



## 基于环境亮度动态感知的红外与可见光航拍图像融合方法

田枫, 韩卓翰, 刘芳, 张岩, 解红涛, 高帆, 黄彬, 刘宗堡

引用本文:

田枫, 韩卓翰, 刘芳, 等. 基于环境亮度动态感知的红外与可见光航拍图像融合方法[J]. 智能系统学报, 2026, 21(5): 1249–1259.

TIAN Feng, HAN Zhuohan, LIU Fang, et al. Infrared and visible aerial image fusion via dynamic environmental brightness perception[J]. CAAI transactions on intelligent systems, 2026, 21(5): 1249–1259.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202509028>

## 您可能感兴趣的其他文章

### 双向特征融合与注意力机制结合的目标检测

Target detection based on bidirectional feature fusion and an attention mechanism  
智能系统学报. 2021, 16(6): 1098–1105 <https://dx.doi.org/10.11992/tis.202012029>

### 多感知兴趣区域特征融合的图像识别方法

Image recognition method based on multi-perceptual interest region feature fusion  
智能系统学报. 2021, 16(2): 263–270 <https://dx.doi.org/10.11992/tis.201906032>

### 隔级融合特征金字塔与CornerNet相结合的小目标检测

Small target detection based on a combination of feature pyramid and CornerNet  
智能系统学报. 2021, 16(1): 108–116 <https://dx.doi.org/10.11992/tis.202004033>

### 多特征融合的异视角目标关联算法

Target association from different perspectives based on multi-feature fusion  
智能系统学报. 2020, 15(5): 847–855 <https://dx.doi.org/10.11992/tis.202006037>

### 一种基于级联神经网络的飞机检测方法

Cascade convolutional neural networks for airplane detection  
智能系统学报. 2020, 15(4): 697–704 <https://dx.doi.org/10.11992/tis.201908028>

### 基于宽度学习方法的多模态信息融合

Multi-modal information fusion based on broad learning method  
智能系统学报. 2019, 14(1): 150–157 <https://dx.doi.org/10.11992/tis.201803022>

DOI: 10.11992/tis.202509028

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20260706.1703.004>

# 基于环境亮度动态感知的红外与可见光 航拍图像融合方法

田枫<sup>1,2</sup>, 韩卓翰<sup>1,2</sup>, 刘芳<sup>1,2</sup>, 张岩<sup>1,2</sup>, 解红涛<sup>1,2</sup>, 高帆<sup>1,2</sup>, 黄彬<sup>1,2</sup>, 刘宗堡<sup>2</sup>

(1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318; 2. 黑龙江省石油大数据与智能分析重点实验室, 黑龙江 大庆 163318)

**摘要:** 为克服现有红外与可见光图像融合方法在无人机航拍环境下的亮度自适应融合及互补特征挖掘不充分的问题, 提出一种基于环境亮度动态感知的红外与可见光航拍图像融合方法。设计环境亮度感知模块, 实现可见光特征的动态重组, 减轻环境亮度变化对网络学习的干扰。构建高低频信息协同注意力, 增强模型对红外图像边缘细节特征的捕捉能力, 突出融合图像中的显著目标。设计离散小波交叉注意力块, 通过让两种模态交叉关注, 实现红外与可见光图像互补特征的高效融合。在公开数据集上使用多种评价指标对所提模型进行实验验证, 实验结果表明, 所提方法能够在无人机航拍环境下实现红外与可见光图像亮度自适应融合, 并在定性和定量评价上取得了较好的融合效果, 此外, 通过实验证明, 所提方法生成的融合图像能有效适用于无人机对地目标检测任务。

**关键词:** 图像融合; 无人机; 环境亮度自适应; 红外图像; 高频信息增强; 特征融合; 交叉注意力; 目标检测

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2026)05-1249-11

中文引用格式: 田枫, 韩卓翰, 刘芳, 等. 基于环境亮度动态感知的红外与可见光航拍图像融合方法 [J]. 智能系统学报, 2026, 21(5): 1249-1259.

英文引用格式: TIAN Feng, HAN Zhuohan, LIU Fang, et al. Infrared and visible aerial image fusion via dynamic environmental brightness perception [J]. CAAI transactions on intelligent systems, 2026, 21(5): 1249-1259.

## Infrared and visible aerial image fusion via dynamic environmental brightness perception

TIAN Feng<sup>1,2</sup>, HAN Zhuohan<sup>1,2</sup>, LIU Fang<sup>1,2</sup>, ZHANG Yan<sup>1,2</sup>, XIE Hongtao<sup>1,2</sup>,  
GAO Fan<sup>1,2</sup>, HUANG Bin<sup>1,2</sup>, LIU Zongbao<sup>2</sup>

(1. School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China; 2. Heilongjiang Provincial Key Laboratory of Oil Big Data &amp; Intelligent Analysis, Daqing 163318, China)

**Abstract:** To address the issues of inadequate brightness-adaptive fusion and insufficient mining of complementary features in existing infrared and visible image fusion methods under UAV aerial conditions, this paper proposes a fusion method based on dynamic environmental brightness perception. First, an environmental brightness perception module is designed to dynamically reorganize visible features, reducing the interference caused by varying illumination conditions on network learning. Second, a cooperative high- and low-frequency attention mechanism is constructed to enhance the model's ability to capture edge details from infrared images and highlight salient targets in the fused results. Finally, a discrete wavelet cross-attention block is introduced to facilitate efficient fusion of complementary features through cross-modal interaction. Extensive experiments on public datasets with multiple evaluation metrics demonstrate that the proposed method achieves adaptive brightness fusion of infrared and visible aerial images under UAV scenarios, yielding superior qualitative and quantitative results. Moreover, it is verified that the fused images generated by our method can effectively enhance ground target detection in UAV applications.

**Keywords:** image fusion; unmanned aerial vehicles; environmental brightness adaptation; infrared image; high-frequency information enhancement; feature fusion; cross attention; object detection

收稿日期: 2025-09-17. 网络出版日期: 2026-07-07.

基金项目: 黑龙江省科技创新基地项目 (JD24A009); 东北石油大学特色科研团队项目 (2023TSTD-04).

通信作者: 刘芳. E-mail: [lfliufang1983@126.com](mailto:lfliufang1983@126.com).

随着无人机技术的快速发展, 利用其获取特定场景全面信息的需求变得愈发重要, 然而, 仅配备可见光或红外传感器的无人机只能获取单方

面的信息,无法提供对环境的全面理解。因此,红外与可见光图像融合技术应运而生,作为多模态图像融合的重要分支,红外与可见光图像融合的重点是提取、融合并重建来自不同传感器的跨模态特征<sup>[1]</sup>。

可见光图像具有分辨率高、纹理细节丰富等优势,符合人眼视觉,但成像质量易受光照和天气影响,导致纹理弱化,进而影响图像融合算法准确性。红外图像通过感知物体热辐射获取信息,受环境亮度制约程度很小;然而,红外传感器通常像元尺寸较大,致使分辨率较低且高频细节缺乏<sup>[2]</sup>。因此,融合红外与可见光图像的目标在于整合可见光图像蕴含的纹理细节与红外图像中的热辐射目标信息,生成信息更全面、视觉效果更佳的融合图像。如何设计先进的融合方法,已成为过去十几年的持续研究热点。

