

基于环境光感知和红外特征分层引导的图像融合网络

王爱侠^{1†}, 胡傲杰¹, 闫爱云¹, 高尚¹, 金硕巍¹, 庞永恒²

(1. 东北大学信息科学与工程学院, 沈阳 110819; 2. 中国刑事警察学院信息科学与工程学院, 沈阳 110854)

摘要: 在红外图像中, 目标物体的突出显示与可见光图像中丰富的纹理细节相结合, 可以有效地增强融合图像的信息熵, 从而为夜间智能驾驶等下游视觉任务提供重要支持. 然而, 现有的主流融合算法对于可见光图像在恶劣光照夜间道路环境下的信息熵低与像素强度高之间的矛盾, 尚缺乏针对性的研究. 因此, 在正常环境下表现良好的融合算法, 在强光干扰下只能生成与可见光图像相似、信息熵较低的融合图像. 对此, 提出一种能够抵抗恶劣光照环境干扰的图像融合网络, 结合信息熵和信息论原理, 增强图像融合的鲁棒性和信息保留能力. 首先, 设计一个在正常光照条件下具备高鲁棒性和优异性能的图像融合网络, 在该融合网络的基础上设计一个环境光感知模块, 以便在极端光照条件下对低信息熵的可见光图像的特征权重进行分析. 然后, 设计一个红外边缘特征分层引导融合模块, 以充分提取红外图像中的有效特征信息. 实验结果表明, 该融合网络能够在夜间恶劣光照条件下充分利用可见光和红外图像的特征信息, 显著提高这种情况下融合图像的质量. 与其他主流算法相比, 所提出方法生成的融合结果包含了更丰富和更有效的信息.

关键词: 图像融合; 夜间恶劣光照感知; 红外特征挖掘; 高级视觉任务

中图分类号: TP391 文献标志码: A

DOI: 10.13195/j.kzyjc.2024.1390

引用格式: 王爱侠, 胡傲杰, 闫爱云, 等. 基于环境光感知和红外特征分层引导的图像融合网络 [J]. 控制与决策, 2025, 40(10): 3177-3189.

An image fusion network based on ambient light awareness and infrared feature layer guidance

WANG Ai-xia^{1†}, HU Ao-jie¹, YAN Ai-yun¹, GAO Shang¹, JIN Shuo-wei¹, PANG Yong-heng²

(1. College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; 2. College of Information Science and Engineering, China Criminal Police University, Shenyang 110854, China)

Abstract: In infrared images, the prominent display of target objects combined with the rich texture details in visible light images can effectively enhance the information entropy of the fused image, thereby providing important support for downstream visual tasks such as nighttime intelligent driving. However, existing mainstream fusion algorithms lack targeted research on the contradiction between the low information entropy and high pixel intensity of visible light images under adverse nighttime road lighting conditions. Therefore, fusion algorithms that perform well under normal conditions often can only generate fused images that are similar to visible light images and have low information entropy under strong light interference. To address this issue, this paper proposes an image fusion network capable of resisting interference from adverse lighting environments, combining information entropy and principles of information theory to enhance the robustness and information retention capability of image fusion. Specifically, we first design an image fusion network with high robustness and excellent performance under normal lighting conditions. Based on this fusion network, we introduce an ambient light perception module to analyze the feature weights of low-information-entropy visible light images under extreme lighting conditions. Additionally, we design an infrared edge feature hierarchical guided fusion module to fully extract effective feature information from infrared images. Experimental results show that this fusion network can fully utilize the feature information of visible and infrared images under adverse nighttime lighting conditions, significantly improving the quality of fused images in such scenarios. Compared with other mainstream algorithms, the fusion results generated by this method contain richer and more effective information.

收稿日期: 2024-11-29; 录用日期: 2025-04-09.

基金项目: 国家自然科学基金项目 (62406342); 辽宁自然科学基金项目 (2024-BS-260); 上海市法医学重点实验室暨司法部司法鉴定重点实验室开放课题项目 (KF202415).

†通信作者. E-mail: wangaixia@ise.neu.edu.cn.

Keywords: image fusion; adverse nighttime illumination awareness; infrared feature extraction; advanced visual tasks

0 引言

红外和可见光传感器获得的不同模态图像从不同的角度描述和解释成像场景. 可见光传感器成像结果纹理细节丰富, 视觉效果好, 但对光强敏感; 而红外传感器可以忽略光照变化, 产生目标突出的图像, 但细节模糊, 易受噪声影响.

在熵度量和信息论的指导下, 融合两种模态的关键信息, 实现有用信息最大化和噪声最小化, 极大地提高了融合图像中场景描述的信息熵和完整性. 因此, 红外与可见光图像的融合能够增强场景视觉效果, 为后续高水平任务提供实质性的辅助, 在智能车辆辅助驾驶^[1]、道路安全监控^[2]、现代卫星遥感^[3]等领域得到了广泛的应用.

在可见光-红外图像融合研究中, 不断涌现出许多新的方法. 随着方法的不断改进, 融合效果得到了提升, 并在高级视觉任务中展现出优异表现. 红外与可见光图像的融合方法可以分为传统方法和深度学习方法. 传统方法具体包括基于稀疏字典训练的方法^[4]、基于多尺度变换的方法^[5]、基于低秩表示学习的方法^[6]、基于分割或显著性的方法^[7], 以及混合建模方法^[8]. 近年来, 具有良好训练效果的广义神经网络结构因其卓越的特征提取能力被广泛应用于大规模数据集, 并可轻松转化为当前的融合任务, 相较于传统融合算法实现了更优的融合性能. 基于深度学习的方法主要根据网络结构的不同进行分类, 例如自编码器 (AE) 网络^[9]、生成对抗网络 (GANs)^[10]、卷积神经网络 (CNNs)^[11] 以及 Transformer 网络^[12].

尽管当前深度神经网络的融合结果优于传统方法, 并克服了一些传统方法的缺点, 如在夜间道路等场景中缺乏对可见光图像信息分布在恶劣光照干扰下的针对性研究, 但基于深度学习的方法在同时处理恶劣光照背景和正常环境时仍难以取得令人满意的结果. 首先, 大多数现有的融合方法都基于一个先验假设, 即红外图像目标像素的强度突出, 可见光图像的梯度富含纹理细节, 可有选择地从可见光和红外图像中提取特征. 然而, 这一假设忽视了目标边界信息在目标检测和语义分割等高级视觉任务中的重要性, 导致融合结果中目标边界模糊, 影响了融合效果. 其次, 虽然主流的先进融合方法在良好环境中能够获得较好的结果, 但在恶劣光照干扰条件下, 可见光图像中低信息熵和高像素值的矛盾使得大量目标信息被无效的恶劣光信息遮蔽, 阻碍了满意融合结

果的获得. 最后, 大多数基于深度学习的方法使用卷积层从图像中提取多样的特征, 然后在粗略连接后将其传递到下一层. 然而, 这种方法忽略了特征层次之间的差异, 降低了特征信息的利用率. 因此, 在融合过程中, 目标边界等高级视觉任务所需的重要特征部分丢失, 影响了后续高级视觉任务中的融合效果. 这些挑战导致了夜间恶劣光照条件下图像融合的独特难题.

基于此, 本文提出一种抗强光干扰的目标边缘信息引导特征融合网络, 以应对强光干扰环境下的图像融合挑战. 首先, 结合图像信息熵和信息分布, 网络引导融合任务, 确保在正常和恶劣光照条件下均能生成高质量的融合图像, 从而提升后续高级视觉任务的表现. 接着, 设计 3 大关键模块以增强融合效果: 1) 提出红外显著目标边缘特征提取模块, 结合高级语义交互融合 (HSIF) 和边缘特征增强块 (EFEB), 有效提取边缘的显著语义信息, 避免了融合结果中的边缘细节模糊; 2) 设计强光环境感知 (HLEA) 模块, 根据强光可见光图像信息的分布, 为分层特征融合模块赋予合适的权重, 解决恶劣光照条件下因低信息熵和高像素值区域导致的融合效果下降问题; 3) 提出边缘引导的分层特征融合 (EGHF) 模块, 结合目标边缘特征引导和强光感知权重分配, 在强光干扰下有效提取关键语义特征, 避免目标边界模糊和强光遮挡影响. 该网络在处理强光干扰时, 能够保留关键细节信息, 提升融合图像的质量.

1 相关工作

近年来, 红外与可见光图像融合技术取得了显著进展. 传统方法以多尺度变换为核心, 如金字塔变换与小波变换, 虽能保留一定图像细节, 但在边缘表达和复杂特征融合方面仍存在局限. 后续研究引入梯度传递、稀疏表示和显著性引导等策略以提升细节质量, 但在动态环境或恶劣光照下效果不佳.

随着深度学习的发展, 基于神经网络的融合方法成为研究热点. 自编码器类方法 (如 DeepFuse、DenseFuse 和 NestFuse) 借助密集连接和注意力机制改善了语义与细节表达; 卷积网络 (如 DRF) 通过解耦提取与融合过程进一步提升了效果. 生成对抗网络 (如 FusionGAN) 引入自动学习机制, 增强了融合稳定性, 但其训练收敛仍面临挑战. 近期, Transformer 结构被引入图像融合, 如 SwinFuse 通过全局建模显著提升融合质量, 但在强光等特殊场景下仍表现不稳.

