

# 基于长程依赖建模与动态特征融合的森林火灾检测<sup>①</sup>



李建东<sup>1,2</sup>, 高兴淇<sup>1</sup>

<sup>1</sup>(辽宁工程技术大学 软件学院, 葫芦岛 125105)

<sup>2</sup>(辽宁工程技术大学 矿业学院, 阜新 123000)

通信作者: 高兴淇, E-mail: [gao\\_xingqi@163.com](mailto:gao_xingqi@163.com)

**摘要:** 针对森林场景下火灾目标的特征衰减、背景干扰和实时性瓶颈问题, 提出基于长程依赖建模与动态特征融合的森林火灾检测算法. 首先, 通过融合条带池化模块 Strip Pooling 与金字塔池化的多尺度感知能力, 构建长程-局部双模态特征增强机制, 强化目标特征的全局形态表征与局部细节提取; 其次, 设计空间增强注意力检测头 Detect-SEAM, 通过通道-空间双重注意力协同机制抑制背景噪声干扰, 增强遮挡目标的空间特征响应; 最后, 在颈部网络引入动态上采样算子 DySample, 基于输入特征自适应性调整采样策略, 减少特征信息损失并平衡检测精度与实时性. 实验结果表明: 改进模型在森林火灾数据集 Wildfire 上的 *mAP* 值达到 86.5%, 提升 3.7%, 精度达到 85.1%, 提升 2.2%, 召回率达到 78.2%, 提升 3.4%, 推理速度达到 68.86 f/s. 该模型实现了检测精度与推理效率的协同优化, 为森林火灾检测提供有效解决方案.

**关键词:** 森林火灾; 火灾检测; 计算机视觉; 图像识别; 模型优化

引用格式: 李建东, 高兴淇. 基于长程依赖建模与动态特征融合的森林火灾检测. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10134.html>

## Forest Fire Detection Based on Long-range Dependency Modeling and Dynamic Feature Fusion

LI Jian-Dong<sup>1,2</sup>, GAO Xing-Qi<sup>1</sup>

<sup>1</sup>(College of Software, Liaoning Technical University, Huludao 125105, China)

<sup>2</sup>(College of Mining, Liaoning Technical University, Fuxin 123000, China)

**Abstract:** To address the challenges of feature attenuation, background interference, and real-time bottlenecks in forest fire detection, this study proposes a novel algorithm based on long-range dependency modeling and dynamic feature fusion. First, a long-range-local dual-mode feature enhancement mechanism is constructed by integrating the multi-scale perception capabilities of the strip pooling and pyramid pooling, which strengthens the global morphological representation and local detail extraction of target features. Second, a spatially enhanced attention head (Detect-SEAM) is designed to suppress background noise through a channel-spatial dual-attention coordination mechanism, thereby enhancing the spatial feature response of occluded targets. Finally, the dynamic upsampling operator (DySample) is introduced into the neck network to adaptively adjust the sampling strategy based on input features, reducing feature information loss while balancing detection accuracy with real-time performance. Experimental results on the Wildfire dataset show that the improved model reaches 86.5% *mAP* with a 3.7% increase, 85.1% precision with a 2.2% increase, 78.2% recall with a 3.4% increase, and an inference speed of 68.86 f/s. The proposed model realizes a synergistic optimization of detection accuracy and inference efficiency, offering an effective solution for forest fire detection.

① 基金项目: 辽宁省教育厅基本科研项目 (JYTMS20230804); 辽宁工程技术大学学科创新团队 (LNTU20TD-23)

收稿时间: 2025-09-24; 修改时间: 2025-10-14, 2025-11-11; 采用时间: 2025-11-24; csa 在线出版时间: 2026-03-09

**Key words:** forest fire; fire detection; computer vision; image recognition; model optimization

近年来,由于全球气候变化、森林资源过度开发以及人为疏忽等因素,森林火灾发生频率和严重程度显著增加。据联合国粮农组织(FAO)发布的《2024年世界森林状况》报告统计,在2021年,北方因干旱导致林火强度突破历史记录,碳排放量占全球林火排放总量的近1/4;2023年,全球林火导致的二氧化碳排放量约66.87兆吨;仅北方林火就占了全球林火二氧化碳排放量的10%。这不仅造成了大量森林资源和生物多样性的损失,也对人类社会的经济发展和生命安全构成严重威胁。因此,如何快速、准确地检测森林火灾并采取有效的防控措施,已成为亟待解决的重要问题。

早期的森林火灾检测主要依赖于烟雾探测器或温度传感器<sup>[1]</sup>,此类设备覆盖范围有限,受地形和天气条件的影响较大,且成本较高。通常需要提取手工设计的特征<sup>[2]</sup>,如果特征选取不当,火点的检测效果将显著下降。为了克服这些局限性,经典机器学习算法开始被应用于目标检测领域。杨海川等<sup>[3]</sup>提出一种基于支持向量机(support vector machine, SVM)的森林火灾检测框架,通过对森林遥感图像的像素点进行分析来提高火灾识别准确率。Wang等<sup>[4]</sup>在第1级决策层采用朴素贝叶斯算法和D-S证据理论算法,并在第2级决策层引入模糊综合评价法,最终利用百分比评级值对当前环境状态进行判定,提高了火灾预测的准确性。然而,机器学习算法也存在一些不足,如计算复杂度较高,导致在大规模数据集上训练和推理速度较慢以及目标出现漏检或误检等问题。