现有红外与可见光图像融合方法可分为传统方法和深度学习方法两大类<sup>[3]</sup>。传统方法根据采用的理论可分为基于稀疏字典训练的方法<sup>[4]</sup>、基于多尺度变换的方法<sup>[5]</sup>、基于低秩表示学习的方法<sup>[6]</sup>、基于分割或显著性的方法<sup>[7]</sup>,以及混合建模方法<sup>[8]</sup>。传统方法的融合结果在目标辨识度方面有一定提升,但融合规则常采用手工方法设计,难以满足复杂环境融合需求。目前,基于深度学习的红外与可见光图像融合方法成为主流,该方法主要根据网络结构的不同进行分类<sup>[9]</sup>。例如基于自编码器(autoencoder, AE)的融合方法 DenseFuse<sup>[10]</sup>、基于卷积神经网络(convolutional neural network, CNN)的融合方法 IFCNN<sup>[11]</sup>、基于 Transformer 的融合方法 CrossFuse<sup>[12]</sup>以及基于生成对抗网络(generative adversarial network, GAN)的融合方法 TC-GAN(texture conditional generative adversarial network)<sup>[13]</sup>等。这类方法在评价指标以及特征提取能力上优于传统方法,但由于卷积感受野的限制、特征融合方式的单一,进而忽视多模态特征互补性。此外,红外图像通常分辨率较低,自身边缘细节等高频信息较弱,上述方法常包含下采样以提取低频全局信息,这一过程又会主动损失高频细节,难以突出目标边缘信息。

近年来,涌现出一批新的红外与可见光图像融合思路,例如首个利用语义驱动的图像融合方法 SeAFusion<sup>[14]</sup>,该方法在融合网络后级联语义分割网络,通过梯度反向传播为融合网络提供语义需求反馈;同样, Liu 等<sup>[15]</sup>用目标检测网络代替分割网络,提出 TarDAL(target-aware dual adversarial learning),从目标检测的角度约束融合网络保留丰富的语义信息。Gao 等<sup>[16]</sup>提出一种新的多层次跨

模态融合网络用于无人机航拍图像融合,该方法有效提高了后续无人机目标检测的性能。Li 等<sup>[17]</sup>提出了 CMF(confidence measurement fusion)模块,通过评估红外与可见光模态间的交互置信度,刻画模态特征在当前环境下的可靠性,并据此自适应地调整多模态特征的融合比例,为动态融合提供了重要启示。Tang 等<sup>[18]</sup>通过弱光增强的方式解决昏暗环境下的图像融合。然而,上述方法大都选择性提取亮度稳定状态下的目标与纹理细节特征进行图像融合,未考虑光照波动状态下的亮度自适应图像融合。程一航等<sup>[2]</sup>通过亮度解耦等方法,有效解决了亮度自适应图像融合问题,但将可见光图像转换为 HSV(Hue, Saturation, Value)空间可能会导致信息丢失,缺乏对可见光特征的动态处理。王爱侠等<sup>[9]</sup>通过光感知计算与红外特征分层引导,提出一种夜间恶劣光照下的自适应图像融合方法,但未考虑环境亮度对模态特征质量的影响。Cao 等<sup>[19]</sup>提出一种包含局部到全局专家的亮度自适应融合网络 MoE-Fusion,为图像动态融合领域提供了新的解决方案,但专家数量为经验性选择,会影响模型在不同场景下的泛化能力。

为解决上述问题,本文提出一种基于环境亮度感知的红外与可见光航拍图像融合方法。主要工作体现在以下 4 个方面:

1) 针对受亮度变化影响极大的可见光图像特征,设计了环境亮度感知模块(environment brightness perception module, EBPM)。对可见光特征进行亮度相关通道归一化,结合模态特征质量评估,重组可见光特征通道,消除亮度变化对可见光特征与模型学习的干扰。

2) 为增强模型对红外图像的特征提取能力,设计高低频信息协同注意力(high and low frequency information coordinates attention, HLIC)。在有效提取其边缘特征信息的同时,保留低频全局信息,使融合结果突出显著目标。

3) 为充分挖掘并利用红外与可见光成像互补性,构建离散小波交叉注意力块(discrete wavelet cross-attention block, DWTC)。通过让两种模态交叉学习,结合离散小波公共查询,实现红外与可见光特征的高效融合。

4) 通过实验验证,本文方法的融合结果在环境亮度变化下稳定保留了纹理与边缘细节,并在定性和定量评价上取得了较好的融合效果,实现了在无人机航拍环境下的亮度自适应融合,此外,本文还通过下游目标检测实验证明,所提方

法生成的融合图像能有效适用于无人机对地目标检测任务。

## 1 本文方法

本章首先介绍网络模型的整体结构。随后将详细解释所提出的3个核心模块:环境亮度感知模块、高低频信息协同注意力和离散小波交叉注

意力块。在本章最后,将阐述网络训练涉及的损失函数。

### 1.1 网络结构概述

为应对无人机航拍环境下的亮度自适应图像融合,本文提出了基于环境亮度动态感知的红外与可见光航拍图像融合网络。网络的总体框架如图1所示。

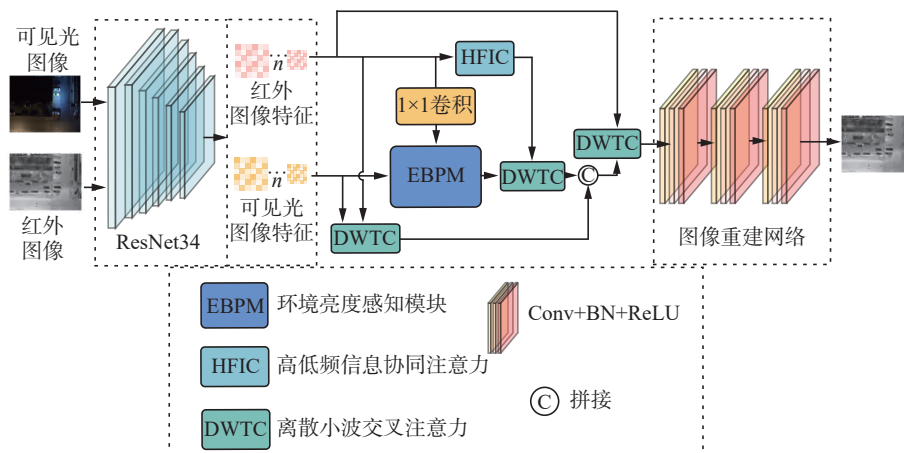


图1 基于环境亮度动态感知的红外与可见光航拍图像融合网络

Fig. 1 Infrared and visible aerial image fusion network based on ambient brightness perception

对于已配准好的红外与可见光图像对,利用ResNet34<sup>[20]</sup>骨干网络对其进行编码,得到不同层级的红外与可见光模态特征集,将包含丰富语义信息与纹理细节的多层级特征进行环境亮度自适应调整以应对环境亮度干扰。首先,衡量当前环境下可见光特征在空间维度上的亮度相关成分,评估源可见光特征质量,基于此进行通道二元化筛选,最终实现可见光特征的重组,减轻环境亮度对可见光图像的干扰,实现网络的环境亮度感知。其次,为增强网络对红外图像的高频细节捕捉能力,同时保留低频信息,本文通过多维度Scharr核与多尺度空间金字塔(spatial pyramid pooling, SPP)<sup>[21]</sup>分别提取高低频特征,通过全局注意力实现二者的信息协同,实现对红外图像的有效特征提取。在特征融合方面,为充分挖掘利用多模态特征的互补信息,本文设计了离散小波交叉注意力块,对红外与可见光特征分别计算其键值对,通过离散小波变换与逆变换得到丰富的层级频域特征共用查询,并基于此指导红外与可见光特征交叉关注,实现高效的互补特征融合。在网络结构的最后,通过解码器完成融合图像的重建。

### 1.2 环境亮度感知模块

在无人机航拍场景中,环境亮度变化是影响红外与可见光图像融合稳定性的关键因素之一。

受光照条件影响,可见光图像在强光、逆光或昏暗环境下容易出现亮度波动、对比度下降以及纹理细节退化等问题,从而导致其特征质量在不同场景下存在较大不确定性。相比之下,红外图像基于目标热辐射成像,其结构信息对环境亮度变化不敏感,在不同光照条件下具有较为稳定的表现。因此,如何在融合过程中准确感知环境亮度对可见光特征的影响,并据此动态调整可见光特征的参与方式,是实现亮度自适应融合的关键问题。