恶劣光照条件对融合算法影响显著. EV-Fusion 结合图像增强改善低光表现, 而 PIAFusion 通过昼夜分类增强了光照适应性, 但融合策略复杂, 适应性有限. 针对强光干扰问题, 现有研究如 AT-GAN 虽尝试提升鲁棒性, 却缺乏有效的模块设计与验证.

为此, 本文提出一种强光感知融合模块, 通过光照分类与分解网络动态调整融合策略, 有效抑制强光干扰并增强语义表达, 为复杂光照环境下图像融合任务提供新思路.

2 提出的方法

首先, 介绍网络模型的整体结构; 然后, 详细解释网络模型的 3 个核心组成部分: 边缘特征提取、恶劣光环境感知和分层特征融合; 最后, 阐述网络训练所使用的损失函数.

2.1 网络结构概述

为了在强光干扰下实现图像融合, 避免热目标信息被强光遮蔽, 同时提升后续高级视觉任务中的融合效果, 本文设计一个新颖的可见光和红外图像融合网络. 融合网络的具体框架如图 1 所示. 网络主干包括 4 个部分: 多模态特征提取、边缘特征提取、强光环境感知以及图像解码与重构.

1) 边缘特征提取. 该模块旨在从红外图像中提取丰富的边缘信息以引导图像融合. 首先, 通过 3 种扩张因子的卷积路径提取多尺度特征, 并结合 Sobel 算子生成不同尺度的边缘图. 随后, 浅层 (Sobel) 与深层多尺度边缘特征通过通道拼接融合, 形成层次化的边缘表征. 为了增强融合效果, 引入基于 NSNP 神经元构建的 CNSNP 块, 在保留细节与轮廓的同时提升非线性建模能力与鲁棒性. 因此, 这里由不同通

道组合的方式生成从低层次到高层次的边缘特征 E_3 、 E_2 和 E_1 进行层次化前馈, 进而得到 3 个不同层次的显著目标边缘特征. 该模块通过多尺度、多层次边缘特征提取与融合, 有效引导图像融合过程, 增强了融合结果的边缘和语义信息. 为了加速计算边缘特征提取模块的权重, 提前优化并不参与整个模型权重更新的过程.

2) 强光环境感知. 设计一个强光环境感知 (HLEA) 模块来应对背景强光干扰. 首先, 使用预训练的亮度分类子网络和亮度分解子网络预测可见图像的亮度概率和亮度图, 并结合两者计算红外-可见特征权重; 然后, 使用无参照图像空间质量评估器 (BRISQUE)^[13] 对权重进行修正, 以获得强光环境感知权重. 这里, 由 HLEA 计算的权重集表示为 $W = \{W_{ir}, W_{vi}\}$.

3) 图像特征提取. 设计一种双分支编码结构, 用于从源图像中提取特征. 在每个分支中, 使用 3 个梯度残差密集块 (GRDB)^[14] 作为主要组件, 从输入图像中提取细粒度和深层特征. 双分支对可见光图像和红外图像提取的特征集分别表示为 $F_{ir} = \{F_{ir}^1, F_{ir}^2, F_{ir}^3\}$ 和 $F_{vi} = \{F_{vi}^1, F_{vi}^2, F_{vi}^3\}$. 在合并从可见光和红外源提取的特征过程中, 没有采用简单的拼接融合, 而是受到自注意力机制的启发, 使用边缘特征 E_i 引导融合特征, 从而将显著目标的边缘特征嵌入到融合特征中, 以便于后续传输.

4) 图像重建. 对于由可见光和红外融合的双分支编码器提取的特征, 使用解码器来完成图像重建. 解码器为单分支结构, 由 5 个卷积层组成, 前 4 个卷积层使用 ReLU 激活函数, 最后 1 个卷积层使用 Tanh

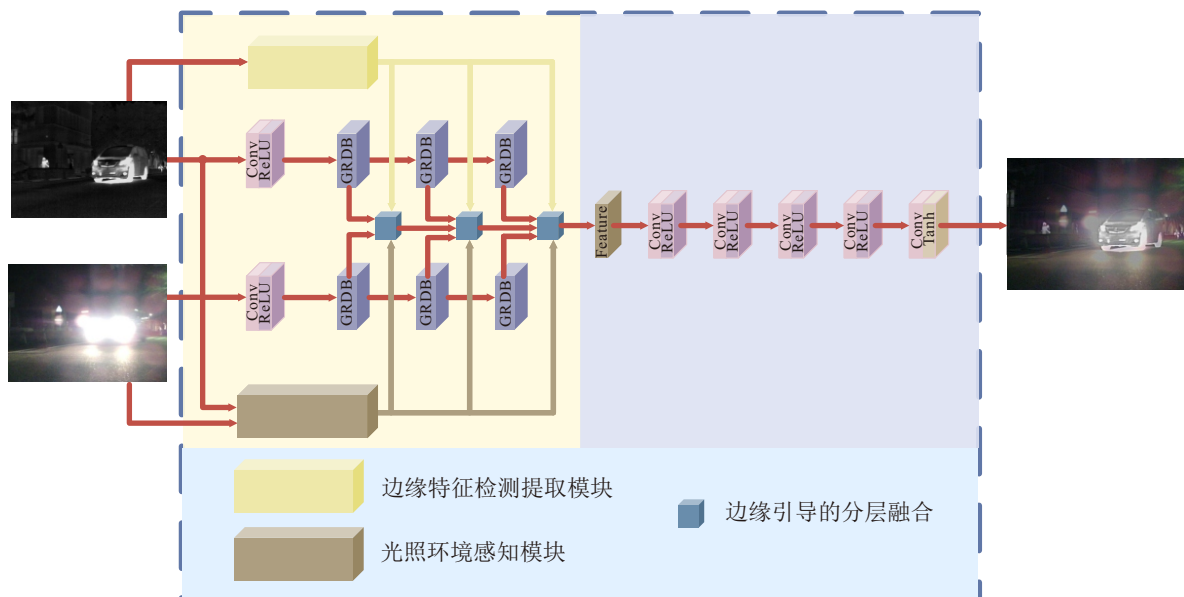


图1 融合网络体系结构

激活函数.

2.2 边缘特征提取

在计算机视觉任务中,边缘特征不仅刻画物体结构,也是高级语义信息的重要组成.然而,传统边缘检测方法依赖像素级梯度,易受噪声干扰,难以捕捉全局上下文,常导致边缘模糊或断裂.为此,引入多尺度特征聚合与上下文感知机制,将局部梯度与深层语义信息相融合,使边缘提取兼具精细定位与语义一致性.高层语义增强了对真实边界的辨识能力,细粒度边缘特征则保证了结构的准确还原.基于此,提出一种融合语义信息的边缘检测框架,有效提升边缘的连贯性与鲁棒性,为分割与识别等高层视觉任务提供更优支持.

为了充分挖掘不同尺寸下的红外图像语义特征和,利用3种不同尺寸扩张因子的卷积路径对红外图像分别进行两次特征提取后,利用sobel算子计算不同尺度下的边缘特征后按照通道做拼接操作,具体过程如下.

设输入红外图像为 $I \in \mathbb{R}^{H \times W \times C}$,通过不同扩张因子(dilation rate)的卷积核提取多尺度特征

$$F = \text{Conv}(I, K) \in \mathbb{R}^{H \times W \times C_{\text{out}}}. \quad (1)$$

其中: F 为红外图像经过空间特征提取层转换得到的空间特征,该提取层利用 1×1 卷积对通道信息进行重构,以增强特征表达能力.

$$F_s = \text{Conv}(\text{Conv}(F, K_s^{d_s}), K_s^{d_s}) \in \mathbb{R}^{H \times W \times C_s}, \quad (2)$$

$$F_m = \text{Conv}(\text{Conv}(F, K_m^{d_m}), K_m^{d_m}) \in \mathbb{R}^{H \times W \times C_m}, \quad (3)$$

$$F_l = \text{Conv}(\text{Conv}(F, K_l^{d_l}), K_l^{d_l}) \in \mathbb{R}^{H \times W \times C_l}. \quad (4)$$

其中: K_s, K_m, K_l 为不同尺寸的卷积核, $d_s < d_m < d_l$.大尺寸的卷积核具有更大的感受野,能够充分捕捉大尺寸边缘特征,而小尺寸的卷积和更加注重小尺寸边缘特征挖掘,最后通过通道拼接的操作

可以得到丰富的具有不同尺寸三通道边缘特征图.

