

融合可见光与热红外图像的无人机 早期森林火灾深度学习检测

杨靛青, 吴俊杰

福州大学 经济与管理学院, 福州 350108

摘要: 早期森林火灾检测方法可以及早发现火情并为救援工作争取宝贵时间, 而近年来无人机遥感虽然凭借高空分辨率和快速响应能力在早期森林火灾监测中展现出重要应用前景, 但仍面临目标尺度小、背景复杂、计算资源受限等关键挑战。针对这些问题, 本文提出了一种融合可见光与热红外双模态遥感图像的早期森林火灾深度学习检测模型 EFFNet (Early Forest Fire Net)。该模型针对遥感成像特性与火灾目标特征进行了4项关键设计: (1) 采用早期融合策略 EF (Early Fusion) 实现可见光与热红外信息的有效融合, 增强模型在复杂环境下的识别适应性; (2) 设计异构全局路由特征提取网络 HGRNet (Heterogeneous Global Routing Network) 增强多尺度感知能力并抑制关键特征丢失, 降低漏检率; (3) 引入特征聚合网络 FAN (Feature Aggregation Network), 采用双重自顶向下融合路径提高对小型火焰与烟雾目标的定位精度; (4) 设计高效半耦合检测头 ESCHead (Efficient Semi-Coupled detection Head) 降低模型计算复杂度与误报率。在 RGBT-3M (RGB-thermal multi-scene, multi-target, and multimodal forest fire aerial photography) 数据集上的实验结果表明, EFFNet 的 mAP50 达到 97.2%, 查准率达 94.9%, 查全率达 94.1%, 模型参数量仅为 1.84 M, 计算量仅为 5.6 GFLOPs。相较于当前先进目标检测模型, EFFNet 在实现更高检测精度的同时, 还具有轻量化和实时性优势, 能够满足无人机遥感平台对早期森林火灾实时监测的应用需求。

关键词: 无人机, 遥感图像, 深度学习, 目标检测, 早期森林火灾, 目标检测流程, 数据集, 性能评估指标

中图分类号: S762.3+2/TP751.1/P2

引用格式: 杨靛青, 吴俊杰. 2026. 融合可见光与热红外图像的无人机早期森林火灾深度学习检测. 遥感学报, 30(6): 1647-1663
Yang D Q and Wu J J. 2026. A deep learning-based early forest fire detection method integrating visible light and thermal infrared images from unmanned aerial vehicles. National Remote Sensing Bulletin, 30(6): 1647-1663 [DOI: 10.11834/jrs.20265227]

1 引言

森林火灾作为一种严重的自然灾害, 对生态系统、人类社会及经济活动具有显著的破坏性影响 (Mohapatra 和 Trinh, 2022)。它不仅直接烧毁大量植被、破坏森林生态系统的结构和功能 (Dos Santos 等, 2021), 还释放大量二氧化碳和有害气体, 加剧全球变暖与空气污染 (Coogan 等, 2019), 更对林区居民的生命财产安全构成严重威胁 (Bu 和 Gharajeh, 2019)。

森林火灾的发展通常经历初起、蔓延、发展、衰退4个阶段。初起阶段是指火灾刚刚发生、尚未形成大面积燃烧的阶段, 其特征包括火势规模小、热

量释放低、烟雾稀薄且对能见度影响有限 (Honary 等, 2025)。初起阶段是森林火灾防控的关键窗口期, 此时火势尚未形成自持燃烧, 扑救相对容易且成本较低。及时发现并扑灭初起阶段的火灾, 能够有效防止火势进入蔓延、发展阶段, 从而避免其演变成大规模毁灭性野火 (Sudhakar 等, 2020; Bouguettaya 等, 2022)。因此, 早期森林火灾检测成为当前研究的重点领域, 旨在通过先进的检测技术实现对初起阶段火灾的快速识别和准确定位。

目前, 针对森林火灾的主要监测方法包括: (1) 依靠护林员的人工检查方法, 但其监测范围有限且耗时较长; (2) 利用烟雾、温度和红外传感器进行自动监测, 但其易受浓雾、阳光直射等自然状

收稿日期: 2025-06-26; 预印本: 2026-03-31

基金项目: 国家社科基金后期资助项目(编号:25FGLB019);福建省自然科学基金(编号:2025J01451, 2025J011137)

第一作者简介: 杨靛青, 研究方向为智能决策与算法优化。E-mail: jgxy_ydq@fzu.edu.cn

况的干扰，误报率较高；(3) 利用卫星遥感技术进行大范围监测，但其时效滞后，对早期的森林火灾并不灵敏；(4) 利用地面固定摄像头进行实时监测，但单个摄像头覆盖范围有限，监测整片林区需要部署大量设备，成本高昂。近年来，随着5G、物联网等技术的发展，基于无人机的空中巡查平台为森林火灾监测提供了新的解决方案（Yang和Wang，2025）。如图1所示，无人机搭载可见光与热红外传感器进行低空巡查，可实时采集高分辨率遥感图

像；机载边缘计算设备结合深度学习模型对图像进行实时分析，识别地表火、林冠火等可见森林火灾类型；一旦发现火情，系统立即将位置信息上传至地面指挥中心，以便快速部署救援力量。与卫星遥感相比，无人机遥感具有更高的空间分辨率和更快的响应速度；与地面监控相比，无人机覆盖范围广且部署成本低，因此在早期森林火灾预警中展现出重要应用前景（Titu等，2024）。

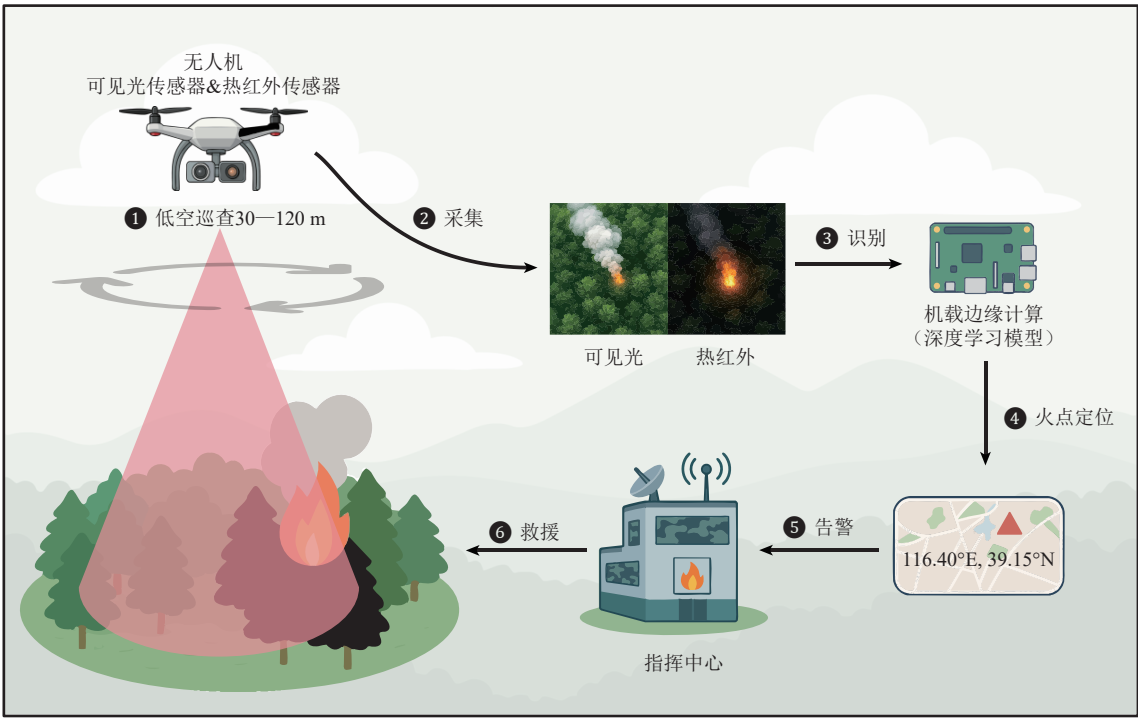


图1 无人机森林火灾监测系统架构
Fig.1 The architecture of the UAV-based forest fire monitoring system

然而，当前基于无人机遥感的早期森林火灾检测仍面临诸多挑战：(1) 小型目标检测困难——初期森林火灾的目标小，在高分辨率遥感图像中仅占很小的像素区域，往往缺乏足够的细节特征，容易被背景噪声所掩盖；(2) 目标存在遮挡情形——森林火灾易受茂密植被遮挡，使得暴露的特征信息减少，识别难度增加；(3) 目标边界模糊且形态各异——其形状与颜色易受光照、风向等环境因素影响，导致边界不清晰、纹理特征丧失，影响检测精度。此外，烟雾与火焰具有类内多样性和类间相似性，即在不同的环境条件及火灾发展阶段下，火焰与烟雾的形态和色调差异较大，而两者之间又存在相似性，如火焰的照亮使得烟雾的色调近似于火焰本身。加之拍摄视角与拍摄距离的变化，以及