为此,本文提出EBPM,其核心思想是:在红外特征的稳定参考约束下,评估当前环境亮度条件下可见光特征的可靠性,并对可见光特征通道进行动态重组,从而减弱亮度变化对融合过程的干扰。如图2所示,EBPM主要由3个阶段构成:可见光特征亮度归一化、跨模态特征质量差异评估及可见光特征动态重组。

首先,为降低环境亮度变化对可见光特征分布的直接影响,本文对可见光特征 $F_{vi}$ 进行通道级亮度归一化处理。具体而言,分别计算每个特征通道在空间维度上的均值 $\chi_{b,c}$ 与标准差 $\sigma_{b,c}$ ,计算公式分别为

$$\chi_{b,c} = \frac{1}{H \times W} \sum_{h,w} F_{vi}$$

$$\sigma_{b,c} = \sqrt{\frac{1}{H \times W} \sum_{h,w} (F_{vi} - \chi_{b,c})^2}$$

式中:  $b \in [1, B]$ ,  $c \in [1, C]$ ,  $h$  与  $w$  分别为特征高度维度索引与宽度维度索引,  $B$  与  $C$  分别代表样本数量与特征通道数。随后, 利用均值与标准差对可见光特征进行归一化, 得到一组亮度归一化后的特征  $F_{vi}^*$ , 公式表示为

$$F_{vi}^* = \frac{F_{vi} - \chi_{b,c}}{\sigma_{b,c}}$$

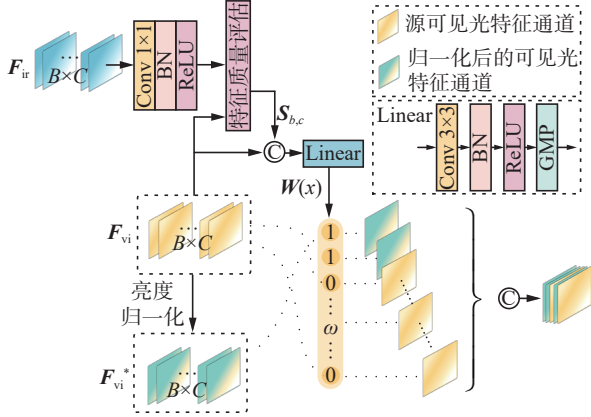


图 2 环境亮度感知模块结构

Fig. 2 Structure of ambient brightness perception module

经过该处理后, 可见光特征在不同环境亮度条件下具有更加稳定的分布形式, 更加侧重于反映结构与纹理信息, 而非亮度强弱本身。然而, 仅通过亮度归一化并不能判断当前环境下应优先保留原始可见光特征, 还是采用归一化后的特征。

因此, 受 CMF<sup>[17]</sup> 启发, 本文进一步引入红外特征  $F_{ir}$  作为亮度无关的参考信息, 对当前环境下 2 种模态特征的质量差异进行评估。具体地, 采用余弦相似度在通道级别衡量可见光特征与红外特征在空间结构分布上的一致性, 其公式化表达为

$$S_{b,c} = \frac{\sum_{h,w} F_{vi} \cdot F_{ir}}{\sqrt{\sum_{h,w} (F_{vi})^2} \cdot \sqrt{\sum_{h,w} (F_{ir})^2}}$$

式中  $S_{b,c} \in [0, 1]$ 。该计算结果可被视为当前环境

亮度条件下可见光特征质量的间接度量: 当环境亮度较高时, 此时可见光图像的纹理与结构信息较为完整,  $S_{b,c}$  值更高; 而在低亮度环境下, 可见光图像纹理退化,  $S_{b,c}$  值相应降低。

由于离散的  $S_{b,c}$  值难以直接用于稳定的通道选择, 本文进一步通过二值化对  $S_{b,c}$  结果进行二值化映射, 生成位于  $[0, 1]$  区间内的通道选择权重, 并据此实现可见光特征的动态重组, 公式表示为

$$\omega = W(F_{vi,s}, S_{b,c}) = \frac{F_{vi}^2 + \lambda \cdot S_{b,c}}{F_{vi}^2 + \lambda \cdot S_{b,c} + e}$$

式中:  $F_{vi,s}$  为经过卷积与池化后的中间特征,  $\lambda$  为权重系数,  $e$  为极小正数。当环境亮度较强,  $S_{b,c}$  值较高时, 二进制权重  $\omega$  更倾向于取 1, 此时模块倾向于选用亮度归一化后的可见光特征, 以消除亮度波动带来的干扰, 反之, 则选择亮度影响较弱的源可见光特征通道, 保留其稳定的结构和纹理信息。重组  $F_{vi}$  的过程可表示为

$$F_{vi} = \omega \odot F_{vi}^* + (1 - \omega) \odot F_{vi}$$

式中  $\odot$  为逐元素相乘操作。基于二进制通道选择的可见光特征动态重组不仅消除了可见光特征中与环境亮度相关的成分, 还通过动态调整特征通道的方式, 保留了源可见光图像中的结构表示, 确保无论是在亮度动态变化的环境中, 还是在较暗的环境中, 最终融合图像都能有效地保留关键的纹理和目标信息, 实现了网络的环境亮度自适应。

### 1.3 高低频信息协同注意力

为增强融合网络对红外特征高频细节捕捉能力并保留低频上下文信息, 突出融合图像中的显著目标信息, 本文提出了高低频信息协同注意力, 其结构如图 3 所示。具体来说, 对于输入的红外特征  $F_{ir} \in \mathbb{R}^{B \times C \times H \times W}$ , 首先通过  $1 \times 1$  卷积对其进行通道重组, 得到嵌入特征  $F'_{ir} \in \mathbb{R}^{B \times C' \times H \times W}$ , 以丰富原特征的表达维度。

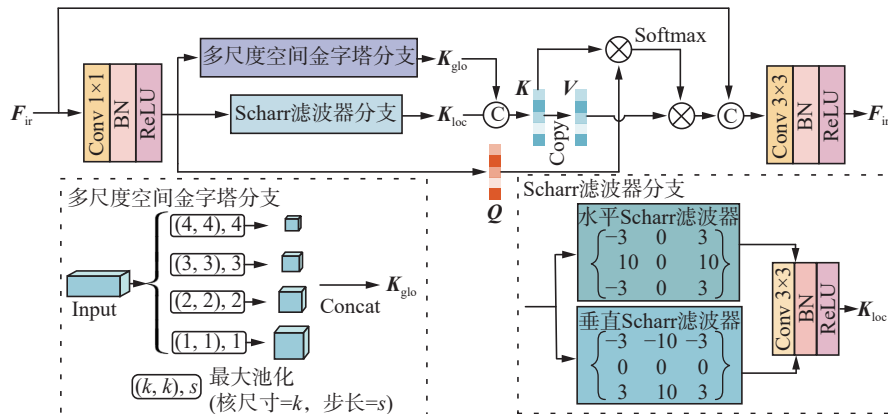


图 3 高低频协同注意力结构

Fig. 3 High and low frequency collaborative attention structure

在图3中,本文采用多尺度空间金字塔分支通过不同尺度的池化结构对 $F'_{ir}$ 进行聚合,该分支的输出作为特征的低频信息。多尺度空间金字塔分支主要用于提取目标的整体结构和上下文信息,对池化结果经过卷积以及双线性插值恢复,避免了单一尺度池化导致的低频信息丢失,有效保留全面的低频特征。整个过程可表示为

$$K_{glo} = \text{Concat}(\text{Interp}(\text{Pool}_i(F'_{ir})))_{i=1,2,3}$$

式中: $K_{glo} \in \mathbb{R}^{B \times C \times H \times W}$ 表示为拼接后的低频信息, $\text{Pool}_i(\cdot)$ 代表不同尺度的全局池化, $\text{Interp}(\cdot)$ 代表双线性插值恢复操作。另一方面,为弥补注意力机制对局部细节建模能力有限的问题,本文引入Scharr滤波器分支对特征进行提取。具体而言,Scharr滤波器分支由水平方向和垂直方向的Scharr滤波器构成,其核大小均为 $3 \times 3$ ,用于显式提取特征中的边缘和梯度信息。该设计能够在不引入额外可学习参数的前提下,有效增强目标轮廓与结构边界的表达能力,并提升网络对细节信息的感知能力。具体表示为

$$K_{loc} = \text{Conv} \left( \sqrt{(\text{Conv}(F_{ir}, S_x))^2 + (\text{Conv}(F_{ir}, S_y))^2} \right)$$