为了使得到的边缘特征图同时具有不同尺寸边缘特征和不同深度的边缘特征,将红外图像直接与sobel算子进行运算得到一张浅层的边缘特征图与深层次下的多尺度边缘特征图进行通道拼接,拼接过程如下所示:

$$F_{\text{edge}} = \text{Cat}(F_s^{\text{edge}}, F_m^{\text{edge}}, F_l^{\text{edge}}, F^{\text{edge}}). \quad (5)$$

其中: $F_s^{\text{edge}}, F_m^{\text{edge}}, F_l^{\text{edge}}$ 由 F_s, F_m, F_l 三种不同尺度特征图通过sobel算子边缘特征提取得到,而 F^{edge} 直接由红外图像进行边缘特征提取得到.将4种边缘特征图按照通道维度进行拼接,可获得不仅同时具备不同尺度还同时具备不同层次的红外边缘特征图像.

最后,引用一种由NSNP神经元模型^[15]组合的CNSNP块结构用于对边缘特征进行融合,CNSNP块结合非线性激活和卷积操作,使模型能够更好地处理复杂的特征分布,尤其是在边缘检测中,可以同时保留低层次的细节信息和高层次的轮廓信息.与传统的卷积块相比,CNSNP块在多层特征的融合中表现出了更高的灵活性和鲁棒性,在边缘特征提取任务中表现出了卓越的性能,而CNSNP块对于边缘特征提取部分的应用可以用如下公式表示:

$$E_s = \text{Conv}(\text{ReLU}(\text{BN}(\text{Cat}(F^{\text{edge}}, F_s))), \quad (6)$$

$$E_m = \text{Conv}(\text{ReLU}(\text{BN}(\text{Cat}(F^{\text{edge}}, F_m))), \quad (7)$$

$$E_l = \text{Conv}(\text{ReLU}(\text{BN}(\text{Cat}(F^{\text{edge}}, F_l)))). \quad (8)$$

其中: E 为边缘特征提取模块从红外图像中提取的主要边缘特征信息,且一定程度上尽可能地包含了红外图像中可挖掘的纹理信息,按照边缘特征图从小尺度到大尺度的顺序,去引导红外光图像和可见光图像逐步融合.

综上所述,边缘特征提取模块如图2所示,可以

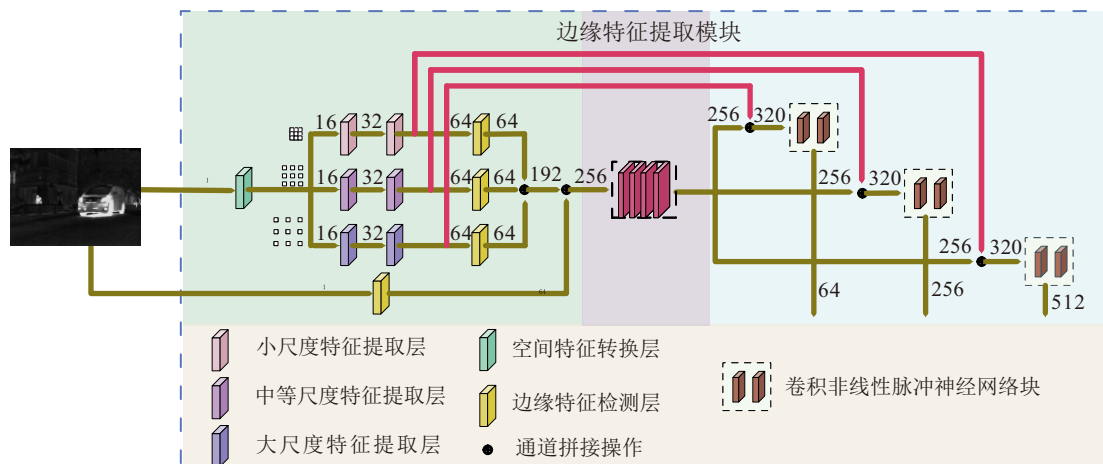


图2 边缘特征提取模块结构

做到对红外图像边缘特征的充分挖掘, 并将注重不同层次的边缘特征图引入图像融合结构, 去引导调整可见光图像和红外光图像的融合过程。

2.3 恶劣环境光感知

提升融合网络的抗干扰能力一直是研究重点。尽管已有方法在低光照、烟雾遮挡等场景中展现出一定鲁棒性^[10,16], 但针对强光干扰的研究仍较为空白, 尤其在夜间道路等高亮背景下, 融合图像质量显著下降。现有部分方法虽在强光条件下表现尚可, 但缺乏针对性设计和消融验证, 导致其抗干扰能力缺

乏理论支撑, 也限制了相关研究的进一步发展。

为解决上述问题, 本文提出了一种强光环境感知模块 (HLEA), 结构如图 3 所示。HLEA 集成了亮度分类网络^[17]、亮度分解网络^[18] 及无参图像质量评估器 BRISQUE, 旨在感知并适应强光场景下的信息分布差异。受渐近式亮度感知方法启发, HLEA 通过亮度分类判定可见光图像所处光照条件, 并动态调整红外与可见光特征的融合策略, 从而实现对抗强光干扰的有效抑制, 增强融合图像的质量与稳定性。

$$[P_d, P_n] = N_{IC}(I_{vi}). \quad (9)$$

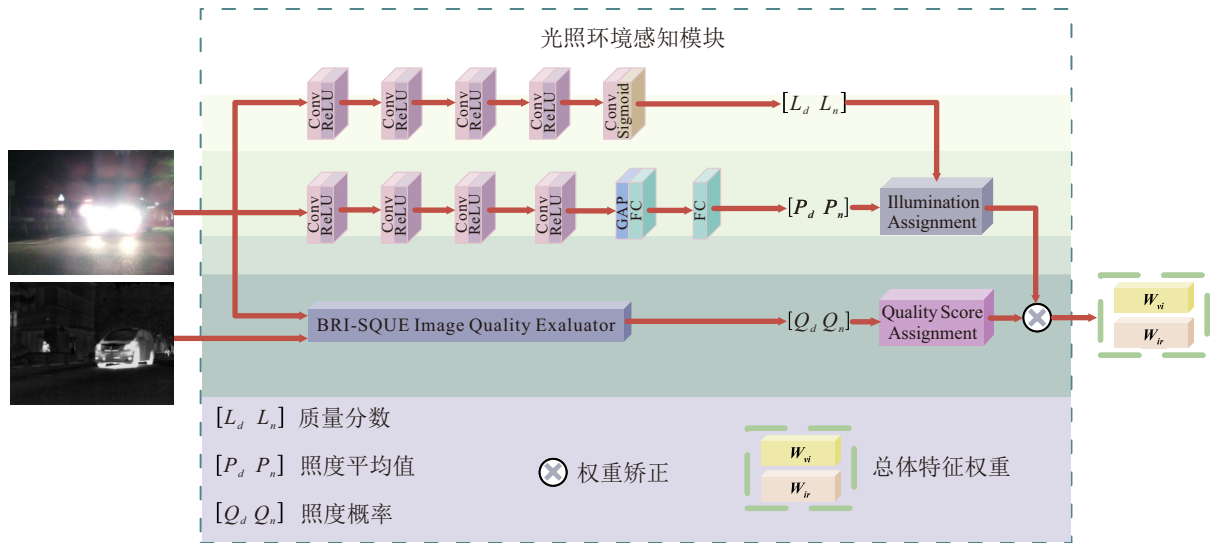


图3 强光环境感知模块 (HLEA) 结构

仅依靠 N_{IC} 来计算感知权重存在局限性。在强光条件下, 可见光图像的整体亮度较高, 导致预测概率中偏差增加。为此, 引入亮度分解网络, 将输入图像分解为亮度图和细节图, 其中亮度图有效地反映了可见图像的亮度分布。在强光背景下, 由 N_{IC} 计算的 P_d 值往往较高。同时, 亮度图在前 50% 的像素中也表现出较高的平均值。因此, 本文使用亮度图的像素均值来修正由 N_{IC} 预测的感知权重。修正后的红外特征权重增加, 可见光特征权重减少。这增强了红外热目标信息的融合, 减少了可见光图像中无意义强光信息的融合, 从而提高了图像在强光条件下的融合效果。具体过程如下:

$$L = N_{ID}(I_{vi}), L_{day} = \frac{P_d}{L_{pixl}}, L_{night} = \frac{P_n}{L_{pixl}}; \quad (10)$$

$$W'_{vi} = \frac{L_{day} \cdot P_d}{(L_{day} + L_{night})(P_d + P_n)}; \quad (11)$$

$$W'_{ir} = \frac{L_{night} \cdot P_n}{(L_{day} + L_{night})(P_d + P_n)}; \quad (12)$$

$$W_{vi}^c = \frac{W'_{vi}}{W'_{vi} + W'_{ir}}, W_{ir}^c = \frac{W'_{ir}}{W'_{vi} + W'_{ir}}. \quad (13)$$

其中: L 表示可见光图像的亮度图, N_{ID} 为亮度分解网络, L_{pixl} 和 L_{pixr} 分别表示亮度图前 50% 和后 50% 像素的灰度均值, L_{day} 和 L_{night} 分别为亮度图的白天和夜晚校正系数, W_{vi}^c 和 W_{ir}^c 表示亮度图校正后的可见光和红外的修正感知权重。

考虑到可见光和红外图像在强光干扰下不同程度的退化, 使用图像质量评估算法 BRI-SQUE 来评估源图像的质量, 并基于评估质量得分计算感知权重。具体的计算过程如下:

$$W_{vi}^q = \frac{Q_{ir}}{Q_{vi} + Q_{ir}}, \quad (14)$$

$$W_{ir}^q = \frac{Q_{vi}}{Q_{vi} + Q_{ir}}. \quad (15)$$

其中: Q_{vi} 和 Q_{ir} 分别表示使用无参数质量评估算法 BRI-SQUE 计算得到的可见光和红外图像的质量得分。较高的得分反映了较差的图像质量, 而较低的得分则表示较好的图像质量。 W_{vi}^q 和 W_{ir}^q 分别表示从图像质量得分中计算得到的可见光和红外数据的感知

权重.

