

基于扩散重建与小波频域感知的机载光电图像融合*

孔令兵¹ 陈 昕² 连思铭² 宋美琦² 闻映红¹ 刘少鹏²

(1. 北京交通大学自动化与智能学院 北京 100044; 2. 北京航天控制仪器研究所 北京 100039)

摘 要: 远距离侦察任务中多模态图像融合功能旨在生成兼具各模态显著特征且真实有效的融合图像。为解决远距离目标特征表达真假混杂、纹理结构失真的难题,提出一种扩散重建与小波频域感知融合架构 BalanceFusion,旨在提升远距离下红外与可见光图像融合的细节信息表征能力及真实性。该方法首先采用高效图像复原网络增强目标的特征表示能力;然后引入小波条件驱动融合模块,融合频域感知强调目标热辐射轮廓特性,降低因超分辨率过度聚焦于目标而忽略背景结构的影响;最后,逐层聚合模态差异显著的多尺度高频特征信息,确保多模态融合图像包含充分有效的细节信息,在增强特有特征的基础上保持共有特征的真实性和完整性。实验结果表明,算法在实飞数据上显著优于现有框架,结构保留与感知质量(Qabf)指标和空间频率(SF)指标分别提升33%和21%,展现了远距离多模态图像融合能力。

关键词: 红外可见光融合;小波频域感知;扩散重建

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510

Airborne electro-optical image fusion based on diffusion reconstruction and wavelet frequency-domain perception

Kong Lingbing¹ Chen Xin² Lian Siming² Song Meiqi² Wen Yinghong¹ Liu Shaopeng²

(1. School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China;

2. Beijing Institute of Aerospace Control Device, Beijing 100039, China)

Abstract: In long-range reconnaissance missions, multimodal image fusion aims to generate fused images that preserve salient features from each modality while maintaining photorealism. To address the challenges of ambiguous feature representation and structural-textural distortion in distant targets, this paper proposes BalanceFusion, a novel fusion architecture integrating diffusion-based reconstruction with wavelet-domain frequency-aware perception, specifically designed to enhance detail representation and fidelity in infrared and visible image fusion under long-range conditions. The proposed method first employs an efficient image restoration network to strengthen target feature representation. Subsequently, a wavelet-conditioned fusion module is introduced to incorporate frequency-domain awareness, emphasizing thermal radiation contours of targets while mitigating the adverse effects of super-resolution methods that overly prioritize foreground targets at the expense of background structural integrity. Finally, modality-specific high-frequency features exhibiting significant differences are hierarchically aggregated across multiple scales to ensure the fused image retains both sufficient discriminative details and authentic shared structures. Experimental results on real-flight datasets demonstrate that the proposed algorithm outperforms state-of-the-art methods, achieving a 33% improvement in the structure-preserving perceptual quality metric Qabf and 21% gain in spatial frequency (SF).

Keywords: infrared-visible image fusion; wavelet frequency-domain perception; diffusion reconstruction

0 引言

在远距离侦察与监视任务中,机载光电吊舱作为态势感知关键装备,其融合红外与可见光成像能力获取多模态情报信息对于提升复杂场景理解能力和任务执行成功率具有重要意义^[1]。然而,在数百公里乃至更远的观测距离下,成像系统面临严峻挑战:一方面,红外成像虽具备全天候、抗干扰能力强的优势,但受限于大气衰减、热辐射扩散及传感器分辨率等因素,目标的热特征严重退化,导致细节信息缺失、边缘模糊,难以支撑高精度识别^[2];另一方面,可见光成像在长焦状态下虽能保留丰富的纹理与结构细节,却极易受光照条件、云雾遮挡及大气湍流影响,稳定性与鲁棒性显著下降。因此,有效融合多模态图像,在保留可见光高分辨细节的同时,增强红外图像的目标语义与热特性,克服二者在空间对齐、信噪比失衡及特征互补机制等方面的固有矛盾,成为当前远距离多模态图像融合亟待突破的核心技术难点。

目前,多模态图像融合方法随深度学习的不断发展,分为传统方法^[3-5]和深度学习方法。其中传统方法多采用空间域融合或变换域融合的方法。Wang等^[6]提出基于多分辨率分析的自适应权重调制方法,动态计算多光谱图像的比例融合系数,实现高保真图像融合。Liu等^[7]构建基于上下文增强的多图像融合评估框架,揭示异构图像融合差异,增强融合结果的物理可解释性。Ma等^[8]引入基于滚动引导滤波和多尺度分解策略,增强融合图像细节信息表现力。传统融合方法为图像融合奠定了坚实理论基础,以任务需求为导向的应用中取得显著成效。然而,这类方法强依赖于手工设计的融合规则,缺乏对复杂场景的自适应能力,在动态多变的实际应用环境中往往难以满足日益增长的性能需求^[9-10]。

近年来,由于深度学习方法强大的特征表现能力在异构图像融合领域受到了广泛关注。基于卷积神经网络的融合方法和基于Transformer架构的图像融合方法逐渐成为主流^[11-13]。Zhang等^[14]提出一种基于跨模态条件可控视觉增强器,可在保证纹理细节信息的同时实现可见光亮度的连续可控。Zeng等^[15]则聚焦异构图像频谱特征,构建信息引导模块增强多模态信息感知能力。Chen等^[16]设计了自适应跨模态融合网络,设计动态可调门控选择模块,实现无需人工干预的自适应特征融合。Helvig等^[17]利用交叉注意力特征捕捉异构图像之间语义关联,构建了通用性融合框架。Hu等^[18]提出了一种融合交叉注意力机制与小波卷积的多光谱网络,动态调整各模态特征权重,增强融合图像的整体特征表现力。上述方法在通用异构图像融合领域均具备较强优势,而针对远距离复杂场景下异构图像特征极度不平衡场景则难以有效

发掘并协同利用不同模态间的互补特性,限制了融合图像的情报表征能力^[19]。

针对远距离侦察任务中目标甄别困难的问题,本文结合远距离侦察场景下红外与可见光成像的互补特性,提出了一种面向高保真超分辨率融合的BalanceFusion模型。该模型的核心创新在于通过“细节复原—特征对齐—频域驱动”三级协同机制,系统性解决远距离红外图像细节缺失、异构模态空间感知不一致以及融合结果保真度不足3大关键问题。