云雾等类烟物体的干扰，进一步增加了检测难度。因此，早期森林火灾检测属于极具挑战性的复杂场景下的小型目标检测任务。

针对以上难点，现有研究主要从多尺度特征增强、多模态图像融合与多源数据融合等方向展开探索。在小型目标检测方面，Li等（2024）采用多维窗口注意力机制扩大感受野并利用截断卷积减少特征冗余，从而提高了模型对小型目标的检测精度，但该方法参数量大、计算开销显著，难以满足无人机平台的实时性要求。为兼顾精度与计算效率，Titu等（2024）采用知识蒸馏技术，将轻量化YOLOv8n（You Only Look Once version 8 nano）模型部署于无人机边缘计算平台，实现实时火灾检测，但其所用数据集涵盖车辆火灾、工业火灾

等多种场景,森林火灾的样本比例不足,限制了模型在森林火灾检测中的针对性与有效性。Wang等(2025)结合多分支膨胀卷积与合成图像进行训练,提升了模型对小型目标的感知与学习能力,但合成图像难以准确复现真实森林环境中的复杂光照、植被遮挡和细节纹理,导致模型泛化能力受限。刘惠临等(2025)通过改进YOLOv5s(You Look Only Once V5)骨干网络加快特征提取速度,实现模型轻量化,但在植被遮挡条件下仍难以有效捕获小型火焰的目标特征。针对目标遮挡问题,Rui等(2023)提出融合可见光与热红外图像的野火语义分割模型,通过自适应模态学习策略提升了昼夜野火目标的识别能力,然而受限于数据集规模较小、场景单一,模型在复杂森林环境中的泛化能力不足。Zhang等(2025)通过渐进式学习策略挖掘可见光与热红外图像的互补特征,改善了植被遮挡条件下的检测效果,但融合过程显著增加了计算开销,难以满足实时应用需求。

烟雾识别是早期森林火灾检测中的另一大难点。由于烟雾形态多变、边界模糊且易与云雾混淆,其检测面临更高的不确定性和误检风险。针对这一问题,Hu等(2022)利用多方向检测与值转换注意力机制,结合混合非极大值抑制NMS(Non-Maximum Suppression)方法,有效减少了类烟物体误检,提升了倾斜烟雾的检测效果,但对边界模糊、浓度较低的稀薄烟雾仍难以准确识别。Sheng等(2021)采用图像超像素分割结合聚类的方法为卷积神经网络提供更聚焦的输入,提高了烟雾目标检测的准确率,但误报率也较高。Zhan等(2022)通过并行两个相同结构的残差神经网络ResNet(residual neural network)模型增强烟雾特征的提取能力,但并行结构导致计算量大幅增加且产生大量特征冗余,反而加大了误检率和漏检率。此外,还有研究尝试整合多源异构数据以提高森林火灾检测的可靠性。Phan等(2020)通过时空数据集成算法将气象数据与卫星遥感图像对齐以实现野火定位,但该方法对早期火灾不够灵敏,难以满足早期预警需求。Bhamra等(2023)通过融合卫星、气象和光学图像信息,建立多源数据融合模型以模拟火灾的时序动态特征,但该方法仅针对烟雾目标检测,且数据集主要来自固定点位监控,场景单一,模型鲁棒性受限。

总的来说,尽管现有研究在森林火灾检测方

面取得了一定进展,但仍存在以下不足:(1)模型复杂度高,在算力有限的无人机平台上难以实现实时监测;(2)小型目标识别能力不足,难以捕获初期森林火灾的微弱特征信号;(3)泛化能力有限,缺乏在大规模森林火灾数据集上的充分训练与验证;(4)研究重点多局限于单一模态的单一目标检测,未充分利用多模态信息融合与火焰-烟雾多目标协同检测的优势。

为解决上述的检测局限,需融合可见光与热红外传感器的互补信息。两种传感器的成像原理及特性对比如下。

可见光传感器工作在380—760 nm波段,通过捕获物体反射的可见光进行成像,能够提供关于物体纹理、形状和颜色的详细信息,但在夜间或浓烟遮挡时性能受限。热红外传感器工作在8—14 μm 波段,通过捕获物体自身辐射的红外能量成像,可直接反映物体表面温度分布,对火灾引起的高温异常极为敏感,并能在一定程度上穿透遮挡障碍,但对初期火灾产生的稀薄烟雾不敏感。因此,融合两种模态信息则能够发挥互补优势,同时捕获火焰的温度特征与烟雾的纹理特征,实现对火情的协同检测。

在此背景下,本文提出了一种融合无人机可见光与热红外图像的早期森林火灾检测模型EFFNet(Early Forest Fire Net),实现对火焰与烟雾的联合检测。该模型针对无人机遥感成像特性及森林火灾目标特点进行了针对性设计以提升检测精度,同时采用轻量化设计降低计算复杂度,使其适用于计算资源受限的无人机平台。在RGBT-3M(RGB-thermal multi-scene, multi-target, and multimodal forest fire aerial photography)数据集上的实验结果表明,该模型具有有效性。

2 数据准备

2.1 数据集

RGBT-3M数据集(Zhang等,2025)是针对森林火灾检测任务构建的可见光与热红外双模态无人机航拍遥感数据集。该数据集的采集工作在安徽省、云南省和内蒙古自治区的多个森林区域进行,采用了大疆经纬M300RTK无人机搭载H20T云台相机以及大疆Mavic 2行业版无人机搭载一体化相机两种设备。采集场景包括真实森林火灾现

场、户外控制实验以及无火对照场景，时间跨度涵盖日间和夜间，保证了数据集场景的多样性和代表性。在数据采集过程中，可见光和热红外图像采用相同帧率同时获取，以保证时间同步。针对不同传感器间的空间位置差异，采用M-RIFT方法对双模态图像进行配准，实现了像素级的精确对齐。

RGBT-3M数据集共包含22440张图像（即11220对可见光—热红外图像），所有图像的大小均为640像素×480像素。数据集按照7：3的比例划分为训练集（7854对）和验证集（3366对）。数据集标注使用LabelImg工具完成，标注类别包括火焰、烟雾和人员3类目标，对应标注数量分别为11315个、13574个、和5888个。由于火焰和烟雾是森林火灾检测的主要目标，因此后续分析仅针对这两类目标展开。

表1给出了数据集中不同类别图片数量的分布情况，其中同时包含火焰和烟雾的图像共有5433对，体现了森林火灾场景中多目标共存的复杂性。

图2（a）对所有火焰和烟雾实例的边界框高宽进行了标准化统计，结果显示大多数实例框的高度和宽度集中在0—0.1之间。图2（b）按面积大小对所有火焰和烟雾实例的边界框分布进行了统计，根据像素级目标区域大小与图像总面积的比值，目标被分为小型目标（即面积占比<1%）、中型目标（1%≤面积占比<5%）和大型目标（面积占比≥5%）3类（Rui等，2023）。结果表明，小型目标占数据集的主要部分，这有助于模型学习早期森林火灾的目标特征。

表1 不同类别图片数量的分布情况
Table 1 Distribution of image counts across different categories

类型	图像对数
仅包含火焰	3258
仅包含烟雾	2529
同时包含火焰和烟雾	5433
总计	11220

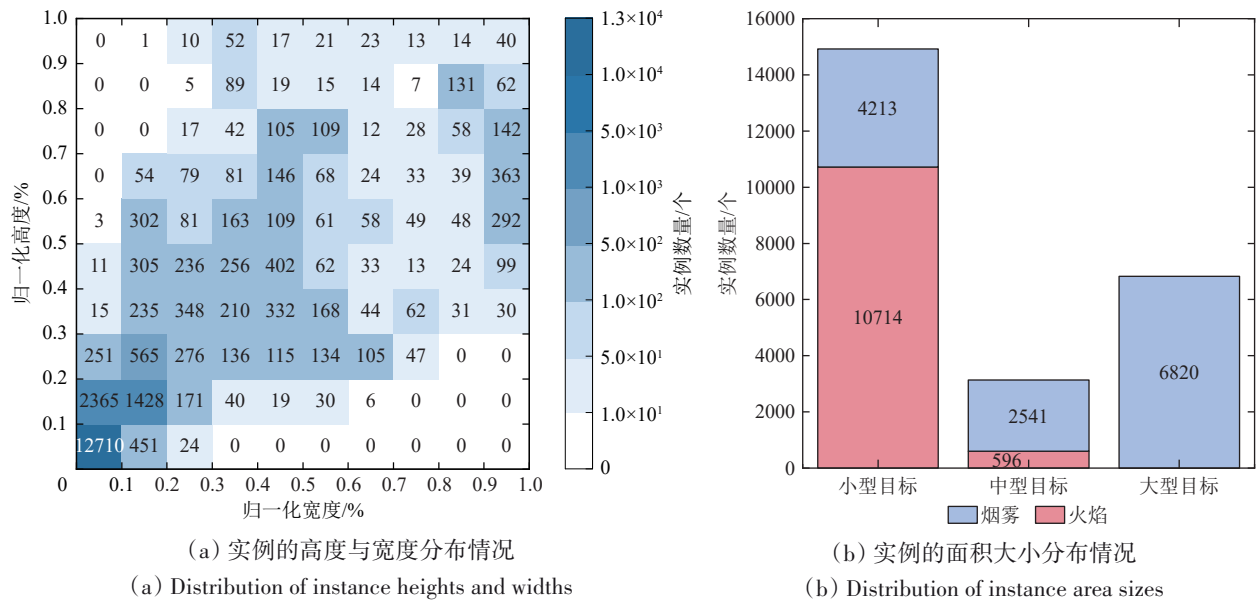


图2 实例的分布情况
Fig.2 Instance distributions

2.2 数据预处理

为提升训练数据的多样性与模型对复杂环境的鲁棒性，本研究在模型训练阶段采用了多种在线数据增强策略。各增强策略按预设概率相互独立地随机触发，在不改变物理样本规模的前提下扩展输入特征分布，使每个训练轮次的等效样本量约为原始训练集的4倍，进而提高模型在多场景下