最后,使用 N_{IC} 和 N_{ID} 计算的感知权重通过BRI-SQUE计算的感知权重进行修正,以获得最终结果.具体的修正过程如下所示:

$$W_{ir} = (1 - a)W_{ir}^c + aW_{ir}^q, \quad (16)$$

$$W_{vi} = (1 - a)W_{vi}^c + aW_{vi}^q. \quad (17)$$

其中: W_{ir} 和 W_{vi} 分别是由HLEA模块输出的红外和可见光的感知权重.值得注意的是,这里的 a 是一个校正系数,用于平衡从亮度评估获得的权重和通过图像质量评分评估计算的权重.在本文的实验中,经验性地将其设为0.2,以平衡这两个因素.

2.4 边缘引导分层融合

通过边缘特征提取模块获取显著目标的边缘特征后,直接将边缘特征与红外-可见光特征通道进行特征融合(例如直接连接或逐元素相加)容易导致边缘噪声或冗余信息干扰,使得边缘特征难以被充分利用.为了解决该问题并增强显著目标边缘特征的引导作用,提出一个边缘引导的分层融合(EGHF)模块,用于处理不同层次的红外-可见光特征和显著目标边缘特征的交互与融合.与直接将边缘特征与红外-可见光特征进行连接或相加融合不同,这里将边缘特征作为红外-可见光特征的引导信息,从而实现全面的特征融合.该模块的详细结构如图4所示.

受Transformer中多头自注意力机制^[19]的启发,

该机制能够有效捕捉长距离依赖关系.本文捕捉了边缘特征与红外-可见光特征之间的关联,并基于此完成了边缘特征引导的融合.首先将边缘特征 $E_{edge}^i \in \mathbb{R}^{C_i \times H_i \times W_i}$ 上采样到 $E_{edge'}^i \in \mathbb{R}^{C_i \times H \times W}$,以匹配可见光-红外特征的尺度.其中: H 和 W 分别表示输入图像的高度和宽度, C_i 为每一层的通道数.通过卷积层的线性投影获得查询矩阵 $E_q^i \in \mathbb{R}^{HW \times C'}$ 和键矩阵 $E_k^i \in \mathbb{R}^{C' \times HW}$,其中 $C' = 2C_i$.具体过程如下:

$$E_{edge'}^i = Up(E_{edge}^i), \quad (18)$$

$$E_i^q = \nu_1(\text{Conv}_1(E_{edge'}^i)), \quad (19)$$

$$E_i^k = \nu_2(\text{Conv}_1(E_{edge'}^i)). \quad (20)$$

其中: $\nu_1(\cdot)$ 表示将矩阵维度从 $\mathbb{R}^{A_1 \times A_2 \times A_3}$ 转换为 $\mathbb{R}^{A_2 \times A_3 \times A_1}$ 的变换算子, $\nu_2(\cdot)$ 表示将矩阵维度从 $\mathbb{R}^{A_1 \times A_2 \times A_3}$ 转换为 $\mathbb{R}^{A_1 \times A_2 \times A_3}$ 的变换算子; E_i^{edge} 表示第 i 个EFEB模块输出的边缘特征, $E_i^{edge'}$ 表示上采样后的边缘特征, E_i^q 和 E_i^k 表示对应于边缘特征的查询矩阵和键矩阵.

在获得与边缘特征对应的查询矩阵 E_i^q 和键矩阵 E_i^k 之后,需要从红外-可见特征中提取数值矩阵,以建立边缘特征与红外-可见特征之间的关联.首先,使用从HLEA模块获得的感知权重 W_{ir} 和 W_{vi} 控制特征输入比例来进行抗强光干扰融合,实现红外特征与可见特征的融合.处理后的红外和可见特征的感知权重分别为 $F_{ir'}^i \in \mathbb{R}^{C \times H \times W}$ 和 $F_{vi'}^i \in \mathbb{R}^{C \times H \times W}$.

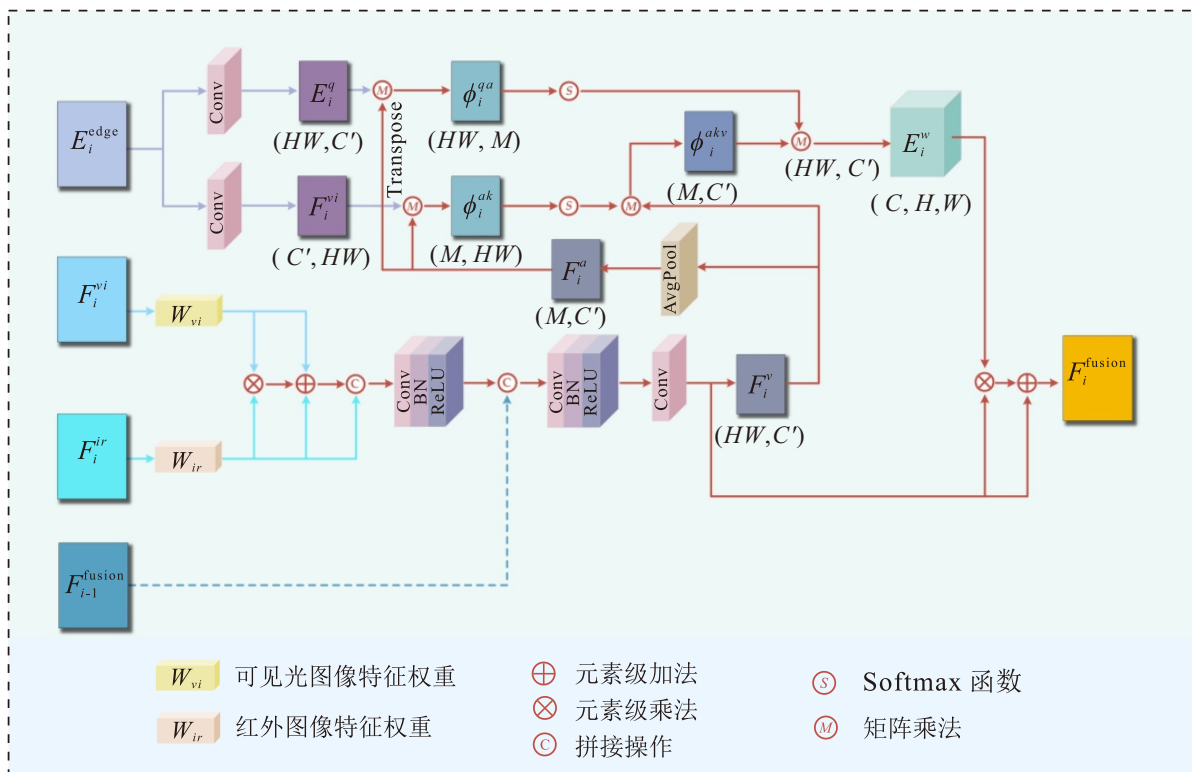


图4 边缘引导分层融合模块结构

然后, 通过逐元素相乘和相加来分别突出两种模式的差异和共同特征. 最后, 将相乘和相加的结果与输入特征通道连接后, 通过 $\text{CBR}(\cdot)$ 层获得融合特征 $F_i^{ir-vi} \in \mathbb{R}^{HW \times C'}$. 在获得红外-可见融合特征后, 通过卷积层的线性投影获得数值矩阵 $F_i^v \in \mathbb{R}^{HW \times C'}$. 需要注意的是, 在第2层和第3层的可见和红外特征的融合过程中, 引入了上一层的融合特征 $F_{i-1}^{\text{fusion}} \in \mathbb{R}^{C \times H \times W}$. 因此, 在第2层和第3层的融合模块 EGHF 与第1层的 EGHF 之间的特征输入存在差异. 初始层的融合特征的数值矩阵的推导过程如下:

$$F_1^{ir'} = W_{ir'} \cdot F_1^{ir}, F_1^{vi'} = W_{vi'} \cdot F_1^{vi}; \quad (21)$$

$$F_1^{ir-vi} = \text{CBR}(\text{CBR}(\text{Cat}(((F_1^{ir'} \otimes F_1^{vi'}) \oplus F_1^{ir'} \oplus F_1^{vi'}), F_1^{ir'}))); \quad (22)$$

$$F_1^v = \nu_1(\text{Conv}_1(F_1^{ir-vi})). \quad (23)$$

其中: F_1^{ir} 和 F_1^{vi} 表示第1层的红外和可见特征, 而 $F_1^{ir'}$ 和 $F_1^{vi'}$ 表示经过感知权重处理后的红外和可见特征; F_1^{ir-vi} 表示融合红外和可见图像后的集成特征, F_1^v 对应于红外-可见特征的数值矩阵.