1 理论分析

目前大多数融合算法倾向通过多源信息强化融合图像的细节特征,但将其直接迁移至单一源图像信息时往往存在噪声、图像模糊与纹理丢失等退化问题。因此,本文提出基于扩散重建机制的小波融合网络,基于该架构的图像融合算法,既重现了图像真正的原始细节,又通过多源信息弥补了单一源图像在侦察任务中的短板。

1.1 总体架构

本文提出的BalanceFusion模型整体架构如图1和2所示。该模型旨在面向远距离侦察场景,充分利用红外与可见光成像的互补特性,实现高保真、高分辨率的多模态图像融合。其核心由3部分构成,基于扩散机制的红外图像重建模块、小波域多尺度分解模块,以及小波条件驱动的融合模块。

在训练阶段,扩散机制重建模块聚焦于对低质量远距离红外图像的高保真超分辨率重建。如图1所示,该过程以低分辨率红外图像及其配准的可见光图像作为双模态输入,采用两阶段训练策略,一阶段首先通过像素反重组层将空间信息嵌入通道维度以增强细节建模能力;然后由紧凑型图像复原先验提取网络(CPEN)提取结构化先验并生成潜在表示 Z ,为扩散过程提供语义引导;最后,在此基础上,结合动态变换注意力模块(DMTA)与扩散变分自编码器(DGFN),在恢复过程中将潜在表示 Z 融入特征图以增强细节重建。二阶段在扩散去噪迭代中利用条件向量与预训练潜变量联合优化重建路径,生成带噪声预估 Z ,然后执行完整反向往去噪过程,逐步恢复出细节丰富、噪声抑制的高分辨率红外图像。训练阶段引入基于大气散射模型的合成退化增强(包括高斯模糊、对比度衰减、热辐射扩散模拟),并在损失函数中加入一致性掩码引导,使模型聚焦物理可信区域,有效提升高空场景鲁棒性。该训练机制可有效克服远距离成像中的模糊与热噪声问题,为后续多模态融合奠定了高质量红外源基础。

在推理阶段,BalanceFusion模型实现红外与可见光图像的高保真深度融合,整体流程如图2(b)所示。首

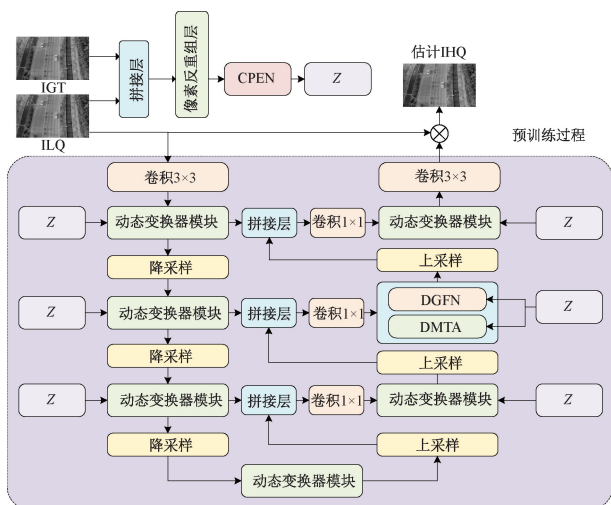


图1 BalanceFusion 模型一阶段训练过程

Fig.1 The Pretraining Process of BalanceFusion model

先,利用训练阶段获得的扩散重建子网络对输入的低分辨率红外图像进行超分辨率复原,生成细节增强、噪声抑制的高分辨率红外图像;然后,将复原图像与上采样对齐的可见光图像分别进行多级离散小波变换,分解为低频近似分量和多方向高频细节分量;在此基础上,通过小波条件驱动融合模块(WCFM),以可见光图像的结构信息为指导,依据梯度能量、结构显著性及目标物理特性自适应计算融合权重,对低频分量保留红外热辐射语义,对高频分量则选择性融合更具判别性的纹理细节,有效滤除阴影、热漂移等非目标伪影;最终,经小波逆变换重构出全分辨率融合图像。该推理流程在频域中实现了跨模态特征的多尺度协同聚合,显著提升了远距离侦察场景下目标的可辨识性与融合结果的空间一致性。