的泛化能力。

（1）Mosaic数据增强。遥感图像中背景往往占据较大面积，导致目标容易被复杂背景所淹没。为增强模型对小型目标及复杂背景的学习能力，训练阶段以100%概率应用Mosaic数据增强，即从训练集中随机采样4张图像并按2×2网格拼接，对子图施加随机缩放与平移后统一裁剪为640像素×

640 像素，并对对应标注框应用一致的仿射变换。该策略通过跨图拼接与几何扰动，在单个样本中引入多尺度与跨场景背景干扰，从而在不增加样本总量的前提下扩展目标与背景的联合分布，提升模型在复杂场景下的鲁棒性与泛化能力。

(2) 色调、饱和度、亮度 HSV (Hue, Saturation, Value) 随机变换。遥感成像受光照、天气、传感器差异等条件影响显著，色调、饱和度和亮度在不同时间段内存在较大差异，为增强模型在不同环境条件下的适应性，以 100% 概率对训练图像在 HSV 空间进行随机扰动，得到 H' 、 S' 、 V' ：

$$H' = (H + r_H \times 180) \bmod 180 \quad (1)$$

$$S' = \text{clip}(S \times (1 + r_S), 0, 255) \quad (2)$$

$$V' = \text{clip}(V \times (1 + r_V), 0, 255) \quad (3)$$

式中， $\bmod$ 为取模运算， clip 为截断函数，将数值限定在指定区间内，其中 $r_H \sim U(-0.015, 0.015)$ 、 $r_S \sim U(-0.7, 0.7)$ 、 $r_V \sim U(-0.4, 0.4)$ 为相互独立的变化系数，分别对应 $\pm 2.7^\circ$ 的色调偏移、0.3—1.7 倍的饱和度缩放以及 0.6—1.4 倍的亮度缩放。

(3) CLAHE (contrast limited adaptive histogram equalization)。在森林火灾无人机遥感中，常见烟雾遮挡、阴影与局部欠曝等问题，导致目标局部细节对比度下降或信息丢失。为增强暗弱区域细节并抑制噪声放大，以 1% 的概率应用 CLAHE 算法 (Pizer 等, 1990)，即将整幅图像按 8×8 网格分块，对各块分别执行直方图均衡化，并设置对比度限制 $\text{ClipLimit} \sim U(1, 4)$ ，以提升暗部细节与局部对比度。

(4) 随机雾化增强。无人机低空遥感易受大气散射与烟霾影响，为提高模型在低能见度天气下的检测能力，以 1% 概率对训练图像进行雾化模拟 (Buslaev 等, 2020)。通过在图像中随机生成若干雾斑并进行高斯平滑，将原始无雾图像 $I(x, y)$ 转换为合成雾图像 $I_{\text{fog}}(x, y)$ ：

$$I_{\text{fog}}(x, y) = (1 - \beta)I(x, y) + \beta \alpha \left(\sum_{i=1}^N G_i(x, y) \right) * K_{\sigma_i} \quad (4)$$

$$G_i(x, y) = \exp \left(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma_i^2} \right) \quad (5)$$

式中， $\alpha = 0.08$ ，为透明度系数； $\beta \sim U(0.3, 1.0)$ 为雾强度系数； N 为随机生成的雾斑数量； (x_i, y_i) 为第 i 个雾斑的中心位置，均匀采样于图像范围内；

$\sigma_i = \beta \times r_i$ ， $r_i \sim U(0.02 \times \min(H, W), 0.08 \times \min(H, W))$ ，为雾斑的扩散标准差； $G_i(x, y)$ 为二维高斯核函数， H 、 W 分别为图像的高度和宽度； K_{σ_i} 为标准差 $\sigma_i = 0.05 \times \min(H, W)$ 的高斯模糊核； $*$ 表示卷积运算。该方法能够模拟真实雾天中的能见度退化效果，从而提升模型在烟雾、晨雾等低能见度场景下的鲁棒性。

3 EFFNet 网络结构

为解决无人机遥感在早期森林火灾检测中面临的目标尺度小、背景复杂、检测精度不足等问题，本文提出了一种双模态目标检测网络 EFFNet，其结构如图 3 所示。该网络以 YOLO11 (You Only Look Once 11) 为基础框架，结合无人机遥感与森林火灾检测任务的特点，对输入层、骨干网络、颈部和检测头 4 个模块进行了针对性改进。通过这些设计，EFFNet 能够显著降低森林火灾的误检率与漏检率，有效提升对小型火焰与烟雾目标的检测精度，同时具备轻量化和实时性的优势，有效满足无人机平台对准确检测与实时监测的需求。

3.1 YOLO11

YOLO (You Only Look Once) (Redmon 等, 2016) 是一种快速且准确的一阶段目标检测器，广泛应用于计算机视觉任务。YOLO11 是 YOLO 系列的最新版本，由骨干网络、颈部和检测头 3 部分组成，分别负责特征提取、多尺度特征融合以及目标的定位与分类。本文采用 YOLO11 参数量最少、速度最快的轻量化配置版本 YOLO11n 作为基准模型，该模型特别适用于无人机遥感实时监测森林火灾的应用场景。

3.2 早期融合策略 EF

传统 YOLO11 模型仅支持单一模态的图像输入，无法发挥多模态遥感图像的互补优势。单独使用可见光遥感图像虽然能提供丰富的纹理细节信息，但由于遥感图像中目标尺寸较小，且易受低能见度、烟雾遮挡和复杂背景等因素的干扰，模型常难以有效区分目标与背景，导致漏检和误检；而单独使用热红外遥感图像虽然对火灾引起的温度差异更为敏感，能够在复杂背景中快速定位高温区域，但其目标边界往往不够清晰，且在森林火灾初期温度变化不显著时，检测精度仍然十分有限。

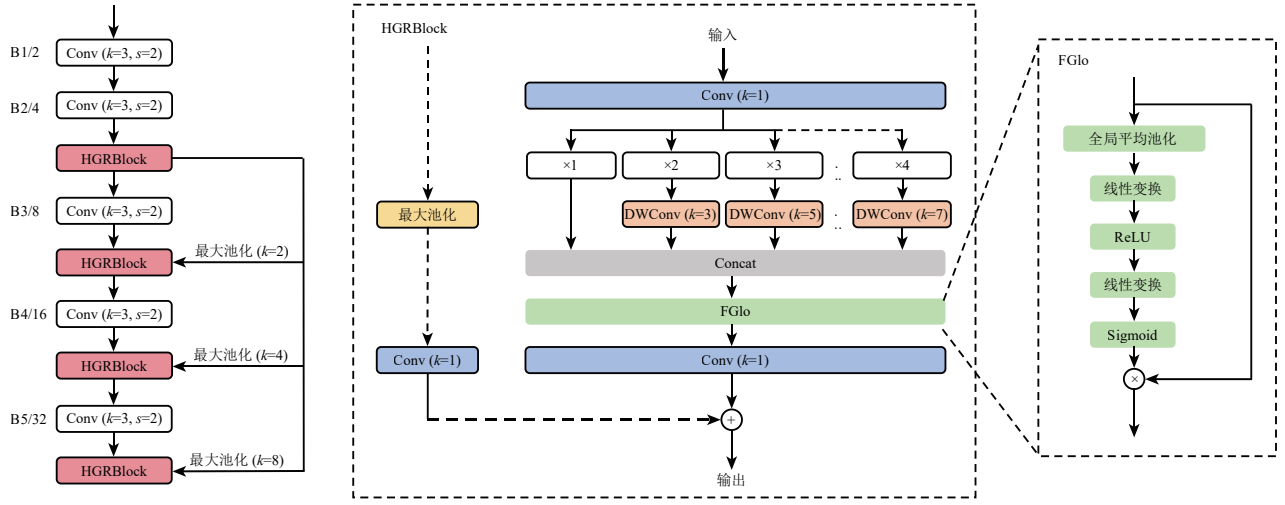


图4 HGRNet网络结构

Fig.4 Network structure of HGRNet

$$[F_1, F_2, \dots, F_N] = \text{Split}(\text{Conv}_{1 \times 1}(F_{in})) \quad (6)$$

式中， F_{in} 为输入特征图， F_i 为第 i 个分支的输出特征图， N 为分支数量。

第一个分支直接传递原始特征，其他分支则采用不同卷积核尺寸的DWConv提取多尺度特征：

$$F_i = \begin{cases} F_1, & i = 1 \\ \text{DWConv}_{k_i \times k_i}(F_i), & i = 2, 3, \dots, N \end{cases} \quad (7)$$

