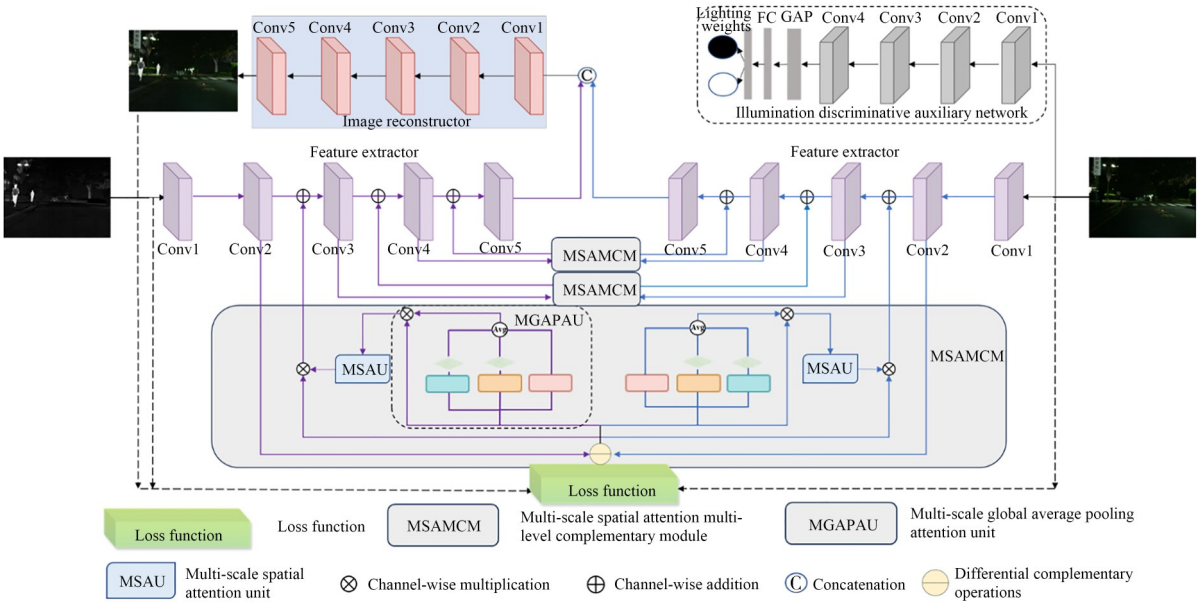

图1　基于多尺度空间注意力互补的红外和可见光图像融合网络总体框架

Fig. 1　Overall framework of progressive infrared and visible light image fusion network based on multi-level differential complementary multi-scale spatial attention

基于上述关键组成部分进行特征提取、融合和图像重建，本文提出的基于多尺度空间注意力互补的红外和可见光图像融合方法能够充分利用红外和可见光多层级融合过程中的中间层特征信息，并实现多层级跨层信息融合，有效避免图像特征信息的损失。

2.2　特征提取器与图像重建器

本文采用两独立分支的5个连续卷积层作为特征提取器进行特征提取，图像重建器采用单分支的5个连续卷积层进行图像重建。如表1所示，特征提取器中，第一个卷积层使用1×1的卷积核且步幅为1，输入通道为1，输出通道为16，不进行填充。这层主要用于通道数转换，减少空间信息的影响。第2～5个卷积层使用3×3的卷积核且步幅为1，均使用反射填充保持输出特征图尺寸减少边缘信息的损失。从第1至第5个卷积层，后一个卷积层的输入通道与前一个卷积层的输出通道相同。第2～5个卷积层的输出通道

为 16,32,64 和 128。特征提取器通过连续的卷积操作,随着通道数的增加可以提取越来越深层次的特征,捕捉到更多复杂的图像细节。各卷积层均采用 Leaky Relu 作为激活函数,网络在训练过程中的梯度变化会更加平滑,从而提升稳定性和收敛速度。此外,第 2~5 个卷积层共享权重,减少参数减轻计算成本。双分支卷积网络每一层的卷积操作都作用于可见光图像和红外图像,保证两类图像的特征提取过程相对独立,提取两种不同源的图像特征不受干扰,进而为后续的融合操作提供高质量的特征信息,最终用于融合。特征提取器和图像重建器以级联方式连接,意味着特征提取器提取的红外和可见光图像特征进行通道拼接,然后输入图像重建器。通道拼接是将特征图进行堆叠,即将所有特征全部堆叠,在通道数上表示为特征提取器提取的红外图像特征和可见光图像特征的通道数之和。

表 1　特征提取器和图像重建器中所有卷积层的核大小、输出通道和激活函数

Tab. 1　Kernel size, output channel, and activation function of all convolutional layers in feature extractor and image reconstructor

Layer	Feature extractor			Image reconstructor		
	Kernel size	Output channel	Activation function	Kernel size	Output channel	Activation function
Conv 1	1×1	16	LeakyRelu	3×3	256	Leaky Relu
Conv 2	3×3	16	LeakyRelu	3×3	128	Leaky Relu
Conv 3	3×3	32	LeakyRelu	3×3	64	Leaky Relu
Conv 4	3×3	64	LeakyRelu	3×3	32	Leaky Relu
Conv 5	3×3	128	LeakyRelu	1×1	1	Tanh

图像重建器将融合特征恢复为融合图像。图像重建部分包含 5 个卷积层,将输入图像重建器的目标总特征逆向恢复为融合图像。除最后一层卷积核尺寸为 1×1 外,所有层的卷积核尺寸均为 3×3,与特征提取器的卷积层对应。图像重建器第 1 个卷积层的输出通道为特征提取器双分支的最后一个卷积层之和为 256,第 2~5 层依次与特征提取部分对应,通道数为 128,64,32 和 1。除最后一层的激活函数为 Tanh 外,图像重建器中所有卷积层均采用 Leaky Relu 作为激活函数。在生成融合图像时,负值像素在实际应用中通常没有意义,因此采用 Tanh 激活函数,使输出范围符合实际需求。

2.3　多尺度空间注意力多级互补模块

在红外和可见光图像融合任务中,CNN 的融合方法未能有效利用中间卷积层提取的特征信息,导致有用特征信息丢失。因此,MSAMCM 关注互补特征信息并与红外和可见光特征实现相加融合,有效利用卷积网络的中间层信息,如图 2 所示。MSAMCM 对互补信息进行多尺度处理和注意力机制增强。注意力机制在处理图像时能够聚焦在最相关的部分,通过给每个输入部分分配不同的权重,让网络模型关注最重要的信息,自适应地选择哪些输入部分最重要,从而显著提升性能。MSAMCM 使显著特征信息得到增强,冗余特征信息被注意力权重削弱。显著特征与冗余特征的判断并非由人工直接定义,而是通过模块内部的操作和动态权重赋予机制自动实现的。该模块参与网络架构组成,根据输入特征的不同,动态计算注意力权重,并通过这种动态调整机制有针对性地增强显著特征,同时抑制冗余特征。最终,网络在训练过程中不断优化,自动拟合出最优的注意力增强策略。

MGAPAU 先对红外和可见光图像的互补特征进行第一级处理,通过在不同尺度上细化输入的互补特征来关注全局和局部信息,从而在融合过程中更为全面地利用不同尺度的特征。不同尺度的 GAP 机制均相同。MGAPAU 通过 1×1,2×2,4×4 等不同尺度的全局平均池化层,对输入的互补特征在不同尺度下计算每个特征(特征图)像素的平均值,并压缩成包含全局与局部信息的更小的多尺度特征。然后对这些特征逐元素应用 Sigmoid 函数,将特征的每个像素值归一化到(0,1)内。最后,对归一化后的特征进行插