在第2层和第3层的红外和可见特征融合中, 与第1层的区别在于第1个 $\text{CBR}(\cdot)$ 层之后、第2个 $\text{CBR}(\cdot)$ 层之前, 连接了上一层的融合输出 F_{i-1}^{fusion} . 详细过程如下:

$$F_i^{ir-vi} = \text{CBR}(\text{Cat}(\text{CBR}(\text{Cat}(((F_i^{ir'} \otimes F_i^{vi'}) \oplus F_i^{ir'} \oplus F_i^{vi'}), F_i^{ir'})))); \quad (24)$$

$$i = \{2, 3\}. \quad (25)$$

其中: $F_i^{ir'}$ 和 $F_i^{vi'}$ 分别表示经过感知权重处理后的第 i 层的红外和可见特征, F_i^{ir-vi} 表示第 i 层红外和可见图像的集成特征, i 取值为 2 和 3 分别对应于第2层和第3层.

在获得查询矩阵 E_i^q 、键矩阵 E_i^k 和数值矩阵 F_i^v 后, 边缘注意引导的权重矩阵值的计算方法遵循 Transformer 中的自注意机制. 然而, 该方法在通过 E_i^q 和 E_i^k 的矩阵乘法进行计算时, 会产生大小为 (HW, HW) 的注意力权重. 例如, 对于一个 640×480 的输入图像, 注意力权重的大小为 $(307, 200, 307, 200)$, 这对于硬件计算资源来说是不切实际的. 本文借鉴 Agent 注意力机制^[20], 引入一个中间量 F_i^a 以减少计算复杂度. 具体计算过程如下:

$$F_i^a = \text{GAP}(F_i^{ir-vi}), \quad (26)$$

$$\phi_i^{qa} = \text{mul}(E_i^q, T(F_i^a)), \quad (27)$$

$$\phi_i^{ak} = \text{mul}(F_i^a, E_i^k), \quad (28)$$

$$\phi_i^{akv} = \text{mul}(\phi_i^{ak}, F_i^v), \quad (29)$$

$$E_i^w = \nu_3(\text{mul}(\phi_i^{qa}, \phi_i^{akv})). \quad (30)$$

其中: $F_i^a \in \mathbb{R}^{M \times C'}$ 表示通过对红外-可见特征进行全局平均池化后获得的中间量, $\text{mul}(\cdot)$ 表示矩阵乘法, $T(\cdot)$ 表示矩阵转置算子, $\phi_i^{qa} \in \mathbb{R}^{HW \times M}$ 表示中间查询矩阵, $\phi_i^{ak} \in \mathbb{R}^{M \times HW}$ 表示中间键矩阵, $\phi_i^{akv} \in \mathbb{R}^{M \times C'}$ 表示中间权重矩阵, $\nu_3(\cdot)$ 表示将矩阵维度从 $\mathbb{R}^{A_1 A_2 \times A_3}$ 转换为 $\mathbb{R}^{A_3 \times A_1 \times A_2}$ 的操作, 且 $E_i^w \in \mathbb{R}^{C' \times H \times W}$ 表示边缘特征权重矩阵. 引入中间变量将权重矩阵的计算过程分解为多个步骤, 将空间复杂度转化为时间复杂度, 从而避免了硬件资源不可接受的计算复杂度.

在建立边缘特征与红外-可见特征之间的关联后, 获得了边缘特征的注意力权重矩阵. 利用该矩阵, 将权重分配到红外-可见特征上, 实现了由边缘特征引导的特征融合. 详细过程如下所示:

$$F_i^{\text{fusion}} = (E_i^w \otimes F_i^{ir-vi}) \oplus F_i^{ir-vi}, \quad (31)$$

$$i = \{1, 2, 3\}. \quad (32)$$

其中: $F_i^{\text{fusion}} \in \mathbb{R}^{C \times H \times W}$ 表示第 i 层 EGHF 模块的输出融合特征, i 的取值分别表示不同层次的特征融合.

2.5 损失函数

为了提高网络在定性视觉感知和定量无参考评估指标方面的性能, 并在强光照条件下获得高质量的融合图像, 本文模型选择内容损失和语义损失来形成联合损失函数, 即

$$L_{\text{fusion}} = L_{\text{con}} + \lambda L_{\text{sem}}. \quad (33)$$

其中: L_{con} 表示内容损失, L_{sem} 表示语义损失, λ 表示融合网络训练中的语义损失引导水平. 内容损失

$$L_{\text{con}} = L_{\text{int}} + \alpha L_{\text{texture}}. \quad (34)$$

其中: L_{int} 是像素强度损失, 反映了从网络输入到最终输出的像素值差异; L_{texture} 是细节纹理损失, 反映了从网络输入到最终输出在纹理细节上的差异; α 是平衡因子. 像素强度损失 L_{int} 的定义如下:

$$L_{\text{int}} = \frac{1}{HW} \|I_{\text{fused}} - \max(I_{\text{vis}}, I_{\text{ir}})\|_1. \quad (35)$$

其中: H 是图像的高度, W 是图像的宽度, $\|\cdot\|_1$ 表示 L_1 范数计算, I_{fused} 是融合输出的像素强度值, I_{ir} 是红外光源的像素强度值, I_{vis} 是可见光源的像素强度值.

此外, 为了在保留额外有意义的纹理结构信息的同时突出显著目标, 定义细节纹理损失

$$L_{\text{texture}} = \frac{1}{HW} \|\nabla I_{\text{fused}} - \max(|\nabla I_{\text{vis}}|, |\nabla I_{\text{ir}}|)\|_1. \quad (36)$$

其中: $\|\cdot\|_1$ 表示 L_1 范数计算, $|\cdot|$ 表示绝对值操作, ∇ 表示图像纹理的 Sobel 梯度算子.

为了增强融合结果中场景的可理解性, 引入如下语义分割网络和语义损失^[14]:

$$L_{\text{sem}} = L_{\text{sem_main}} + \beta L_{\text{sem_aux}}, \quad (37)$$

$$L_{\text{sem_main}} = -\frac{1}{HW} \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W L_5^{(c,h,w)} \log(I_{\text{sem}}^{(c,h,w)}), \quad (38)$$

$$L_{\text{sem_aux}} = -\frac{1}{HW} \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W L_5^{(c,h,w)} \log(I_{\text{sem_a}}^{(c,h,w)}). \quad (39)$$

其中: C 表示语义分割的类别数, $I_{\text{sem}} \in \mathbb{R}^{C \times H \times W}$ 和 $I_{\text{sem_a}} \in \mathbb{R}^{C \times H \times W}$ 分别是分割结果和辅助分割结果, $L_5 \in (1, C)^{H \times W}$ 表示从单一独热编码标签获得的结果表示. 语义损失的使用不仅可以辅助分割网络, 还可以辅助融合网络的训练.

3 实验分析

为展示本文所提出的融合方法的优越性与先进性, 通过实验在多个红外与可见光数据集上进行不同方法的对比及模块消融实验. 首先, 详细阐述了实验配置的具体细节和评估指标的选取; 其次, 在融合任务和高层视觉检测任务中, 将所提出的融合方法与 9 种主流方法进行了对比, 包括 3 类传统方法 (CBF^[21]、NSST^[22] 和 GTF^[7])、2 种基于自编码器的方法 (DIDFuse^[23] 和 NestFuse^[24])、2 种基于生成对抗网络的方法 (FusionGAN^[25] 和 GANMcC^[26]), 以及 2 种基于卷积神经网络的方法 (PIAFusion^[16] 和 DRF^[27]).

3.1 实验细节

3.1.1 联合训练策略

对于本文提出的网络模型, 选择 MFNET 和 TNO 数据集进行训练, 以及从 MFNet 数据集中筛选并处理得到的 MSRS 数据集进行训练, 共有 1 083 对可见光和红外图像的训练数据.

为了防止因样本量小而导致的过拟合, 对原始训练样本应用了数据增强方法, 如随机裁剪、颜色抖动、随机缩放和水平翻转. 此外, 引入学习率预热、学习率衰减以及 Adam 优化策略, 以防止由于学习率不当而导致的训练发散问题.

为了在融合图像中包含更丰富的语义信息, 引入分割网络并在融合损失中添加分割语义损失. 因此, 设计一个用于融合和分割网络的协同训练算法, 使优化融合效果变得至关重要. 为此, 本文设计了一种协同训练策略, 如算法 1 所示.