式中: $K_{loc} \in \mathbb{R}^{B \times C \times H \times W}$ 表示为融合后的高频信息, $S_x$ 和 $S_y$ 分别表示水平和垂直Scharr核。将高频信息 $K_{loc}$ 与低频信息 $K_{glo}$ 拼接后调整空间维度大小,与 $F_{ir}$ 进行注意力计算,以此强化红外特征内部的关联协同性,使网络同时具备全局结构建模能力与局部细节刻画能力,从而提高整体特征表示的鲁棒性,使融合结果蕴含丰富边缘信息,突出显著目标。最终将注意力图与源红外特征进行拼接并通过 $3 \times 3$ 卷积,得到高低频信息协同增强后的红外特征 $F'_{ir}$ ,其计算过程表示为

$$K = V = \text{Conv}(\text{Concat}(K_{loc}, K_{glo}))$$

$$F'_{ir} = \text{Conv}(\text{Concat}(\text{Softmax}(\text{Conv}(F_{ir}) \cdot K^T) V, F_{ir}))$$

#### 1.4 离散小波交叉注意力块

红外与可见光模态特征具备极强的互补性。本文就如何有效挖掘红外与可见光模态间互补信息特征,提出了离散小波交叉注意力块,其详细结构如图4所示。

受交叉注意力机制<sup>[22]</sup>的启发,本方法分别提取红外与可见光特征键值对,通过离散小波正逆变换得到公共查询,实现同时对2种差异特征的关联,基于此完成互补特征交叉学习。在交叉注意力机制中,公共查询向量 $Q$ 的质量直接影响其对不同模态特征的学习效果,进而影响融合结果的清晰度与稳定性。在传统方法中,公共查询 $Q$

通常直接由CNN或自编码器从特征中学习得到,该类方法的结构特性,会导致查询向量更偏向整体语义而忽略局部结构细节。本文通过离散小波变换与卷积联合构建公共查询,更好地保留源图像中的纹理与边缘信息,为后续的交叉学习提供了更稳定的源图特征参考。具体而言,对于输入的可见光特征 $F_{vi} \in \mathbb{R}^{B \times C \times H \times W}$ 和红外特征 $F_{ir} \in \mathbb{R}^{B \times C \times H \times W}$ ,经过多层卷积处理以初步保留整体的纹理细节与边缘信息特征。传统的并通过离散小波变换整合图像核心信息,得到丰富的层级频域特征,该过程可表示为

$$F_D = \text{DWT}(\text{Conv}(\text{Conv}(\text{Concat}(F_{vi}, F_{ir}))))$$

式中 $F_D$ 是经过离散小波变换得到丰富的多层级频域特征。随后,对其进行权重分配与小波逆变换,最后完成空间维度重塑,得到了富含源图像互补特征信息的查询结果 $Q$ ,该过程可表示为

$$Q = \text{Reshape}(\text{IWT}(\text{Conv}(\text{Conv}(F_D))))$$

式中 $Q \in \mathbb{R}^{B \times C \times (H \times W)}$ 。对于原始输入的可见光与红外特征,通过一系列通道特征融合与模态内聚合,得到各自键值对,并通过之前得到的查询结果,得到各自的注意关注图,该过程可表示为

$$K_j = V_j = \text{PointConv}(\text{DepthConv}(F_j))$$

$$A_j = \text{Softmax}(QK_j^T)$$

式中 $j$ 用于区分红外特征与可见光特征,当 $j=vi$ 时,此时 $K_j$ 、 $V_j$ 和 $A_j$ 分别表示为可见光特征的键值对和注意力图。随后,将模态各自的注意力图与值相乘,得到具有全局交互信息的特征,将该特征与另一模态原始特征进行残差拼接,得到互补特征 $F_{iratt}$ 与 $F_{viatt}$ ,最终通过拼接得到红外与可见光特征交叉融合特征 $F_{fusion}$ ,该过程表示为

$$F_{fusion} = \text{Conv}(\text{Concat}(F_{viatt}, F_{iratt}))$$

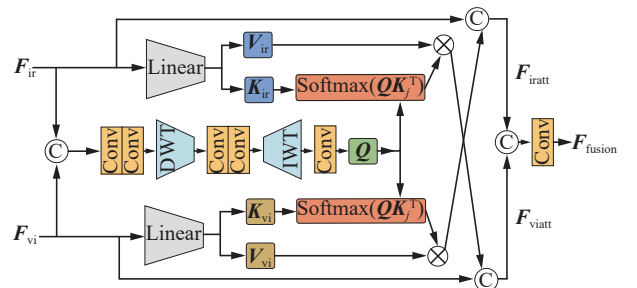


图4 离散小波交叉注意力块结构

Fig. 4 Discrete wavelet cross-attention block structure

通过这种特征交叉融合机制,充分提升了网络在特征融合阶段对2种模态特有信息的挖掘能力,实现红外与可见光特征的高效融合,使融合结果蕴含丰富纹理细节与热辐射信息。

### 1.5 损失函数

为提高融合网络在亮度变化环境下的定性主观感知与定量评估性能, 本文通过环境感知损失、细节纹理损失以及模态相关性损失形成联合损失函数, 指导融合网络生成高质量的环境自适应融合图像, 其表达式为

$$L_{\text{all}} = L_{\text{env}} + L_{\text{tex}} + L_{\text{sdmc}}$$

式中:  $L_{\text{env}}$  代表环境感知损失,  $L_{\text{tex}}$  为引入的细节纹理损失,  $L_{\text{sdmc}}$  为模态相关性损失。为了增强融合结果中环境的可理解性, 约束自适应环境感知门控, 本文定义环境感知损失  $L_{\text{env}}$  为

$$L_{\text{env}} = \frac{\|I_{\text{vi}} - I_{\text{ir}}\|_2}{\|I_{\text{vi}}\|_2 + \|I_{\text{ir}}\|_2 + e}$$

式中:  $I_{\text{vi}}$  和  $I_{\text{ir}}$  分别代表可见光与红外图像,  $\|\cdot\|_2$  表示  $L_2$  范数计算。为了使融合结果保留有效的纹理结构细节与边缘信息, 本文引入细节纹理损失  $L_{\text{tex}}$ <sup>[23]</sup>, 表达式为

$$L_{\text{tex}} = \frac{1}{HW} \|\nabla I_f - \max(|\nabla I_{\text{vi}}|, |\nabla I_{\text{ir}}|)\|_1$$

式中:  $I_f$  表示融合结果;  $\|\cdot\|_1$  表示  $L_1$  范数计算;  $|\cdot|$  表示取绝对值操作;  $\nabla$  代表图像的 Sobel 梯度算子, 计算图像在  $x$  和  $y$  各自方向上的梯度。为了突出 2 种模态与融合图像相关性的协同约束, 有效保留 2 种模态的互补信息, 本文定义了模态相关性损失  $L_{\text{sdmc}}$ , 其表达式为

$$L_{\text{sdmc}} = \alpha - (\text{corr}(I_f, I_{\text{ir}}) + \text{corr}(I_f, I_{\text{vi}}))$$