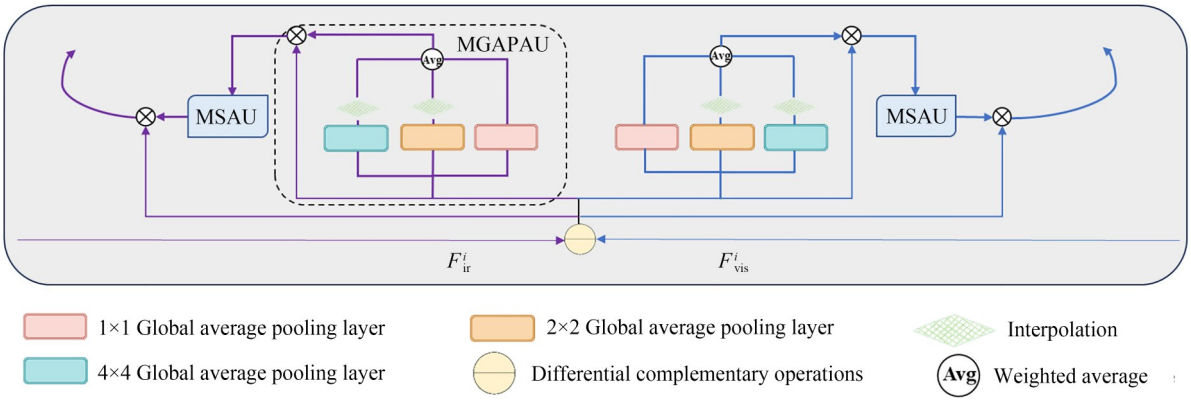

图 2　多尺度空间注意力多级互补模块

Fig. 2　Multi-scale spatial attention multi-level complementary module(MSAMCM)

值处理，并恢复到与原始互补特征相同的维度再求平均，得到捕获多尺度信息的注意力特征，如下：

$$\begin{cases}\phi_i^1=\sigma\left(GAP_1(\Delta F_{ir})\right)\\ \phi_v^1=\sigma\left(GAP_1(\Delta F_{vis})\right)\end{cases},\tag{2}$$

$$\begin{cases}\phi_i^2=\mathrm{Interp}\left(\sigma\left(GAP_2(\Delta F_{ir})\right)\right)\\ \phi_v^2=\mathrm{Interp}\left(\sigma\left(GAP_2(\Delta F_{vis})\right)\right)\end{cases},\tag{3}$$

$$\begin{cases}\phi_i^4=\mathrm{Interp}\left(\sigma\left(GAP_4(\Delta F_{ir})\right)\right)\\ \phi_v^4=\mathrm{Interp}\left(\sigma\left(GAP_4(\Delta F_{vis})\right)\right)\end{cases},\tag{4}$$

$$\begin{cases}\phi_i'=\left(\phi_i^1+\phi_i^2+\phi_i^4\right)/3\\ \phi_v'=\left(\phi_v^1+\phi_v^2+\phi_v^4\right)/3\end{cases},\tag{5}$$

其中：GAP_1表示1×1全局平均池化层，GAP_2表示2×2全局平均池化层，GAP_4表示4×4全局平均池化层，分别对输入的两种互补特征进行不同尺度的池化操作，ϕ_i^1，ϕ_i^2，ϕ_i^4和ϕ_v^1，ϕ_v^2，ϕ_v^4分别表示经1×2，2×2，4×4全局平均池化，归一化和插值后的注意力特征，σ表示Sigmoid函数，Interp表示插值；ϕ_i'和ϕ_v'分别表示最终加权平均后的红外注意力特征和可见光注意力特征；ΔF_{ir}和ΔF_{vis}分别为原红外互补特征和原可见光互补特征。

捕获多尺度信息的注意力特征，分别与对应的红外和可见光互补特征进行通道乘法，每个通道的特征图进行逐元素相乘操作，实现对不同通道的特征进行加权，控制每个通道的重要性。每个通道特征图的每个像素点都被对应的多尺度注意力权重重新加权，实现对互补特征的第一级处理，即有：

$$\begin{cases}\Delta F_{ir}'=\phi_i'\otimes(\Delta F_{ir})\\ \Delta F_{vis}'=\phi_v'\otimes(\Delta F_{vis})\end{cases},\tag{6}$$

其中：$\Delta F_{ir}'$和$\Delta F_{vis}'$分别表示经第一级MGAPAU处理后的红外互补特征和可见光互补特征，$\otimes$表示通道乘法；ΔF_{ir}和ΔF_{vis}分别为原红外互补特征和原可见光互补特征；ϕ_i'和ϕ_v'分别表示最终加权平均后的红外注意力特征和可见光注意力特征。

MGAPAU利用多尺度的注意力权重捕获不同尺度像素空间位置的重要性，体现注意力空间加权，从而使融合网络在特征选择时区分通道中像素位置的重要性，能够更好地聚焦关键信息。经第一级MGAPAU处理的互补特征进入MSAU实现第二级处理，对于给定的输入特征，MSAU先将输入的特征进行分组，目的是避免过多通道的相似特征混合造成信息冗余，同时降低计算复杂度。MSAU结构如图3所示，它采用不同尺度的多并行分支结构对特征进行处理和聚合，这种并行结构可以避免更多的顺序处理深度，通过在不同尺度上分别聚合信息来丰富特征聚合。具体来说，1×1尺度分支的输出处理全局信息，将特征在高度和宽度方向上进行自适应平均池化，然后进行通道拼接并输入1×1卷积核中，通过卷积核与Sigmoid函数进行加权整合，使得高度特征和宽度特征之间产生相互补充和交互的关系，形成一种融合高度和宽度的全局信息表达的特征。GroupNorm对每个组的特征进行标准化，避免特征值差异过大。而3×3尺度分支的输出增强局部空间感受野，补充局部细节信

息。接着，采用并行的交叉互补方式，实现1×1尺度分支的特征和3×3尺度分支的特征充分互补，通过矩阵和点积操作聚合并行路径的输出，对互补特征进行空间注意力增强。最终，每个分组组内的输出特征通过叠加和聚合成为未分组之前的形状，捕获像素级配对关系，突出全局和局部的信息。

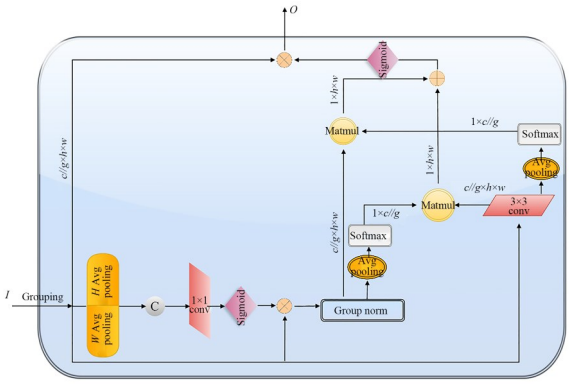

图3　多尺度空间注意力单元

Fig. 3　Multi-scale spatial attention unit

MSAU并行路径上高度H和宽度W的自适应平均池化将整个空间维度（高度和宽度）上的所有特征值进行平均，对于每个通道所有空间位置的高度和宽度均变为1，压缩成1×1的特征图，从而获得每个通道的全局特征信息。然后，Softmax函数根据每个通道的全局特征信息计算其权重，并在聚合时用这些权重控制不同通道特征的贡献，最终指导其他并行路径信息的加权聚合。并行路径上从1×1尺度分支和3×3尺度分支输出的特征，代表每个通道在空间维度上的信息（简称空间信息），经过自适应平均池化与Softmax函数的平行路径1×1尺度分支特征和不经处理的3×3尺度分支特征通过矩阵相乘进行形状匹配及运算加权，实现每个通道的全局信息与其他通道的空间信息结合。同理，与之对应的另一并行路径也实现了加权运算。这种并行的交叉互补方式下，对局部特征信息赋予全局信息权重，对全局特征信息赋予局部信息权重，实现了全局特征和局部特征的交互，增强关键特征同时抑制冗余特征。接着，将两尺度分支的各自平行路径的矩阵相乘结果逐元素相加，经Sigmoid函数归一化为空间注意力权重，并与输入到MSAU的原始特征相乘，最终得到每个分组的捕获全局和局部信息的特征。MSAU结合全局和局部空间信息的聚合建立远程依赖关系，通过融合不同尺度的网络关联信息，实现了深层次特征的像素级关注。此外，本文采用的多尺度卷积核的并行化设计方案能够捕获更多的上下文信息，从而显著增强融合网络的性能。

