

基于多尺度对比度增强和跨维度交互注意力机制的 红外与可见光图像融合

邸敬, 梁婵, 任莉, 郭文庆, 廉敬
(兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070)

摘要: 针对目前红外与可见光图像融合存在特征提取不足、融合图像目标区域不显著、细节信息缺失等问题, 提出了一种多尺度对比度增强和跨维度交互注意力机制的红外与可见光图像融合方法。首先, 设计了多尺度对比度增强模块, 以增强目标区域强度信息利于互补信息的融合; 其次, 采用密集连接块进行特征提取, 减少信息损失最大限度利用信息; 接着, 设计了一种跨维度交互注意力机制, 有助于捕捉关键信息, 从而提升网络性能; 最后, 设计了从融合图像到源图像的分解网络使融合图像包含更多的场景细节和更丰富的纹理细节。在 TNO 数据集上对提出的融合框架进行了评估实验, 实验结果表明本文方法所得融合图像目标区域显著, 细节纹理丰富, 具有更优的融合性能和更强的泛化能力, 主观性能和客观评价优于其他对比方法。

关键词: 红外与可见光图像融合; 多尺度对比度增强; 跨模态交互注意力机制; 分解网络
中图分类号: TP391.9 **文献标志码:** A **文章编号:** 1001-8891(2024)07-0754-11

Infrared and Visible Image Fusion Based on Multi-Scale Contrast Enhancement and Cross-Dimensional Interactive Attention Mechanism

DI Jing, LIANG Chan, REN Li, GUO Wenqing, LIAN Jing

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

Abstract: Addressing the issues of inadequate feature extraction, lack of saliency in fused image regions, and missing detailed information in infrared-visible image fusion, this paper proposes a method for infrared-visible image fusion based on multi-scale contrast enhancement and a cross-modal interactive attention mechanism. The main components of the proposed method are as follows. 1) Multi-scale contrast enhancement module: Designed to strengthen the intensity information of target regions, facilitating the fusion of complementary information from both infrared and visible images. 2) Dense connection block: Employed for feature extraction to minimize information loss and maximize information utilization. 3) Cross-modal interactive attention mechanism: Developed to capture crucial information from both modalities and enhance the performance of the network. 4) Decomposition network: Designed to decompose the fused image back into source images, incorporating more scene details and richer texture information into the fused image. The proposed fusion framework was experimentally evaluated on the TNO dataset. The results show that the fused images obtained by this method feature significant target regions, rich detailed textures, better fusion performance, and stronger generalization ability. Additionally, the proposed method outperforms other compared algorithms in both subjective performance and objective evaluation.

Key words: infrared and visible image fusion, multi-scale contrast enhancement, cross-dimensional interactive attention mechanism, decomposition network

0 引言

由于技术限制和拍摄环境的影响, 同一设备拍摄的图像往往无法全面描述整个场景。因此, 图像融合

收稿日期: 2023-08-15; 修订日期: 2023-09-25.

作者简介: 邸敬 (1979-), 女, 甘肃兰州人, 硕士, 副教授, 硕士生导师, 主要从事图像检测识别、信号处理技术和宽带无线通信方面的研究。

通信作者: 梁婵 (2001-), 女, 甘肃金昌人, 硕士研究生, 主要从事图像融合, 通信技术方面的研究。E-mail: 2431413505@qq.com

基金项目: 国家自然科学基金 (62061023); 甘肃省科技计划资助项目 (22JR5RA360); 甘肃省杰出青年基金 (21JR7RA345)。

技术应运而生，它能从不同源图像中提取有意义的信息，并将其融合到包含丰富信息的单幅图像中^[1]。其中红外与可见光图像融合是应用最广泛的图像融合技术。红外与可见光图像融合可以减少数据冗余，产生具有红外显著对比度和可见光丰富背景纹理细节的高质量融合图像。红外与可见光图像融合技术在目标检测^[2]、自动驾驶^[3]、视频监控^[4]、医学成像^[5]、军事应用^[6]等领域具有重要价值。

目前，现有的融合算法可以分为两类，传统融合算法^[7]和基于深度学习的融合算法^[8]。传统融合算法包括多尺度变换融合算法^[9]、稀疏表示融合算法^[10]、基于子空间方法^[11-12]等。然而，采用传统方法进行图像融合任务时通常需要复杂的活动水平测量和手工设计的融合规则，这些方法无法有效地对图像进行特征提取并应用于复杂场景。基于深度学习的算法能够利用神经网络强大的拟合能力，可以很好地弥补传统方法的不足，获得纹理细节更丰富、质量更高的融合图像。Liu 等人^[13]首次利用卷积神经网络联合生成活动水平测量和融合规则，克服了传统方法手动设置融合规则的问题，但卷积神经网络 (Convolutional Neural Network, CNN) 简单，特征提取不充分，从而造成融合图像信息丢失。Zhao 等人^[14]提出了一种克服传统 CNN 红外与可见光图像融合模型缺点的网络 AUIF，该方法用两个传统优化模型建立完成两尺度分解，提取特征信息更充分，但忽视了背景纹理细节，融合图像不够清晰。Li 等人^[15]提出了一种端到端的残差融合网络 RFN-Nest，利用自编码器的嵌套连接提取图像特征，保留了可见光图像纹理细节，但融合图像目标边缘模糊。Tang^[16]等人提出了一种可以自适应地保持显著目标强度分布的图像融合网络 PIAFusion，融合图像目标边缘清晰，但存在的问题是图像对比度不足。

Tang 等人^[17]提出了一种无黑暗红外与可见光图像融合方法 DIVFusion，该方法利用图像增强融合特征图像的对比度，融合图像对比度明显，但该方法忽略了图像边缘纹理细节，背景纹理不够清晰。

综上，针对目前红外与可见光融合图像深层特征提取不足，纹理细节不够清晰，融合图像对比度低等问题，本文提出了一种多尺度对比度增强和跨模态交互注意力机制的红外与可见光图像融合方法。在融合网络中设计了多尺度对比度增强模块，利用多尺度增大感受野，对比度块计算特征对比度，为对比度较高的特征分配较高的权重，从而实现对对比度增强。同时，利用通道注意力机制增强重要特征通道，空间注意力机制增强对重要空间位置中信息的注意，构造通道和空间跨维度交互的新注意力机制，增强特征提取信息，生成具有更高对比度、更丰富纹理细节的融合图像。最后设计了分解网络，不仅考虑了从源图像到融合图像的融合过程，而且还考虑了从融合图像到源图像的分解过程，分解结果与源图像一致会尽可能减少信息损失，使融合图像纹理细节更清晰。