随着深度学习技术的快速发展,基于深度学习的火灾检测技术凭借其强大的特征提取和模式识别能力,已经成为该领域的研究热点。其中,YOLO系列算法<sup>[5]</sup>在目标检测任务中表现突出,经过不断迭代优化,在火灾检测领域取得显著成效。研究者通过引入注意力模块、优化模型结构、嵌入可分离卷积与空洞卷积、改进损失函数等方式实现了检测能力的提升。马跃等<sup>[6]</sup>提出RF-YOLOv8s林火检测算法,引入RFCBAMConv卷积模块,实现模型轻量化的同时提升了检测精度。杨国为等<sup>[7]</sup>采用EDBAN模块替代YOLOv8中的C2f模块,并提出LOTT检测模块。Park等<sup>[8]</sup>在YOLOv3的骨干网络中引入弹性卷积(elastic convolution)模块,通过动态调整卷积核感受野增强特征表征能力,实现对复

杂场景下火灾信息的有效探测。李宝民等<sup>[9]</sup>提出基于CenterNet2的林火检测算法,将VoVNetV2与非对称卷积核结合,并引入注意力机制eSE进行特征融合。张立国等<sup>[10]</sup>设计了GC-C3模块替换C3模块,采用Focal-EIoU损失函数并修改网络空间金字塔池化结构,减少了模型参数量和计算量。张彬彬等<sup>[11]</sup>提出改进的YOLOv3火焰目标检测算法,采用完全交并比和Focal loss作为损失函数,满足火焰检测的实时性需求。Rahmi等<sup>[12]</sup>通过SGLI的可见光-短波红外波段提取烟雾光谱反射率特征,结合归一化燃烧指数增强火灾区域识别能力。Fralenko<sup>[13]</sup>补充并测试了实际的神经网络方法和算法,以改进利用无人机图像和视频流进行森林火灾早期检测的方法。Hoang等<sup>[14]</sup>提出了一种基于CNN的森林烟雾检测模型,该模型在提高检测精度的同时降低了计算负载。然而,这些改进策略在很大程度上受目标特征的影响,难以直接应用于火灾监测场景。

通过以上分析,现有火灾检测仍面临3项关键技术瓶颈。在空间依赖建模方面,现有检测方法多依赖于局部卷积操作,其有限的感受野难以捕捉火焰在蔓延过程中形成的细长形态。这种特征信息的缺失,是导致远距离小目标漏检的根本原因之一;同时,在森林环境中,火焰目标常被树木、地形严重遮挡。而一些主流注意力机制(如SE、CBAM)多为静态或局部响应,缺乏对遮挡区域的动态特征补偿能力,导致目标在遮挡情况下特征响应衰减;最后,在特征金字塔融合阶段,传统的上采样方法(如最近邻插值)采用固定采样核,难以适应火灾区域复杂的纹理变化,导致重建后的特征图细节丢失,限制检测精度的进一步提升。

然而在实际应用中,火灾检测系统不仅需要具备较高的检测精度,还需兼顾模型的实时性,以满足应急响应需求。然而,许多高精度模型计算复杂度较高,难以部署在资源受限的设备上,如无人机或边缘计算终端。因此,本研究聚焦以下3大核心问题。

(1) 目标特征衰减: 由于火焰面积较小( $<0.1 \text{ m}^2$ ),传统检测方法易漏检,导致火灾难以及时发现。

(2) 复杂背景干扰: 火焰特征易与环境中的高亮区域混淆,影响检测的稳定性和准确性。

(3) 目标特征损失与精度平衡: 在减少特征损失条

件下,兼顾检测精度,以满足实时监测需求.

### 1 模型与研究方法

#### 1.1 YOLO11 概述

YOLO11 是 Ultralytics 推出的集目标检测、语义分割、图像分类等任务于一体的神经网络,核心架构改进聚焦于主干网络、特征融合机制与检测头设计的协同优化.其核心改进包括主干网络中的 C3k2f 模块、颈部网络的 C2PSA 注意力机制以及解耦检测头结构. C3k2f 模块通过双卷积核堆叠以增强特征提取能力, C2PSA 注意力机制旨在强化跨通道的特征交互,而解耦检测头则将目标分类与边界框回归任务分离,以

提升检测精度.本文以 YOLO11 为基线模型,针对森林火灾检测任务进行模型改进.

#### 1.2 LDM-DFE 模型

针对基线模型在森林火灾检测的一系列问题,本研究提出 LDM-DFE 模型,通过特征协同优化机制进行改进创新.其网络结构如图 1 所示,基于基线模型主要有 3 个改进部分.