红外和可见光互补特征经MGAPAU第一级处理后输入到MSAU进行第二级处理，得到的互补特征与原红外和可见光图像特征进行相乘运算后得到新的目标特征，作为MSAMCM的输出红外和可见光补充特征，并最终回归到特征提取器中进行通道加法运算（即对每个通道对应的特征进行相加运算）。值得注意的是，可见光补充特征回归相加到特征提取器提取的红外特征中，红外补充特征回归相加到特征提取器提取的可见光特征中，得到红外和可见光总特征，分别如下：

$$F_{\mathrm{ir}}=F_{\mathrm{ir}}^{i}\oplus(\Delta F_{\mathrm{vis}})\otimes MSAU(\Delta F_{\mathrm{vis}}'),\quad(7)$$

$$F_{\mathrm{vis}}=F_{\mathrm{vis}}^{i}\oplus(\Delta F_{\mathrm{ir}})\otimes MSAU(\Delta F_{\mathrm{ir}}'),\quad(8)$$

其中：F_{ir}和F_{vis}分别表示MSAMCM输出和特征提取器输出相加融合的红外总特征和可见光总特征；$\oplus$表示通道加法，$\otimes$表示通道乘法；MSAU表示经多尺度空间注意力单元MSAU处理；ΔF_{ir}和ΔF_{vis}分别为原红外互补特征和原可见光互补特征；$\Delta F_{\mathrm{ir}}'$和$\Delta F_{\mathrm{vis}}'$分别表示经第一级MGAPAU处理后的红外互补特征和可见光互补特征；F_{ir}^{i}和F_{vis}^{i}分别为经特征提取器提取的红外图像特征和可见光图像特征。

MSAMCM处理后产生的补充特征回归到特征提取器中，与提取的红外和可见光特征相加实现特征融合，形成总特征，这样充分提取和融合多源图像中的互补信息。

总体而言，MSAMCM以红外与可见光特征差分结果为输入，突出多源图像间的差异性，并通过多尺度全局平均池化、分组归一化和卷积运算等，在多个空间尺度上捕获图像特征之间的相关性。此外，结合不同尺度下的全局特征信息和空间特征信息，通过不断地赋予注意力权重增强差异特性提高显著信息的特征表示能力，消除冗余信息，从而增强融合效果，避免融合时多源信息的失衡，获得高质量的融合图像。

2.4　损失函数与光照判别辅助网络

在基于深度学习的红外和可见光图像融合任务中，损失函数通过对融合网络生成的融合图像质量进行损失计算，根据源图像强度和梯度等信息决定融合图像信息的保留程度。在融合网络输入多源图像到输出融合图像的不断训练过程中，损失函数不断计算损失并迭代优化融合网络，损失最小时表明融合网络训练后的拟合能力最好。本文根据红外和可见光融合图像中的强度和梯度信息来设计损失函数，所采用的损失函数包括强度损失、辅助强度损失和梯度损失，以实现视觉质量高的融合效果。

强度损失衡量融合图像与源图像之间的像素强度差异，目的是确保融合图像在强度上与红外和可见光图像尽可能接近，使得融合后的图像在保持红外和可见光图像的强度信息方面达到平衡。强度损失如下：

$$\begin{aligned} L_{\text{int}}^{\text{ir}} &= \frac{1}{HW}\left\| I_f - I_{\text{ir}} \right\|_1 \\ L_{\text{int}}^{\text{vis}} &= \frac{1}{HW}\left\| I_f - I_{\text{vis}} \right\|_1 \end{aligned}, \tag{9}$$

$$L_{\text{int}} = w_{\text{ir}} L_{\text{int}}^{\text{ir}} + w_{\text{vis}} L_{\text{int}}^{\text{vis}}, \tag{10}$$

其中：H 和 W 分别为输入图像的高度和宽度；$\|\cdot\|_1$ 为 L1 范数；$L_{\text{int}}^{\text{ir}}$ 和 $L_{\text{int}}^{\text{vis}}$ 表示红外图像强度和可见光图像强度；I_{ir}，I_{vis}，I_f 分别表示红外图像、可见光图像、融合图像；w_{ir} 和 w_{vis} 表示红外图像强度和可见光图像强度的权重系数，由光照判别辅助网络提供；L_{int} 表示总强度损失。

光照判别辅助网络[33]组成部分包括 4 个卷积层、1 个全局平均池化层和 2 个完全连接层。具体来说，各卷积层的卷积核尺寸为 4×4，步幅为 2，激活函数为 LeakyRelu。卷积层从压缩输入的可见光图像中提取光照信息，经全局平均池化层整合光照信息后，最终由两个完全连接层根据光照信息计算光照概率。通过光照判别辅助网络可以计算输入的可见光图像是黑夜还是白天的光照概率，光照概率从侧面反映了源图像的信息丰富程度。红外图像强度损失的权重系数 w_{ir} 由黑夜概率与总概率比提供，可见光图像强度损失的权重系数 w_{vis} 由白天概率与总概率比提供，如下：

$$\begin{cases} w_{\text{ir}} = P_1/(P_1+P_2) \\ w_{\text{vis}} = P_2/(P_1+P_2) \end{cases}, \tag{11}$$

其中：P_1 表示图像属于黑夜的概率，P_2 表示图像属于白天的概率。

光照判别网络本质上作为一种分类器，可采用交叉熵损失 L_{CE} 约束光照判别网络的训练过程，如下：

$$L_{\text{CE}} = -m\log\sigma(n) - (1-m)\log\big(1-\sigma(n)\big), \tag{12}$$

其中：m 表示输入图像的昼夜光照标签，其值为 0 或 1，分别表示黑夜和白天；$n=[P_1,P_2]$ 表示输出是黑夜概率还是白天概率，σ 表示 softmax 函数，此函数将黑夜和白天的概率归一化为 [0,1]。

通过计算融合图像与红外和可见光图像中较强部分的最大值之间的 L1 损失，辅助强度损失 L_{as} 进一步保留重要的强度信息，有助于引导融合网络优先选择高质量的红外或可见光图像信息，从而提高融合图像的整体质量。辅助强度损失为融合图像保持最优的强度分布，其表达式如下：

$$L_{\text{as}} = \frac{1}{HW}\left\| I_f - \max(I_{\text{ir}}, I_{\text{vis}}) \right\|_1, \tag{13}$$

其中：$\max(\cdot)$ 表示逐元素的最大选择；H 和 W 分别为输入图像的高度和宽度；$\|\cdot\|_1$ 为 L1 范数；I_{ir}，I_{vis}，I_f 分别表示红外图像、可见光图像和融合图像。