式中， k_i 表示第 i 个分支的卷积核尺寸。

接着，采用Wu等(2021)提出的全局上下文提取模块 $f_{\text{glo}}(*)$ (global context extractor)，对多尺度特征进行自适应加权，以强化关键分支特征并抑制冗余信息：

$$z = L\left(R\left(L\left(\text{GAP}\left(\text{Concat}(F_1, F_2, \dots, F_N)\right)\right)\right)\right) \quad (8)$$

$$F_{\text{gc}} = \sigma(z) \otimes \text{Concat}(F_1, F_2, \dots, F_N) \quad (9)$$

式中，GAP为全局平均池化， L 为线性变换， R 为ReLU (Rectified Linear Unit) 激活函数， z 表示全局上下文中间向量， σ 为Sigmoid激活函数， $\otimes$ 表示逐元素相乘， F_{gc} 表示经全局上下文加权后的特征图。

最后，将多尺度加权特征与浅层局部最大池化操作提取的关键特征进行融合输出：

$$F_{\text{HGR}} = \text{Conv}_{1 \times 1}(F_{\text{gc}}) + \text{Conv}_{1 \times 1}(F_{\text{shallow}}) \quad (10)$$

式中， F_{shallow} 为浅层池化特征， F_{HGR} 为HGRBlock的最终输出。

3.4 特征聚合网络FAN

骨干网络的低层特征图分辨率高、定位能力强，高层特征图语义丰富、利于关系建模，两者的有效融合是提升模型检测性能的关键。YOLO11模

型的颈部采用自顶向下再自下而上的路径聚合特征金字塔网络PAFPN (path aggregation feature pyramid network) 结构，该结构在通用目标检测任务中表现良好。然而，在无人机遥感监测早期森林火灾的特定场景下，受限于远距离遥感成像和早期火灾特征的影响，目标尺寸往往较小。此时，模型的检测性能更依赖于高层语义信息对低层空间信息的指导，而自下而上路径的贡献相对有限 (Jing等, 2024)。此外，YOLO11仅融合骨干网络下采样8倍、16倍、32倍的三层输出，缺少高分辨率浅层特征，这种设计易导致跨尺度特征耦合 (Deng等, 2022)，即不同尺寸目标的特征在同一层级中混合，使小型目标的特征容易受到中等或大型目标语义信息的干扰。

为此，本文引入特征聚合网络FAN (dual top-down Feature Aggregation Network) (Jing等, 2024) 作为颈部的特征融合网络，如图5所示。FAN采用双重自顶向下的特征融合路径，并在每一层执行先融合后传递的策略，使高层语义信息能够稳定地向低层传递并逐步增强特征表达。此外，FAN还引入了来自骨干网络的4倍下采样浅层特征，利用更高分辨率实现细节与语义的对齐，从而有效缓解跨尺度特征耦合问题，提升小型目标的定位能力。

具体而言，本文将骨干网络下采样4倍、8倍、16倍、32倍的B2、B3、B4、B5层输出作为FAN的输入。FAN首先通过 1×1 卷积对各层输出进行降维，并统一通道数：

$$P_i = \text{Conv}_{1 \times 1}(C_i), i = 2, 3, 4, 5 \quad (11)$$

式中, C_i 为骨干网络第 i 层的输出特征图, P_i 为降维后的特征图。

随后 FAN 采用自顶向下的方式逐层融合多尺度特征, 并利用 C3k2 模块进一步提取特征:

$$P_i = \begin{cases} P_i + \text{UpSample}(P_{i+1}), & i = 2, 3, 4 \\ P_i, & i = 5 \end{cases} \quad (12)$$

$$P_i = \text{C3k2}(P_i) \quad (13)$$

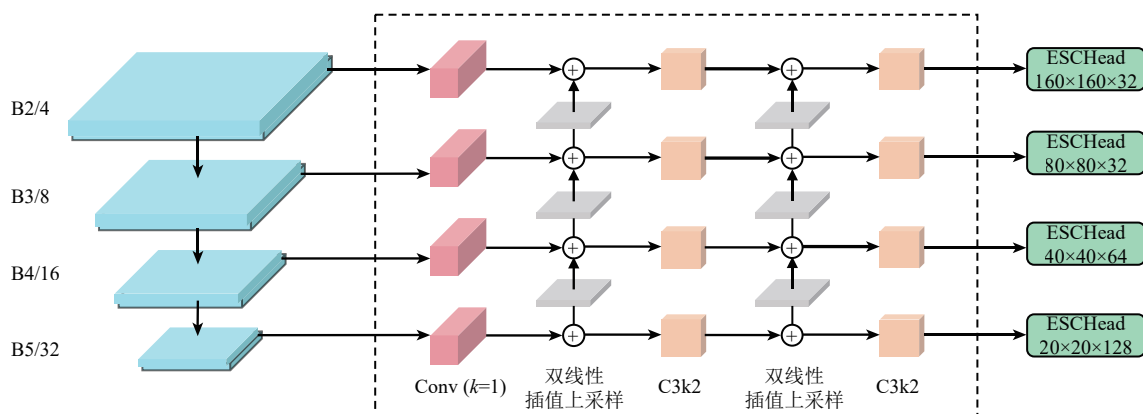


图5 FAN网络结构

Fig.5 Network structure of FAN

3.5 高效半耦合检测头 ESCHead

检测头基于颈部输出的多尺度特征进行目标的定位与分类。YOLO11的检测头采用解耦式结构, 将定位任务与分类任务分离, 独立优化各自参数。然而, 这种设计也导致两个任务之间缺乏特征交互, 容易出现任务不对齐的问题, 如预测框的分类置信度高但定位精度低等 (Feng等, 2021)。在早期森林火灾检测中, 这种不对齐会增加误报的风险, 影响检测结果的可靠性。此外, 在3.4节中引

入的高分辨率检测层虽然能够捕获更细粒度的特征, 但高分辨率特征图的卷积计算也显著增加了模型的计算开销, 从而影响实时性能。

针对上述问题, 本文提出高效半耦合检测头 ESCHead (Efficient Semi-Coupled detection Head), 其结构如图6所示。ESCHead采用半解耦式设计, 通过共享主干建立分类任务与定位任务之间的交互路径, 并在各分支中引入 1×1 卷积进行通道压缩, 以减少特征冗余并降低计算复杂度。

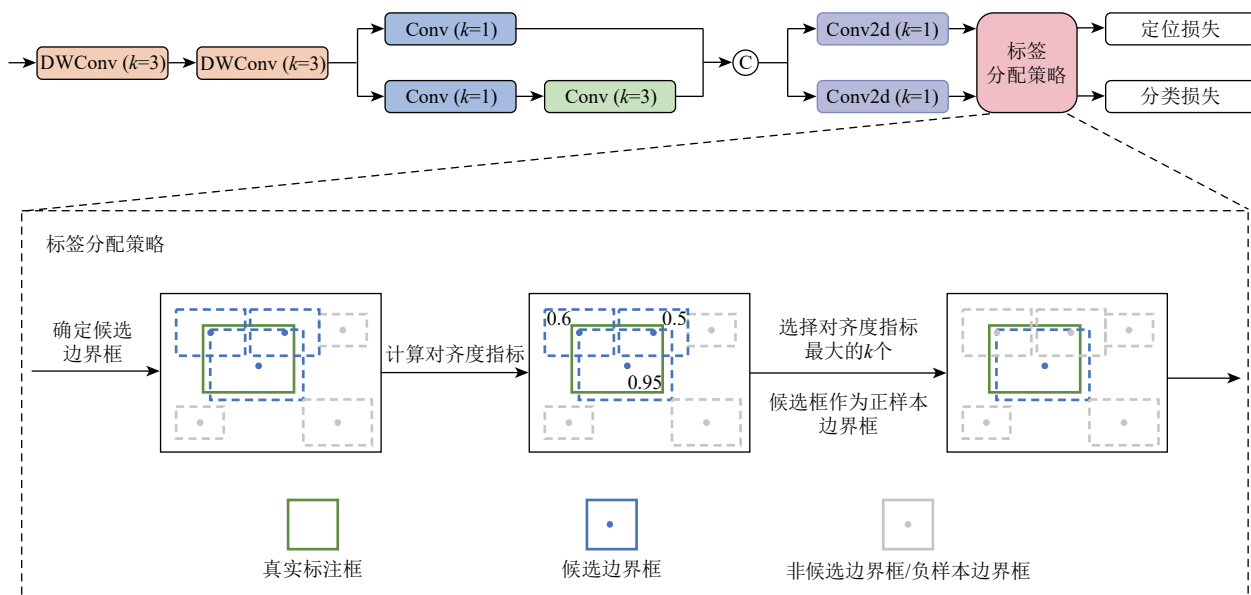


图6 ESCHead结构

Fig.6 Structure of ESCHead

具体而言，ESCHead采用连续的DWConv提取共享特征，为分类与定位任务提供一致的特征表示： $Q_i = \text{DWConv}_{3 \times 3}(\text{DWConv}_{3 \times 3}(P_i))$, $i = 2, 3, 4, 5$ (14) 式中， P_i 为FAN输出的特征图， Q_i 为共享特征图。

随后，定位分支通过1×1卷积压缩通道数以保留空间细节；分类分支则通过3×3卷积增强判别能力，再通过1×1卷积压缩通道数，最后拼接两分支特征以增强任务协同：

$$D_i = \text{Concat}(\text{Conv}_{1 \times 1}(Q_i), \text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3}(Q_i))), \quad i = 2, 3, 4, 5 \tag{15}$$