骨干网络:在基线模型的 C3k 与 Bottleneck 模块中嵌入了 Strip Pooling<sup>[15]</sup>,形成了 C3k\_SP 与 Bottleneck\_SP 模块(蓝色、红色框示意).该路径(红色箭头示意)负责提取融合了长程轴向上下文与局部细节的增强特征.

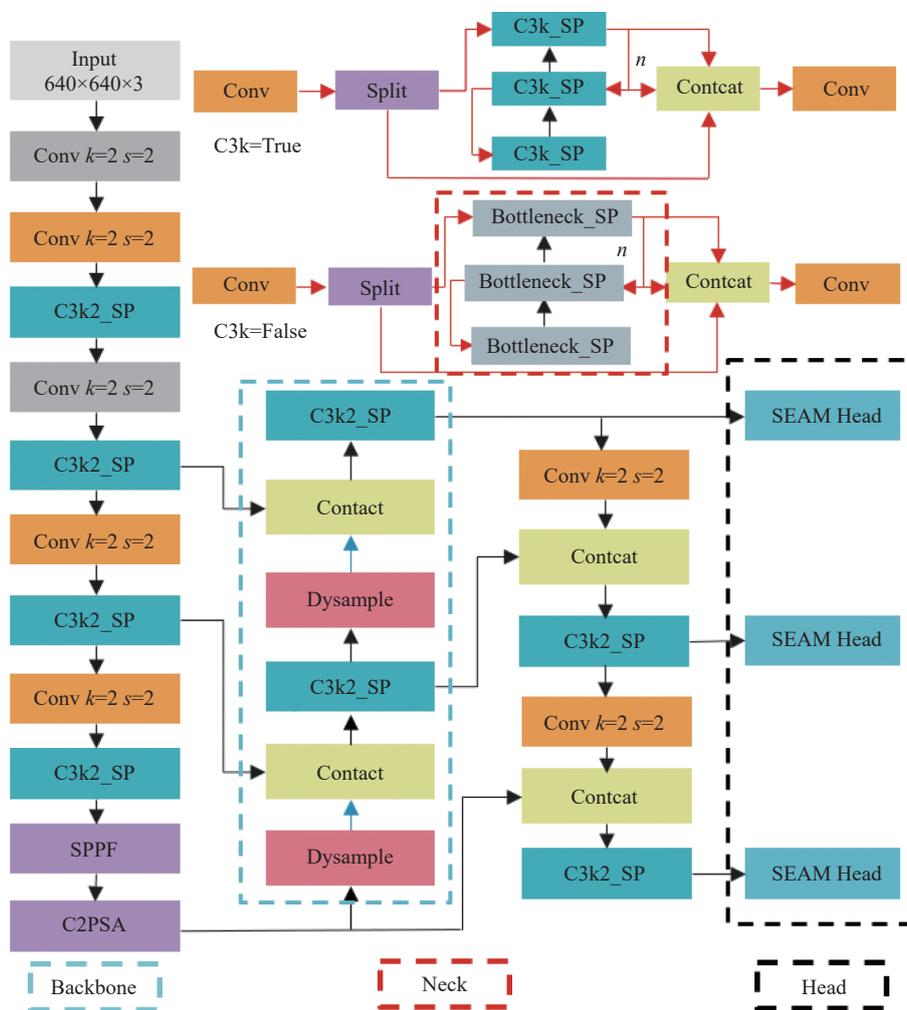


图 1 LDM-DFE 结构图

颈部网络:在特征金字塔部分,使用 DySample 动态上采样算子替代了传统的最近邻上采样(蓝色框示意).该操作(蓝色箭头示意)根据输入特征自适应地进

行分辨率重建,以减少细节丢失.

检测头:原检测头被替换为本文设计的 Detect-SEAM(黑色框线示意).主要接收来自颈部的多尺度特征,并

通过其内部的 SEAM 模块进行特征优选, 最终完成边界框预测与分类。

具体而言, Strip Pooling 模块输出的长程特征与局部细节特征在颈部网络中经由 DySample 进行动态上采样, 避免分辨率损失; Detect-SEAM 在此基础上对多尺度特征进行空间-通道双重筛选, 强化目标区域响应, 形成“特征增强-分辨率重建-特征优选”的闭环优化流程, 整体架构具备明确的层次化分工与信息流协同, 显著提升模型的检测性能。3 阶段分工如下。

#### (1) 特征增强阶段 (Backbone→Neck)

骨干网络中的 Strip Pooling 双路径模块对输入的特征图进行并行处理。其中, 局部卷积路径 (C3k) 保留火焰纹理等细节, 而长程池化路径 (C3k\_SP) 则生成表征烟雾扩散形态的轴向上下文特征。二者通过通道拼接进行自适应权重融合, 确保模型能同时关注局部细节与全局形态。