梯度损失 L_{texture} 通常用于强调图像中如边缘等梯度变化区域，通过比较融合图像与红外图像和可见光图像的梯度，使融合图像的梯度信息尽可能接近这两种图像的最大梯度，从而减少融合图像中的噪声，使融合图像包含更多的纹理信息。融合图像的最优纹理细节是红外和可见光图像纹理细节的最大聚合，这对于图像的清晰度保持和细节保留非常重要。梯度损失表达式如下：

$$L_{\text{texture}} = \frac{1}{HW}\left\| |\nabla I_{\text{f}}| - \max\big(|\nabla I_{\text{ir}}|, |\nabla I_{\text{vis}}|\big) \right\|_1, \tag{14}$$

其中：∇ 为 Sobel 梯度算子，用于测量图像的细粒度纹理信息；$|\cdot|$ 表示绝对运算，$\max(\cdot)$ 表示逐元素的最大选择，H 和 W 分别为输入图像的高度和宽度，$\|\cdot\|_1$ 为 L1 范数；I_{ir}，I_{vis}，I_f 分别表示红外图像、可见光图像和融合图像。

最终将总强度损失 L_{int}、辅助强度损失 L_{as} 与梯度损失 L_{texture} 进行加权叠加，得到的总损失

如下：

$$L_f=\lambda_1\cdot L_{\text{int}}+\lambda_2\cdot L_{\text{as}}+\lambda_3\cdot L_{\text{texture}},\quad(15)$$

其中：λ_1，λ_2，λ_3分别表示强度损失、辅助强度损失、梯度损失重要性的权重参数，用来平衡不同损失项对融合结果的影响。

总体来说，损失函数的设计考虑了红外图像和可见光图像融合任务中的多个关键因素，旨在平衡强度保留、细节保留和边缘清晰度，确保融合图像既能有效保留红外图像的热信息，又能保留可见光图像的细节和边缘信息，从而提升图像的整体质量。

3　实　验

3.1　实验设置

为全面评估所提出的方法，本文在MSRS和RoadScene数据集上进行了实验验证，并将结果与目前主流应用的已有方法进行对比，包括基于自动编码器（AE）的Densefuse、基于生成对抗网络（GAN）的FusionGAN，以及基于卷积神经网络（CNN）的IFCNN，LRRNet，SDNet和PIAFusion。其中，MSRS和RoadScene数据集包含红外和可见光图像，这里将数据集裁剪成64×64的小尺寸图像，形成2万余对源图像数据用于训练本论文提出的融合网络，使用光照标签作为光照判别网络的参考，将白天和黑夜场景标签分别设置为向量[1,0]和[0,1]，按照先训练光照判别网络再训练融合网络的顺序进行数据训练。

训练过程中，通过预先训练好的光照判别网络计算光照概率，并进一步为红外和可见光的强度损失分配相应的权重。设定批量大小为b，它代表神经网络每次前向传播和反向传播所使用的样本数量。此外，训练需要的epoch设为M，即整个训练数据集完整地输入融合网络一次的次数。为了使融合网络能够更充分地学习数据分布，需要经过多个epoch（即对整个训练集进行多次遍历）。训练步长设定为p，表示在一个epoch内模型权重更新的次数。对于光照判别网络，上述参数经验设定为b=128，M=100，p=400；对于融合网络，将批量大小b设为64，训练epoch M为40，步长p为800。对于损失函数的权重参数，设置λ_1为3，λ_2为7，λ_3为50。模型参数的优化由Adam优化器完成，初始学习率设定为0.001，并采用指数衰减策略逐步降低学习率。实验验证在Pytorch平台上实现，所有实验均在NVIDIA GeForce RTX 3050 GPU和11th Gen Intel(R) Core(TM) i7-11800H@2.30 GHz CPU上进行。

在评价图像融合效果方面，本文采用通用性较强的互信息（Mutual Information，MI）[38]、视觉信息保真度（Visual Information Fidelity，VIF）[39]等指标作为评价依据。其中，MI主要用于评估从源图像传输到融合图像的信息量，通过量化源图像传递到融合图像的信息量，从信息论的角度评估融合性能；VIF则基于人类视觉系统的特性，衡量融合图像的信息保真度。一般来说，对于相同数据源，融合图像具有更大的MI和VIF，则说明图像融合方法的性能更好。

3.2　基于MSRS数据集的对比实验

3.2.1　融合图像对比

从MSRS数据集中选取包括道路、建筑、行人等目标特征的典型应用场景，并覆盖白天和黑夜等光照强烈变化的情况，采用多种图像融合方法进行图像融合实验，通过视觉效果对比和定量化指标对比验证本文融合方法的有效性和可行性。图4和图5分别展示了MSRS数据集中白天场景和夜晚场景下的图像融合效果。从图4中可以看出，白天场景下可见光图像具有较为丰富的目标和场景反射信息，图像分辨率高，颜色层次丰富，是图像融合的主要特征来源。但对于与周围色彩一致或反射率较低的目标，如行人等，却难以获取有效的目标特征，而红外图像可以提供目标热辐射维度的特征信息，对可见光特征进行有效补充。如图5所示，黑夜类低照度场景中，红外图像包含目标热辐射特征和部分场景特征信息，成为图像融合的主要特征来源，而可见光图像的场景特征信息有限，则实现对红外图像特征信息的有效补充。

如图5所示，在现有主流融合方法的实验结果中，采用FusionGAN方法得到的融合图像能较为清晰地显示黑暗背景中的井盖，但存在场景模糊的情况，无法保留可见光图像的场景特征信息；Densefuse方法能够保留可见光图像的场景特征信息，但行人红外特征稍有削弱并且背景中井盖的特征显示度不足；IFCNN，LRRNet和SDNet等方法虽然能够将可见光图像的场景特征信

息与红外图像目标的热辐射信息相结合，但 IFCNN，LRRNet 方法显著削弱了红外热目标特征，致使融合图像中两个行人的红外特征显著度较弱，同时 IFCNN 和 SDNet 方法对建筑物和告示牌的特征显示度不够清晰；PIAFusion 方法能够保留红外热目标特征显著度，清晰地显示建筑物和告示牌等图像特征，融合图像总体质量较好，但同样存在黑暗背景中井盖的特征显示度不足的情况。相比之下，本文方法生成的融合图像无论在建筑物和告示牌等纹理场景特征，还是在行人、黑暗背景中的井盖等红外特征方面均具有良好的特征显示度和清晰度，同时，抗红外背景辐射干扰能力也明显优于其他方法，融合图像质量和特征信息更为显著。

综合图 4 和图 5 所示的图像融合实验结果可以看出，本文提出的 IVF-MSAC 方法在保留目标热辐射特征、突显场景特征、抗噪声干扰等方面表现出更优良的性能，能够有效地展示目标及所在场景特征信息，并实现不同光照条件下红外和可见光源图像特征融合的平衡互补，使得融合图像具有更优的特征展示效果。

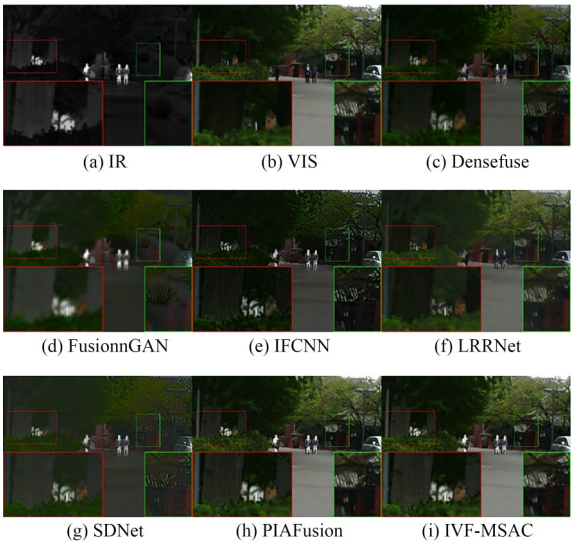

图 4　基于 MSRS 数据集 559 样本白天场景对比实验结果（红框区域为融合图像中树木的放大显示，绿框区域为房屋与窗户的放大显示，彩图见期刊电子版）

Fig. 4　Comparison experimental results of sample 559 based on MSRS dataset for daytime scence (The red-framed area is a zoomed-in display of trees, and the green-framed area is a zoomed-in display of houses and windows in the fused image, color figures is showed in PDF file)