式中:  $\alpha$  为可学习参数;  $I_f$  表示融合结果;  $\text{corr}(\cdot)$  表示皮尔逊相关系数, 用来衡量融合结果与单模态源图像之间的线性相关度,  $\text{corr}(I_f, I_j)$  的具体计算过程为

$$\text{corr}(I_f, I_j) = \frac{\sum_1^n (I_f - \bar{I}_f)(I_j - \bar{I}_j)}{\sqrt{\sum_1^n (I_f - \bar{I}_f)^2} \sqrt{\sum_1^n (I_j - \bar{I}_j)^2}}$$

式中:  $I_j$  代表红外图像或可见光图像,  $\bar{I}_f$  代表融合图像向量均值,  $n$  为向量长度。

## 2 实验结果与分析

本章在公开数据集上评估所提方法, 并与主流先进的红外与可见光图像融合方法进行对比。首先, 说明并提供实验数据集、训练参数细节以及评估指标。其次, 本章从定性定量 2 个维度, 在无人机公开数据集上验证融合性能, 并进行了面向无人机对地任务的目标检测性能评估。最后, 为验证本文所提方法的有效性, 本文进行了

消融实验。

### 2.1 数据集与评价指标

#### 2.1.1 数据集

本文在无人机全时红外与可见光融合检测数据集 RGBT-Tiny<sup>[24]</sup> 上进行实验评估。RGBT-Tiny 数据集是由国防科技大学创建的大型红外-可见光双模态数据集, 该数据集包含 115 对序列、93 000 帧, 以及 120 万个手动标注, 涵盖 7 类目标和 8 种多样场景, 主要用于解决多变亮度下的可见光与热红外图像融合、检测和跟踪问题。数据集创建过程中, 使用了 DJI Mavic 2 无人机进行数据采集, 并通过相机校准和同质变换确保图像对齐。为了更客观地验证本文所提亮度动态融合的有效性, 后续的部分实验将测试数据集按场景亮度高低进行了划分。具体而言, 采用可见光图像的平均亮度作为划分依据: 将可见光图像转换为 HSV 颜色空间, 并计算其 V 通道的平均值作为亮度均值指标。根据该指标对测试样本进行排序, 并将其划分为低亮度与高亮度 2 类亮度场景。

#### 2.1.2 评估指标

本文从定性和定量 2 个方面评估各方法的性能。定性评估是一种主观评估方式, 其依赖于人的主观视觉感受, 好的融合结果应同时包含红外图像的显著对比度和可见光图像的丰富纹理; 定量评估则通过一些统计指标来客观评估融合性能。本文选用了 6 个在图像融合领域被广泛使用的定量指标, 如视觉信息保真度 (visual information fidelity, VIF)<sup>[25]</sup>、信息熵 (entropy, EN)<sup>[26]</sup>、标准差 (standard deviation, SD)<sup>[27]</sup>、空间频率 (spatial frequency, SF)<sup>[28]</sup>、特征互信息 (feature mutual information, FMI)<sup>[29]</sup>、差异信息和 (sum of correlation differences, SCD)<sup>[30]</sup>。VIF 值越高, 意味着融合结果越符合人类视觉感知; EN 值代表图像中包含的信息量, 值越高, 意味着融合图像包含丰富信息; SD 反应图像的对比度及分布, SD 值越高代表融合结果具有更好的对比度; SF 揭示融合图像的梯度信息, SF 值越高代表图像包含丰富的边缘和纹理细节; FMI 度量特征传输程度, FMI 值越高意味着更多的特征转移到了融合图像中; SCD 用于表征融合算法优劣, SCD 值越高代表融合图像包含源图像中的信息越丰富。

### 2.2 实验环境与参数设置

为了验证本文所提图像融合方法的性能, 在无人机全时红外与可见光融合数据集 RGBT-Tiny 上进行实验。所有实验均在 NVIDIA Ge-

Force RTX 3060 上实施。在网络训练过程中, 将输入图像归一化到  $[0,1]$  的范围, 应用随机梯度下降 (stochastic gradient descent, SGD) 来更新网络参数, 初始学习率设置为 0.001, 批处理大小设置为 16, 训练时间为 300 个 epoch。设置二值化函数中的  $\lambda=0.1$ ; 模态相关性损失  $L_{\text{sdmc}}$  中  $\alpha$  取 2。

### 2.3 对比实验

#### 2.3.1 定性评估

图 5 给出了本文方法在 RGBT-Tiny 数据集中 2 个典型图像对的直观定性评估结果以展示所提方法在亮度变换的航拍环境下的亮度自适应融合优势。图中每 2 行代表 1 种环境亮度状态。具体来说, 在第 1 种黄昏亮度环境状态下, 从图 5(e)、(c) 的蓝色框部分可以看出, IFCNN 等方法存在房屋纹理细节模糊等问题, 说明该融合方法对红外与可见光模态互补特征信息的提取学习能力较弱, 相比之下本文的融合结果呈现出更清晰的纹理与目标细节, 同时保留了一定的颜色细节, 保证融合图像的主观视觉质量, 如图 5(h) 中蓝色框部分。

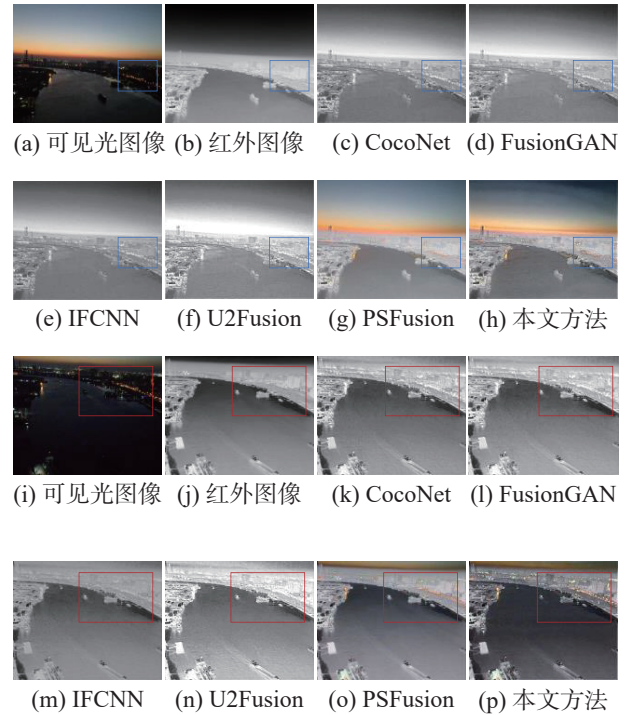


图 5 各种融合算法在 RGBT-Tiny 数据集上的定性结果  
Fig. 5 Qualitative results of various fusion algorithms on the RGBT-Tiny datasets