基于上述框架,BalanceFusion 模型能够在扩散过程中有效保留图像原始轮廓和细节纹理特征,在实际飞行

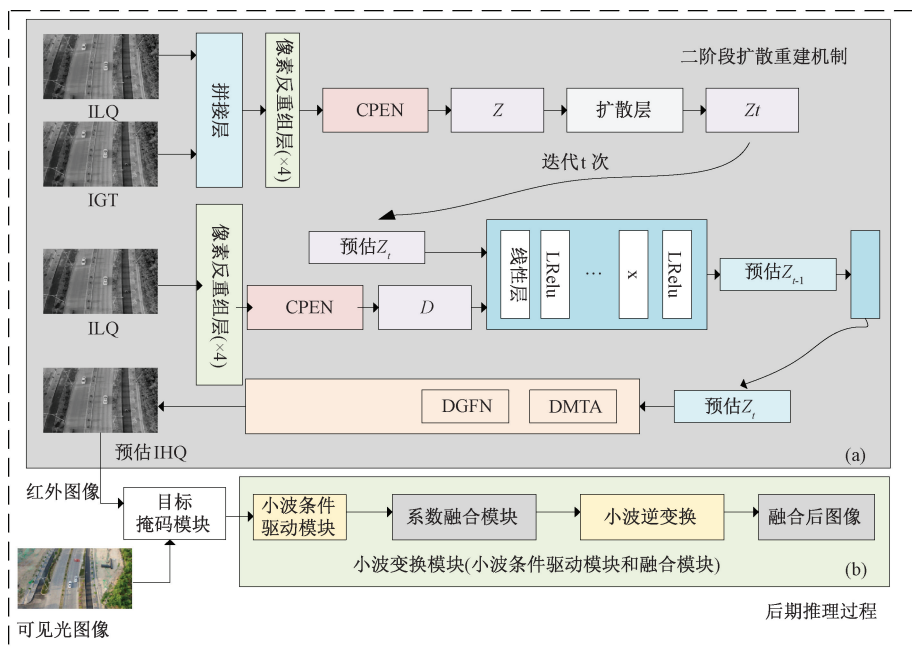


图2 BalanceFusion 模型二阶段训练及推理过程

Fig.2 The Training and Inferring Process of BalanceFusion model

数据集上进行测试,效果优于大多数算法。接下来,对各模块的具体细节进行介绍。

1.2 扩散重建机制

扩散重建机制主要由扩散图像恢复网络(diffusion model for image restoration, DiffIR)构成。扩散图像恢复网络主要有两个阶段,一阶段是预训练阶段,主要包含CPEN和动态传导机制;二阶段是扩散重建和推理阶段。其中CPEN具体结构主要由一层卷积层、激活层、残差网络、一层卷积层、平均池化层、线性层以及激活层组成;动态传导机制是由动态传导模块、卷积层、拼接层、下采样

层、动态门控前馈网络和动态多头注意力机制组成,主要作用是将重建细节的深度信息引入特征网络中,CPEN、DMTA和DGFN具体流程可参考文献[20]。

扩散重建机制首先将真实高清的红外数据与低分辨率的图像通过拼接层整合,然后通过(PixelUnshuffle)像素反重组层对拼接结果进行下采样后输入至CPEN,根据如下公式可得到潜变量Z的表达式:

$$Z = \text{CPEN}(\text{PixelUnshuffle}(\text{Concat}(I_{\text{GT}}, I_{\text{LQ}}))) \quad (1)$$

潜变量Z为动态门控前馈网络和动态多头注意力机制的动态调制参数。训练过程的损失函数如下:

$$L = \| I_{GT} - \hat{I}_{HQ} \|_1 \quad (2)$$

其中, $\| \cdot \|_1$ 代表 L_1 范数, 该函数主要通过真实图像与预估的高质量图像的误差来约束重建图像的质量。

扩散重建和推理阶段利用扩散过程得到迭代 t 次的潜变量 Z , 公式如下:

$$q(Z_t | Z) = N(Z_t; \sqrt{\alpha_t} Z, (1 - \bar{\alpha}_t) I) \quad (3)$$

式中: $\alpha_t = 1 - \beta_t$; $\bar{\alpha}_t = \prod_{i=0}^t \alpha_i$; β_i 代表预先设定的尺寸系数。对于逆过程, DiffIR 根据第 t 次得到的预估潜变量 $\hat{Z}$ 推演出第 $t-1$ 次的预估潜变量:

$$\hat{Z}_{t-1} = \frac{1}{\sqrt{\alpha_t}} (\hat{Z}_t - \varepsilon \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}}) \quad (4)$$

其中, ε 代表噪声, 该噪声服从高斯分布。因此, 根据上述流程可得最终的目标函数为:

$$L = \| I_{GT} - \hat{I}_{HQ} \|_1 + \frac{1}{4C'} \sum_{i=1}^{4C'} |\hat{Z}(i) - Z(i)| \quad (5)$$

式中: C' 为预估潜变量的通道数。

1.3 目标显著性掩码生成

为克服红外与可见光图像在语义表征与几何结构上的固有异构性, 本文提出一种跨模态联合特征分解策略, 构建目标导向的一致性掩码, 以甄别并抑制与真实目标物理特性不符的虚假细节 (如可见光中的阴影伪影或红外图像中的热漂移区域), 从而实现跨模态空间感知的一致性对齐。

设 $\hat{I}_{IR, HQ} \in R^{H \times W}$ 为经复原网络增强后的红外图像, $I_{VIS} \in R^{H \times W}$ 为原始可见光图像。首先, 分别从两模态提取梯度幅值、边缘响应与局部对比度显著图。

$$G_{IR} = \|\nabla \hat{I}_{IR, HQ}\|_2, G_{VIS} = \|\nabla I_{VIS}\|_2$$

$$E_{ir} = \text{canny}(\hat{I}_{IR, HQ}), E_{VIS} = \text{canny}(I_{VIS}) \quad (6)$$

$$S_{ir} = \text{std}_w(\hat{I}_{IR, HQ}), S_{VIS} = \text{std}_w(I_{VIS})$$

式中: ∇ 表示 Sobel 梯度算子; $\text{std}_w(\cdot)$ 为 7×7 窗口计算的局部标准差, 用以度量局部纹理活跃度。

在此基础上, 定义跨模态一致性响应为各特征在双模态下的共现度量。依据物理真实目标应在多特征维度上同时激发双模态响应的先验, 采用逐元素最小值操作融合双模态信息:

$$C_G = \min(G_{IR}, G_{VIS})$$

$$C_E = E_{IR} \odot E_{VIS} \quad (7)$$

$$C_S = \min(S_{IR}, S_{VIS})$$

该策略可有效抑制单模态特有伪影: 例如, 可见光中由光照变化引起的阴影区域虽在 G_{VIS} 或 S_{VIS} 中显著, 但因缺乏红外对应响应而被 C_G 与 C_S 削弱; 同理, 红外传感器热漂移产生的孤立热点因无可见光边缘或梯度支撑, 在 C_E 与 C_G 中亦被抑制。最终, 将 3 类一致性图线性融合并

经非线性增强, 生成目标导向掩码:

$$M' = C_G + C_E + C_S \quad (8)$$

$$M = \sigma(k(M' - \mu_{M'}))$$

式中: $\mu_{M'}$ 为 M' 的均值; $k=5$ 控制 sigmoid 函数的陡峭程度; $\sigma(\cdot)$ 为标准 sigmoid 激活函数。所得掩码 $M \in [0, 1]^{H \times W}$ 高亮跨模态几何与语义一致的区域, 为后续优化提供物理可信的空间权重。