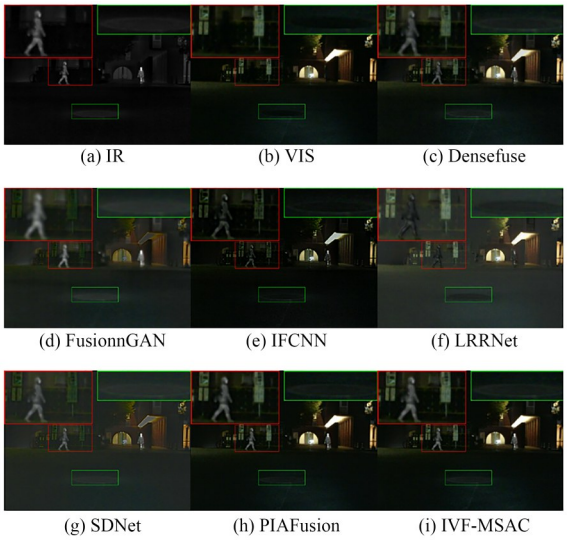

图 5　基于 MSRS 数据集 808 样本夜晚场景对比实验结果（红框区域为融合图像中行人和告示牌的放大显示，绿框区域为井盖的放大显示，彩图见期刊电子版）

Fig. 5　Comparison experimental results of sample 808 based on MSRS dataset for nighttime scene (The red framed area is a zoomed-in view of pedestrians and notice boards, and the green framed area is a zoomed-in view of manhole covers in the fused image, color figures is showed in PDF file)

3.2.2　融合图像性能

为进一步验证所提方法的性能和适用性，本文基于 MSRS 数据集对不同方法的融合效果进行了指标定量化实验和对比分析。从 MSRS 数据选择 11 组红外和可见光图像，用上述方法对这些具有丰富特征信息的源图像进行融合。通过计算源图像融合后每个融合图像的 MI 和 VIF 指标来表征融合图像对源图像信息的保留程度和视觉效果，最终对得到的 MI 和 VIF 指标值计算平均值和标准差来消除个例影响，增强评估的客观性，结果如表 2 所示。更高的 MI 度量意味着融合图像从源图像中获得最多的信息，更高的 VIF 表明融合图像具有更高质量的视觉效果。从表 2 可以看出，本文方法的 MI 和 VIF 指标最高，分别达到 3.515 2±1.013 6 和 1.051 4±0.082 2，比其他方法指标的最好结果分别提高约 4.1% 和 5.0%。

表 2　基于 MSRS 数据集本文方法在 MI, VIF 指标的平均值和标准差上的对比

Tab. 2　Comparison of mean and standard deviation of MI and VIF indicators between existing methods and proposed method on MSRS dataset

Method	MI	VIF
Densefuse	2.392 8±0.486 2	0.887 9±0.176 0
FusionGAN	1.979 0±0.280 1	0.498 4±0.142 4
IFCNN	1.736 8±0.506 4	0.731 5±0.061 7
LRRNet	2.264 2±0.807 7	0.489 5±0.054 9
SDNet	1.915 9±0.331 6	0.577 2±0.059 7
PIAFusion	3.370 7±1.009 1	0.999 3±0.067 3
IVF-MSAC	**3.515 2±1.013 6**	**1.051 4±0.082 2**

本文采用累积分布函数对每个融合图像指标值 MI 和 VIF 进行累积，可以更直观展示不同方法性能的优越性，结果分别如图 6 和图 7 所示。不同方法对应的 MI 和 VIF 取值越大，表示在相同累积概率下该方法得到的融合图像从源图像获得的信息越多，融合图像的视觉质量更高。可以看出，本文方法的曲线始终保持最高。

综合表 2 和图 6～图 7 结果可以看出，本文提出方法可以将源图像的有效信息更多保留到融合图像中，融合结果的视觉质量更高，对应的 MI 和 VIF 指标也更好。此外，本文提出方法在基于

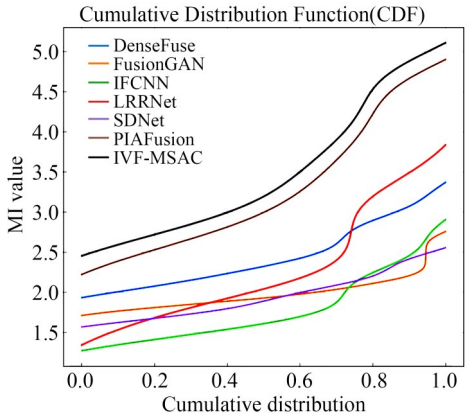

图 6　基于 MSRS 数据集的不同图像融合方法在不同累计概率下 MI 指标对比

Fig. 6　Comparison of MI index for different image confusion methods under different cumulative propagation on MSRS dataset

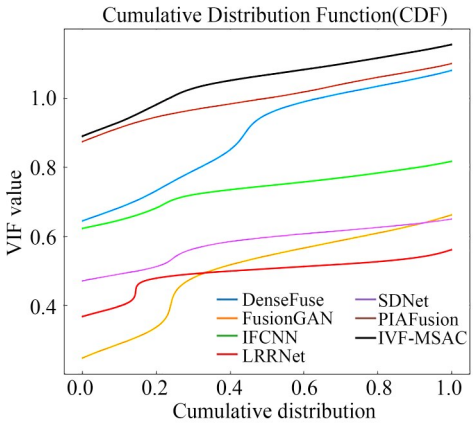

图 7　基于 MSRS 数据集的不同图像融合方法在不同累计概率下 VIF 指标对比

Fig. 7　Comparison of VIF index for different image confusion methods under different cumulative propagation on MSRS dataset

MSRS 数据集的图像融合过程中的平均耗时约为 0.43 s，相对于其他 6 种方法来说处于中等水平（0.1～4.5 s）。虽然增加多个处理模块会增加算法的复杂度，但显著提升了图像融合质量，并可以满足大多数场景中可见光与红外图像融合处理的实时性要求。

3.3　基于 RoadScene 数据集的对比实验

3.3.1　融合图像对比

为进一步验证本文提出融合方法的普适性，本文在另一数据集 RoadScene 上选取包含道路、建筑、行人等目标特征的典型应用场景，采用多种图像融合方法进行图像融合实验，通过视觉效果对比和定量化指标对比验证本论文提出方法的有效性和可行性。图 8～图 11 展示了多种融合方法在 RoadScene 数据集不同应用场景中的视觉效果对比。

从图 8～图 11 可以看出，采用 FusionGAN 方法可以保留红外图像目标的热辐射信息，但得到的融合图像场景模糊，无法保留可见光图像场景的特征信息。Densefuse, IFCNN, LRRNet 和 SDNet 等方法虽然能够将可见光图像的场景信息与红外图像目标的热辐射信息相结合，但融合图像背景区域会受到红外图像背景辐射干扰，显著降低视觉效果和对比度。其中，Densefuse, LRRNet, SDNet 方法削弱了红外热目标特征，致使融合图像中 3 个行人红外特征显著度较弱，如绿框区域中的行人（彩图见期刊电子版）。上述 4