#### (2) 分辨率重建阶段 (Neck 内部)

经由骨干网络增强的特征在进入颈部网络进行多尺度融合时, 面临着上采样导致细节丢失的瓶颈。为此, 本阶段采用 DySample 动态上采样算子替代传统的固定核插值。DySample 能根据输入特征内容动态预测采样偏移量, 其优势在于能在火焰边缘等高频区域实施更精细的非均匀采样, 同时在平滑背景区域采用高效采样策略。

#### (3) 特征优选与检测阶段 (Neck→Head)

经过增强与重建后的特征图虽富含信息, 但仍存在背景干扰问题。为此, 设计 Detect-SEAM 模块改进检测头, 旨在对融合后的多尺度特征执行“特征优选”操作。它接收来自颈部网络的多尺度特征, 通过可变形卷积实现自适应空间聚焦, 并将生成的空间权重与 SEAM 注意力提供的通道权重进行融合。

### 1.3 Strip Pooling 空间池化模块

传统方形池化核 (大小为  $k \times k$ ) 在处理火焰、烟雾等细长目标时存在固有局限, 这一问题在火灾检测场景中尤为突出: 其一, 方形池化核的感受野  $R_{\text{square}} = k \times k$  局限在于其各向同性结构难以适配火焰沿垂直方向蔓延、烟雾沿水平方向扩散的轴向延伸特征, 无法有效捕捉目标的全局形态信息; 其二, 在检测图像边缘的细长目标时, 方形池化核易突破图像边界, 裹挟背景像素进入池化计算, 造成目标特征污染, 直接影响目标精确定位精度。因此, 针对这一核心问题, 结合火焰与烟雾

的轴向扩散特性, 本研究选用 Strip Pooling<sup>[15]</sup>构建长程依赖建模机制, 重新设计了基线模型中的 Bottleneck、C3k 和 C3k2 模块。

#### 1.3.1 相关方法对比与适配性分析

为定量与定性验证 Strip Pooling 在火焰、烟雾轴向特征建模中的优势, 将其与 Transformer、LSTM 等主流长程依赖建模方法进行系统对比, 核心差异聚焦于目标形态适配性、计算效率及特征表达能力: Transformer 通过自注意力机制计算全局特征依赖, 核心公式如式 (1) 所示:

$$A_{ij} = \text{Softmax}\left(\frac{Q_i K_j}{\sqrt{d}}\right) \quad (1)$$

其中,  $A_{ij}$  表示位置  $i$  和  $j$  之间的注意力权重,  $Q_i$  和  $K_j$  是查询 (Query) 和键 (Key) 向量,  $d$  是向量的维度。该方法虽能建模全局关联, 但计算复杂度高达  $O((HW)^2)$ , 且全局无差别的注意力分配模式, 难以聚焦火焰、烟雾沿特定方向扩散的轴向核心特征, 易受背景信息干扰。LSTM 需将二维特征图展平为一维序列, 通过单向递推更新隐藏状态如式 (2) 所示:

$$h_t = \text{LSTM}(x_t, h_{t-1}) \quad (2)$$

其中,  $x_t$  是当前时间步的输入,  $h_t$  是当前隐藏状态,  $h_{t-1}$  是前一时刻的隐藏状态。该方法受限于序列建模的一维特性, 无法有效捕捉目标在二维空间中的形态关联, 易导致特征响应断裂, 难以保持火焰、烟雾形态的连续性。Strip Pooling 采用  $H \times 1$  (垂直) 与  $1 \times W$  (水平) 非对称池化核, 其感受野为  $R_{\text{strip}} = H + W - 1$ 。在与方形池化保持相当计算复杂度的前提下, 该模块可沿轴向维度建立长程空间依赖关系, 同时在另一维度完整保留局部细节, 更适用于火焰细长形态与烟雾扩散路径的建模, 解决了传统方法在细长目标建模中的适配不足问题。

#### 1.3.2 Strip Pooling 结构

Strip Pooling 采用水平与垂直池化的双路径并行结构, 通过两个空间维度进行编码, 并融合对应的空间位置信息, 进行特征细化<sup>[16]</sup>, Strip Pooling 模块如图 2 所示。

首先, 输入特征图  $X \in R^{C \times H \times W}$  ( $C$  为通道数、 $H$  为高度、 $W$  为宽度), 经两路  $1 \times 1$  卷积降维至  $C/4$  通道, 分别得到水平池化分支输入  $X^h$  与垂直池化分支输入  $X^v$ , 维度均为  $C/4 \times H \times W$ 。接着, 采用平均池化策略分别聚合水平与垂直方向长程上下文信息, 适配烟雾与火焰的差异化扩散特性。