1.4 小波变换模块

由于扩散机制在不同时间步对频域信息关注存在差异, 为了能够充分利用频域信息融合红外图像和可见光图像, 本文引入小波分解, 将输入特征转化为多尺度频域特征, 实现多频率解耦, 保证频域信息提取的空间一致性。

输入经扩散机制增强红外图像 $x \in R^{m \times n}$, 通过构建低通滤波器 $h[k] = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ 和高通滤波器 $g[k] = [\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$ 对输入红外图像 x 进行变换, 行变换可得:

$$a(k) = \sum_n x[n] \cdot h[2k - n] = (x \times h[-n])[2k]$$

$$d(k) = \sum_n x[n] \cdot g[2k - n] = (x \times g[-n])[2k] \quad (9)$$

其中, $2k$ 代表下采样操作, 对列进行同样操作可得到 4 种小波子带, 这 4 种小波子带分别代表近似图像、垂直边缘、水平边缘和对角线细节。同理可得, 可见光的 4 种小波子带。

通过对输入红外图片的小波子带和可见光图片的小波子带分别进行拼接并编码, 编码后特征经过双三次插值下采样, 使得特征图维度一致, 具体流程如图 3 所示。

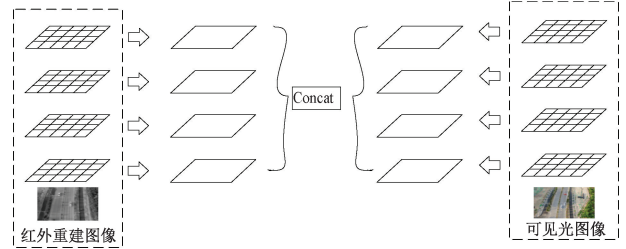


图 3 小波分解示意图

Fig. 3 The structure of wavelet model

在小波与神经网络特征的交互过程中, 红外与可见光的小波子带(LL、LH、HL、HH)分别同尺度神经网络特征图进行通道拼接, 再通过 1×1 卷积实现自适应加权融合; 同时采用小波逆变换替代反卷积完成上采样, 避免棋盘格伪影与高频信息丢失。

为验证小波基函数对融合性能的影响, 本文在自建

远距离机载测试集上对比了3种常用正交小波基,分别是 Haar、Daubechies-4 (Db4) 与 Symlets-4 (Sym4)。3种小波基处理后结构相似性 (SSIM)、平均梯度 (AG) 及推理耗时 (RTX 3090, 1 024×768 pixels) 平均值如表 1 所示,图 4 所示为 3 种小波基融合的部分结果。

表 1 小波基函数敏感性分析结果

Table 1 Sensitivity analysis of wavelet basis functions

小波基	SSIM	AG	推理时间/ms
Haar	0.915	8.72	120
Db4	0.913	8.95	162
Sym4	0.908	8.81	158



图 4 小波基融合结果图 (部分)

Fig. 4 Fusion results with different wavelet bases

从表 1 可以看出, Haar 小波在 SSIM 和计算效率上综合表现最优; Db4 小波仅在极弱纹理场景下 AG 提升 2.6%, 但推理耗时增加 35%, 难以满足机载平台实时性要求, 因此本文最终选择 Haar 小波作为默认基函数。

1.5 融合模块

依据小波模块, 得到可见光和红外图像编码结果, 将编码结果划分为多个子网格, 计算每个网格下幅值和平均幅值, 低于平均幅值的视作低频分量, 高于平均幅值的视作高频分量。考虑到以往研究中分层策略的优越性^[21], 采用拉普拉斯金字塔分层融合策略分别对高频分量和低频分量进行加权融合。针对不同分量特性, 对小波子带编码结果进行融合。

首先将各小波子带结果分解为 N 层拉普拉斯金字塔, 将各小波子带分解结果在每层进行加权融合, K 层的加权融合公式如下:

$$f_k(pi) = \sum_{i=1}^4 f_{k,i}(pi) \times G_{k,i}(pi) / \sum_{i=1}^4 G_{k,i}(pi) \quad (10)$$

式中: $f_k(pi)$ 表示第 i 幅特征图像在金字塔第 K 层的像素 pi 的值; $G_{k,i}(pi)$ 代表第 i 幅特征图像在金字塔第 K 层的高斯滤波图, 得到拉普拉斯金字塔分层融合的重构图像。金字塔分解层数为 $K=4$, 利用小波变换融合频域细节特征, 在保持轮廓结构的同时, 有效提升细节纹理特征, 使得融合结果在多尺度下得到有效增强。

1.6 损失函数

基于一致性掩码, 本文构建加权复合损失函数, 引导

整体模型在可信区域实现高保真融合, 同时避免对虚假细节的过度拟合。总损失定义为:

$$L_{total} = \lambda_1 L_{pixel} + \lambda_2 L_{ssim} + \lambda_3 L_{grad} \quad (11)$$

其中各项损失均以掩码 M 作为空间调制因子。

1) 一致性区域像素损失

约束融合图像 I_{fused} 在可信区域与可见光图像的亮度一致性:

$$L_{pixel} = \| M \odot (I_{fused} - I_{VIS}) \|_1 \quad (12)$$

2) 一致性区域结构相似性损失

保持高可信区域的结构完整性, 避免感知失真:

$$L_{ssim} = \| M \odot (1 - SSIM(I_{fused}, I_{VIS})) \|_1 \quad (13)$$

3) 一致性区域梯度损失

强化真实目标边缘的锐利度, 确保几何对齐精度:

$$L_{grad} = \sum_{d \in \{x, y\}} \| M \odot (\nabla_d I_{fused} - \nabla_d I_{VIS}) \|_1 \quad (14)$$

通过式 (12) ~ (14) 的联合优化, 模型在训练过程中自适应聚焦于红外与可见光共同支持的物理真实区域, 有效实现“保留红外热语义、继承可见光结构、剔除单模态伪影”的融合目标, 达成跨模态空间感知的一致性对齐。

2 实验验证

本文实验在 NVIDIA GeForce RTX 3090 上进行, 网络模型基于 PyTorch 框架进行实现, 优化器采用 Adam 算法等权重配置, 使得结构保真、空间频率及感知质量取得最优平衡。训练过程中设置 batch size 为 2, epoch 总数为 50。从训练结果观察, 损失函数在第 50 个 epoch 附近趋于平稳, 表明模型已基本收敛, 具备较优的融合能力和稳定性。