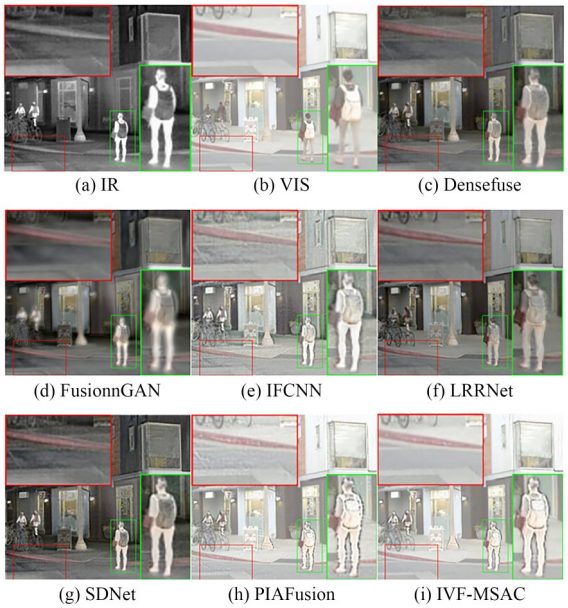

图 8　基于 RoadScene 数据集 08835 样本对比实验结果（红框区域为融合图像中道路和自行车的放大显示，绿框区域为行人的放大显示，彩图见期刊电子版）

Fig. 8　Comparison experimental results of sample 08835 on RoadScene dataset (The red-framed and green-framed areas are zoomed-in display of roads and bicycles, and pedestrians respectively in the fused image, color figures is showed in PDF file)

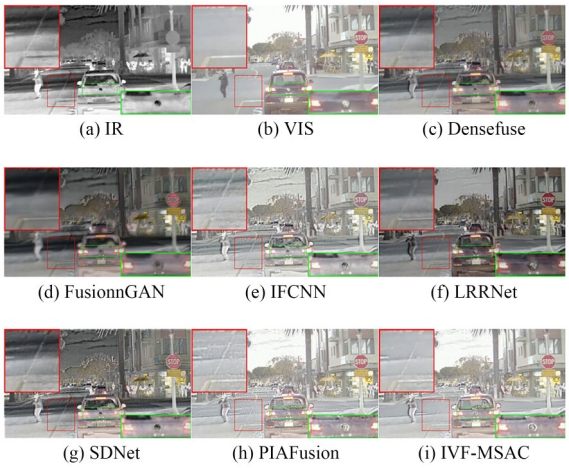

图 9　基于 RoadScene 数据集 08749 样本对比实验结果（红框区域为融合图像中道路的放大显示，绿框区域为汽车后门的放大显示，彩图见期刊电子版）

Fig. 9　Comparison experimental results of sample 08749 on RoadScene dataset (The red-framed and green-framed areas are zoomed-in views of road and rear door of car respectively in the fused image, color figures is showed in PDF file)

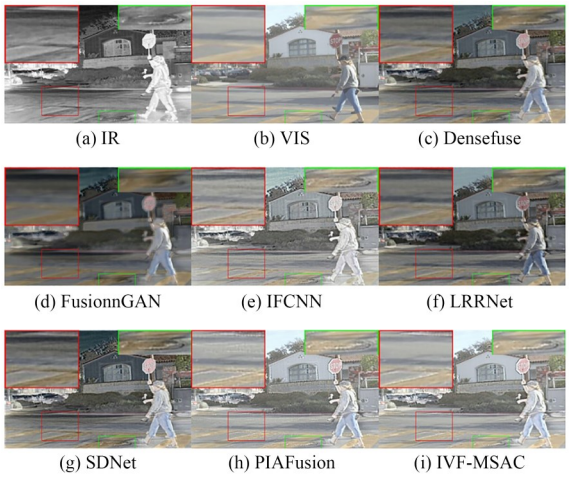

图 10　基于 RoadScene 数据集 05027 样本对比实验结果（红框区域为融合图像中道路的放大显示，绿框区域为井盖的放大显示，彩图见期刊电子版）

Fig. 10　Comparison experimental results of sample 05027 based on RoadScene dataset (The red-framed and green-framed areas are zoomed-in views of road and manholecover respectively in the fused image, color figures is showed in PDF file)

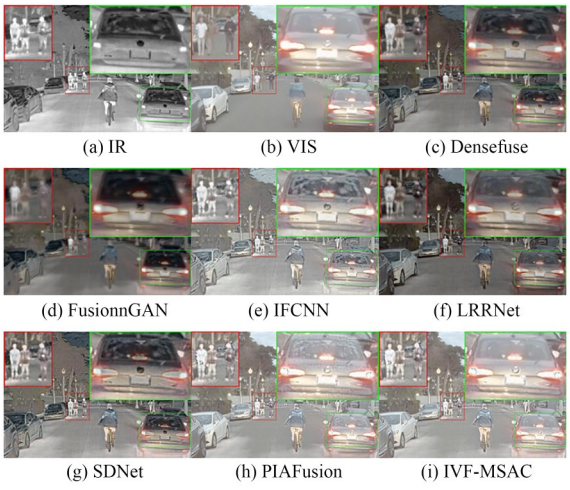

图 11　基于 RoadScene 数据集 06832 样本对比实验结果（红框区域为融合图像中行人的放大显示，绿框区域为汽车整体的放大显示，彩图见期刊电子版）

Fig. 11　Comparison experimental results of sample 06832 based on RoadScene dataset (The red-framed and green-framed areas are zoomed-in views of pedestrians and car as a whole in the fused image, color figures is showed in PDF file)

种方法生成的融合图像中，自行车、道路、井盖、远处建筑和近处汽车等均存在特征较为模糊、难以分辨的情况。PIAFusion方法则能够保留红外热目标特征显著度，清晰地显示自行车、树木、建筑等图像特征，融合图像总体质量较好，但存在引入红外冗余信息的情况，如红框中道路的白色伪影、绿框中车辆的白色伪影等，红外热目标边缘的清晰度不足。相比而言，本文方法生成的融合图像无论在自行车、建筑、树木和道路等可见光场景特征，还是在行人等红外特征方面均具有良好的特征显示度和清晰度，同时抗红外背景辐射干扰，去除噪声也显著比其他方法更优，融合图像质量和特征信息更为显著。

对比图像融合结果可知，本文提出的IVF-MSAC方法在保留目标热辐射特征、突显可见光场景特征、抗噪声干扰等方面的性能表更优，能够有效地展示目标及所在场景特征信息，在红外与可见光图像融合方面具有明显的性能优势。

3.3.2 融合图像性能

为进一步验证所提方法的性能和适用性，本文基于RoadScene数据集对不同方法融合效果进行了指标定量化实验和对比分析。表3给出了15组RoadScene数据集中具有代表性的源图像融合结果指标值MI和VIF的平均值和标准差，可以看出，本文方法的MI和VIF指标最高，分别达到3.678 1±0.378 5和0.582 3±0.085 5，比其他方法指标的最好结果仍分别提高约4.3%和2.3%。

表3 基于RoadScene数据集本文方法在MI，VIF指标的平均值和标准差上的对比

Tab. 3 Comparison of average and standard deviation of MI and VIF indicators between existing methods and proposed method on RoadScene dataset

Method	MI	VIF
Densefuse	2.522 9±0.419 7	0.479 7±0.040 8
FusionGAN	2.492 6±0.306 3	0.320 2±0.023 2
IFCNN	2.347 1±0.391 6	0.473 3±0.038 6
LRRNet	2.608 1±0.411 7	0.437 2±0.064 4
SDNet	2.953 5±0.432 3	0.493 0±0.033 5
PIAFusion	3.519 7±0.331 9	0.568 9±0.077 7
IVF-MSAC	**3.678 1±0.378 5**	**0.582 3±0.085 5**