1 融合方法

红外与可见光图像融合目标是获得具有完整场景表达能力的高质量融合图像。由于目前红外与可见光融合图像有深层特征提取不足，纹理细节不清晰，融合图像对比度低等问题，因此本文提出了一种多尺度对比度增强和跨维度交互注意力机制的红外与可见光图像融合方法。该方法主要由融合网络和分解网络两部分构成，融合网络是实现图像融合任务的目标网络，旨在将源图像融合成单幅图像；分解网络专用于分解该融合结果以获得与源图像尽可能一致的图像。图 1 描述了该网络的整体结构。

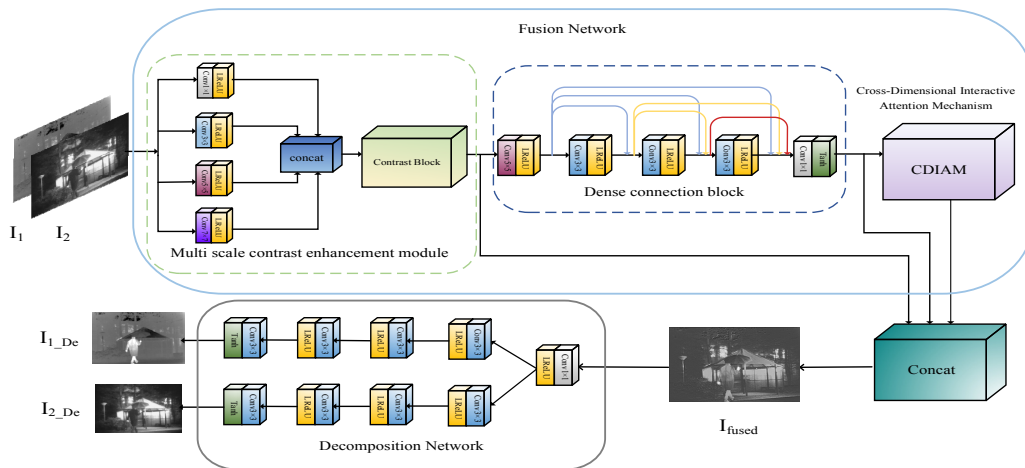


图 1 多尺度对比度增强和跨维度交互注意力机制网络整体框架

Fig.1 Overall network framework of multi-scale contrast enhancement and cross-dimensional interactive attention mechanisms

融合网络的目的是将来自不同传感器具有不同性质的源图像融合生成单幅图像,主要由图像增强、特征提取、显著特征提取3部分组成。在训练阶段,同时输入大小都为 120×120 两种模态的图像到融合网络中,采用多尺度块能够捕获原始图像的平移不变性,增强图像的多尺度卷积结构。在这个基础上,通过对比度块计算特征的对比度,特别是对具有较高对比度的特征赋予更高权重,从而有效提升融合图像的对比度。将多尺度对比度增强块提取的信息通过密集连接块进行深度特征提取,进一步加强特征表示。再将深度特征提取得到的信息通过跨维度交互注意力机制,从通道和空间两方面对红外和可见光图像的显著信息聚焦,并抑制无用信息确保在融合时所有显著特征都可以被利用。最后将提取的特征采用Concat的策略进行融合。而分解网络的目的是分解融合图像得到与源图像尽可能一致的图像,迫使融合图像包含更丰富的场景信息。为此本文设计了双分支网络以分解融合图像,主要通过公共卷积层从融合图像中提取特征,再从两个分支进行分解,以获得与源图像一致的图像。

1.1 融合网络

1.1.1 多尺度对比度增强

为解决单一尺度特征提取不充分问题,并增强图像中不同区域之间的亮度差异,使图像细节更加明显突出,本文设计了多尺度对比度增强模块。多尺度有助于提取深度特征,而对比度增强则专注于增强目标区域的强度信息。本文采用了4个不同卷积核大小分别为 1×1 、 3×3 、 5×5 、 7×7 的卷积层用于捕获不同尺度的深度特征。然后将获得的多尺度特征在通道维度上进行级联,再传送到对比度块中。在对比度块中,多尺度特征信息会被进一步加强,从而最大程度地减少信息丢失。该方法能够整合互补信息并增强对比度,进而获得更为准确和可靠的空间信息。这些增强后的特征对于提升网络模型的精度至关重要。

具体来说,在多尺度特征提取的基础上,引入了用于计算特征的对比度块,为对比度较高的特征分配更大权重,从而实现对对比度的增强。其中,对比度块主要由最大池化层和全连接层组成,而卷积层的激活函数采用了Leaky ReLU (LReLU)。在对比度块中,对残差流特征的每个位置计算标准差(i, j),计算公式如下:

$$\sigma_{i,j} = \frac{1}{2r+1} \sqrt{\sum_{-r \leq p, q \leq r} (\phi_m(i+p, j+q) - \mu_{i,j})^2} \quad (1)$$

式中: r 是窗口的半径。 $\mu_{i,j}$ 和 $\sigma_{i,j}$ 分别是以 (i, j) 为中心

的半径为 r 的窗口内的平均值和标准差。 ϕ_m 是多尺度特征提取信息。因此,在计算标准差之后输出对比度图,接着是最大池化层,分别使用激活函数为LReLU的全连接层和激活函数为sigmoid的全连接层。残余流的最终输出是表示特征的对比度权重的激活向量。随后,将主流中的特征和残余流的激活向量通过通道相乘以生成增强的特征。如图2为对比度增强模块。

多尺度对比度增强模块的设计着重考虑了将多尺度提取深度特征集成在一起,以实现特征级别的对比度增强。将获得的多尺度特征在通道维度上级联并传送至对比度块,从而在对比度块内包含多尺度信息的特征得到进一步增强。

1.1.2 密集连接块

卷积神经网络强大的特征提取能力为图像融合提供了多种新方法。然而,普通卷积神经网络在图像特征提取中只利用最后一层输出作为图像特征,忽略中间层提取的信息,造成信息丢失,尤其在深层网络中更明显。密集连接网络(DenseNet)^[18]克服了这一局限,将所有先前特征层作为当前层的输入,确保信息充分传递和利用,为图像融合奠定了基础。密集连接块有效解决了梯度消失和模型退化问题,使网络更易训练,增强了特征保存,减少了参数数量。因此,密集连接网络显著强化了卷积神经网络在图像融合中的表现,有效地改善了特征提取效果,使网络能够更好地保留图像关键信息。