表 1 RGBT-Tiny 数据集上对比试验的定量结果

Table 1 Quantitative results of comparative experiments on the RGBT-Tiny dataset

| 模型                         | 年份   | EN           | SD            | VIF          | SF            | FMI          | SCD          |
|----------------------------|------|--------------|---------------|--------------|---------------|--------------|--------------|
| FusionGAN <sup>[31]</sup>  | 2018 | 5.785        | 40.530        | 0.632        | 7.075         | 1.047        | 1.419        |
| IFCNN <sup>[11]</sup>      | 2019 | 5.974        | 36.528        | 0.493        | 7.128         | 0.904        | 1.398        |
| U2Fusion <sup>[32]</sup>   | 2020 | 4.968        | 44.837        | 0.584        | 10.147        | 0.963        | 1.428        |
| Coconet <sup>[33]</sup>    | 2023 | 5.910        | 43.940        | 0.739        | 10.963        | 1.075        | 1.693        |
| PSFusion <sup>[23]</sup>   | 2023 | 6.539        | <b>46.816</b> | 0.954        | 11.267        | 1.167        | 1.668        |
| CoMofusion <sup>[34]</sup> | 2024 | 6.544        | 45.962        | 0.738        | 11.379        | 1.098        | 1.671        |
| CrossFuse <sup>[12]</sup>  | 2025 | 6.576        | 45.821        | 0.817        | 11.561        | 1.170        | 1.685        |
| 本文方法                       | —    | <b>6.647</b> | 46.367        | <b>0.983</b> | <b>12.493</b> | <b>1.239</b> | <b>1.713</b> |

注: 加粗表示该结果在本列效果最好。

在第 2 种同场景视角微变的昏暗亮度环境状态下, 可见光图像存在成像模糊且纹理信息不均匀等问题, 这使得大多数融合方法很难在此类场景中产生令人满意的结果, 本文提方法展现了基于环境亮度感知的融合优势。从图 5(l)、(m)、(o) 红色框部分可以看到, 在面对环境质量差的情况下, 该类图像融合方法对源可见光图像的纹理细节利用不够, 部分方法产生了过渡曝光问题。相比之下, 本文的自适应融合方法在差环境下依旧保留了红外与可见光图像的互补特征, 有效减轻了环境亮度对图像融合的干扰, 从而获得具有高信息的融合图像。

#### 2.3.2 定量评估

为了更加客观地评估本文所提方法在 RGBT-

Tiny 数据集上的性能, 同时防止个别融合图像特例对定量评估的影响, 本文选取了 30 张具有代表性的融合图像进行定量分析。所选样本的源可见光图像在环境亮度上具有较大差异, 能够体现不同环境亮度条件下融合方法的性能表现。与其他融合方法的各项评估指标对比结果如表 1 所示。可以看出本文所提方法在 EN、VIF、SF、FMI、SCD5 项指标上实现了最佳性能, 在 SD 方面获得了次佳性能, 这表明, 本文所提的融合方法有效保留了源红外与可见光图像各自的互补信息, 总之, 本文的方法在确保足够的热辐射信息的同时, 保留了丰富的目标与纹理细节。

#### 2.3.3 计算效率分析

考虑到本文方法面向无人机航拍应用场景,

对计算资源和实时性具有较高要求,本文进一步对近两年融合效果较为突出的融合方法进行计算效率对比分析。实验选取 RGBT-Tiny 数据集中环境亮度差异较大的 40 张图像对作为测试样本,评价指标分别为 FPS、Params 和 FLOPs,分别表示推理速度、模型参数量和浮点运算次数。所有方法均在相同硬件环境和输入条件下测试,评估结果如表 2 所示。

表 2 各模型在测试数据上的计算效率对比

Table 2 Comparison of computational efficiency of different models on test data

| 模型                         | FPS/(帧/s)    | FLOPs         | Params        |
|----------------------------|--------------|---------------|---------------|
| Coconet <sup>[25]</sup>    | 32.70        | 176.17        | 44.940        |
| PSFusion <sup>[23]</sup>   | 40.28        | 117.63        | 47.723        |
| CoMofusion <sup>[34]</sup> | 37.52        | 194.72        | 63.890        |
| CrossFuse <sup>[12]</sup>  | 42.73        | 329.82        | 51.970        |
| 本文方法                       | <b>43.39</b> | <b>115.79</b> | <b>42.637</b> |

注:加粗表示该结果在本列效果最好。

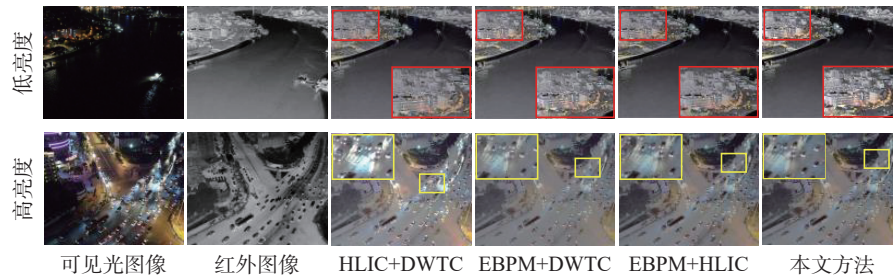


图 6 强亮度环境下消融实验定性对比结果

Fig. 6 Qualitative comparison results of ablation experiments

对比结果显示,由于源可见光图像受不同程度的亮度干扰,在缺少 EBPM 的情况下,融合图像缺乏纹理细节与语义边缘信息;在加入 EBPM 后,得益于通道级亮度动态选择,削弱了亮度对网络融合的影响,融合后的图片保留了稳定的源可见光纹理细节。同时,由于仅去除了 HLIC,第 4 列的融合结果存在车辆等红外目标边界不清晰问题;仅去除 DWTC 的情况下,第 5 列的融合结果虽然目标边缘特征相对明显,但出现了可见光信息对比度弱化问题,这说明此时网络未充分利用 2 种模态的互补信息进行特征融合。本文所提方法融合效果在 2 种亮度环境干扰情况下,整体保持了较高的边缘清晰度,在可见光纹理细节不均匀情况下展现了亮度自适应融合的稳定性和很好地保留了红外与可见光图像的互补特征,融合效果极佳,更符合人眼视觉感知。

为进一步客观验证本文所提各模块的有效性,仍使用对比实验中的客观指标进行定量分析,结果如表 3 所示。可以看出,相较于去除环境亮度感知模块,本文方法在各项指标均有提升,

实验结果表明,在保持较高融合性能的同时,本文方法在计算效率方面具有明显优势。具体而言,本文方法在 Params 和 FLOPs 2 个指标上均取得了最小值,表明模型整体结构更加紧凑、对硬件算力的需求更低;同时,其推理速度约为 43 帧/s,在所对比方法中表现最优。上述效率优势主要得益于本文对可见光特征采用通道级动态处理策略,避免了复杂的空间注意力建模,从而有效控制了整体计算复杂度。综上所述,在当前实验配置下,本文方法能够在连续视频流中实时完成融合推理,验证了其在实际无人机平台部署中的可行性与实用价值。

## 2.4 消融实验

为验证本文方法中提出的各模块的有效性,本文在 RGBT-Tiny 数据集上进行了消融实验。在高、低 2 种亮度环境下共设置 3 组实验,从定性和定量 2 种角度对本文所提模块有效性进行评估。定性评估结果如图 6 所示。

尤其是在 SD 方面提升了 5.44%,表明模块有效保留了当前光照不均匀环境下的纹理细节;相较于去除小波交叉注意力块,本文方法在 EN、SD、FMI 指标上分别明显提升了 0.457%、3.236%、0.449%,这说明该模块能够充分利用源图像的互补信息,显著提升融合图像的整体质量;相较于去除高低频信息协同注意力模块,本文方法在 SF 指标上显著提升了 2.717%。进一步证明了该模块能有效保留目标边缘细节,提升融合图像的梯度信息。

表 3 消融实验定量对比结果

Table 3 Quantitative comparison of ablation experiments

| 模型             | EN           | SD            | VIF          | SF           | FMI          | SCD          |
|----------------|--------------|---------------|--------------|--------------|--------------|--------------|
| HLIC+DWTC      | 5.695        | 39.394        | 0.910        | 7.315        | 1.258        | 1.385        |
| EBPM+DWTC      | 6.118        | 44.029        | 0.931        | 7.036        | 1.283        | 1.492        |
| EBPM+HLIC      | 5.830        | 41.598        | 0.916        | 9.180        | 0.922        | 1.487        |
| EBPM+HLIC+DWTC | <b>6.287</b> | <b>44.834</b> | <b>0.937</b> | <b>9.753</b> | <b>1.371</b> | <b>1.510</b> |