图12和图13给出了基于RoadScene数据集的不同图像融合方法在不同累计概率下MI及VIF指标的对比分析结果。从图中可以看出，本文方法在不同累计概率下对应的MI和VIF指标均最高，即性能最优。这表明本文方法能够有效保留源图像中的更多关键信息，从而生成质量更高的融合图像。此外，本文方法在基于RoadScene数据集的图像融合过程中的平均耗时约为

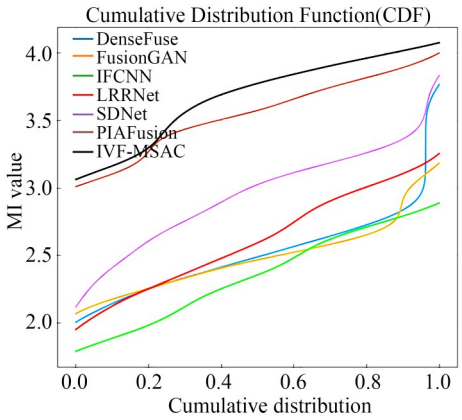

图12 基于RoadScene数据集不同图像融合方法在不同累计概率下的MI指标对比

Fig. 12 Comparison of MI index for different confusion methods under different cumulative propagation based on RoadScene dataset

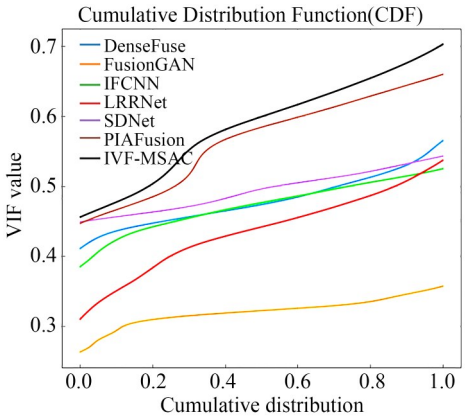

图13 基于RoadScene数据集不同图像融合方法在不同累计概率下的VIF指标对比

Fig. 13 Comparison of VIF index for different confusion methods under different cumulative propagation on RoadScene dataset

0.25 s，相对于其他6种对比方法处于中等水平（0.04～3.96 s），可以满足大多数场景中可见光与红外图像融合处理的实时性要求。

3.4　消融实验

考虑到本文方法通过设计MSAMCM实现多源图像特征信息平衡提取的同时，兼顾复杂场景中细节特征提取，抑制冗余特征。为了验证该模块的合理性，进行了消融实验，即去掉该模块的网络和原网络的融合效果进行对比，结果如图14所示。从图中可以明显看出，将MSAMCM消融后，融合网络无法将红外和可见光图像中的目标特征信息进行良好融合，融合结果受红外冗余场景信息影响明显，信息失衡且细节特征不显著。然而，未消融的融合图像目标的细节特征突出，信息比较平衡，对冗余特征进行了有效抑制。该结果进一步验证了本文方法的有效性和实用性。

图14　典型场景下融合图像的消融实验结果

Fig. 14　Ablation results of fusion images in three typical scenes

4　结　论

本文提出了一种基于多尺度空间注意力互补的红外和可见光图像融合方法，采用双分支卷积网络提取目标在红外和可见光波段特征信息，并差异互补后输入多尺度空间注意力多级互补模块处理得到输出特征，最终回归叠加至原图像实现目标特征中途叠加的图像融合。该方法在保留源图像信息特征的同时，将赋予注意力权重的互补信息聚集和融合，充分利用信息的跨层连接，并与不同层次特征融合，实现多源图像特征信息的平衡提取，兼顾图像细节特征提取，抑制冗余特征信息，实现像素强度分布和复杂场景特征方面的平衡，达到良好的图像融合效果。在MSRS和RoadScene数据集上进行实验对比，结果表明：相比于Densefuse，PIAFusion等主流融合方法，本文提出方法在通用性较强的互信息（MI）方面分别提升了4.1%和4.3%，在视觉信息保真度（VIF）方面分别提升了5.0%和2.3%。相比传统图像融合方法，本文方法具备自动特征学习、端到端训练、多尺度信息融合、噪声冗余抑制、结构信息保持等特点，图像融合自动化和适应性强，且关注目标多层次、多尺度特征，融合结果真实性和抗噪性能优越，能够克服传统方法在复杂环境适应性弱、目标特征保持不佳、冗余特征识别及消除能力不高等局限，获取高质量的图像融合结果。此外，本文方法虽然一定程度上会牺牲计算效率，计算实时性并非最优，但可以满足绝大多数应用场景要求。未来将在简化计算复杂度、提升计算实时性等方面继续深入研究。

作者贡献声明：

张永兴：论文方法的提出及验证，论文构思和撰写；

连博文：实验数据管理，论文修改；

顾乃庭：获取资助，提供思路，论文指导、审核与编辑；

李方召：论文审核与编辑；

李杨：研究工作管理，论文审核与编辑。

参考文献：

[1] ZHANG H, XU H, TIAN X, *et al*. Image fusion meets deep learning: a survey and perspective[J]. *Information Fusion*, 2021, 76: 323-336.

[2] 马鑫, 喻春雨, 童亦新, 等. 基于二次图像分解的红外图像与可见光图像融合[J]. 光学 精密工程, 2024, 32(10): 1567-1581.
MA X, YU CH Y, TONG Y X, *et al*. Infrared image and visible image fusion algorithm based on secondary image decomposition [J]. *Opt. Precision Eng.*, 2024, 32(10): 1567-1581. (in Chinese)

[3] TAHARA T. Adding dimensions with Lucy - Richardson-Rosen algorithm to incoherent imaging [J]. *Opto-Electronic Advances*, 2023, 6(5): 230047.

[4] FANG A Q, LI Y. Dynamic and static fusion mechanisms of infrared and visible images [J]. *Pattern Recognition*, 2024, 155: 110689.

[5] TANG L, ZHANG H, XU H, *et al*. Deep learning-based image fusion: A survey[J]. *Journal of Image and Graphics*, 2023, 28(1): 3-36.

[6] 龙志亮, 邓月明, 谢竞, 等. 改进多尺度结构化融合的红外与可见光图像融合[J]. 光学 精密工程, 2024, 32(7): 1101-1110.
LONG ZH L, DENG Y M, XIE J, *et al*. Infrared and visible images fusion based on improved multi-scale structural fusion [J]. *Opt. Precision Eng.*, 2024, 32(7): 1101-1110. (in Chinese)

[7] 徐少平, 周常飞, 肖建, 等. 基于预训练固定参数和深度特征调制的红外与可见光图像融合网络[J]. 电子与信息学报, 2024, 46(8): 3305-3313.
XU SH P, ZHOU CH F, XIAO J, *et al*. A fusion network for infrared and visible images based on pre-trained fixed parameters and deep feature modulation [J]. *Journal of Electronics & Information Technology*, 2024, 46(8): 3305-3313. (in Chinese)

[8] 龙志亮, 邓月明, 谢竞, 等. 改进多尺度结构化融合的红外与可见光图像融合[J]. 光学 精密工程, 2024, 32(7): 1101-1110.
LONG Z L, DENG Y M, XIE J, *et al*. Infrared and visible images fusion based on improved multi-scale structural fusion [J]. *Opt. Precision Eng.*, 2024, 32(7): 1101-1110. (in Chinese)

[9] WANG Z G, ZHAN J, LI Y, *et al*. A new scheme of vehicle detection for severe weather based on multi-sensor fusion [J]. *Measurement*, 2022, 191: 110737.