密集连接块如图3所示,输出是之前每一层的输入的密集连接,每一层的输出如式(2)所示:

$$x_t = H(x_0, \dots, x_t) \quad (2)$$

式中: x_t 表示第 t 层的输出; H_t 是一个组合函数操作表示非线性转化函数,本文方法采用的密集连接块结构为Conv 3×3 +LReLU。

1.1.3 跨维度交互注意力机制

多尺度对比度增强模块主要用来实现图像增强,密集连接块进行深度特征提取,但这两大模块都适用于局部特征提取,为实现全局显著特征提取本文采取了注意力机制。

通道注意力机制主要通过增加权重来关注显著信息,提高网络的泛化能力,但不能体现空间维度信息。空间注意力机制能够增强空间位置中重要信息但未能充分提取通道信息。因此本文提出了一种通道-空间跨维度交互注意力机制来融合深层特征,提取包含丰富的细节和结构信息,更有效地保留红外与可见光图像的关键信息。如图4所示。

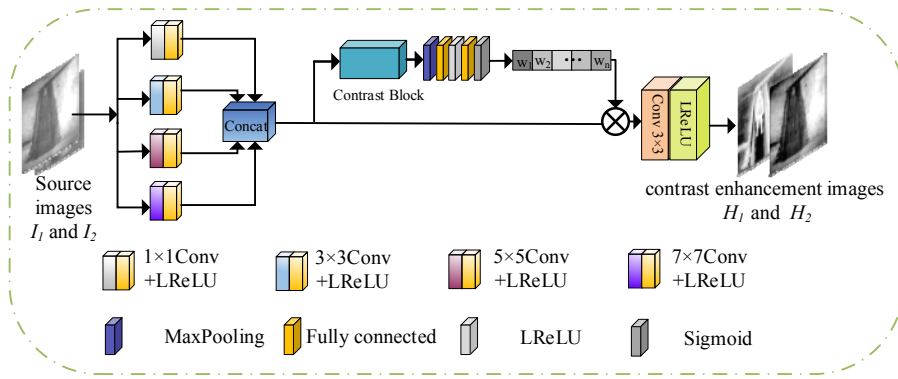


图 2 多尺度对比度增强模块

Fig.2 Multi-scale contrast enhancement module

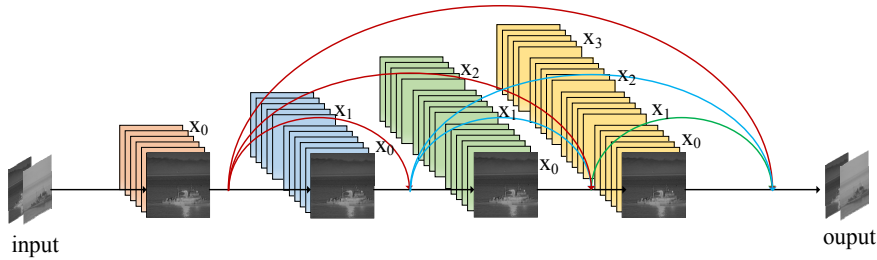


图 3 密集连接块

Fig.3 Dense connection block

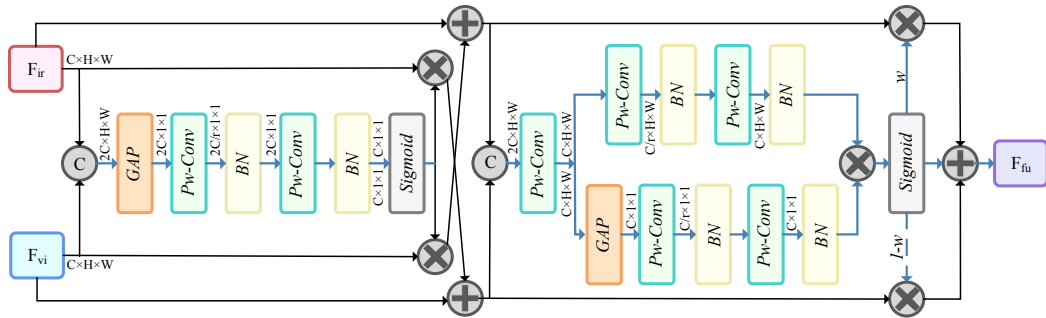


图 4 跨维度交互注意力机制

Fig.4 Cross-dimensional interactive attention mechanisms

从密集连接提取的红外和可见光深度特征，首先在通道维度将它们送入一个通道注意力模块，该模块由卷积和池化操作生成注意力权重。然后通过逐元素相乘将这些权重应用于原始特征，并与原始特征进行加权。随后，将加权后的特征与来自另一分支的原始特征相加，从而增强它们的对应性特征。这一特征增强过程如式(3)~(4)所示：

$$F_{ir}^i = F_{ir} \oplus \left\{ F_{vi} \otimes \left[\text{Pw-Conv}^n \left(\text{GAP} \left(C(F_{ir}, F_{vi}) \right) \right) \right] \right\} \quad (3)$$

$$F_{vi}^i = F_{vi} \oplus \left\{ F_{ir} \otimes \left[\text{Pw-Conv}^n \left(\text{GAP} \left(C(F_{ir}, F_{vi}) \right) \right) \right] \right\} \quad (4)$$

式中： \oplus 是指逐元素求和； \otimes 表示逐元素相乘；Pw-

$\text{Conv}^n(\cdot)$ 代表 n 个级联逐点卷积层； $C(\cdot)$ 表示通道维度中的级联操作； $\delta(\cdot)$ 和 $\text{GAP}(\cdot)$ 分别表示 sigmoid 函数和全局平均池化。之后将增强的特征在通道维度中级联并运送到并行通道注意力和空间注意力模块以生成最终的融合权重。融合权重生成过程如式(5)~(7)所示：

$$A_C^i = \text{Pw-Conv}^n \left\{ \text{GAP} \left[\text{Pw-Conv} \left(C(F_{ir}^i, F_{vi}^i) \right) \right] \right\} \quad (5)$$

$$A_S^i = \text{Pw-Conv}^n \left\{ \text{Pw-Conv} \left[C(F_{ir}^i, F_{vi}^i) \right] \right\} \quad (6)$$

$$W^i = \delta(A_C^i \otimes A_S^i) \quad (7)$$

由于红外和可见光特征是互补的，因此将生成的权重用于其中一种模态，而另一种模态的融合权重可

以表示为 $1 - W^i$ 。表面特征的融合过程如式(8)所示:

$$F_{fu}^i = (W^i \otimes F_{ir}^i) \oplus ((1 - W^i) \otimes F_{vi}^i) \quad (8)$$

1.2 分解网络

分解网络用于对融合图像进行分解,以获得与源图像尽可能一致的图像。首先使用一个公共卷积层从融合图像中提取特征,然后构建双分支网络得到与源

图像相似的图像,其中每个分支包含4个卷积层。第一个公共卷积层使用大小为 1×1 的卷积核, LReLU 的激活函数,中间3个卷积层均使用大小为 3×3 的卷积核, LReLU 作为激活函数,最后一个卷积层使用大小为 3×3 的卷积核, tanh 作为激活函数。分解网络框架如图5所示。