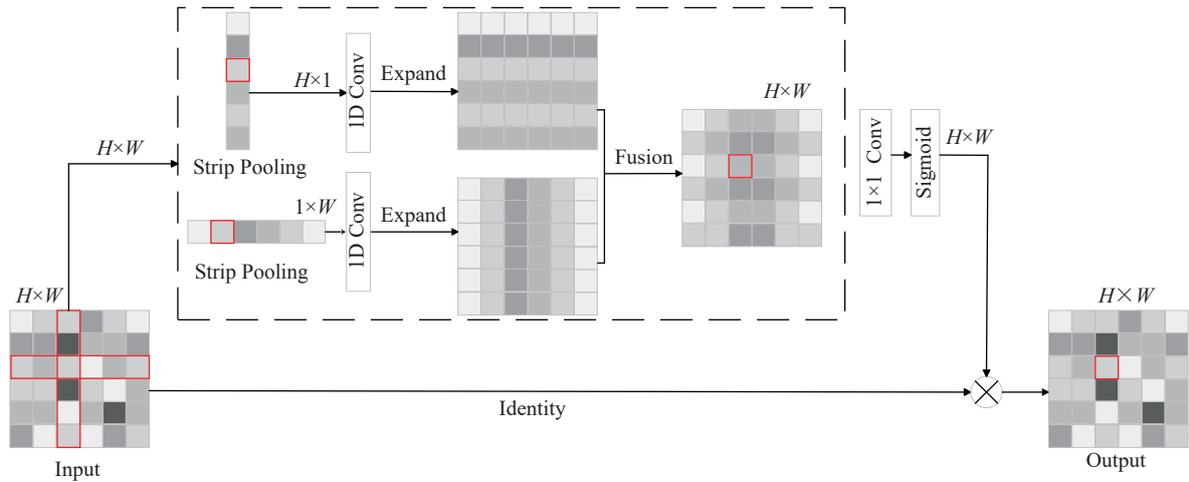


图2 Strip Pooling 结构图

水平条带池化 (适配烟雾水平扩散): 对  $X^h$  每行所有特征值求平均, 输出  $y^h \in R^H$ . 如式 (3) 所示:

$$y_i^h = \frac{1}{W} \sum_{0 \leq j < w} x_{i,j} \quad (3)$$

垂直条带池化 (适配火焰垂直蔓延) 后的输出  $y^v \in R^W$ . 如式 (4) 所示:

$$y_j^v = \frac{1}{H} \sum_{0 \leq i < h} x_{i,j} \quad (4)$$

为了融合双路信息以捕获更丰富的空间上下文, 将水平与垂直池化输出的特征图进行融合与变换. 具体实现为, 将  $Y^h$  和  $Y^v$  通过外积求和进行结合, 然后使用一个  $1 \times 1$  卷积层进行融合与降维, 再经过一个  $3 \times 3$  卷积层进行特征变换, 最后通过双线性上采样恢复分辨率. 这一融合过程的输出可表示为式 (5):

$$Z = Conv_{3 \times 3}(Conv_{1 \times 1}(Y^h \oplus Y^v)) \quad (5)$$

其中,  $\oplus$  表示矩阵的外积求和操作. 最终, 通过 Sigmoid 激活函数生成空间注意力权重, 并与原始输入特征进行加权融合, 即  $X_{out} = X \otimes \sigma(Z)$ , 从而增强对目标形态的建模能力.

### 1.3.3 基于 Strip Pooling 的基线模型重构

为充分发挥 Strip Pooling 的方向建模优势, 将其深度嵌入基线模型的核心模块, 完成 Bottleneck、C3k 及 C3k2 模块的针对性重构, 具体设计如下.

(1) Bottleneck\_SP 模块: 将 Strip Pooling 模块嵌入 Bottleneck 结构中, 替代传统标准卷积层, 得到 Bottleneck\_SP, 增强空间上下文建模能力, 如图 3 所示. 该模

块采用多方向长条池化与特征融合策略, 通过自适应平均池化实现 3 种池化操作: 全局池化、水平长条池化、垂直长条池化. 然后通过  $1 \times 1$  卷积降维、 $3 \times 3$  卷积特征变换及双线性上采样恢复分辨率, 最终将多分支特征融合以捕捉跨尺度上下文信息.

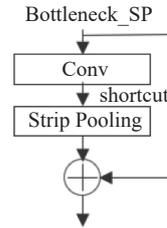


图3 Bottleneck\_SP 结构图

(2) C3k\_SP 模块: 基于 Bottleneck\_SP 模块对基线模型中的 C3k 模块进行完整重构, 将长程依赖建模能力与 C3k 的局部特征提取优势相结合, 强化模块对火焰、烟雾全局形态与局部细节的协同捕捉能力.

(3) C3k2\_SP 模块: 将 C3k\_SP 模块和 Bottleneck\_SP 以异构分支形式结合, 构建得到 C3k2\_SP 模块 (图 1). 该模块采用双路协同机制: 一路保留 C3k 的局部卷积操作, 通过小尺度卷积核聚焦火焰边缘、烟雾纹理等细粒度特征, 避免背景噪声干扰; 另一路通过 C3k\_SP 的长程池化策略, 捕获烟雾扩散路径的全局形态特征与空间分布规律, 解决传统方形池化核在细长目标形态中适配不足问题.