Algorithm 1 自适应协同训练策略.

network input: 可见光图像 I_{vi} 和红外图像 I_{ir} ;
network output: 融合图像 I_{fused} .
选取 a 组可见光图像及标签 $\{(I_{vi}^1, L_{vi}^1), \dots, (I_{vi}^a, L_{vi}^a)\}$;
使用 Adam 优化器训练光照分类网络 N_{IC}
 $\nabla_{N_{\text{IC}}} \{L_{\text{class}}(N_{\text{IC}})\}$;
选择 b 组可见光图像 $\{I_{vi}^1, I_{vi}^2, \dots, I_{vi}^b\}$;
使用 Adam 优化器训练光照分解网络 N_{ID}
 $\nabla_{N_{\text{ID}}} \{L_{\text{split}}(N_{\text{ID}})\}$;
for $m \leq \max \text{iterations } M$ do
 for p steps do
 选取 c 组可见光图像 $\{I_{vi}^1, I_{vi}^2, \dots, I_{vi}^c\}$;
 选取 c 组红外光图像 $\{I_{ir}^1, I_{ir}^2, \dots, I_{ir}^c\}$;
 更新平衡参数
 $\lambda = \tau(1 + e^{-(m-1)})^{-1}$;
 使用 Adam 优化器更新融合网络 N_f 的权重参数
 $\nabla_{N_f} \{L_{\text{fusion}}(N_f)\}$;
 end for
 利用融合网络生成融合图像, 用于后续分割网络训练;
 for q steps do
 从融合网络输出中选取 d 组融合图像 $\{I_f^1, I_f^2, \dots, I_f^d\}$
 使用 Adam 优化器更新分割网络 N_s 的权重参数:
 $\nabla_{N_s} \{L_{\text{sem}}(N_s)\}$;
 end for
end for

在算法 1 的训练过程中, 首先对 HLEA 模块中的光照分类与分解子网络进行预训练, 随后进行融合网络与分割网络的协同训练. 每一训练阶段中, 先训练融合网络生成融合图像, 再输入分割网络获取语义分割结果, 并反向引导融合网络优化, 从而逐步增强融合图像中的语义表达.

考虑到分割网络在不同阶段的准确性差异, 引入语义损失平衡参数 λ 控制其对融合过程的影响. 初期 λ 较小, 避免误导融合网络, 后期随着分割准确性的提升, 逐步增大该参数. 平衡函数定义为 $\lambda = \tau(1 + e^{-(m-1)})^{-1}$. 其中: $\tau = 3.15$, m 为当前训练阶段, 最大阶段数 $M = 4$. 该函数控制语义引导力度非线性增长, 贴合模型性能提升节奏.

训练中, 融合网络最大迭代步数设为 $p = 2\,700$, 分割网络为 $q = 20\,000$. 各网络分别表示为 N_{IC} (光照分类)、 N_{ID} (光照分解)、 N_f (融合) 和 N_s (分割); 对应损失为 L_{class} 、 L_{split} 、 L_{fusion} 和 L_{sem} . 此外, 设置纹理

损失系数 $\alpha = 10$, 辅助语义损失系数 $\beta = 0.1$.

所有实验在 PyTorch 框架下完成, 使用 Adam 优化器进行训练, 硬件平台为 NVIDIA RTX 4080 GPU 和 Intel i9-13900K CPU.

3.1.2 融合评价指标

融合结果的定量评估可从融合任务本身及其对高级视觉任务的影响两个角度展开. 针对融合图像质量, 本文采用多种无参考评价指标, 对所提出模型在 TNO 和 MFNet 数据集上的表现进行评估, 包括 EN、FMI_{pixel}、SD、Qabf、MS-SSIM、VIF 和 PSNR. 这些指标从不同维度评估图像质量: EN 和 FMI_{pixel}从信息理论角度度量图像的信息量及像素级互信息; MS-SSIM 和 Qabf评估融合图像与源图像在结构层面的相似性, 前者侧重多尺度结构一致性, 后者关注信息保留程度; SD 和 PSNR 分别反映图像对比度和与源图像的重建误差; VIF 则从人类视觉感知角度估计图像质量, 强调感知信息的保真度. 通过引入多维度、广泛认可的无参考指标, 力求实现对融合性能的全面、稳健和客观的定量评估.

3.1.3 检测评估细节

融合任务的重要目标之一是提升后续高级视觉任务的性能. 然而, 融合图像的无参考质量评估指标往往无法完全反映其在实际任务中的表现, 导致单纯从融合质量角度的评估不够全面. 为此, 进一步在

目标检测任务中验证所提出方法在强光干扰下的实用性. 本文采用 YOLOv5 检测算法, 利用包含 500 张手动标注的可见光与红外图像构建训练集, 并在 MFNet 数据集中标注“人”与“车辆”两个类别. 随后, 在 MFNet 中的 40 个正常场景、25 个强光背景场景, 以及 RoadScene 中的 15 个高曝光场景 (共计 40 个特殊场景) 上进行检测实验, 全面评估模型在复杂光照条件下的表现. 在检测结果的评估方面, 本文采用 F1 分数、AP@50、AP@70、AP@90 以及 AP@[50 : 95] 等 5 项指标, 从不同置信度阈值下分析检测精度与召回率, 以全面衡量融合图像对下游任务的支持能力.

3.2 在 TNO 数据集上的对比实验

为了验证本文所提出方法在一般场景中的性能, 首先从 TNO 数据集中选择 26 个不同的场景. TNO 数据集是一个包含常规场景下可见光和红外图像的标准数据集, 用于从融合任务的角度对本文模型和对比方法进行定性和定量评估.

3.2.1 定性比较

来自 TNO 数据集的两个典型图像对的定性结果显示在图 5 中. 值得注意的是, 图中的红色框用于放大图像中的关键目标对象和纹理背景区域, 以提供更清晰的视角.

在第 1 个场景的定性结果展示中, 使用红色框

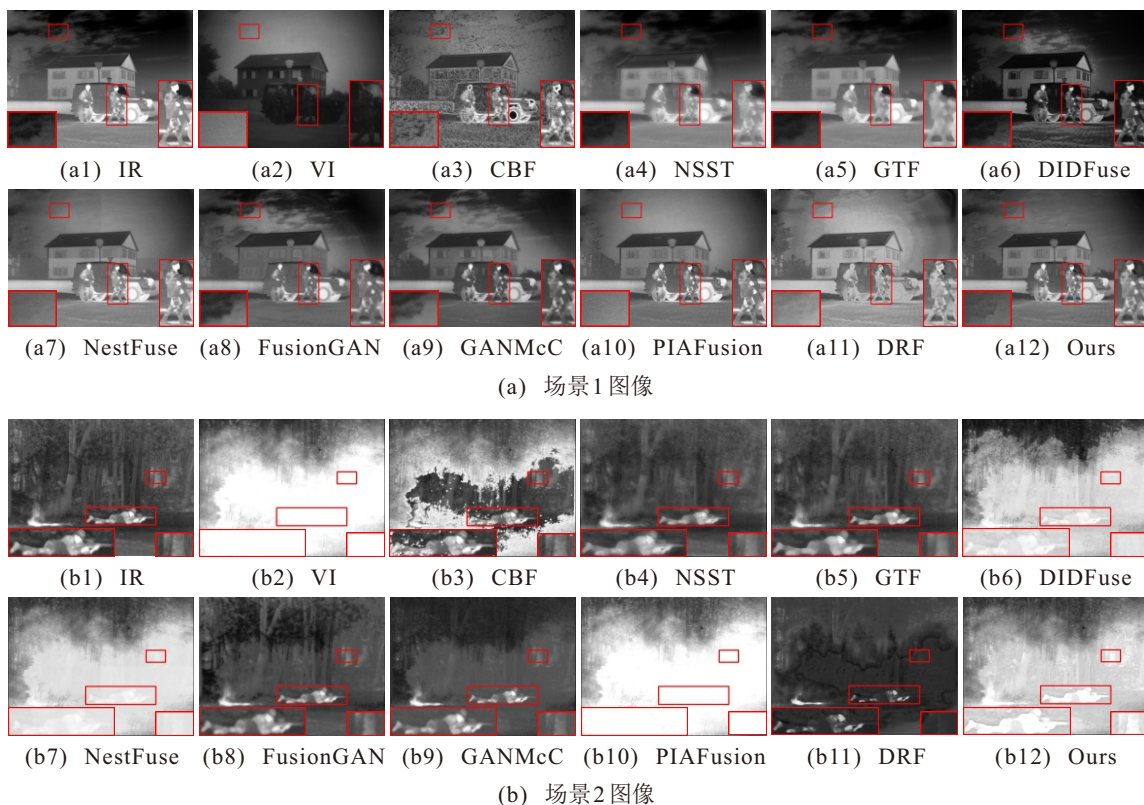


图5 TNO 数据集图像的定性结果

选择了显著的人体目标区域和天空云层纹理区域. 对于显著人体目标区域的融合结果, FusionGAN 和 GTF 方法严重模糊了人体的边缘和纹理细节; DRF 和 DIDFuse 方法保留了人体目标的整体显著信息, 但仍然存在由于部分显著特征丢失和人体目标显著性减弱而导致的无用信息干扰问题. 只有 NestFuse 和本文方法能够较好地保留人体目标的纹理, 同时避免显著信息的丢失. 但是, NestFuse 在处理红外图像纹理方面存在严重缺陷, 如天空云层纹理区域所示, 其纹理细节完全丢失. 因此, 只有本文方法能够在保留复杂纹理特征的同时突出显著的红外目标细节.