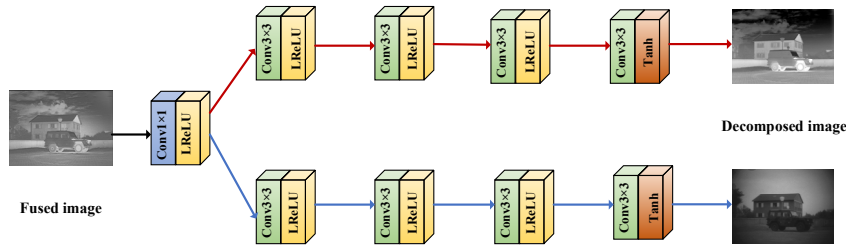


图5 分解网络框架

Fig.5 Decomposition network framework

1.3 损失函数

融合网络通过提取和重构强度和梯度信息来生成单幅融合图像,而分解网络则是对强度和梯度信息进行分解,生成尽可能与源图像一致的图像。强度损失约束融合图像保持与源图像相似的强度分布,梯度损失迫使融合图像包含丰富的纹理细节,分解一致性损失可以使融合结果包含更丰富的场景内容,从而达到更好的融合效果。

本文的损失函数由强度损失、梯度损失和分解一致性损失3种类型的损失项组成,如公式(9)所示:

$$L = \beta L_{grad} + L_{int} + L_{dc} \quad (9)$$

式中: L_{grad} 表示梯度损失; L_{int} 表示强度损失; L_{dc} 表示分解一致性损失; β 用于平衡强度和梯度项。

1.3.1 梯度损失

本文采用了一种自适应梯度损失 L_{grad} ,旨在使融合图像包含丰富的纹理特征。在梯度损失中引入了自适应权重块。自适应权重块首先利用高斯低滤波进行线性平滑滤波,最大程度减少噪声对权重块的影响,之后通过基于梯度丰富度对源图像相应像素的优化目标进行评估。这个过程的目标是引导融合图像的纹理与源图像在相应位置的最强纹理尽可能地保持一致。其原理图如图6所示。

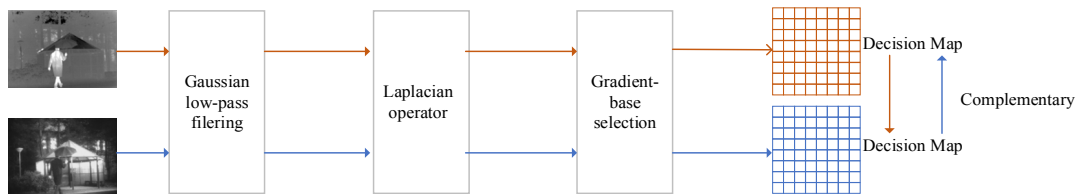


图6 自适应权重原理图

Fig.6 Schematic diagram of adaptive weighting

梯度损失公式如下:

$$L_{grad} = \frac{1}{HW} \sum_i \sum_j S_{1_{i,j}} \cdot \left(\nabla I_{fused_{i,j}} - \nabla I_{1_{i,j}} \right)^2 + S_{2_{i,j}} \cdot \left(\nabla I_{fused_{i,j}} - \nabla I_{2_{i,j}} \right)^2 \quad (10)$$

$$S_{1_{i,j}} = \text{sign} \left(\left| \nabla \left(L(I_{1_{i,j}}) \right) \right| \right) - \min \left(\left| \nabla \left(L(I_{1_{i,j}}) \right) \right|, \left| \nabla \left(L(I_{2_{i,j}}) \right) \right| \right) \quad (11)$$

$$S_{2_{i,j}} = 1 - S_{1_{i,j}} \quad (12)$$

式中: i 和 j 表示像素在第 i 行和第 j 列权重图; H 和 W 代表图像的高度和宽度; I_1 和 I_2 为源图像; I_{fused} 为融合图像; $|\cdot|$ 表示绝对值函数; $\nabla(\cdot)$ 为拉普拉斯算子; $L(\cdot)$ 表示高斯低通滤波器函数; $\min(\cdot)$ 表示最小函数; $\text{sign}(\cdot)$ 为符号函数。

1.3.2 强度损失

强度损失 L_{int} 的目的是约束融合图像保持与源图像相似的强度分布,有助于融合图像保留像素强度所代表的对比度等有用信息,同时使融合图像的整体场景风格更加自然真实。

强度损失可以形式化如下:

$$L_{\text{int}} = \frac{1}{HW} \sum_i \sum_j \left(I_{\text{fused}_{i,j}} - I_{1_{i,j}} \right)^2 + \alpha \left(I_{\text{fused}_{i,j}} - I_{2_{i,j}} \right)^2 \quad (13)$$

1.3.3 分解一致性损失

分解一致性损失 L_{dc} 要求来自融合图像的分解结果尽可能类似于源图像，其被定义为：

$$L_{\text{dc}} = \frac{1}{HW} \sum_i \sum_j \left(I_{1_{\text{de}_{i,j}}} - I_{1_{i,j}} \right)^2 + \left(I_{2_{\text{de}_{i,j}}} - I_{2_{i,j}} \right)^2 \quad (14)$$

式中： $I_{1_{\text{de}}}$ 和 $I_{2_{\text{de}}}$ 是融合图像的分解结果。

2 实验分析

2.1 实验设置

本文选用来自公开数据库 TNO 的数据集进行红外和可见光图像融合的实验。在训练阶段，采用裁剪分解的方法来扩充数据集。具体而言，对源图像进行裁剪，生成了 21242 个大小为 120×120 的图像块用于训练。在测试阶段，选择了 10 对试数据。值得注意的是，与训练数据不同，测试是在源图像的原始大小上执行。

本实验的软件环境为 Windows11, Python3.7, 硬

件配置环境为 13th Gen Intel(R) Core(TM) i5-13500HX, 2.50 GHz, RTX 4060 GPU。主要采用 tensorflow、imageio 库在 Pycharm 编译器中训练和测试网络性能。实验设置 epoch=15、batch_size=32、learning_rate=1e-4。在优化算法方面，选择了适应性强、收敛性好的自适应优化算法 Adam。在公式(13)中 α 为 0.5，使网络从红外图像中获取主要的强度信息，以保持显著的对比度。此外，在总损失中包含的梯度损失、强度损失、分解一致性损失的比例经过重复实验后设置为 80、1、1。

2.2 消融实验

为进一步验证本文所提各个模块的有效性，设计了 4 组融实验。实验一将融合网络的密集连接块替换为普通卷积，实验二去掉融合网络的多尺度对比度增强模块，实验三去掉融合网络的跨模态交互注意力机制，实验四去掉分解网络。