### 1.3.4 性能验证与融合机制优势

传统方形池化、Transformer 检测、LSTM 检测与 Strip Pooling 轴向池化检测结果对比如图 4 所示. 轴向

池化检测两类目标的平均精度达到 0.91, 较方形池化、Transformer 检测、LSTM 检测分别提高了 0.18、0.08、0.16。

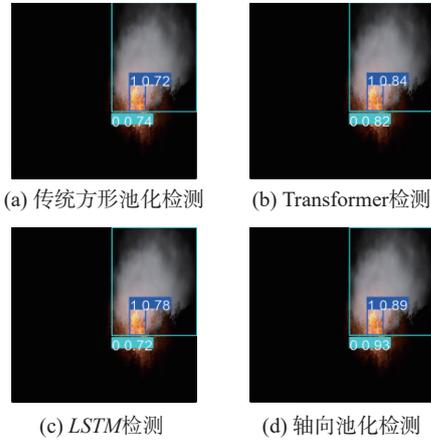


图 4 不同检测方法可视化对比

传统方形池化感受野有限, 仅激活目标局部区域, 无法覆盖细长目标完整形态, 且易受背景干扰. Strip Pooling 轴向池化特征响应连续且集中, 能精准对齐火焰垂直形态与烟雾水平扩散路径, 有效覆盖目标完整区域, 特征表达更具针对性. Strip Pooling 的核心优势在于方向性长程依赖建模: 通过分离空间维度的池化操作, 使模型在一个维度上具备长程上下文聚合能力, 同时在另一维度保留局部细节. 这一设计有效解决了传统方形池化核的局限性, 不仅扩大了感受野, 还提高了模型对目标形状和分布的适应性.

#### 1.4 Detect-SEAM 模块

在火灾场景中, 火焰的视觉特征常因树木遮挡导致特征响应弱化, 传统检测头在复杂遮挡场景下特征提取能力不足, 易出现漏检或误检情况, 针对这一技术瓶颈, 本研究提出面向火灾检测的多维特征强化架构, 对检测头进行了重新设计, 将 SEAM (spatially enhanced attention module)<sup>[17]</sup>注意力机制嵌入到传统检测头, 如图 5 所示, Detect-SEAM 检测头包括两个独立的分支, 分别负责边界框回归与目标分类. 每个分支都拥有独立的 SEAM 注意力模块, 这使得模型能够为定位与分类这两个不同任务, 自适应地学习并强化最相关的特征. 在每个分支内部, SEAM 模块放置在两个卷积层之间. 其流程为: 输入特征首先经过一个标准卷积层进行基础特征变换. 随后, 特征被送入 SEAM 模块. SEAM 通过其空间与通道注意力协同机制, 对特征进行动态

重加权, 显式增强被遮挡火焰区域的可见部分特征, 并抑制复杂背景干扰. 最后, 优化后的特征再经由一个逐深度卷积进一步提取特征, 最终输出用于分类或回归的预测结果. 有效增强被遮挡目标的特征表达.

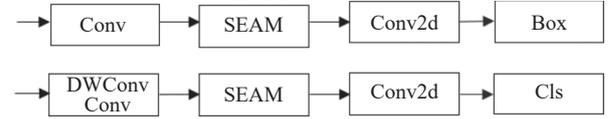


图 5 SEAM 检测头结构图

Detect-SEAM 的核心是其所实现的动态特征补偿机制, 该机制通过可变形卷积与通道重校准的协同工作, 具体实现如下: 可变形卷积通过在标准卷积核上附加一个偏移量学习网络, 使采样点能够根据输入特征动态调整位置. 给定卷积核采样位置  $R = \{(-1, -1), (-1, 0), \dots, (1, 1)\}$ , 对于输出特征图上的每个位置  $p_0$ , 可变形卷积的计算如式 (6) 所示:

$$Y(p_0) = \sum_{p_n \in R} W(p_n) X(p_0 + p_n + \Delta p_n) \quad (6)$$

其中,  $\Delta p_n$  是由一个轻量级网络预测的偏移量. 在遮挡场景下, 网络会学习将采样点  $\Delta p_n$  偏移至目标火焰的可见部分, 从而动态地绕开遮挡物, 实现对有效区域的聚焦.

在通道注意力路径中, 引入指数归一化, 通过全局平均池化与全连接层生成初始通道权重  $\delta \in [0, 1]^C$  ( $C$  为通道总数), 随后通过指数函数进行非线性映射, 如式 (7) 所示:

$$\hat{\delta}_c = \frac{e^{\lambda \delta_c}}{\sum_{j=1}^C e^{\lambda \delta_j}} \quad (7)$$

其中,  $\lambda$  为尺度因子. 该操作放大了关键通道与次要通道间的权重差异, 在特征因遮挡而整体减弱时, 确保模型能更依赖判别性强的特征通道.

上述空间与通道路径的输出通过一个特征重建模块进行融合. 该模块将空间路径提供的动态空间权重与通道路径提供的增强型通道权重相结合, 对原始特征进行重校准, 显著地补偿了因遮挡而丢失的特征响应. 经实验验证, 相较于通道注意力 SE<sup>[18]</sup>、混合注意力 CBAM<sup>[19]</sup>、SimAM<sup>[20]</sup>等注意力机制, SEAM 注意力机制能更精准地定位被遮挡目标的局部特征.