[10] SUN L, LI Y H, MUHAMMAD G. Soft computing-driven infrared and visible image fusion network for security application service [J]. *Applied Soft Computing*, 2024, 165: 112114.

[11] 郑海君, 葛斌, 夏晨星, 等. 多特征聚合的红外-可见光行人重识别[J]. 光电工程, 2023, 50(7): 110-123.
ZHENG H J, GE B, XIA C X, *et al*. Infrared-visible person re-identification based on multi feature aggregation [J]. *Opto-Electronic Engineering*, 2023, 50(7): 110-123. (in Chinese)

[12] LUO X, ZHU H, ZHANG Z L. IR-YOLO: real-time infrared vehicle and pedestrian detection [J]. *Computers*, Materials & Continua, 2024, 78(2): 2667-2687.

[13] LIU Y P, JIN J, WANG Q, *et al*. Region level based multi-focus image fusion using quaternion wavelet and normalized cut[J]. *Signal Processing*, 2014, 97: 9-30.

[14] LIU X B, MEI W B, DU H Q. Structure tensor and nonsubsampled shearlet transform based algorithm for CT and MRI image fusion [J]. *Neurocomputing*, 2017, 235: 131-139.

[15] ZHANG Q, MALDAGUE X. An adaptive fusion approach for infrared and visible images based on NSCT and compressed sensing[J]. *Infrared Physics & Technology*, 2016, 74: 11-20.

[16] CHEN J, LI X J, LUO L B, *et al*. Infrared and visible image fusion based on target-enhanced multi-scale transform decomposition[J]. *Information Sciences*, 2020, 508: 64-78.

[17] LI H, WU X J, KITTLER J. MDLatLRR: a novel decomposition method for infrared and visible image fusion[J]. *IEEE Transactions on Image Process-*

ing, 2020. DOI: 10.1109/TIP.2020.2975984.

[18] LIU Y, CHEN X, WARD R K, *et al*. Image fusion with convolutional sparse representation[J]. *IEEE Signal Processing Letters*, 2016, 23(12): 1882-1886.

[19] CVEJIC N, BULL D, CANAGARAJAH N. Region-based multimodal image fusion using ICA bases[J]. *IEEE Sensors Journal*, 2007, 7(5): 743-751.

[20] MOU J, GAO W, SONG Z X. Image fusion based on non-negative matrix factorization and infrared feature extraction[C]. 2013 6*th International Congress on Image and Signal Processing* (*CISP*). *December* 16-18, 2013. *Hangzhou*, *China*. IEEE, 2013: 1046-1050.

[21] FU Z Z, WANG X, XU J, *et al*. Infrared and visible images fusion based on RPCA and NSCT[J]. *Infrared Physics & Technology*, 2016, 77: 114-123.

[22] LIU C H, QI Y, DING W R. Infrared and visible image fusion method based on saliency detection in sparse domain[J]. *Infrared Physics & Technology*, 2017, 83: 94-102.

[23] MA J L, ZHOU Z Q, WANG B, *et al*. Infrared and visible image fusion based on visual saliency map and weighted least square optimization[J]. *Infrared Physics & Technology*, 2017, 82: 8-17.

[24] WU S T, YANG Z C, MA C G, *et al*. Deep learning enhanced NIR-II volumetric imaging of whole mice vasculature[J]. *Opto-Electronic Advances*, 2023, 6(4): 220105.

[25] 马鑫, 喻春雨, 童亦新, 等. 基于二次图像分解的红外图像与可见光图像融合[J]. 光学 精密工程, 2024, 32(10): 1567-1581.
MA X, YU CH Y, TONG Y X, *et al*. Infrared image and visible image fusion algorithm based on secondary image decomposition[J]. *Opt. Precision Eng.*, 2024, 32(10): 1567-1581. (in Chinese)

[26] LU Y X, ZHANG W J, ZHAO D W, *et al*. PTP-Fusion: a progressive infrared and visible image fusion network based on texture preserving[J]. *Image and Vision Computing*, 2024, 151: 105287.

[27] LU G S, FANG Z L, TIAN J J, *et al*. GANHA: a generative adversarial network with a novel heterogeneous dual-discriminator network and a new attention-based fusion strategy for infrared and visible image fusion[J]. *Infrared Physics & Technology*, 2024, 142: 105548.

[28] LONG Y Z, JIA H T, ZHONG Y D, *et al*. RXDNFuse: a aggregated residual dense network for infrared and visible image fusion[J]. *Information Fusion*, 2021, 69: 128-141.

[29] ZHANG H, XU H, XIAO Y, *et al*. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12797-12804.

[30] 杨艳春, 高晓宇, 党建武, 等. 基于WEMD和生成对抗网络重建的红外与可见光图像融合[J]. 光学 精密工程, 2022, 30(3): 320-330.
YANG Y CH, GAO X Y, DANG J W, *et al*. Infrared and visible image fusion based on WEMD and generative adversarial network reconstruction[J]. *Opt. Precision Eng.*, 2022, 30(3): 320-330. (in Chinese)

[31] LI H, WU X J. DenseFuse: a fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2019, 28(5): 2614-2623.

[32] ZHANG Y, LIU Y, SUN P, *et al*. IFCNN: a general image fusion framework based on convolutional neural network[J]. *Information Fusion*, 2020, 54: 99-118.

[33] LI H, XU T Y, WU X J, *et al*. LRRNet: a novel representation learning guided fusion network for infrared and visible images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(9): 11040-11052.

[34] ZHANG H, MA J Y. SDNet: a versatile squeeze-and-decomposition network for real-time image fusion[J]. *International Journal of Computer Vision*, 2021, 129(10): 2761-2785.

[35] TANG L F, YUAN J T, ZHANG H, *et al*. PIAFusion: a progressive infrared and visible image fusion network based on illumination aware[J]. *Information Fusion*, 2022, 83: 79-92.

[36] 祁艳杰, 侯钦河. 多尺度和卷积注意力相结合的红外与可见光图像融合[J]. 红外技术, 2024, 46(9): 1060-1069.
QI Y J, HOU Q H. Infrared and visible image fusion combining multi-scale and convolutional attention[J]. *Infrared Technology*, 2024, 46(9): 1060-1069. (in Chinese)

[37] MA J Y, YU W, LIANG P W, *et al*. Fusion-

GAN: a generative adversarial network for infrared and visible image fusion[J]. *Information Fusion*, 2019, 48: 11-26.

[38] QU G H, ZHANG D L, YAN P F. Information measure for performance of image fusion[J]. *Electronics Letters*, 2002, 38(7): 313-315.

[39] HAN Y, CAI Y Z, CAO Y, *et al*. A new image fusion performance metric based on visual information fidelity[J]. *Information Fusion*, 2013, 14(2): 127-135.

作者简介：

张永兴(1999—),男,河北保定人,硕士研究生,主要从事基于人工智能的多源图像融合处理技术的研究。E-mail: zhangyongxing22@mails. ucas. ac. cn

通讯作者：

顾乃庭(1984—),男,安徽怀远人,博士,教授,博士生导师,2012年于中国科学院大学获得博士学位,主要从事光电望远镜、自适应光学以及偏振光场探测成像等研究。E-mail: gnt7328@163. com

通讯作者：

李　杨(1990—),男,陕西宁强人,博士,副研究员,2017年于中国科学院大学获得博士学位,主要研究方向为自适应光学、波前探测及相关技术研究。E-mail: liyang@ioe. ac. cn