本文随机选择“Street”和“Kaptein”两组场景的源图像融合结果做主观对比展示，选取了 10 组场景的融合结果均值作为客观评价指标对比。图 7 展示了不同模块生成的两组融合结果的消融实验结果。红色实线框表示红外显著信息，黄色虚线框标记纹理信息。

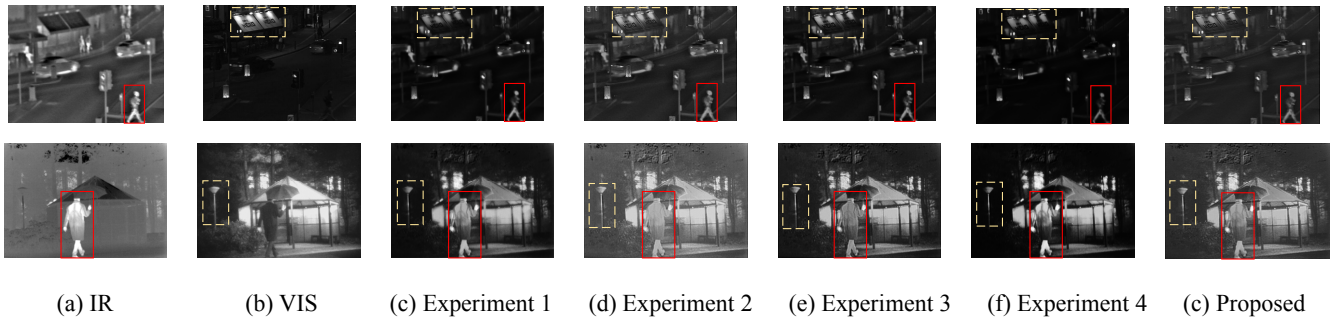


图 7 消融实验结果

Fig.7 Results of ablation experiments

从实验结果可以明显看出：实验一将融合网络的密集连接块替换为普通卷积块，导致融合图像细节不够充分且亮度信息不明显。实验二去掉融合网络的多尺度对比度增强模块，图像对比度不足，同时背景纹理细节丢失严重。实验三去掉融合网络的跨维度交互注意力机制，导致融合图像未能充分展现更多的图像特征信息，也无法有效关注显著特征。实验四去掉分解网络，明显融合图像较暗，红外显著信息缺失严重。与上述实验相比，本文所提出的多尺度对比度增强和跨维度交互注意力机制在图像融合中表现出色，有效地保留了红外与可见光图像特征信息，纹理细节更加丰富，同时增强了图像的对比度。消融实验的客观指标选取了表示融合方法效果的平均梯度（Average Gradient, AG）、空间频率（Spatial Frequency, SF）、

边缘保持度（ $Q^{AB/F}$ ）。表 1 为消融实验的 10 组场景客观评价指标均值数据。由表可知本文所提方法指标均有所提升，这说明本文方法得到的融合图像红外显著目标明显、可见光细节纹理清晰，更符合人眼视觉。

2.3 主观评价

使用 TNO 数据集对本文方法进行测试验证，随机选取 6 组不同场景图像作为对比实验，选择了红外-可见光图像融合与零相位分量分析（ResNet^[19]）、基于巢连接和空间/通道注意力模型的红外-可见光图像融合架构（Nest connection image Fusion, Nest Fusion^[20]）、红外和可见光图像的贝叶斯融合（Infrared and Visible Images Bayesian Fusion, IVIFBF^[21]）、基于模型的高效红外-可见光图像融合算法（Algorithm Unrolling Image Fusion, AUIF^[14]）、用于红外和可见光图像的端到端残

差融合网络 (Residual Fusion Network, RFN-Nest^[15])、基于特征流和边缘补偿的快速轻量级红外-可见光图像融合网络 (Fast and Lightweight Fusion Network, FL fusion^[22])、通过耦合特征学习实现多模态图像融合 (Comparative Coupled Feature Learning, CCFL^[23])、基于光照感知的渐进式图像融合 (Progressive Infrared and Visible image fusion network, PIAF^[16]) 8 种先进且具有代表性的方法与本文融合方法进行比较。主观视觉描述是从清晰度、图像细节、对比度等方面对融合图像质量进行主观评价。实验结果如图 8 所示。

表 1 10 组场景消融实验客观评价指标均值

Table 1 Mean values of objective evaluation indexes in 10 groups of ablation experiment scenes

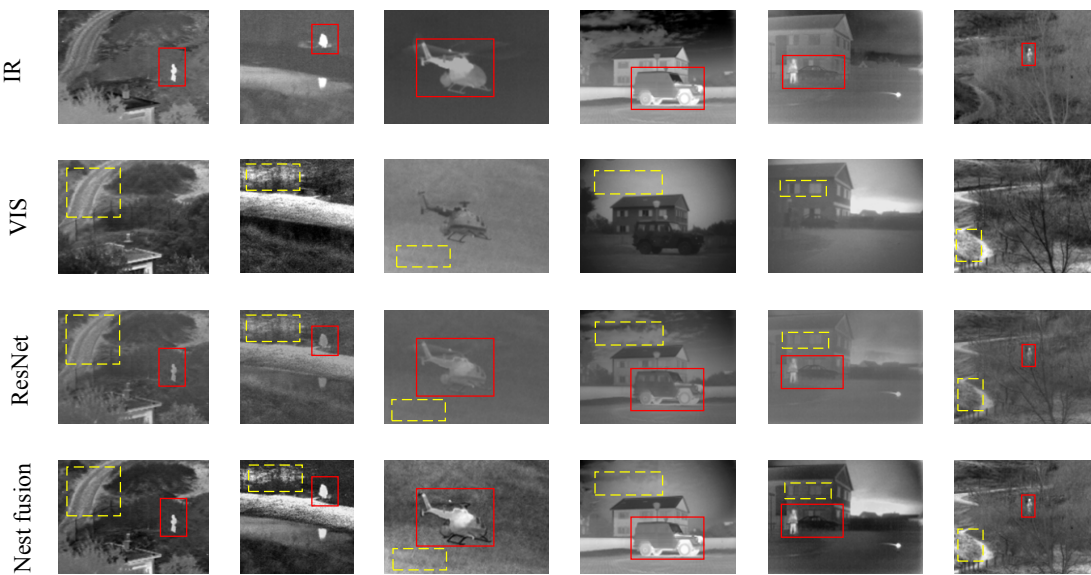
Experiments	Index		
	AG	SF	$Q^{AB/F}$
Experiment 1	3.8619	16.9474	0.3435
Experiment 2	5.2188	19.4560	0.4054
Experiment 3	5.0563	18.0114	0.4096
Experiment 4	3.9564	17.5561	0.3078
Proposed	5.4976	19.7360	0.4288