在火灾检测任务中, 遮挡问题会导致部分目标特征丢失, 从而影响模型的识别效果. Detect-SEAM 检测头能够增强 Neck 层输出后被遮挡区域的响应能力, 通

过通道注意力增强和空间注意力增强两种策略协同作用,提高模型对遮挡场景的适应性.通道注意力增强通过不同通道间的依赖关系,动态调整特征图的通道权重,并削弱冗余信息;空间注意力增强则侧重于分析特征图的局部和全局信息,提高模型对目标位置的感知能力,从而增强遮挡区域的特征响应.图6展示了火焰被烟雾遮挡场景下注意力机制可视化对比结果. SEAM 检测方法在两类目标上的平均检测精度均达到 0.9,显著高于基线注意力检测方法.

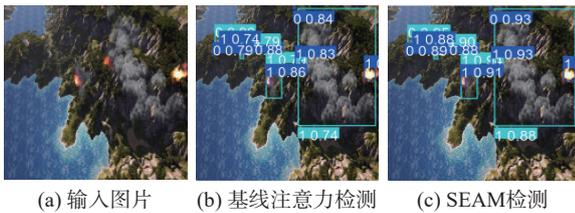


图6 遮挡场景下注意力机制可视化对比

SEAM 采用两层全连接网络来融合各通道的信息,从而增强通道间的信息流动性,如图7所示,采用指数归一化层通过全连接网络学习通道间的权重<sup>[21]</sup>,并使用指数函数将权重值从[0, 1]映射到[1, e],以增强通道特征的区分度,提高模型在目标位置偏移情况下的适应能力.

同时引入 Patch 嵌入机制,将输入图像或特征图划分为小块,并将每个小块转化为向量表示,便于局部特征的提取和全局信息的融合.同时,多尺度融合策略结合不同尺度的特征信息,利用 SiLU 激活函数捕获上下文信息,不仅增强了特征表达能力,还有效避免了梯度消失问题,使计算更高效.综合来看,SEAM 通过通道和空间注意力增强、遮挡区域特征补偿以及多尺度融合等策略<sup>[22]</sup>,提升了模型在被遮挡场景下的检测能力,使其更具泛化性和鲁棒性.

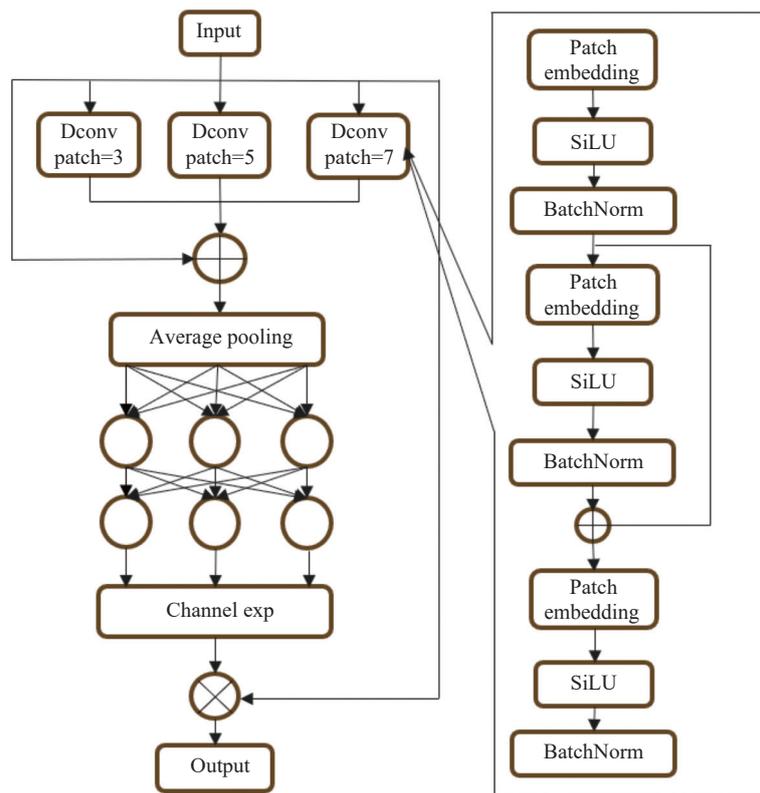


图7 SEAM 结构图

### 1.5 DySample 动态上采样算子

传统插值法在上采样过程中的固定采样策略可能会导致细节丢失和实时性差等问题,具体而言,固定核插值对所有区域采用统一的采样模式,难以适应图像

中复杂的纹理变化在高频细节区域因固定权重聚合导致的边缘模糊问题.此外,传统方法的上采样过程与输入特征无关,无法根据局部特征动态调整采样策略,导致上采样后的特征图出现语义不连贯或细节模糊,严

重影响多尺度特征融合精度。

针对上述问题, 本研究将 DySample<sup>[23]</sup>这种动态上采样方法引入到颈部网络中以替代最近邻插值上采样方法, 构建多尺度动态核预测模块, 实现自适应特征分辨率提升. 该模块的核心机制如下: 首先, 将输入特征图通过卷积层、特征增强层等处理后送入 DySample 模块. 然后, 在上采样过程中, DySample 会动态地根据输入的特征图分析不同区域的需求, 通过

插值方法将特征图的尺寸增大, 根据区域的特征不同调整插值参数或采样策略. 最后, 经过 DySample 上采样处理后, 特征图的分辨率会增加, 同时保留更多有用的特征信息.