2.1 实验数据集和定量评价指标说明

为验证本文算法的有效性, 通过机载光电吊舱采集实飞数据构建实飞数据集, 在实飞数据上进行了定性和定量实验。

为全面评估 BalanceFusion 模型在非均衡分辨率可见光和红外图像融合任务的有效性和先进性, 选取近年来具有代表性的 9 种异构图像融合方法进行对比实验。9 种异构图像融合方法从不同技术路线出发的, 较为全面覆盖传统算法和深度学习的多个研究方向, 可有效地反映当前异构图像融合技术的发展趋势和性能水平, 包括梯度迁移方法 (GTF)^[22]、基于 AE 方法 DenseFuse 方法^[23]、基于卷积神经网络 (CNN) 的 U2Fusion^[24] 和 CFNet 方法^[25]、基于生成对抗网络 (GAN) 的 FreqGAN 方法^[26]、基于 Transformer 的 DATFuse 方法^[27]、CS2Fusion 方法^[28]、基于记忆单元的 MemoryFusion 方法^[29] 以及与 Mamba-UNet 和小波变换相结合的 F2Fusion 方法^[30]。

选用 12 个指标来客观评价融合图像的性能结果,包括信息熵(EN)、视觉保真度(VIF)、AG、SSIM、空间频率(SF)等。

2.2 对比实验及分析

为评估设计模型的表现性能,本文在真实飞行场景中采集的 8 组红外与可见光图像对上进行了测试。

1) 实际飞行数据定性比较

本文实验对多种红外与可见光图像融合方法进行了定性分析,旨在评估所提出方法在细节保留、边缘清晰

度、纹理一致性以及整体视觉感知质量方面的综合表现。

如图 5 所示,对比了 GTF、DenseFuse、U2Fusion、CFNet、CS2Fusion、DATFuse、F2Fusion、FreqGAN、MemoryFusion 等主流融合算法与本文所提 BalanceFusion 模型的差异。从视觉效果来看,原始可见光图像具有丰富的纹理和结构信息,但缺乏热辐射特征;而红外图像则能突出热源分布,但在细节表现上较为模糊。理想的融合结果应同时保留可见光的精细结构与红外的热信息,且无明显伪影或失真。

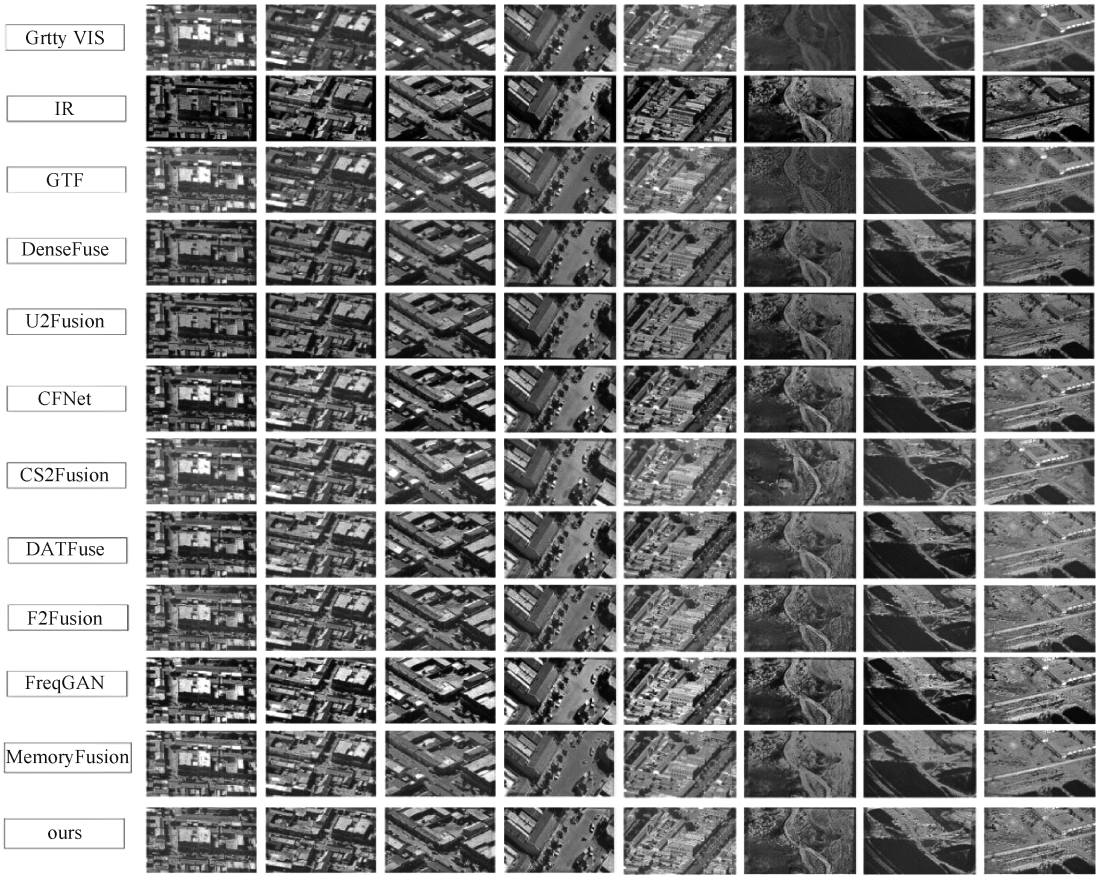


图 5 实飞数据融合结果对比

Fig. 5 Comparison of fusion results on real flight data

从图 5 可以看出,传统方法(如 GTF)能在一定程度上融合多源信息,但存在边缘模糊、细节丢失等问题,在建筑物轮廓和道路边界处表现不佳。深度学习基线模型(如基于 AE 方法 DenseFuse、基于 CNN 的 U2Fusion、CFNet、基于 Transfomer 的 DATFuse、CS2Fusion)在结构保持方面有所提升,但仍存在局部过曝光或欠融合现象,部分区域出现色彩不一致或纹理混叠。频域类方法(如 FreqGAN、F2Fusion)虽能较好地分离高低频成分,但在低频区域易产生平滑效应,导致细微纹理丢失,且在复杂场景下会出现振铃效应。MemoryFusion 方法虽然引入了记忆机制来增强特征表征的一致性,但在多尺度特征整合