从实验结果可以看出: ResNet 方法保留了一定的可见光信息,但是纹理细节不够充分,如云朵、草丛等一些背景纹理信息丢失,同时红外显著信息不够明显。Nest Fusion 方法很好地保留了红外图像显著信息,但是在背景纹理方面表现欠佳,生成的融合图像对比度低。IVIFBF 方法突出红外信息,但纹理细节不够清晰,背景信息严重丢失。AUIF 方法保留了一定的纹理细节信息,同时显著目标明显,但融合图像整体较暗。RFN-Nest 方法生成的融合图像纹理细节清晰度高、亮度明显,但目标边缘较模糊。FL fusion 方法所

得融合图像具有噪声并且带有伪影。CCFL 方法在突出红外辐射信息方面表现较好,但仍存在少量伪影。PIAF 方法更倾向于可见光图像,在背景纹理细节方面信息丢失严重。相比上述方法,本文所提出的方法能够更好地平衡可见光图像的背景信息和红外图像的显著目标,在保留更多的细节信息的同时保持了红外图像的对比度。

2.4 客观评价

客观评价中,选取 7 个客观评价指标与其他典型的融合方法进行对比,其中包含了平均梯度 (Average Gradient, AG)、信息熵 (Entropy, EN)、相关系数 (Correlation Coefficient, CC)、多尺度结构相似度 (Multi-scale structural similarity, MS_SSIM)、标准差 (Standard Deviation, SD)、视觉保真度 (Visual Information Fidelity, VIFF)、空间频率 (Spatial Frequency, SF)。AG 是反映融合图像清晰度的指标,其值越大图像越清晰;EN 是度量图像包含信息量的多少,数值越大代表融合结果包含信息越丰富;CC 用于衡量融合图像与源图像的空间线性相关程度,其值越接近 1 或者 -1 表示融合图像包含源图像信息越多,融合效果越好;MS_SSIM 则衡量图像块之间的结构一致性,值越大表示融合图像与源图像结构越一致;SD 是衡量亮度信息的指标,值越大表示图像离散程度大,融合效果好;VIFF 指标从人类视觉系统出发,衡量融合图像提取信息的多少,数值越大表示与人眼主观特性一致,图像融合效果质量越高;SF 反映图像灰度变化快慢,评价融合图像细节纹理信息,值越大说明纹理细节越清晰,融合效果越好。本文从 TNO 数据集中随机选取 6 组不同场景的图像作为对比实验,6 组图像的客观评价指标均值如表 2 所示。



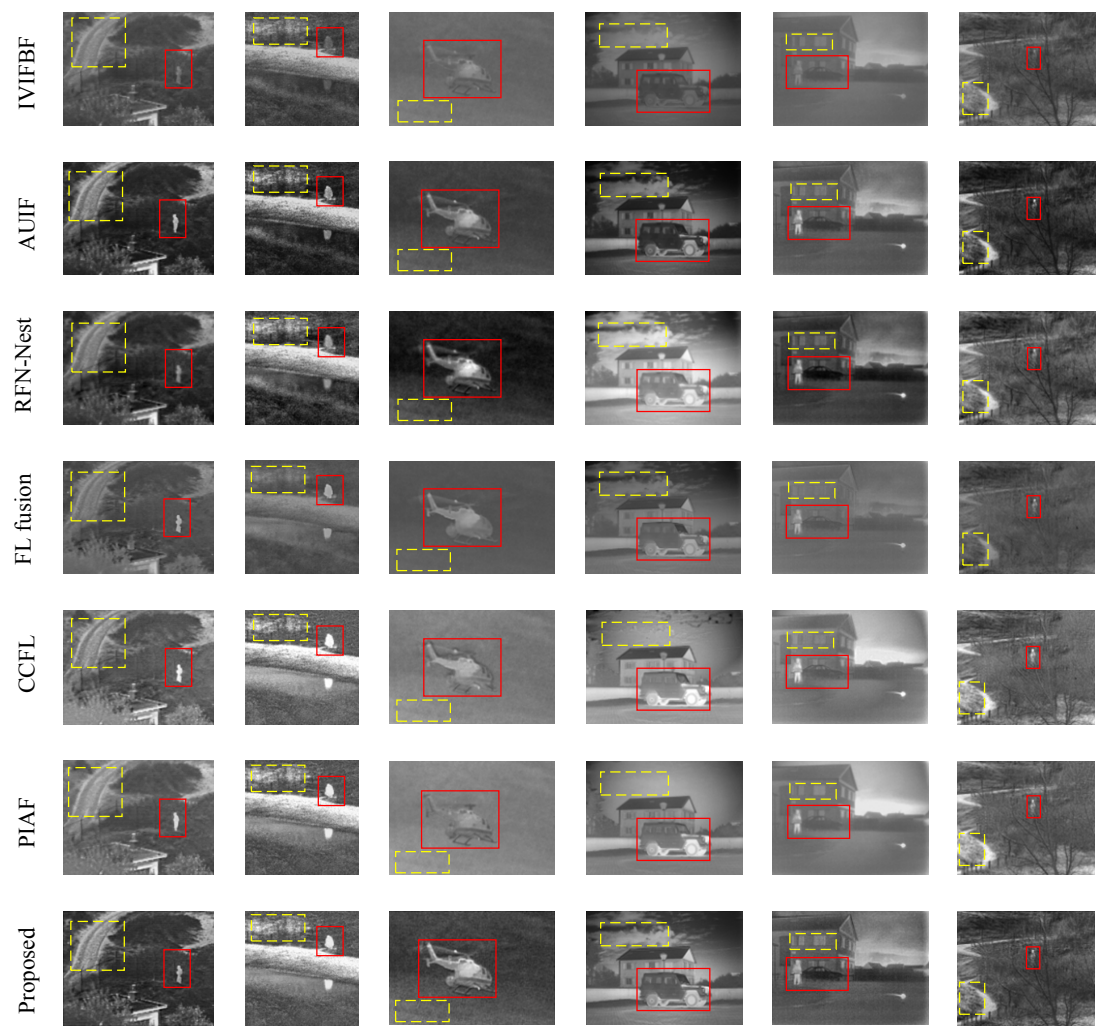


图 8 六组场景的融合结果

Fig.8 Fusion results for six groups of scenes

表 2 6 组图像的客观评价指标均值

Table 2 Mean value of objective evaluation index for 6 groups of images

Methods	Evaluation indicators						
	AG	EN	CC	MS-SSIM	SD	VIFF	SF
ResNet	3.5661	5.9922	0.5708	0.8681	4.3923	0.2375	9.2758
Nest Fusion	5.8888	6.9124	0.5982	0.8529	6.3274	0.4076	16.8701
IVIFBF	4.3408	6.1802	0.6755	0.8676	4.7672	0.1780	11.2212
AUIF	6.2377	6.6338	0.6576	0.9201	5.7228	0.5296	17.1550
RFN-Nest	4.2341	6.9792	0.6744	0.9177	6.1445	0.5784	11.4769
FL fusion	3.3678	5.9282	0.4349	0.7823	4.2301	0.2926	9.1551
CCFL	6.1237	6.7197	0.4392	0.8857	5.6265	0.3542	16.9962
PIAF	5.6041	6.6416	0.6891	0.8681	5.7763	0.6056	15.4443
Proposed	7.3095	6.8959	0.7972	0.9207	6.0486	0.6501	20.2498