式中， D_i 为拼接后的检测特征图。

最后，通过两个独立的1×1卷积分别输出边界框回归和分类预测结果：

$$B_i = \text{Conv}_{1 \times 1}(D_i), i = 2, 3, 4, 5 \tag{16}$$

$$S_i = \text{Conv}_{1 \times 1}(D_i), i = 2, 3, 4, 5 \tag{17}$$

式中， B_i 表示边界框回归预测输出， S_i 表示分类预测输出。

此外，在模型训练阶段，检测头完成预测后，还将执行标签分配与损失计算，以实现梯度的反向传播。其中，标签分配是根据预测框与真实框的匹配程度，将预测框划分为正负样本的过程。YOLO采用综合分类置信度和定位交并比的方式来计算预测框与真实框的匹配度，具体计算公式为

$$\text{align_metric} = \text{predict_score}^\alpha \times \text{overlap}^\beta \tag{18}$$

式中，predict_score为预测框对应真实类别的置信度得分，overlap为预测框与真实框的交并比， $\alpha=1$ ， $\beta=6$ 。

然而，在无人机遥感监测早期森林火灾的场景中，火焰与烟雾的尺寸往往较小，而小型目标的交并比对位置偏差极为敏感，细微的坐标偏移即可导致交并比大幅下降，使得小型目标难以被划分为正样本，进而影响模型对早期火灾特征的学习。

因此，本文提出改进的匹配度计算方式：

$$\text{align_metric} = \text{scale_factor} \times \text{predict_score}^\alpha \times \text{overlap_trans}^\beta \tag{19}$$

$$\text{overlap_trans} = \begin{cases} \text{overlap}^2, & \text{overlap} < 0.3 \\ \text{overlap}^{1/2}, & \text{overlap} \geq 0.3 \end{cases} \tag{20}$$

式中，scale_factor为真实边界框面积与64×64的比值，范围限定为1—5。

该策略通过引入尺寸自适应因子scale_factor来提高小型目标的匹配权重，增强网络对小型目标

样本的响应；同时，通过分段非线性变换增强高质量预测框的权重，抑制低质量预测框的影响，从而有效缓解小型目标匹配难题，提升模型对早期森林火灾特征的学习能力与训练效率。

4 实验结果及分析

本节通过实验验证了EFFNet在早期森林火灾检测任务中的优越性，内容包括实验环境介绍、评价指标说明、消融实验、方法应用性能分析、与其他目标检测方法的对比及可视化结果的比较。

4.1 实验环境介绍

本文所有消融实验均在相同的硬件环境和软件环境下进行，具体环境参数如表2所示。

表2 硬件环境与软件环境配置
Table 2 Hardware environment and software environment configuration

环境	类型	配置
硬件环境	CPU	Intel Xeon Gold 6148
	GPU	NVIDIA GeForce RTX 3090
软件环境	Pytorch	2.3.0
	Python	3.10.14
	OS	Ubuntu 22.04.3
	CUDA	12.1

注：CPU为中央处理器(central processing unit)、GPU为图形处理单元(graphics processing unit)、OS为操作系统(operating system)、CUDA为计算统一设备体系结构(compute unified device architecture)。

本文的所有消融实验均在相同的模型参数设置下进行，具体参数设置如表3所示。

表3 模型参数设置
Table 3 Model parameter settings

参数	含义	值
Image size	图像尺寸	640
Training epochs	训练的总轮数	200
Batch size	批次大小	32
Close mosaic	最后停用马赛克增强的训练轮数	20
Initial learning rate	初始学习率	0.01
Final learning rate	最终学习率	0.01
Optimizer	优化器	SGD
Momentum	优化器的动量因子	0.937
Weight decay	权重衰减系数	0.0005
Seed	随机种子	0

4.2 评价指标说明

多分类问题的混淆矩阵是多种评价指标的计算基础,能够直观展示模型在各类别上的表现。以二分类类混淆矩阵为例,其形式如表4所示。

表4 二元分类混淆矩阵

Table 4 Confusion matrix for binary classification

真实类别	预测类别	
	正类	负类
正类	TP(true positive)	FN(false negative)
负类	FP(false positive)	TN(true negative)

查准率 P (Precision)、查全率 R (Recall)、 $F1$ 分数 $F1$ ($F1$ -Score) 是用于评估模型效果的关键指标,其计算方法如式 (21)、式 (22)、式 (23) 所示 (Padilla等, 2020)。其中, P 表示所有预测为正类的实例中有多少是真正的正类,体现模型预测结果的可靠性或模型的误判水平; R 表示所有真正为正类的实例中有多少被正确预测出来,体现模型对正类样本预测的覆盖能力或模型的漏检水平; $F1$ 是 P 和 R 的调和平均值,表现模型查准查全的综合性能:

$$P = \frac{TP}{TP + FP} \quad (21)$$

$$R = \frac{TP}{TP + FN} \quad (22)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (23)$$

平均精度均值 mAP (mean Average Precision) 是评估模型效果的另一项关键指标,用于综合衡量模型在多类别目标检测任务中的整体表现,其计算方法如式 (24) 所示 (Padilla等, 2020)。首先,根据不同置信度阈值计算相应的 P 与 R ,绘制 PR 曲线,并计算曲线下的面积以获得每一类别的平均精度 AP (Average Precision),其计算方法如式 (25) 所示 (Padilla等, 2020)。然后,对所有类别的 AP 取平均,得到 mAP。常用的 mAP 指标包括 mAP50 与 mAP50—95。其中, mAP50 表示交并比阈值为 0.5 时的 mAP,反映模型在较为宽松的定位精度要求下的表现;而 mAP50—95 则表示交并比从 0.5 到 0.95,以 0.05 为步长的多个阈值下的 mAP 的平均值,评价标准更加严格。

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (24)$$

$$AP = \int_0^1 P(R) dR \quad (25)$$

式中, $P(R)$ 为 R 对应的 P , N 为类别个数, AP_i 为第 i 类的 AP 值。

模型参数量 Params (model parameter quantity) 与浮点运算次数 FLOPs (floating-point operations) 是衡量模型复杂度的关键指标。Params 指的是模型中所有参数的总数,反映了模型的规模;而 FLOPs 则表示模型完成一次前向推理所需的浮点运算次数,反映了模型的计算复杂度。

4.3 消融实验

在表6所示的消融实验中,本文以 YOLO11n 为基准模型,通过对比不同模态输入和逐步替换、加入不同模块来系统评估所提方法的有效性,验证了每个模块或模块组合对整体检测框架的贡献。

对比1、2、3组实验结果可见,可见光单模态 (RGB) 在烟雾识别上表现较好,而对火焰的 R 仅为 85.1%;热红外单模态 (IR) 虽然对火焰的温度异常敏感,但对烟雾的 R 仅为 83.6%;采用 EF 策略后,火焰和烟雾的 R 分别提升至 89.6% 和 93.5%。与两种单模态基准相比,EF 策略使总体 mAP50 分别提升了 2.5% 和 4.2%, mAP50—95 分别提升了 4.7% 和 5.6%。这表明 EF 策略能够有效发挥双模态信息的互补优势,显著提升火焰和烟雾两类目标的检测性能。

对比3、4组实验结果可见,引入 HGRNet 后,总体 R 提升 1.1%, mAP50—95 提升 1.4%。其中,火焰和烟雾的 R 分别提升 1.4% 和 0.7%。这表明, HGRNet 能够有效增强模型对不同尺度目标的感知能力,从而有效降低漏检率。

对比3、5、7组实验结果可见,单独引入 FAN 后,总体各项指标均有所提升。其中,火焰的 R 由 89.6% 提升至 92.2%,提高了 2.6%; mAP50 由 94.4% 提升至 96.1%,提高了 1.7%。然而,烟雾的 R 从 93.5% 下降至 92.2%,减少了 1.3%。这是因为,虽然烟雾中也包含小型目标,但其整体尺度分布较火焰更大, FAN 引入的高分辨率特征在缺少 HGRNet 提供的更大感受野支持时,局部信息难以有效判别烟雾目标。而在同时引入 HGRNet 和 FAN 后,各项指标均实现一致性提升,表明两者的协同作用能够有效满足不同尺度目标的检测需求。

对比3、6组实验结果可见,引入 ESCHead 后,总体 P 提高了 1.0%,表明 ESCHead 能够有效抑制森林复杂背景下相似目标干扰以及多目标共存易引发的误检现象。

表6 验证集上火焰、烟雾及全部检测精度的消融实验结果

Table 6 The ablation experiment results of the detection accuracy of flame, smoke and overall on the verification set