过程中未能有效平衡模态间差异,导致部分区域出现特征混叠或失真。

本文提出的融合框架通过三阶段协同优化显著提升了融合性能。该方法不仅完整保留了可见光图像中的建筑轮廓、街道纹理等高分辨率结构信息,还融入了红外图像中的热目标(如车辆、建筑物热源),并在弱纹理区域表现出更强的细节一致性与自然过渡。尤其在复杂城市场景中,本文方法的融合结果边缘锐利、层次分明,既避免了过度增强带来的噪声放大,又未出现明显的亮度失衡或信息遮蔽。此外,在几何结构复杂的区域,展现出良好的空间连续性,说明其具备更强的语义感知能力和多

模态信息整合能力。

综上所述,本文提出的融合方法在视觉质量、结构保真度和信息完整性方面均优于现有主流技术。

2) 实际飞行数据定量分析

主要对所提出的融合方法与当前主流的多模态图像融合方法(包括 GTF、DenseFuse、U2Fusion、CFNet、CS2Fusion、DATFuse、F2Fusion、FreqGAN 和 MemoryFusion)在多个客观评价指标上进行了全面评估。

表 2 实飞数据融合评价指标比较

Table 2 Comparison of fusion evaluation metrics on real flight data

方法	EN	MI	SF	AG	SD	VIF	MSE	PSNR	Qabf	Nabf	SSIM	MS-SSIM
GTF	7.346	2.436	11.722	5.198	50.951	0.513	0.048	61.35	0.338	0.008	0.734	1.308
DenseFuse	7.269	3.049	12.75	5.324	50.262	0.686	0.025	64.083	0.455	0	0.93	1.443
U2Fusion	7.285	2.627	15.709	6.883	58.061	0.594	0.029	63.581	0.422	0.021	0.847	1.403
CFNet	7.544	3.143	14.279	5.629	61.577	0.652	0.036	62.579	0.31	0.021	0.709	1.316
CS2Fusion	7.239	0.808	12.559	4.622	50.729	0.088	0.089	58.723	0.076	0.034	0.074	0.181
DATFuse	7.479	3.993	19.055	6.932	57.26	0.708	0.035	62.681	0.464	0.03	0.931	1.363
F2Fusion	7.521	2.519	18.528	8.073	54.899	0.566	0.04	62.099	0.463	0.036	0.853	1.331
FreqGAN	7.652	2.82	18.157	7.317	73.427	0.612	0.042	61.856	0.409	0.037	0.86	1.357
MemoryFusion	7.55	2.567	16.393	7.155	56.997	0.559	0.042	61.892	0.488	0.016	0.813	1.346
本文	7.524	2.552	22.02	9.208	55.085	0.643	0.042	61.925	0.617	0.041	0.913	1.347

由于小波变换的多尺度分解能力, BalanceFusion 模型在 SF(22.02)和 AG(9.208)两项指标上均取得最优结果,分别比 FreqGAN(18.157)和 F2Fusion(8.073)方法高出约 21%和 14%。表明通过小波域对高频细节(如纹理、边缘)与低频语义进行分层建模与融合,能够有效增强融合图像的空间清晰度与细节表现力。

因此,本文引入扩散重建机制,在融合过程中通过逐步去噪的方式恢复高质量内容,在保持高 EN(7.524)和 MI(2.552)的同时,有效抑制伪影与信息失真。尽管均方误差(MSE)略高于 DenseFuse(0.025),但其峰值信噪比(PSNR)(61.925 dB)仍处于领先水平,说明重建误差分布更为合理,未牺牲整体图像保真度。

3) 应用性能验证

为验证所提出融合方法在实际复杂场景下的应用有效性,本文基于实飞数据开展目标检测性能评估。该测试集包含 8 组高分辨率航拍图像,覆盖城市典型道路区域,所有图像均标注车辆类目标(共 1 247 个),且仅包含车辆类别以聚焦于融合效果对单一目标类别的提升能力。采用 YOLOv8s 作为目标检测器,在可见光、红外及融合图像上分别进行车辆检测。检测过程中保持模型参数一致,仅输入图像来源不同,确保对比公平性。图 6 所示为 3 类图像上的检测结果对比,图 6(a)为可见光图像识别结果;图 6(b)为红外图像识别结果;图 6(c)为融合后图像识别结果。

从图 6 可以看出,可见光图像在白天条件下能较好地保留纹理细节,但在阴影区或低对比度区域存在漏检

实验结果充分验证了本文方法在融合性能上的优越性。

从表 2 可以看出,在结构保留与感知质量方面, BalanceFusion 模型在 Qabf(0.617)和 SSIM(0.913)等关键指标上显著优于现有方法,其中在指标 Qabf 上相比于 DATFuse 方法(0.464)提升约 33%,表明所提方法能更有效地保留源图像中的显著边缘与结构信息。同时, SSIM 接近 DenseFuse(0.93)和 DATFuse(0.931)的水平,说明融合图像具有优异的结构一致性与视觉自然度。

现象;红外图像虽对热源敏感,可有效识别部分隐蔽车辆,但整体分辨率较低,易受环境温度干扰导致误检;而经 BalanceFusion 融合后的图像综合了可见光的空间细节与红外的热特征信息,显著增强了小目标和弱对比度目标的可辨识度,尤其在密集建筑群与复杂阴影环境下表现出更优的鲁棒性。在相同检测阈值下, BalanceFusion 融合结果在 mAP@0.5 指标上相比原始可见光图像提升了 12.6%,相比原始红外图像提升了 9.3%,验证了融合策略在增强目标检测性能方面的有效性。这一结果说明,通过合理融合可见光与红外模态的信息,不仅弥补了单模态感知的局限性,还显著提升了远距离机载平台下车辆目标的检测精度与稳定性。