根据 6 组红外与可见光图像的融合结果表明，本文提出的多尺度对比度增强和跨纬度交互注意力机制的红外与可见光图像融合方法更符合人眼视觉特

性的融合图像。在这些结果中，红外显著信息被充分地提取，而可见光背景纹理细节也更丰富，整体对比度更明显。从表 2 可以看出，5 项评价指标 AG、

CC、MS-SSIM、VIF、SF 均取得最优，而 EN 和 SD 较低的主要原因是可见光图像通常包含大量噪声，本文方法通过噪声抑制在一定程度上降低了融合结果的熵，而且 RFN-Nest 和 ResNet 均利用残差网络提取信息，RFN-Nest 使用了两个阶段的训练策略，保留了大量的可见光信息，但其 AG、SF 值较低，表明未能充分包含较多的红外显著目标。因此，与其他对比方法相比，本文方法更能突出红外显著信息，并且能充分保留丰富的边缘与背景纹理细节。客观评价实验的数据结果与主观评价结果一致。

将选取的 6 组不同场景的红外与可见光图像进行可视化分析，图 9 为不同场景不同方法的融合结果客观评价指标折线图。本文提出的方法客观评价指标

AG、EN、CC、MS-SSIM、SD、VIF 和 SF 相较于其他 8 种对比算法分别平均提高了 32.69%、5.76%、25.68%、5.47%、11.16%、38.79%、33.58%。从图 9 可以看出，本文所提方法与主观视觉保持一致，有效证明本文方法在特征提取、纹理细节保存、显著目标突出等方面都优于其他对比方法。

2.5 计算复杂度及运行效率分析

为了更全面地评价本文方法，随机选取 20 组红外与可见光图像进行测试，计算其复杂度和平均运行时间以评估方法复杂度和运行效率。同时，将提出的方法与其他 8 种经典的红外与可见光图像融合方法进行了比较，具体比较结果如表 3 所示。

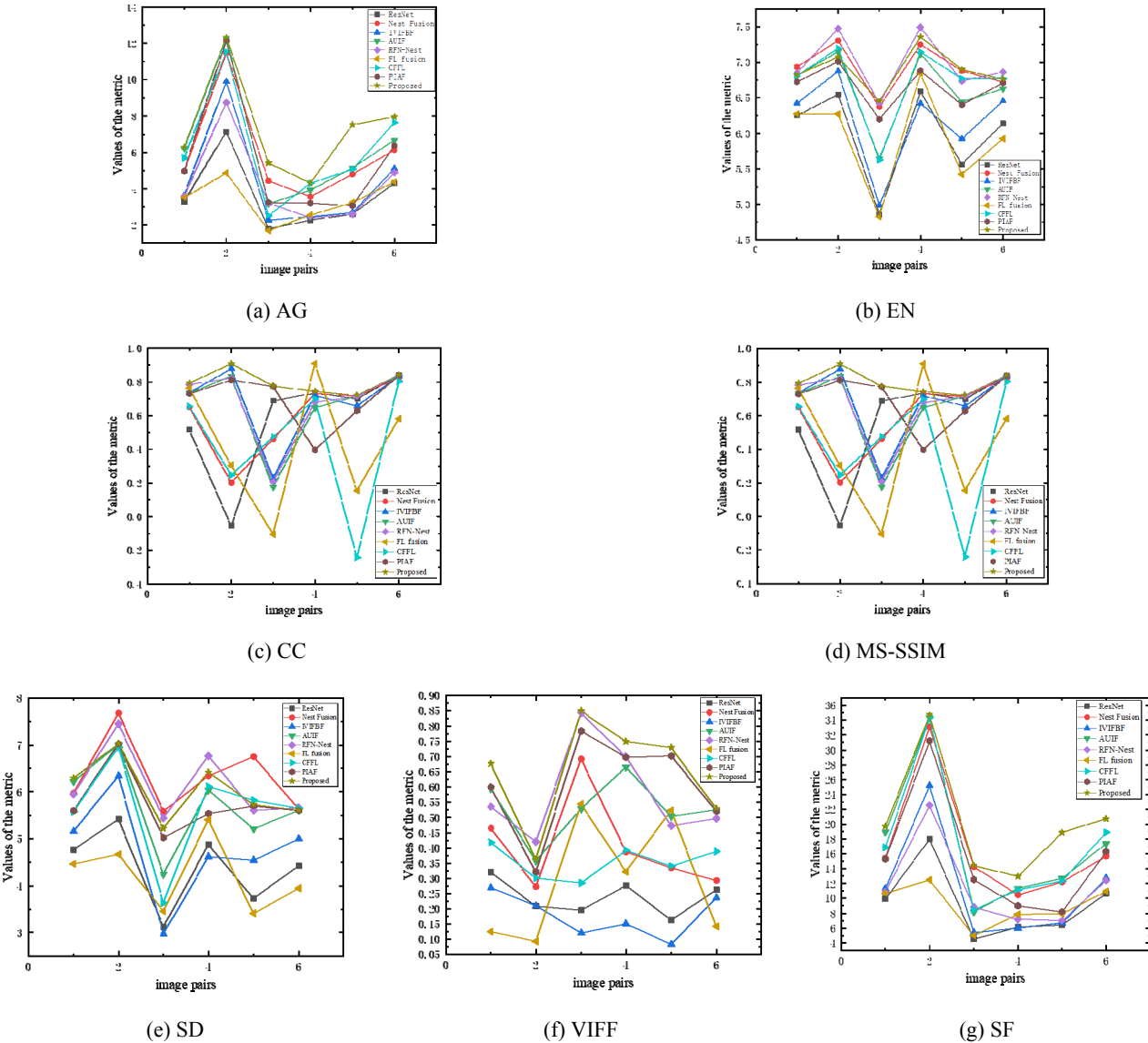


图 9 不同方法的 7 种客观指标对比

Fig.9 Comparison of 7 objective metrics for different algorithms

表 3 20 组图像不同方法的平均运行时间

Table 3 Average running time of different algorithms for 20 sets of images									
Methods	ResNet	NestFuse	IVIFBF	AUIF	RFN-Nest	FL fusion	CCFL	PIAF	Proposed
Time/s	0.1845	0.2512	0.4121	1.1456	0.1789	1.2514	0.6584	0.2816	0.1245
Calculated amount /GFLOPs	2.8741	3.2456	3.8970	4.1257	3.4578	4.2125	3.8974	3.2141	3.1242