3.2.2 定量比较

为了更客观地评估本文方法在 TNO 数据集上的性能, 对方法结果进行了额外的定量评估. TNO 数据集中 26 对图像的定量评估结果如表 1 所示. 本文方法在 EN、MS-SSIM、PSNR、 FMI_{pixel} 和 Q_{abf} 方面表现出优越的性能, 在 SD 指标中排名第 2, 在 VIF 指标上略微落后于 NestFuse. 这些结果表明, 本文设计的 HLEA 模块并没有为了适应强光特殊场景而完全牺牲正常场景下的融合性能. 总而言之, 本文方法保留了更多与输入图像高度相关的有意义信息, 从而在定性和定量上实现了整体最佳性能.

表1 比较方法的 TNO 数据集的定量结果

Method	EN	SD	MS-SSIM	VIF	PSNR	FMI_{pixel}	Q_{abf}
CBF	7.1291	40.4001	0.7084	0.5895	62.9445	0.8646	0.4200
NSST	6.8753	42.4164	0.8237	0.6565	61.7691	0.8921	0.2986
GTF	6.8811	42.7918	0.8238	0.6527	61.7706	0.8897	0.4202
DIDFuse	7.2034	52.4016	0.8796	0.8140	61.9209	0.8645	0.3419
NestFuse	7.1964	44.6529	0.8825	0.9437	62.6227	0.8911	0.5273
FusionGAN	6.5960	30.9010	0.7392	0.6125	60.9479	0.8709	0.2252
GANMcC	6.8340	34.1770	0.8375	0.6736	61.7864	0.8816	0.2646
PIAFusion	6.9998	44.5875	0.8655	0.9345	61.6819	0.8949	0.4854
DRF	6.7745	33.8643	0.7115	0.5633	61.3385	0.8766	0.1977
Ours	7.2103	45.6233	0.9223	0.7144	63.0217	0.8971	0.5361

3.3 在 MFNet 数据集上的对比实验

TNO 数据集中的场景通常较为简单, 包含的物体数量较少. 为了验证本文方法在更复杂场景中的有效性, 进一步从 MFNet 数据集中选择了 25 个场景, 对本文模型和其他方法进行了定性和定量评估. 在 TNO 数据集中, 可见光和红外光谱都以灰度图形式表示, 而在 MFNet 数据集中, 可见光图像为彩色图像. 在融合过程中, 首先将可见光图像从 RGB 色彩空间转换为 YCrCb 色彩空间; 其次, 将亮度通道 Y 与红外图像融合; 然后, 将融合图像与可见光图像的 Cr 和 Cb 通道合并; 最后, 将得到的图像重新转换为 RGB 色彩空间, 以生成彩色融合图像. 值得注意的是, 尽管本文选择了一组在融合领域广泛认可的无参考指标进行多角度图像评估, 这些指标通常仅在正常条件下对融合图像的评估有效, 对于强光条件下的融合图像评估存在局限性.

3.3.1 定性比较

1) 融合结果展示.

图 6 展示了 MFNet 数据集中两个具有代表性的图像对的直观定性结果. 场景 1 显示了具有强光干扰的白天环境, 而场景 2 呈现了正常背景条件下

的夜间环境. 图像中的红色框用于放大重要的目标物体和纹理背景区域.

在图 6 中的两个典型场景的定性视觉性能中, 本文方法生成的结果明显优于其他方法. 这是因为, 与大多数先进的融合算法相比, 在场景中存在强光干扰时, 本文方法能够在融合结果中有效抑制强光, 同时保持热辐射敏感目标的重要性和强光遮挡下的丰富纹理细节. 本文方法处理强光干扰的能力至关重要, 因为去除强光对于提高后续高级视觉任务(如目标检测等)的性能具有决定性作用. 此外, 在典型的道路场景中, 本文的融合方法依然能够显著保留大量纹理细节, 同时保持足够的热辐射信息.

根据两个典型场景的图像融合定性结果, 可以根据去除强光干扰的能力将所选的对比方法分为两组. 第 1 组包括 DIDFuse 和 NestFuse, 这些方法无法有效地去除融合结果中的强光信息. 例如, 在场景 1 中, NestFuse 的融合结果几乎完全被强光信息遮盖, DIDFuse 中大量的车辆纹理被强光淹没. 第 2 组包括 GTF、FusionGAN 和 DRF, 这些方法基本上从融合结果中去除了强光. 然而, 由于缺乏去除强光的研究, 融合结果中强光遮挡下的热目标纹理细节在不

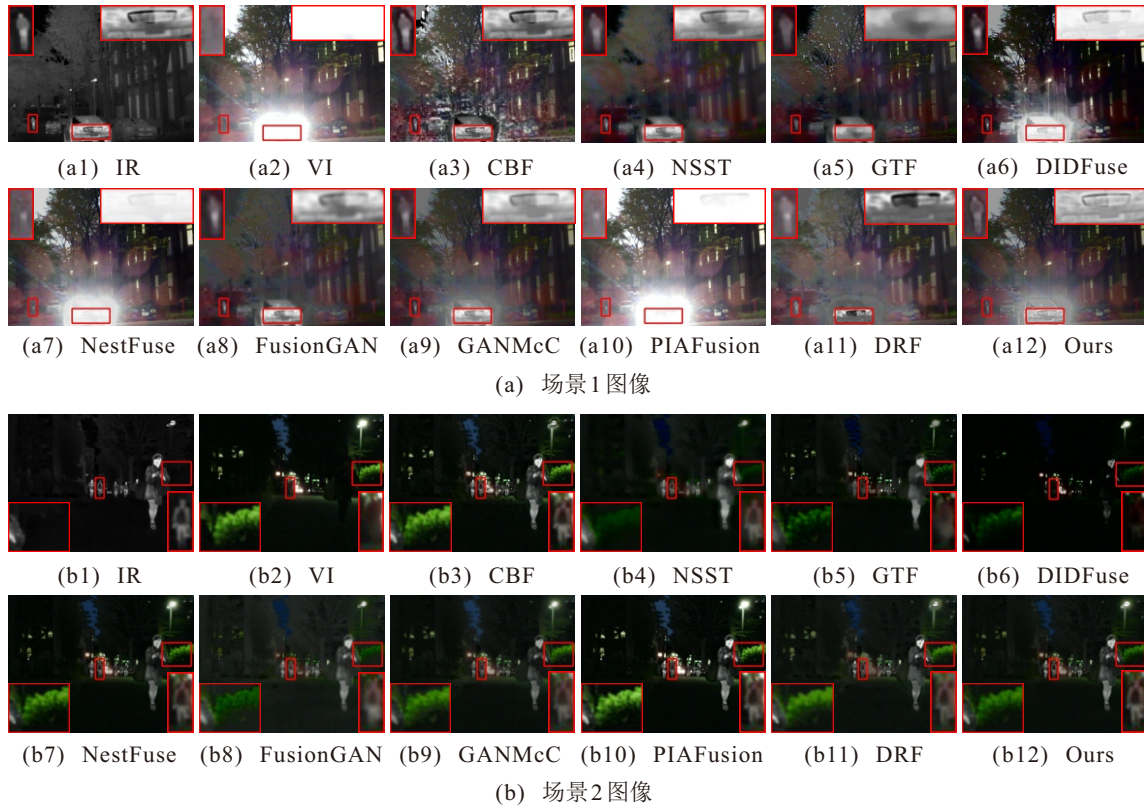


图6 MFNet 数据集图像的定性结果

同程度上出现模糊. 例如, 在场景 1 中, GTF、DRF 和 FusionGAN 等方法强光遮挡下的车辆纹理最为模糊. 只有本文方法在去除强光干扰的同时, 保留了足够的热辐射信息和丰富的纹理细节. 在典型场景中, 第 2 组中只有本文方法在保持与 NestFuse 相同的融合性能的同时, 保留了足够的纹理信息并在融合过程中增强了细节, 使得融合图像看起来更自然. 综上所述, 本文方法结合上述两种方法的优势, 既保留了丰富的纹理细节和热目标的显著性, 又避免了强光背景干扰.

2) MOS 主观质量评价.

MOS 是一种用于评估图像、音频或视频质量的

主观评价方法. 它通过汇总多位观察者的评分来代表媒体的整体感知质量. 本文选择采用该方法来评估高光条件下融合图像的质量. 在实验中, 收集 15 位评审者对 40 张高曝光或亮光条件下的融合图像在 5 个维度上的平均意见评分: 图像清晰度、色彩保真度、抗光性、伪影噪声和整体印象. 然后, 计算相应的总体 MOS. 每个维度在离散的序数尺度上评分, 共 5 分, 对应 5 个等级. 评分从 1 到 5 分别表示该维度的质量从非常差、差、一般、好到优秀. 最终的 MOS 见表 2, 其中每个维度从 1 到 5 进行评级, 得分越高表示在该维度上表现越优秀, 得分越低表现则相反. 红色表示最佳结果, 蓝色表示次优结果.