综上所述,所提方法通过结合扩散模型的高质量重建能力与小波变换的多层级特征融合机制,不仅在定量指标上全面超越现有方法,也为多模态图像融合任务提供了一种兼顾细节保留、结构一致性和感知真实性的新范式。

3 结 论

针对远距离机载侦察场景中可见光与红外图像融合存在的模态感知不一致、细节失真及结构模糊等问题,本文提出了一种基于扩散重建与小波频域感知的多模态融合框架—BalanceFusion。该方法通过如下 3 项改进实现融合性能的大幅提升:1)设计一种面向细节特征均衡的多维度特征分解策略,在融合前有效对齐可见光与红外

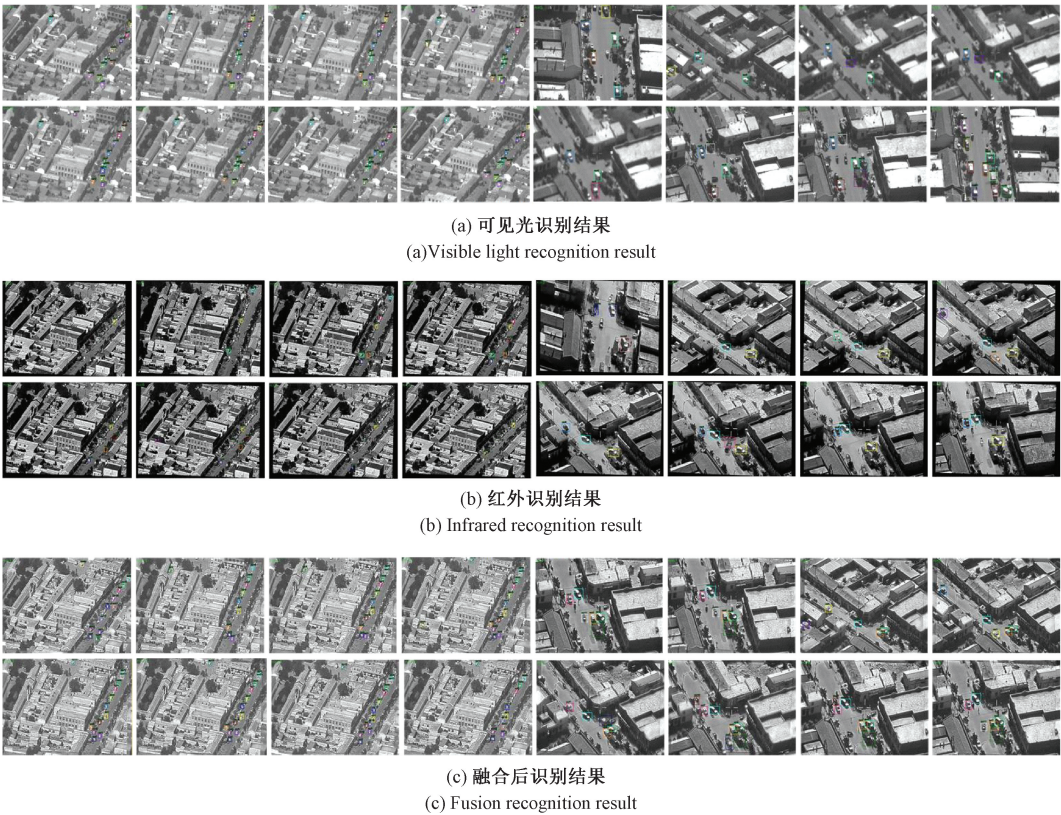


图 6 实飞数据识别结果(部分)

Fig. 6 Actual flight data recognition results (partial)

图像的空间结构表征,保障跨模态特征提取的真实性;2)引入扩散重建机制,在高效去噪的同时恢复图像退化前的高保真细节,显著提升输入特征的质量;3)构建小波条件驱动的频域融合模块,结合金字塔策略协同融合多尺度高频与低频分量,既强化了目标热辐射轮廓的表达,又有效保留了背景纹理与共有结构信息。

在真实飞行数据集上的大量实验表明,所提方法在主观视觉效果与客观评价指标上均显著优于现有主流算法。定量结果显示,结构-感知融合质量指标 Qabf 和 SF 分别提升 33% 和 21%,充分验证了模型在细节保留与结构真实性方面的优势。定性对比进一步表明,BalanceFusion 能够在复杂场景下生成边缘清晰、纹理丰富且热源特征突出的融合图像,显著提升侦察目标的可辨识性。本工作为远距离多模态光电图像融合提供了一种高保真、强鲁棒的技术路径,具有良好的工程应用潜力。

参考文献

[1] 王少敏. 空地观测条件下基于深度学习的目标检测算法研究与实现[D]. 成都: 电子科技大学, 2025.
WANG SH M. Research and implementation of deep learning-based object detection algorithms under air-ground observation conditions [D]. Chengdu: University

of Electronic Science and Technology of China, 2025.
[2] 闫钧华, 张琨, 施天俊, 等. 融合多层级特征的遥感图像地面弱小目标检测[J]. 仪器仪表学报, 2022, 43(3): 221-229.
YAN J H, ZHANG K, SHI T J, et al. Weak small target detection in remote sensing images based on multi-level feature fusion [J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 221-229.
[3] PANG H, ZHU M, GUO L. Multifocus color image fusion using quaternion wavelet transform [C]. 5th International Congress on Image and Signal Processing, 2012: 543-546.
[4] LI H, WU X J, KITTLER J. MDLatLRR: A novel decomposition method for infrared and visible image fusion [J]. IEEE Transactions on Image Processing, 2020, 29: 4733-4746.
[5] CHEN J, LI X, LUO L, et al. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition [J]. Information Sciences, 2020, 508: 64-78.
[6] WANG Z, ZIOU D, ARMENAKIS C, et al. A comparative analysis of image fusion methods [J]. IEEE Transactions on Geoscience and Remote Sensing, 2005,