编号	方法				模态	类型	P/%	R/%	F1/%	mAP50/%	mAP50—95/%
	EF	HRGNet	FAN	ESCHead							
1	×	×	×	×	RGB	全部	91.6	89.2	90.4	93.2	61.4
						仅火焰	90.4	85.1	87.7	89.7	48.9
						仅烟雾	92.8	93.2	93.0	96.8	73.9
2	×	×	×	×	IR	全部	91.4	84.5	87.8	91.5	60.5
						仅火焰	91.7	85.3	88.4	91.4	55.1
						仅烟雾	91.1	83.6	87.2	91.6	65.8
3	√	×	×	×	RGB+IR	全部	93.0	91.5	92.3	95.7	66.1
						仅火焰	92.3	89.6	90.9	94.4	57.7
						仅烟雾	93.8	93.5	93.6	97.1	74.5
4	√	√	×	×	RGB+IR	全部	93.5	92.6	93.0	96.1	67.5
						仅火焰	93.1	91.0	92.1	94.9	59.9
						仅烟雾	93.8	94.2	94.0	97.2	75.2
5	√	×	√	×	RGB+IR	全部	93.7	92.2	93.0	96.4	66.6
						仅火焰	93.1	92.2	92.6	96.1	59.1
						仅烟雾	94.4	92.2	93.3	96.8	74.0
6	√	×	×	√	RGB+IR	全部	94.0	91.6	92.8	95.7	66.2
						仅火焰	93.7	89.4	91.5	94.2	58.1
						仅烟雾	94.3	93.7	94.0	97.2	74.3
7	√	√	√	×	RGB+IR	全部	94.3	93.1	93.7	96.7	68.0
						仅火焰	94.2	92.4	93.3	96.1	60.9
						仅烟雾	94.4	93.8	94.1	97.3	75.2
8	√	√	√	√	RGB+IR	全部	94.9	94.1	94.5	97.2	68.3
						仅火焰	94.7	94.3	94.5	96.7	61.3
						仅烟雾	95.1	93.9	94.5	97.7	75.3

注：“√”表示加入该模块，“×”表示未加入该模块，加粗数字表示所在列指标的最优值。

最后，对比1、2、8组实验结果可见，在整合所有改进模块后，EFFNet的各项指标均实现了显著提升。具体而言，相较于可见光和热红外单模态基准，EFFNet的总体*P*分别提升3.3%、3.5%，*R*分别提升4.9%、9.6%，*F1*分别提升4.1%、6.7%，mAP50分别提升4.0%、5.7%，mAP50—95分别提升6.9%、7.8%。这些结果充分表明，EFFNet通过多模态融合、多尺度特征提取、高分辨率特征融合以及轻量化检测头的协同优化，能够显著提升森林火灾的检测精度和检测结果的可靠性。

4.4 方法应用性能分析

在表7所示的消融实验中，本文以YOLO11n为基准模型，通过逐步替换和加入不同模块来系统评估所提方法的实际应用性能。

对比1—4组实验结果可见，模型检测精度的提升是以一定的参数量和计算量增加为代价的。这主要源于两方面：HGRNet通过多尺度卷积扩展感受野并增强特征表达能力；FAN增加高分辨率

检测层以提升小型目标定位能力。值得注意的是，尽管FAN引入了高分辨率检测层，但其通过1×1卷积进行降维压缩，在一定程度上控制了参数量增长，并有效缓解了直接增加高分辨率检测层所带来的显著计算开销。

对比6、7组实验结果可见，引入ESCHead后，Params减少0.33 M，FLOPs降低5.3 G，有效降低了前述模块带来的计算开销，验证了ESCHead轻量化设计的有效性。最终，相较于单模态基准模型，EFFNet的Params减少0.74 M，FLOPs降低0.7 G，模型大小缩减1.2 MB。

在实际推理性能方面，EFFNet的推理速度达到362.72对图像/s，单对图像检测耗时仅2.757 ms，能够充分满足无人机平台实时检测的要求。综合来看，EFFNet通过多模块的协同优化，在显著提升检测精度的同时实现了整体轻量化，具备在资源受限的无人机平台上高效部署的可行性，能够满足早期森林火灾检测对准确性与实时性的要求。

表7 火焰及烟雾综合检测效率的消融实验结果

Table 7 The ablation experiment results of the detection efficiency of flame and smoke

编号	方法				Params/M	FLOPs/G	模型大小/MB
	EF	HRGNet	FAN	ESCHead			
1	×	×	×	×	2.58	6.3	5.2
2	√	×	×	×	2.58	6.4	5.2
3	√	√	×	×	2.81	7.1	5.7
4	√	×	√	×	1.95	10.2	4.2
5	√	×	×	√	2.47	5.3	5.0
6	√	√	√	×	2.17	10.9	4.6
7	√	√	√	√	1.84	5.6	4.0

注：“√”表示加入该模块，“×”表示未加入该模块，加粗数字表示所在列指标的最优值。

4.5 与其他目标检测方法的对比

为客观评估EFFNet在早期森林火灾检测中的综合性能，本文在相同实验环境下，基于RGBT-3M验证集，将EFFNet与11种主流单模态及多模态

目标检测方法进行了对比。实验结果如表8所示，EFFNet在轻量化优势、检测精度及误检漏检控制3个方面均表现优异。

表8 EFFNet与其他目标检测模型的比较

Table 8 Comparison between EFFNet and other object detection models

方法	输入模态	P/%	R/%	F1/%	mAP50/%	mAP50—95/%	Params/M	FLOPs/G	模型大小
YOLO11n	RGB	91.6	89.2	90.4	93.2	61.4	2.58	6.3	5.2 MB
Faster R-CNN(Ren等,2017)	RGB	92.6	89.6	91.0	94.6	62.6	41.35	178	317.0 MB
Mask R-CNN(He等,2017)	RGB	93.1	89.8	91.3	94.5	62.7	47.38	235	547.6 MB
Cascade R-CNN (Cai和Vasconcelos,2021)	RGB	93.9	90.1	91.9	94.9	65.0	69.17	205	530.4 MB
TOOD(Feng等,2021)	RGB	93.1	90.3	91.7	94.0	62.2	32.02	168	245.2 MB
RTMDet(Lyu等,2022)	RGB	93.2	89.4	91.2	93.0	59.8	4.87	8	96.2 MB
DDQ(Zhang等,2023)	RGB	93.4	90.3	91.8	93.9	57.5	48.30	235	604.0 MB
RT-DETR(Zhao等,2024)	RGB	93.1	89.3	91.1	93.6	60.0	20.09	61.2	307.3 MB
YOLO11n	IR	91.4	84.5	87.8	91.5	60.5	2.58	6.3	5.2 MB
Faster R-CNN(Ren等,2017)	IR	92.0	85.9	88.8	92.4	61.7	41.35	178	317.0 MB
Mask R-CNN(He等,2017)	IR	92.3	86.1	89.1	92.9	62.0	47.38	235	547.6 MB
Cascade R-CNN (Cai和Vasconcelos,2021)	IR	91.8	87.3	89.5	93.4	64.1	69.17	205	530.4 MB
TOOD(Feng等,2021)	IR	92.6	85.4	88.8	91.9	60.8	32.02	168	245.2 MB
RTMDet(Lyu等,2022)	IR	91.5	84.2	87.7	90.8	59.0	4.87	8	96.2 MB
DDQ(Zhang等,2023)	IR	91.4	85.6	88.3	90.1	55.8	48.30	235	604.0 MB
RT-DETR(Zhao等,2024)	IR	91.4	85.3	88.2	91.1	59.3	20.09	61.2	307.3 MB
CALNet(He等,2023)	RGB+IR	83.2	80.7	81.9	86.2	41.4	—	—	391 MB
MoE-Fusion(Cao等,2023)	RGB+IR	92.3	88.3	90.3	94.2	63.4	41.35	178	339.27 MB
RSDet(Zhao等,2026)	RGB+IR	92.1	86.5	89.0	92.9	59.2	139	215	1.0 GB
EFFNet	RGB+IR	94.9	94.1	94.5	97.2	68.3	1.84	5.6	4.0 MB

注：加粗数字表示所在列指标的最优值。

(1) 轻量化优势显著：EFFNet的参数量为1.84 M，计算量为5.6 GFLOPs，远低于多模态模型MoE-Fusion（41.35 M，178 GFLOPs）和RSDet（139 M，215 GFLOPs），也低于单模态模型如两阶段的Faster R-CNN（41.35 M，178 GFLOPs）和一阶段的RT-DETR（20.09 M，61.2 GFLOPs）。该轻

量化设计使 EFFNet 非常适合部署于计算资源受限的无人机边缘计算平台，能够有效满足实时森林火灾检测需求。

(2) 检测精度领先：EFFNet 的 mAP50 达 97.2%，mAP50—95 达 68.3%，在所有对比方法中表现最优。相比于多模态模型 CALNet、MoE-Fusion 和 RSDet，EFFNet 的 mAP50 分别提升 11.0%、3.0% 和 4.3%，mAP50—95 分别提升 26.9%、4.9%、9.1%；与可见光单模态模型表现最优的相比，EFFNet 的 mAP50 和 mAP50—95 分别提升 2.3% 和 3.3%。这得益于 EFFNet 对无人机遥感成像特性及森林火灾目标特点的针对性设计：通过早期融合策略、多尺度卷积和高分辨率分支等模块，EFFNet 能够有效应对目标尺寸小、环境遮挡、背景复杂等问题。

(3) 误检漏检率低：EFFNet 的 P 达 94.9%， R 达 94.1%。相比于多模态模型 CALNet、MoE-Fusion 和 RSDet，EFFNet 的 P 分别提升 11.7%、2.6% 和