表 3 中，1 GFLOPs=10⁹ FLOPs。从表 3 中可以发现，ResNet 由于网络层数较少，仅使用简单的卷积操作来进行特征提取，因此其计算量和运行时间在 8 种方法中最小。然而，这种方法无法有效提取特征信息，导致融合结果较差。相对而言，本文方法在平均运行时间方面表现最佳，在测试阶段所使用的参数数量仅 0.05M，这可谓轻量级。综合考虑总体指标和方法性能，本文方法在各个方面均表现出色，能够高效处理实时融合任务。

3 结论

为了提高红外与可见光融合图像质量，本文提出一种基于多尺度对比度增强和跨维度交互注意力机制的红外与可见光图像融合方法。在融合阶段主要采取对比度增强、特征提取、显著特征提取对红外与可见光图像进行特征提取，并利用 Concat 的融合策略对其进行融合。在分解阶段，构建了双分支网络将融合图像分解为近似于源图像的分解图像。本文提出的方法客观评价指标 AG、EN、CC、MS-SSIM、SD、VIFF 和 SF 相较于其他 8 种对比算法分别平均提高了 32.69%、5.76%、25.68%、5.47%、11.16%、38.79%、33.58%。结果表明，相比于其他融合方法，本文方法整体清晰度更好，提升了视觉表达效果，使得目标对象更加突出，背景纹理信息更加清晰。下一步工作希望能设计一个通用的融合框架，适用于医学图像、多聚焦图像融合，这将有助于进一步提升本文方法的实际应用价值和适用范围。

参考文献：

[1] 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): 3-36.
TANG Linfeng, ZHANG Hao, XU Han, et al. A review of image fusion methods based on deep learning[J]. *Journal of Image and Graphics*, 2023, 28(1): 3-36.

[2] SUN Haijiang, LIU Qiaoyuan, WANG Jiacheng, et al. Fusion of infrared and visible images for remote detection of low-altitude slow-speed small targets[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 2971-2983.

[3] SHUAI Xincheng, JING Zhongliang, TUO Hongya. SAFuseNet: integration of fusion and detection for infrared and visible images[J]. *Aerospace Systems*, 2022, 5(4): 655-661.

[4] LI H, WU X J. CrossFuse: A novel cross attention mechanism based infrared and visible image fusion approach[J]. *Information Fusion*, 2024, 103: 102147.

[5] ZHOU Tao, LI Qi, LU Huiling, et al. GAN review: models and medical image fusion applications[J]. *Information Fusion*, 2023, 91: 134-148.

[6] HU Ke, SUN Wenhao, NIE Zhongbo, et al. Real-time infrared small target detection network and accelerator design[J]. *Integration*, 2022, 87: 241-252.

[7] 孙彬, 诸葛吴为, 高云翔, 等. 基于潜在低秩表示的红外和可见光图像融合[J]. 红外技术, 2022, 44(8): 853-862.
SUN Bin, ZHUGE Wuwei, GAO Yunxiang, et al. Infrared and visible image fusion based on potential low-rank representation[J]. *Infrared Technology*, 2022, 44(8): 853-862.

[8] TAN W, ZHOU H, SONG J, et al. Infrared and visible image perceptive fusion through multi-level Gaussian curvature filtering image decomposition[J]. *Applied Optics*, 2019, 58(12): 3064-3073.

[9] Archana R, Jeevaraj P S E. Deep learning models for digital image processing: a review[J]. *Artificial Intelligence Review*, 2024, 57(1): 11.

[10] WANG Chang, WU Yang, YU Yi, et al. Joint patch clustering-based adaptive dictionary and sparse representation for multi-modality image fusion[J]. *Machine Vision and Applications*, 2022, 33(5): 69.

[11] CHEN Yudan, WANG Yuanbo, HUANG Fuyu, et al. Infrared and Visible Images Fusion base on Wavelet Transform[J]. *Sixth Symposium on Novel Optoelectronic Detection Technology and Applications*, 2020, 11455: 875-882.

[12] LI Hui, WU Xiaojun, Kittler Josef. MDLatLRR: A novel decomposition method for infrared and visible image fusion[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4733-4746.

[13] LIU Yu, CHEN Xun, CHENG Juan, et al. Infrared and visible image fusion with convolutional neural networks[J]. *International Journal of Wavelets, Multiresolution and Information Processing*, 2018, 16(3): 1850018.

[14] ZHAO Z, XU S, ZHANG J, et al. Efficient and model-based infrared and visible image fusion via algorithm unrolling[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 32(3): 1186-1196.

- [15] LI Hui, WU Xiaojun, Kittler Josef. RFN-Nest: An end-to-end residual fusion network for infrared and visible images[J]. *Information Fusion*, 2021, **73**: 72-86.
- [16] 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**: 79-92.
- [17] TANG Linfeng, XIANG Xinyu, ZHANG Hao, et al. DIVFusion: Darkness-free infrared and visible image fusion[J]. *Information Fusion*, 2023, **91**: 477-493.
- [18] Kayadibi İsmail, Güraksın Gür Emre. An explainable fully dense fusion neural network with deep support vector machine for retinal disease determination[J]. *International Journal of Computational Intelligence Systems*, 2023, **16**(1): 28.
- [19] LI H, WU X J, DURRANI T S. Infrared and visible image fusion with ResNet and zero-phase component analysis[J]. *Infrared Physics & Technology*, 2019, **102**: 103039.
- [20] LI Hui, WU Xiaojun, Tariq Durrani. NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(12): 9645-9656.
- [21] ZHAO Zixiang, XU Shuang, ZHANG Chunxia, et al. Bayesian fusion for infrared and visible images[J]. *Signal Processing*, 2020, **177**: 107734.
- [22] XUE Weimin, WANG Anhong, ZHAO Lijun. FLFuse-Net: A fast and lightweight infrared and visible image fusion network via feature flow and edge compensation for salient information[J]. *Infrared Physics and Technology*, 2022, **127**: 104383.
- [23] Veshki Farshad G, Ouzir Nora, Vorobyov Sergiy A, et al. Multimodal image fusion via coupled feature learning[J]. *Signal Processing*, 2022, **200**: 108637.