- 43(6): 1391-1402.
- [7] LIU Z, BLASCH E, XUE Z, et al. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 94-109.
- [8] MA J, ZHOU Z, 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.
- [9] 陈思静, 付志涛, 李梓谦, 等. 基于自适应增强与显著性检测的可见光与红外图像融合算法[J]. *红外技术*, 2023, 45(9): 907-914.
- CHEN S J, FU ZH T, LI Z Q, et al. Visible and infrared image fusion algorithm based on adaptive enhancement and saliency detection [J]. *Infrared Technology*, 2023, 45(9): 907 - 914.
- [10] 王晓娜, 潘晴, 田妮莉. 基于 NSST-DWT-ICSAPCNN 的多模态图像融合算法 [J]. *红外技术*, 2022, 44(5): 497-503.
- WANG X N, PAN Q, TIAN N L. Multimodal image fusion algorithm based on NSST-DWT-ICSAPCNN [J]. *Infrared Technology*, 2022, 44(5): 497 - 503.
- [11] 杨叶君, 刘刚, 肖刚, 等. 基于自适应特征增强和生成器路径交互的红外与可见光图像融合[J]. *激光与光电子学进展*, 2023, 60(14): 1410018.
- YANG Y J, LIU G, XIAO G, et al. Infrared and visible image fusion based on adaptive feature enhancement and generator path interaction [J]. *Laser & Optoelectronics Progress*, 2023, 60(14): 1410018.
- [12] ZHANG X C, DEMIRIS Y. Visible and infrared image fusion using deep learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(8): 10535-10554.
- [13] TANG Y, ZHANG X, LI J, et al. CMTFusion: A cross-modal transformer architecture for efficient visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(8): 10234-10248.
- [14] ZHANG H, TANG L F, XIANG X Y, et al. Dispel darkness for better fusion: A controllable visual enhancer based on cross-modal conditional adversarial learning [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2024: 26477-26486.
- [15] ZENG Y Q, LIANG T F, JIN Y, et al. MMI-det: Exploring multi-modal integration for visible and infrared object detection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(11): 11198-11213.
- [16] CHEN X, XU S, HU S, et al. ACFNet: An adaptive cross-fusion network for infrared and visible image fusion [J]. *Pattern Recognition*, 2025, 159: 111098.
- [17] HELVIG K, ABELOOS B, TROUVÉ-PELOUX P. CAFF-DINO: Multi-spectral object detection transformers with cross-attention features fusion [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 3037-3046.
- [18] HU J, NI L, PENG B, et al. WtCAFNet: A wavelet transform and cross-attention modality-adaptive fusion network for multispectral object detection [J]. *Signal Processing*, 2025, 242: 110446.
- [19] LIU J Y, WU G Y, ZHU L, et al. Infrared and visible image fusion: From data compatibility to task adaption [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, 47(4): 2349-2369.
- [20] XIA B, ZHANG Y, WANG S, et al. Diffir: Efficient diffusion model for image restoration [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 13095-13105.
- [21] LI X L, LI L S, TAN T S, et al. UMCFuse: A unified multiple complex scenes infrared and visible image fusion framework [J]. *IEEE Transactions on Image Processing*, 2025, DOI: 10.1109/TIP.2025.3607623.
- [22] 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.
- [23] LI H, WU X J. DenseFuse: A fusion approach to infrared and visible images [J]. *IEEE Transactions on Image Processing*, 2018, 28.5: 2614-2623.
- [24] XU H, MA J, JIANG J, et al. U2Fusion: A unified unsupervised image fusion network [J]. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 2020, DOI: 10.1109/TPAMI.2020.3012548.
- [25] XING M L, LIU G, TANG H, et al. CFNet: An infrared and visible image compression fusion network [J]. *Pattern Recognition*, 2024, 156(10): 110774.
- [26] WANG Z S, ZHANG Z Q, QI W Q, et al. FreqGAN: Infrared and visible image fusion via unified frequency adversarial learning [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, DOI: 10.1109/TCSVT.2024.3460172.
- [27] TANG W, HE F, LIU Y, et al. DATFuse: Infrared and visible image fusion via dual attention transformer [J]. *IEEE Transactions on Circuits and Systems for Video*

Technology, 2023, DOI: 10. 1109/TCSVT. 2023. 3234340.

[28] WANG X, GUAN Z, QIAN W H, et al. Cs2fusion: Contrastive learning for self-supervised infrared and visible image fusion by estimating feature compensation map[J]. Information Fusion, 2024, 102: 102039.

[29] HE J C, LUO X Q, ZHANG Z C, et al. MemoryFusion: A novel architecture for infrared and visible image fusion based on memory unit[J]. Pattern Recognition, 2026, 170: 112004.

[30] LIU R H, LIU Y WANG H, et al. F2Fusion: Frequency feature fusion network for infrared and visible image via contourlet transform and Mamba-UNet [J]. IEEE Transactions on Instrumentation and Measurement, 2025, DOI:10. 1109/TIM. 2025. 3580829.

作者简介



孔令兵(通信作者),研究员,北京交通大学自动化与智能学院博士研究生,主要研究方向为图像处理、多源图像融合、目标跟踪等。

E-mail: feixiang-f@ 163. com

Kong Lingbing(Corresponding author) is a researcher and a Ph. D. candidate at the School of Automation and Intelligence, Beijing Jiaotong University. His main research interests include image processing, multi-source image fusion, object tracking, and other related areas.



陈昕,博士,现为北京航天控制仪器研究所高级工程师,主要研究方向为机载图像处理、深度学习、目标跟踪。

E-mail: cxzy_paper@ 163. com

Chen Xin, Ph. D. , is a senior engineer at the Beijing Institute of Aerospace Control Devices. Her main research interests include airborne image processing, deep learning, and object tracking.



宋美琦,硕士,现为北京航天控制仪器研究所工程师,主要研究方向为机载图像处理、深度学习、目标跟踪。

E-mail: songmq98@ 163. com

Song Meiqi, M. Sc. , is an engineer at the Beijing Institute of Aerospace Control Devices. Her main research interests include airborne image processing, deep learning, and object tracking.