2.8%， R 分别提升 13.4%、5.8%、7.6%；与可见光单模态模型表现最优的 DDQ 相比， P 和 R 分别提升 1.5% 和 3.8%。在早期森林火灾监测场景中，误报会造成救援资源浪费，漏报则可能错失扑救时机导致火情蔓延。EFFNet 的低误检漏检率表明其检测结果更为可靠，能够为早期森林火灾监测系统提供有力支持。

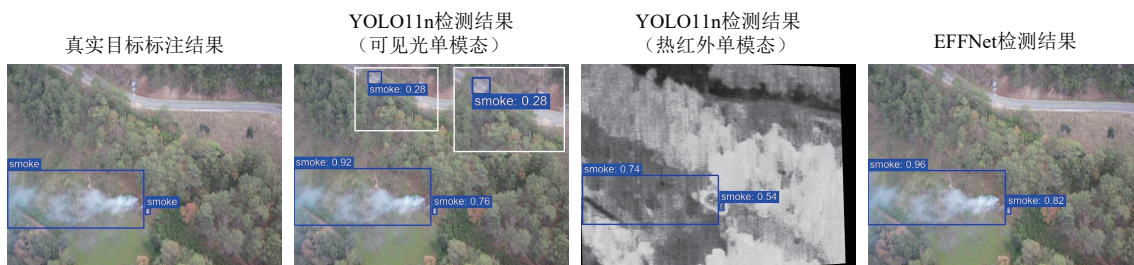
4.6 可视化结果的比较

为更直观地展示 EFFNet 的检测性能，本文在 RGBT-3M 验证集中选取了一组具有代表性的图像进行可视化结果比较，如图 7 所示。图 7 中的矩形框表示模型预测的目标边界框，框上方的“smoke”和“fire”标签表示预测的目标类别——烟雾和火焰，后面的数值为相应的类别置信度。结果表明，EFFNet 在具有复杂背景的森林环境中执行早期森林火灾检测任务时，性能提升显著，主要体现在以下 3 个方面。



(a) 背景干扰下小型目标误检

(a) False detection of small-sized targets under background interference



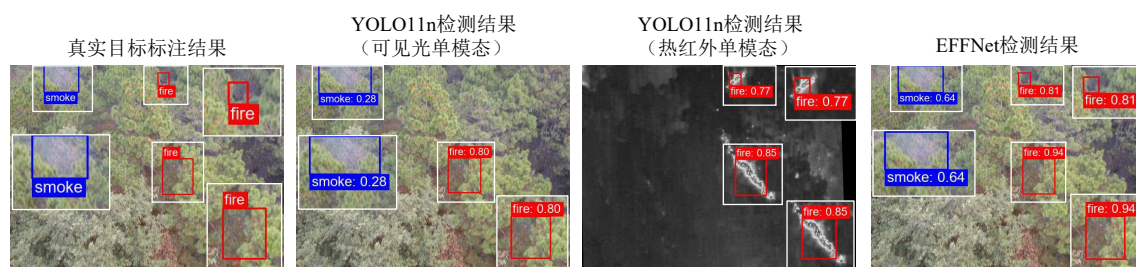
(b) 背景干扰下中型目标误检

(b) False detection of medium-sized targets under background interference



(c) 夜间环境下小型目标漏检

(c) Missed detection of small-sized targets in the night environment



(d) 多燃烧点树木遮挡下中型目标漏检

(d) Medium-sized targets were missed under the obstruction of trees with multiple burning points



(e) 烟雾遮挡下大型目标高精度检出

(e) High-precision detection of large-sized targets under smoke obstruction

图7 检测结果的可视化比较

Fig.7 Visual comparison of detection results

(1) 误报率降低: 如图7(a)、图7(b)所示, 图7(a)中包含小型尺寸烟雾目标且背景模糊; 图7(b)中的烟雾目标在形态和颜色上与被树木遮挡的道路相似, 易产生混淆。针对此类场景, YOLO11n单模态基准模型均出现误检, 而EFFNet准确识别了所有目标, 避免了误报。

(2) 漏报率降低: 如图7(c)、图7(d)所示, 图7(c)为夜间环境下同时存在火焰和烟雾目标的场景, 其中烟雾形态边界模糊; 图7(d)包含多个燃烧点, 但由于树木遮挡导致可见区域有限。在这两类挑战性场景中, YOLO11n的可见光单模态基准模型出现火焰漏检, YOLO11n的热红外单模态基准模型出现烟雾漏检, 而EFFNet实现了完整的目标检测, 未出现漏检情况。

(3) 定位精度与置信度提升: 如图7(e)所示, 该图像同时包含火焰和烟雾目标, 但火焰被其燃烧产生的烟雾部分遮挡。在此场景下, EFFNet相比YOLO11n基准模型在目标置信度和边界框定位精度上均有明显提升。综合图7(a)至图7(e)的检测结果表明, EFFNet不仅在小型目标检测方面表现优异, 在中型、大型目标的检测任务中也实现了精度提升, 体现出良好的多尺度特征提取能力。

5 结论

早期森林火灾检测是一项重要且极具挑战性的任务。近年来, 基于无人机搭载图像传感器并结合深度学习的森林火灾检测方法, 因其成本低、效率高的优势而受到广泛关注。然而, 早期火灾目标尺度小、环境背景复杂以及无人机机载计算资源受限等问题, 使得现有方法的检测性能仍然十分有限。针对这些问题, 本文提出了一种轻量高效的早期森林火灾检测模型——EFFNet。

该模型针对无人机遥感成像特性及森林火灾目标特点, 对输入层、骨干网络、颈部和检测头4个核心模块进行改进: 针对可见光与热红外的互补特性, 采用EF策略实现温度特征与纹理特征的协同提取; 针对火焰和烟雾的多尺度及扩散边界特性, 设计HGRNet骨干网络以扩大感受野并保留浅层关键信息; 针对早期火灾目标尺寸小的特点, 引入FAN特征融合网络以提升小型目标的定位能力; 针对无人机平台计算资源受限的问题, 设计ESCHHead降低计算复杂度与误报率。

在RGBT-3M数据集上的实验结果表明, EFFNet的mAP50达到97.2%, mAP50—95达到68.3%, P 达94.9%, R 达94.1%, 模型参数量仅为1.84 M, 计算量仅为5.6 GFLOPs, 推理速度达到362.72对图像/s,

能够有效满足无人机遥感平台对早期森林火灾实时监测的应用需求。

未来工作将致力于基于EFFNet开发集成系统, 实现森林火灾的实时监测、精准评估与资源高效调度, 从而提升快速应急响应能力。在全球野火频率和强度不断上升的背景下, 这些技术对于实现快速有效的干预、减轻火灾影响具有重要意义。

参考文献(References)

- Bhamra J K, Ramaprasad S A, Baldota S, Luna S, Zen E, Ramachandra R, Kim H, Schmidt C, Arends C, Block J, Perez I, Crawl D, Altintas I, Cottrell G W and Nguyen M H. 2023. Multimodal wildland fire smoke detection. *Remote Sensing*, 15(11): 2790 [DOI: 10.3390/rs15112790]
- Bouguettaya A, Zarzour H, Taberkit A M and Kechida A. 2022. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Processing*, 190: 108309 [DOI: 10.1016/j.sigpro.2021.108309]
- Bu F J and Gharajeh M S. 2019. Intelligent and vision-based fire detection systems: a survey. *Image and Vision Computing*, 91: 103803 [DOI: 10.1016/j.imavis.2019.08.007]
- Buslaev A, Iglovikov V I, Khvedchenya E, Parinov A, Druzhinin M and Kalinin A A. 2020. Albumentations: fast and flexible image augmentations. *Information*, 11(2): 125 [DOI: 10.3390/info11020125]
- Cai Z W and Vasconcelos N. 2021. Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5): 1483-1498 [DOI: 10.1109/TPAMI.2019.2956516]
- Cao B, Sun Y M, Zhu P F and Hu Q H. 2023. Multi-modal gated mixture of local-to-global experts for dynamic image fusion//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE: 23498-23507 [DOI: 10.1109/ICCV51070.2023.02153]
- Coogan S C P, Robinne F N, Jain P and Flannigan M D. 2019. Scientists' warning on wildfire—a Canadian perspective. *Canadian Journal of Forest Research*, 49(9): 1015-1023 [DOI: 10.1139/cjfr-2019-0094]
- Deng C F, Wang M M, Liu L, Liu Y and Jiang Y L. 2022. Extended feature pyramid network for small object detection. *IEEE Transactions on Multimedia*, 24: 1968-1979 [DOI: 10.1109/TMM.2021.3074273]
- Dos Santos S M B, Bento-Gonçalves A and Vieira A. 2021. Research on wildfires and remote sensing in the last three decades: a bibliometric analysis. *Forests*, 12(5): 604 [DOI: 10.3390/f12050604]
- Feng C J, Zhong Y J, Gao Y, Scott M R and Huang W L. 2021. TOOD: task-aligned one-stage object detection//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 3490-3499 [DOI: 10.1109/ICCV48922.2021.00349]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- He X, Tang C, Zou X and Zhang W. 2023. Multispectral object detection via cross-modal conflict-aware learning//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: Association for Computing Machinery: 1465-1474 [DOI: 10.1145/3581783.3612651]
- Honary R, Shelton J and Kavehpour P. 2025. A review of technologies for the early detection of wildfires. *ASME Open Journal of Engineering*, 4: 040803 [DOI: 10.1115/1.4067645]
- Hu Y W, Zhan J L, Zhou G X, Chen A B, Cai W W, Guo K, Hu Y H and Li L J. 2022. Fast forest fire smoke detection using MVMNet. *Knowledge-Based Systems*, 241: 108219 [DOI: 10.1016/j.knosys.2022.108219]
- Jing R D, Zhang W, Li Y Z, Li W L and Liu Y Y. 2024. Feature aggregation network for small object detection. *Expert Systems with Applications*, 255: 124686 [DOI: 10.1016/j.eswa.2024.124686]
- Li R, Hu Y W, Li L, Guan R X, Yang R L, Zhan J L, Cai W W, Wang Y F, Xu H W and Li L J. 2024. SMWE-GFPNet: a high-precision and robust method for forest fire smoke detection. *Knowledge-Based Systems*, 289: 111528 [DOI: 10.1016/j.knosys.2024.111528]
- Liu H L, Fang Q, Jiang Y, Wei H Z, Wang T and Zhang S C. 2025. A lightweight forest fire detection algorithm based on YOLOv5s. *China Safety Science Journal*, 35(1): 75-83 (刘惠临, 方琼, 江宇, 魏华章, 王涛, 张树川. 2025. 基于YOLOv5s的轻量化森林火灾探测算法. *中国安全科学学报*, 35(1): 75-83) [DOI: 10.16265/j.cnki.issn1003-3033.2025.01.0127]
- Lyu C Q, Zhang W W, Huang H A, Zhou Y, Wang Y D, Liu Y Y, Zhang S L and Chen K. 2022. RTMDet: an empirical study of designing real-time object detectors. *arXiv Preprint arXiv*: 2212.07784 [DOI: 10.48550/arXiv.2212.07784]
- Mohapatra A and Trinh T. 2022. Early wildfire detection technologies in practice—a review. *Sustainability*, 14(19): 12270 [DOI: 10.3390/su141912270]
- Padilla R, Netto S L and da Silva E A B. 2020. A survey on performance metrics for object-detection algorithms//Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). Niteroi: IEEE: 237-242 [DOI: 10.1109/IWSSIP48289.2020.9145130]
- Phan T C, Nguyen T T, Hoang T D, Nguyen Q V H and Jo J. 2020. Multi-scale bushfire detection from multi-modal streams of remote sensing data. *IEEE Access*, 8: 228496-228513 [DOI: 10.1109/ACCESS.2020.3046649]
- Pizer S M, Johnston R E, Erickson J P, Yankaskas B C and Muller K E. 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness//Proceedings of the First Conference on Visualization in Biomedical Computing. Atlanta: IEEE: 337-345 [DOI: 10.1109/VBC.1990.109340]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]

- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Sheng D L, Deng J L and Xiang J W. 2021. Automatic smoke detection based on SLIC-DBSCAN enhanced convolutional neural network. *IEEE Access*, 9: 63933-63942 [DOI: 10.1109/ACCESS.2021.3075731]
- Sudhakar S, Vijayakumar V, Kumar C S, Priya V, Ravi L and Subramaniaswamy V. 2020. Unmanned aerial vehicle (UAV) based forest fire detection and monitoring for reducing false alarms in forest-fires. *Computer Communications*, 149: 1-16 [DOI: 10.1016/j.comcom.2019.10.007]
- Titu M F S, Pavel M A, Michael G K O, Babar H, Aman U and Khan R. 2024. Real-time fire detection: integrating lightweight deep learning models on drones with edge computing. *Drones*, 8(9): 483 [DOI: 10.3390/drones8090483]
- Wang G B, Li H Y, Xiao Q, Yu P F, Ding Z S, Wang Z S and Xie S D. 2025. Fighting against forest fire: a lightweight real-time detection approach for forest fire based on synthetic images. *Expert Systems with Applications*, 262: 125620 [DOI: 10.1016/j.eswa.2024.125620]
- Wu T Y, Tang S, Zhang R, Cao J and Zhang Y D. 2021. CGNet: a lightweight context guided network for semantic segmentation. *IEEE Transactions on Image Processing*, 30: 1169-1179 [DOI: 10.1109/TIP.2020.3042065]
- Rui X, Li Z Q, Zhang X Y, Li Z Y and Song W G. 2023. A RGB-Thermal based adaptive modality learning network for day-night wildfire identification. *International Journal of Applied Earth Observation and Geoinformation*, 125: 103554 [DOI: 10.1016/j.jag.2023.103554]
- Yang D Q and Wang W L. 2025. A few-shot target detection method for wildfires: research example of Fujian Province, China. *Journal of Intelligent & Fuzzy Systems*, 49(2): 553-563 [DOI: 10.3233/JIFS-240531]
- Zhan J L, Hu Y W, Zhou G X, Wang Y F, Cai W W and Li L J. 2022. A high-precision forest fire smoke detection approach based on ARGNet. *Computers and Electronics in Agriculture*, 196: 106874 [DOI: 10.1016/j.compag.2022.106874]
- Zhang S L, Wang X J, Wang J Q, Pang J M, Lyu C Q and Zhang W W. 2023. Dense distinct query for end-to-end object detection//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE: 7329-7338 [DOI: 10.1109/CVPR52729.2023.00708]
- Zhang Y L, Rui X and Song W G. 2025. A UAV-based multi-scenario RGB-thermal dataset and fusion model for enhanced forest fire detection. *Remote Sensing*, 17(15): 2593 [DOI: 10.3390/rs17152593]
- Zhao T Y, Yuan M X, Jiang F, Wang N and Wei X X. 2026. Removal then selection: a coarse-to-fine fusion perspective for RGB-infrared object detection. *arXiv Preprint arXiv: 2401.10731*. *IEEE Transactions on Intelligent Transportation Systems*, 27(2): 2504-2519 [DOI: 10.1109/TITS.2025.3638627]
- Zhao Y A, Lv W Y, Xu S L, Wei J M, Wang G Z, Dang Q Q, Liu Y and Chen J. 2024. DETRs beat YOLOs on real-time object detection//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]

A deep learning-based early forest fire detection method integrating visible light and thermal infrared images from unmanned aerial vehicles

YANG Dianqing, WU Junjie

School of Economics and Management, Fuzhou University, Fuzhou 350108, China

Abstract: Forest fires pose serious threats to ecosystems, human life, and property. Detecting fires during the incipient combustion stage is essential for rapid suppression and for preventing escalation into catastrophic events. UAV-based remote sensing provides flexible deployment, broad spatial coverage, and high spatial resolution, making it highly suitable for forest fire monitoring. However, early fire detection remains challenging because (1) incipient flames and smoke are small and easily confused with complex forest backgrounds; (2) dense vegetation can occlude targets; (3) illumination changes and environmental variability produce blurred boundaries and unstable target morphologies; and (4) limited onboard computing resources restrict the use of complex detection models. This study develops a lightweight and efficient detection model that integrates RGB and thermal infrared imagery to enable accurate real-time detection of early-stage flames and smoke on resource-constrained UAV platforms.

We propose EFFNet (early forest fire net), a dual-modal deep learning model for simultaneous flame and smoke detection. The architecture contains four components designed for UAV remote sensing and the visual characteristics of early fires. (1) An EF (early fusion) module at the input layer combines RGB texture cues with thermal infrared temperature patterns, using their complementary information to improve robustness under nighttime conditions, smoke interference, and occlusion. (2) A HGRNet (heterogeneous global routing network) is used as the backbone; it applies parallel convolutions with different kernel sizes to enlarge receptive fields and capture multi-scale diffusion patterns of flames and smoke, while local max-pooling preserves fine-grained shallow features required for small-target detection. (3) A FAN (Feature Aggregation Network) constructs dual top-down fusion paths and adds a high-resolution detection branch to enhance feature

representation and improve localization accuracy for small targets. (4) An ESCHead (efficient semi-coupled detection head) adopts a semi-decoupled design with shared feature extraction and 1×1 convolution-based dimensionality reduction, reducing computational cost while improving alignment between classification and localization branches and thereby decreasing false alarms.

Comprehensive experiments on the RGBT-3M dataset validate the effectiveness of EFFNet. The model achieves mAP50 of 97.2%, mAP50—95 of 68.3%, precision of 94.9%, and recall of 94.1%. It contains only 1.84 M parameters and requires 5.6 GFLOPs of computation. Real-time evaluation yields an inference speed of 362.72 frame pairs per second and a single-pair processing time of 2.757 ms. Visualization results demonstrate that EFFNet accurately detects small-scale flame and smoke targets in complex scenarios, including vegetation occlusion, nighttime imaging, and blurred object boundaries, while maintaining low false-positive and false-negative rates.

Overall, EFFNet achieves a favorable balance among detection accuracy, computational efficiency, and real-time performance, satisfying the operational requirements of UAV platforms for early forest fire monitoring. Its lightweight architecture supports practical deployment on resource-limited edge computing devices, and its high precision and reliability provide technical support for early warning and rapid response in forest fire management systems.

Key words: UAV, remote sensing image, deep learning, object detection, early forest fire, object detection process, dataset, performance evaluation metrics

Supported by Later-stage Funding Project of National Social Science Fund (No. 25FGLB019); General Program of the Natural Science Foundation of Fujian Province (No. 2025J01451, 2025J011137)