注:加粗表示该结果在本列效果最好。

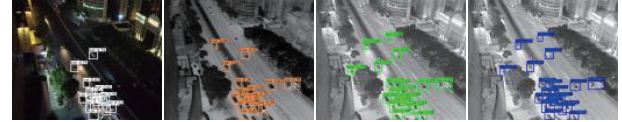
## 2.5 基于融合图像的目标检测实验

无人机对地检测作为一项通用且关键的高级视觉任务,检测性能能够很好地反映航拍融合图像中所蕴含的目标语义信息。为验证所提方法的融合结果能够有效支撑无人机对地检测任务,本文采用目前先进的预训练目标检测器 YOLOv12<sup>[35]</sup> 在 RGBT-Tiny 数据集上进行目标检测实验,该数据集不仅提供了配准的红外与可见光图像对,还包含多尺度的目标边界框标注信息。基于此,本节分别从定性与定量 2 个维度分别对源可见光图像、源红外图像和融合图像的目标检测效果进行分亮度场景评估。

### 2.5.1 定性评估

本文提供了可视化示例,从定性角度来说明所提方法的融合结果能够有效支撑无人机航拍目标检测任务,并选出相对具有代表性的场景检测结果,如图 7 所示。在该场景下,受光照条件干扰,检测器只能从可见光图像中检测出光照正常区域下的目标车辆,即便如此,仍受算法及环境影响呈现低目标置信度。从图 7(c)、(d) 可以看到,该方法的融合结果的目标边缘信息锐化不明显,干扰了检测器对显著目标的判别。相反,本文所提方法在保留了丰富纹理细节的情况下,有明显的高频信息,因此,检测器可以从融合结果中检测出相对较多的物体,并且具有所有方法中

的最高置信度。值得注意的是,在可视化结果中,包括本文方法在内的所有方法,均出现了一定程度上的漏检现象,这可能与本文选用的检测器性能有关。



(a) 可见光图像 (b) 红外图像 (c) CocoNet (d) FusionGAN



(e) IFCNN (f) U2Fusion (g) PSFusion (h) 本文方法

图 7 RGBT-Tiny 数据集中红外、可见光和融合图像的目标检测结果

Fig. 7 Object detection results of infrared, visible and fusion images in the RGBT-Tiny dataset

### 2.5.2 定量评估

本文还从定量角度证明了所提亮度动态融合后的图像能有效支撑亮度变化的无人机目标检测任务。将源红外图像、源可见光图像以及各融合方法生成的融合图像在分别输入 YOLOv12 检测器,按场景亮度高低划分为两组检测结果,每组亮度场景测试数据均为 65 张。评价指标均采用平均精度 (mean average precision, mAP) 作为性能评估指标。评估结果如表 4 所示 (其中 mAP@0.5 表示 IoU 阈值为 0.5 时的 mAP 值)。

表 4 高、低亮度场景下的目标检测性能  
Table 4 Object detection performance in low-light scenes

| 方法                         | 船            | 车            | 行人           | 公交车          | mAP@0.5      | 场景  |
|----------------------------|--------------|--------------|--------------|--------------|--------------|-----|
| Infrared                   | 0.493        | 0.539        | 0.490        | 0.479        | 0.500        | 低亮度 |
| Visible                    | 0.339        | 0.419        | 0.379        | 0.333        | 0.367        |     |
| IFCNN <sup>[11]</sup>      | 0.498        | 0.542        | 0.493        | 0.482        | 0.503        |     |
| FusionGAN <sup>[31]</sup>  | 0.496        | 0.536        | 0.502        | 0.480        | 0.503        |     |
| U2Fusion <sup>[32]</sup>   | 0.509        | 0.544        | 0.536        | 0.513        | 0.525        |     |
| CocoNet <sup>[33]</sup>    | 0.513        | 0.549        | 0.532        | 0.501        | 0.523        |     |
| PSFusion <sup>[23]</sup>   | 0.521        | 0.561        | 0.579        | 0.519        | 0.545        |     |
| CoMofusion <sup>[34]</sup> | 0.511        | 0.553        | 0.573        | 0.510        | 0.536        |     |
| CrossFuse <sup>[12]</sup>  | 0.514        | 0.557        | 0.580        | 0.513        | 0.541        |     |
| 本文方法                       | <b>0.542</b> | <b>0.577</b> | <b>0.601</b> | <b>0.673</b> | <b>0.598</b> |     |
| Infrared                   | 0.502        | 0.541        | 0.479        | 0.273        | 0.448        | 高亮度 |
| Visible                    | 0.487        | 0.402        | 0.258        | 0.280        | 0.356        |     |
| IFCNN <sup>[11]</sup>      | 0.519        | 0.546        | 0.481        | 0.291        | 0.459        |     |
| FusionGAN <sup>[31]</sup>  | 0.518        | 0.544        | 0.484        | 0.281        | 0.456        |     |
| U2Fusion <sup>[32]</sup>   | 0.525        | 0.550        | 0.483        | 0.274        | 0.458        |     |
| CocoNet <sup>[33]</sup>    | 0.519        | 0.549        | 0.498        | 0.279        | 0.461        |     |
| PSFusion <sup>[23]</sup>   | 0.541        | 0.555        | 0.549        | 0.391        | 0.509        |     |
| CoMofusion <sup>[34]</sup> | 0.538        | 0.548        | 0.519        | 0.313        | 0.479        |     |
| CrossFuse <sup>[12]</sup>  | 0.547        | 0.552        | 0.534        | 0.337        | 0.492        |     |
| 本文方法                       | <b>0.598</b> | <b>0.587</b> | <b>0.591</b> | <b>0.473</b> | <b>0.562</b> |     |

注:加粗表示该结果在本列效果最好。

从表 4 中可以看出, 基于本文融合图像的目标检测性能在各检测指标上均达到了最优水平。具体而言, 在 2 种亮度环境下, 源可见光图像目标信息弱化, 检测效果不佳, 与其他融合方法相比, 基于亮度动态融合图像的平均检测精度最高, 这说明所提方法更能充分融合红外与可见光的互补特征, 保留更多有用信息。值得注意的是, 本文所提方法在车类检测效果的领先并不突出, 但在具体的公交车类别的检测精度上远高于其他方法, 这说明所提方法更好地平衡了亮度干扰, 包含更丰富的可见光公交车语义特征。综上所述, 本文所提方法生成的融合图像能够有效提升航拍环境下的检测效果。

### 3 结束语

本文提出了一种基于环境亮度动态感知的红外与可见光航拍图像融合方法。该方法通过构建环境亮度感知模块, 有效减少了环境亮度对可见光图像的干扰; 同时, 通过高低频信息协同注意力提升了网络对红外高频特征的捕捉能力, 最后, 通过离散小波交叉注意力块充分挖掘了红外与可见光图像间的交互特征。实验结果表明, 本文方法在 RGBT-Tiny 数据集上的融合性能优于主流方法, 尤其在环境亮度变化的无人机航拍场景下, 仍能保持融合的稳定性, 实现了高质量的亮度自适应图像融合。同时, 在 RGBT-Tiny 数据集上的目标检测实验进一步证明, 本文方法的融合结果能够有效促进无人机对地目标检测任务, 其平均检测精度显著高于其他方法。消融实验验证了所提 3 个模块的有效性。然而, 由于数据集中包含大量小目标对象, 本文方法及其他对比方法的融合结果在检测时均存在一定程度的漏检现象, 未来工作将侧重于探索基于多模态融合的航拍小目标检测方向的研究, 并提升其在实时应用场景中的竞争力。