表2 在强光条件下融合图像的平均意见得分

Method	Image Clarity	Color Fidelity	Light Resistance	Artifact Noise	Overall Impression	Average Score
CBF	2.0567	1.9917	3.0733	1.6833	2.1033	2.1817
NSST	1.6483	1.8933	3.3917	2.5183	2.0950	2.3093
GTF	2.0133	2.4983	3.2833	2.2317	2.1450	2.4343
DIDFuse	2.4717	2.2650	3.0267	2.5833	2.3283	2.5350
NestFuse	3.1967	3.2783	2.5433	3.4917	3.0367	3.1093
FusionGAN	1.7517	2.3583	3.6433	2.0250	2.1433	2.3843
GANMcC	2.8683	2.8417	3.4017	2.9200	2.9917	3.0047
PIAFusion	2.7567	2.5933	2.2239	3.0733	2.4217	2.6138
DRF	1.7567	2.0967	3.1983	1.7533	1.8833	2.1377
Ours	3.2241	3.2671	3.7211	3.4221	3.1327	3.2033

3.3.2 定量比较

在强光条件下融合图像的主观质量评估中,本文方法在图像清晰度、色彩保真度和整体印象方面表现最佳,而在抗光性和噪声伪影方面表现次优.基于5个维度的平均意见评分,本文方法在所有方法中具有最佳的主观视觉性能.此外,值得注意的是,在抗光性指标上表现较好的方法分别是本文方法和 FusionGAN,而表现较差的是 NestFuse.这与先前展示的强光条件下融合图像的去强光性能一致,表明本文方法有效减轻了强光对观众造成的严重干扰.

为了客观地展示本文方法在复杂场景中的效果,在包含 MFNet 数据集中 25 对图像的数据集上进行了定量无参考评估指标测试,结果在表 3 中列出.可以看出,在典型场景中,本文方法在 EN、MS-SSIM、 FMI_{pixel} 和 Q_{abf} 方面实现了最佳性能,在 SD 和 VIF 方面获得了次佳性能.这表明,本文方法保留了来自源图像的大量纹理和边缘信息.总之,本文方法在确保足够热辐射信息的同时,保持了丰富的纹理细节,达到了客观评估的最高标准.

表3 选定比较方法在 MFNet 数据集上的定量结果

Method	EN	SD	MS-SSIM	VIF	PSNR	FMI_{pixel}	Q_{abf}
CBF	6.5949	36.8343	0.7942	0.7370	64.4616	0.9031	0.5697
NSST	5.8504	23.7725	0.6489	0.6522	63.7954	0.9224	0.2350
GTF	5.8667	23.7152	0.8471	0.5788	64.1360	0.9101	0.4076
DIDFuse	4.8717	39.6267	0.8087	0.4136	63.6994	0.9075	0.2305
NestFuse	6.6019	44.1226	0.9533	0.9247	65.0529	0.9056	0.6275
FusionGAN	5.6645	20.4431	0.7226	0.5559	64.3055	0.9065	0.1476
GANMcC	6.1798	29.4633	0.8767	0.6752	66.0370	0.9088	0.2923
PIAFusion	6.6460	49.8703	0.9549	1.0224	63.6291	0.9255	0.6550
DRF	5.9020	22.6209	0.7820	0.5675	65.2373	0.9060	0.1532
Ours	6.6545	45.1329	0.9554	0.9124	64.3021	0.9311	0.6821

4 结论

为了应对夜间道路环境中强光干扰对融合图像质量的影响,导致融合算法性能急剧下降的问题,本文提出了一种能够有效应对强光干扰的图像融合网络.边缘特征提取模块通过优化信息熵来保留关键特征.HLEA 模块解决了由于图像强光区域低信息熵与高像素值之间的冲突而导致的融合质量下降问题.EGHF 模块通过最小化无关噪声熵和最大化有用信息熵,实现了稳健的特征融合.实验结果表明,与现有的先进融合算法相比,所提出的算法在融合图像质量和后续高级视觉任务的性能方面具有显著优势,有效解决了夜间道路环境中强光干扰对图像融合性能的影响,为夜间车辆辅助驾驶等复合高级视觉任务提供了有力的技术支持.

参考文献 (References)

- [1] 安晓东,李亚丽,王芳.汽车驾驶辅助系统红外与可见光融合算法综述[J].*计算机工程与应用*,2022,58(19):64-75.
(An X D, Li Y L, Wang F. Overview of infrared and visible image fusion algorithms for automotive driving assistance system[J]. *Computer Engineering and Applications*, 2022, 58(19): 64-75.)
- [2] Han Y, Wang B Y, Guan T, et al. Research on road environmental sense method of intelligent vehicle based on tracking check[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(1): 1261-1275.
- [3] 巫震宇,王鹏,马猛.多传感器遥感图像数据融合研究[J].*空间科学学报*,2006,26(1):48-53.
(Wu Z Y, Wang P, Ma M. Study of multisensor image fusion technology[J]. *Chinese Journal of Space Science*, 2006, 26(1): 48-53.)
- [4] 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.
- [5] Liu J Y, Fan X, Jiang J, et al. Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(1): 105-119.
- [6] Kong W W, Lei Y, Zhao H X. Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization[J]. *Infrared Physics & Technology*, 2014, 67: 161-172.
- [7] Ma J Y, Chen C, Li C, et al. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. *Information Fusion*, 2016, 31: 100-109.
- [8] Liu Y, Liu S P, Wang Z F. A general framework for image fusion based on multi-scale transform and sparse representation[J]. *Information Fusion*, 2015, 24: 147-164.
- [9] Ren L, Pan Z B, Cao J Z, et al. Infrared and visible

- image fusion based on variational auto-encoder and infrared feature compensation[J]. *Infrared Physics & Technology*, 2021, 117: 103839.
- [10] Rao Y J, Wu D, Han M N, et al. AT-GAN: A generative adversarial network with attention and transition for infrared and visible image fusion[J]. *Information Fusion*, 2023, 92: 336-349.
- [11] 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.
- [12] Tang W, He F Z, Liu Y. TCCFusion: An infrared and visible image fusion method based on transformer and cross correlation[J]. *Pattern Recognition*, 2023, 137: 109295.
- [13] Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain[J]. *IEEE Trans Image Process*, 2012, 21(12): 4695-4708.
- [14] Tang L F, Yuan J T, Ma J Y. 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: 28-42.
- [15] Xian R H, Xiong X, Peng H, et al. Feature fusion method based on spiking neural convolutional network for edge detection[J]. *Pattern Recognition*, 2024, 147: 110112.
- [16] 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.
- [17] Luo X Q, Gao Y H, Wang A Q, et al. IFSepR: A general framework for image fusion based on separate representation learning[J]. *IEEE Transactions on Multimedia*, 2023, 25: 608-623.
- [18] Ma L, Ma T Y, Liu R S, et al. Toward fast, flexible, and robust low-light image enhancement[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 5627-5636.
- [19] Vig J, Belinkov Y. Analyzing the structure of attention in a transformer language model[J/OL]. 2019, arXiv: 1906.04284.
- [20] Han D C, Ye T Z, Han Y Z, et al. Agent attention: On the integration of softmax and linear attention[C]. *Proceeding of the European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2025: 124-140.
- [21] Shreyamsha Kumar B K. Image fusion based on pixel significance using cross bilateral filter[J]. *Signal, Image and Video Processing*, 2015, 9: 1193-1204.
- [22] Jose J, Gautam N, Tiwari M, et al. An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion[J]. *Biomedical Signal Processing and Control*, DOI: 10.1016/j.bspc.2021.102480.
- [23] Zhao Z X, Xu S, Zhang C X, et al. DIDFuse: Deep image decomposition for infrared and visible image fusion[J/OL]. 2020, arXiv: 2003.09210.
- [24] Li H, Wu X J, Durrani T. 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.
- [25] Ma J Y, Yu W, Liang P W, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. *Information Fusion*, 2019, 48: 11-26.
- [26] Ma J Y, Zhang H, Shao Z F, et al. GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, DOI: 10.1109/TIM.2020.3038013.
- [27] Xu H, Wang X Y, Ma J Y. DRF: Disentangled representation for visible and infrared image fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, DOI: 10.1109/TIM.2021.3056645.

作者简介

王爱侠 (1974-), 女, 讲师, 博士, 主要研究方向为智能信息处理、图像处理, E-mail: wangaixia@ise.neu.edu.cn;

胡傲杰 (2002-), 男, 硕士生, 主要研究方向为计算机视觉、图像处理, E-mail: horacegibbs309@gmail.com;

闫爱云 (1978-), 女, 讲师, 博士, 主要研究方向为计算机视觉、嵌入式系统设计, E-mail: yanaiyun@ise.neu.edu.cn;

高尚 (2001-), 男, 博士生, 主要研究方向为计算机视觉、嵌入式系统设计, E-mail: gaoshang@ise.neu.edu.cn;

金硕巍 (1980-), 男, 副教授, 博士, 主要研究方向为嵌入式系统设计、图像处理, E-mail: jinshuowei@ise.neu.edu.cn;

庞永恒 (1991-), 男, 讲师, 博士, 主要研究方向为深度学习、数据分析, E-mail: yanaiyun@ise.neu.edu.cn.