### 参考文献:

- [1] 赵佳, 辛月兰, 刘冀钊, 等. 脉冲耦合双对抗学习网络的红外与可见光图像融合[J]. *激光与光电子学进展*, 2025, 62(10): 455–466.  
ZHAO Jia, XIN Yuelan, LIU Jizhao, et al. Pulse-coupled dual adversarial learning network for infrared and visible image fusion[J]. *Laser & optoelectronics progress*, 2025, 62(10): 455–466.
- [2] 程一航, 乔正钰, 黄勇, 等. 基于 Retinex 理论的亮度自适应红外与可见光图像融合[J]. *激光与光电子学进展*, 2024, 61(20): 117–130.  
CHENG Yihang, QIAO Zhengyu, HUANG Yong, et al. Luminance-adaptive infrared and visible image fusion based on retinex theory[J]. *Laser & optoelectronics progress*, 2024, 61(20): 117–130.
- [3] 尹海涛, 周昌盛. 基于交叉融合 Transformer 的红外与可见光图像融合方法[J]. *激光与光电子学进展*, 2025, 62(6): 339–349.  
YIN Haitao, ZHOU Changsheng. Cross-fusion transformer-based infrared and visible image fusion method[J]. *Laser & optoelectronics progress*, 2025, 62(6): 339–349.
- [4] LIU Yu, CHEN Xun, WARD R K, et al. Image fusion with convolutional sparse representation[J]. *IEEE signal processing letters*, 2016, 23(12): 1882–1886.
- [5] LIU Jinyuan, FAN Xin, JIANG Ji, 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 Weiwei, LEI Yang, ZHAO Huaixun. 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 Jiayi, CHEN Chen, LI Chang, et al. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. *Information fusion*, 2016, 31: 100–109.
- [8] LIU Yu, LIU Shuping, WANG Zengfu. A general framework for image fusion based on multi-scale transform and sparse representation[J]. *Information fusion*, 2015, 24(C): 147–164.
- [9] 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.  
WANG Aixia, HU Aojie, YAN Aiyun, et al. An image fusion network based on ambient light awareness and infrared feature layer guidance[J]. *Control and decision*, 2025, 40(10): 3177–3189.
- [10] LI Hui, WU Xiaojun. DenseFuse: a fusion approach to infrared and visible images[J]. *IEEE transactions on image processing*, 2019, 28(5): 2614–2623.
- [11] ZHANG Yu, LIU Yu, SUN Peng, et al. IFCNN: a general image fusion framework based on convolutional neural network[J]. *Information fusion*, 2020, 54: 99–118.
- [12] LI Hui, WU Xiaojun. CrossFuse: a novel cross attention mechanism based infrared and visible image fusion approach[J]. *Information fusion*, 2024, 103: 102147.
- [13] YANG Yong, LIU Jiaxiang, HUANG Shuying, et al. Infrared and visible image fusion via texture conditional generative adversarial network[J]. *IEEE transactions on circuits and systems for video technology*, 2021, 31(12): 4771–4783.
- [14] TANG Linfeng, YUAN Jiteng, MA Jiayi. Image fusion in the loop of high-level vision tasks: a semantic-aware real-time infrared and visible image fusion network[J]. *Information fusion*, 2022, 82: 28–42.
- [15] LIU Jinyuan, FAN Xin, HUANG Zhanbo, et al. Target-

- aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 5792–5801.
- [16] GAO Hongwei, WANG Yutong, SUN Jian, et al. Efficient multi-level cross-modal fusion and detection network for infrared and visible image[J]. *Alexandria engineering journal*, 2024, 108: 306–318.
- [17] LI Ruimin, XIANG Jiajun, SUN Feixiang, et al. Multiscale cross-modal homogeneity enhancement and confidence-aware fusion for multispectral pedestrian detection[J]. *IEEE transactions on multimedia*, 2024, 26: 852–863.
- [18] TANG Linfeng, XIANG Xinyu, ZHANG Hao, et al. DIVFusion: Darkness-free infrared and visible image fusion[J]. *Information fusion*, 2023, 91: 477–493.
- [19] CAO Bing, SUN Yiming, ZHU Pengfei, et al. Multi-modal gated mixture of local-to-global experts for dynamic image fusion[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 23498–23507.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [21] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015, 37(9): 1904–1916.
- [22] CHEN C R, FAN Quanfu, PANDA R. CrossViT: cross-attention multi-scale vision transformer for image classification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 357–366.
- [23] TANG Linfeng, ZHANG Hao, XU Han, et al. Rethinking the necessity of image fusion in high-level vision tasks: a practical infrared and visible image fusion network based on progressive semantic injection and scene fidelity[J]. *Information fusion*, 2023, 99: 101870.
- [24] YING Xinyi, XIAO Chao, AN Wei, et al. Visible-thermal tiny object detection: a benchmark dataset and baselines [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2025, 47(7): 6088–6096.
- [25] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. *IEEE transactions on image processing*, 2006, 15(2): 430–444.
- [26] VAN AARDT J. Assessment of image fusion procedures using entropy, image quality, and multispectral classification[J]. *Journal of applied remote sensing*, 2008, 2(1): 023522.
- [27] RAO Yunjiang. In-fibre Bragg grating sensors[J]. *Measurement science and technology*, 1997, 8(4): 355–375.
- [28] ESKICIOGLU A M, FISHER P S. Image quality measures and their performance[J]. *IEEE transactions on communications*, 1995, 43(12): 2959–2965.
- [29] HAGHIGHAT M B A, AGHAGOLZADEH A, SEYEDARABI H. A non-reference image fusion metric based on mutual information of image features[J]. *Computers & electrical engineering*, 2011, 37(5): 744–756.
- [30] ASLANTAS V, BENDES E. A new image quality metric for image fusion: The sum of the correlations of differences[J]. *AEU - international journal of electronics and communications*, 2015, 69(12): 1890–1896.
- [31] MA Jiayi, YU Wei, LIANG Pengwei, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. *Information fusion*, 2019, 48: 11–26.
- [32] XU Han, MA Jiayi, JIANG Junjun, et al. U2Fusion: a unified unsupervised image fusion network[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2022, 44(1): 502–518.
- [33] LIU Jinyuan, LIN Runjia, WU Guanyao, et al. CoCoNet: coupled contrastive learning network with multi-level feature ensemble for multi-modality image fusion[J]. *International journal of computer vision*, 2024, 132(5): 1748–1775.
- [34] MENG Zhiming, LI Hui, ZHANG Zeyang, et al. CoMoFusion: Fast and High-Quality Fusion of Infrared and Visible Image with Consistency Model[C]//Pattern Recognition and Computer Vision. Singapore: Springer, 2025: 539–553.
- [35] TIAN Yunjie, YE Qixiang, DOERMANN D. YOLOv12: attention-centric real-time object detectors[EB/OL]. (2025–02–18)[2025–09–17]. <https://arxiv.org/abs/2502.12524>.

#### 作者简介:



田枫,教授,博士生导师,博士,东北石油大学计算机与信息技术学院院长,主要研究方向为智能油气地质、计算机视觉、智能数据分析处理。E-mail: [tianfeng1980@163.com](mailto:tianfeng1980@163.com)。



韩卓翰,硕士研究生,主要研究方向为低空视觉、多模态图像融合。E-mail: [2869060778@qq.com](mailto:2869060778@qq.com)。



刘芳,副教授,博士,主要研究方向为智能油气地质、智慧教育、多媒体与现代教育技术、计算机视觉。获黑龙江省科技进步二等奖1项、大庆市科技进步二等奖1项,主持和参与国家自然科学基金项目、黑龙江省自然科学基金项目、局级项目多项,发表学术论文20余篇。E-mail: [lfiufang1983@126.com](mailto:lfiufang1983@126.com)。