DySample 动态上采样计算流程如图 8 所示, 点采样根据输入特征图张量  $X$  计算出点采样集  $\delta$ , 点采样集和输入特征图张量  $X$  一并传入上采样模块, 得到输出特征图张量  $X'$ .

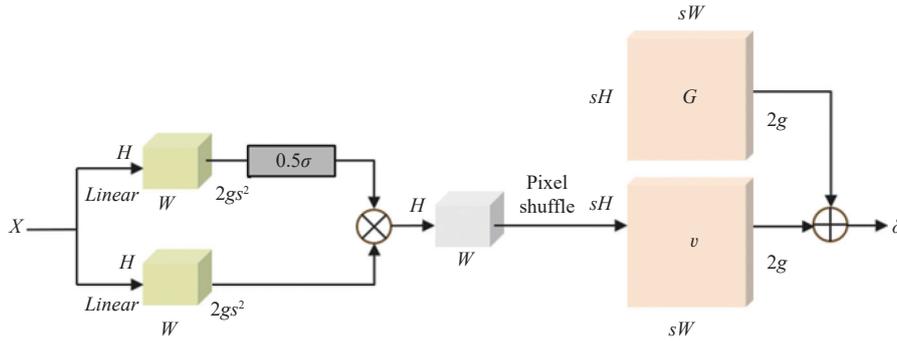


图 8 DySample 模块点采样集计算

给定由网络结构预设的上采样比例因子  $s$  (例如放大 2 倍或 4 倍) 以及大小为  $C \times H \times W$  的特征张量  $X$ , DySample 模块通过一个轻量的线性变换层来生成采样偏移量<sup>[24]</sup>. 具体而言, 该线性层的输入通道数为  $C$ , 输出通道数为  $2s^2$ , 由此生成大小为  $2s^2 \times H \times W$  的偏移量  $v$ . 计算公式如式 (8) 所示, 这一设计使得上采样过程能够根据输入特征内容自适应地确定最优的采样位置, 从而提升特征重建的精度.

$$v = \text{Linear}(X) \quad (8)$$

然后通过像素重塑得到  $2 \times sH \times sW$  的高分辨率特征  $G$ , 那么采样集  $\delta$  是偏移量  $v$  与高分辨率特征  $G$  之和, 计算如式 (9) 所示:

$$\delta = G + v \quad (9)$$

DySample 能自适应地为输入特征图设置偏移量来确定采样位置, 同时引入范围因子限制偏移范围<sup>[25]</sup>. DySample 计算流程如图 9 所示.

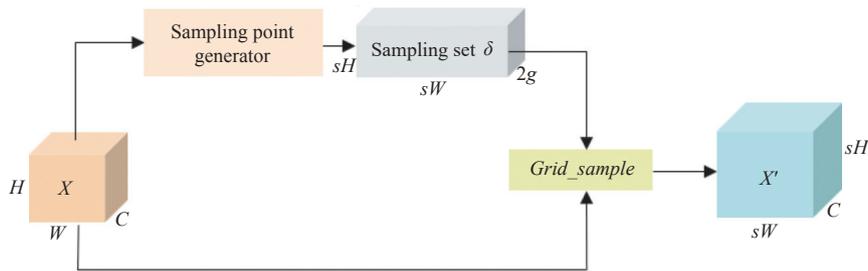


图 9 DySample 上采样模块图

首先传入特征张量  $X$  以及点采样计算得到的点采样集  $\delta$ , 然后  $\text{Grid\_sample}$  函数把形状为  $2g \times sH \times sW$  的点采集中每个位置的元素映射形状为  $2g \times H \times W$  的输入特征张量  $X$  上, 重采样到形状为  $C \times sH \times sW$  的特征张量  $X'$ , 计算公式如式 (10) 所示:

$$X' = \text{Grid\_sample}(X, \delta) \quad (10)$$

与传统插值法相比, DySample 动态地调整采样点的位置与权重分配策略: 在边缘密集的区域实行非均匀的精细采样, 以保留关键细节与结构完整性, 在背景或平滑区域则采用更为高效的采样策略, 优化计算资源分配. 有效缓解了固定插值核常导致的边缘模糊、细节丢失等问题, 在提升特征分辨率的同时, 避免了固

定插值核导致的边缘模糊问题,同时引入了计算量的自适应分配机制,兼顾了模型的实时性需求。

## 2 实验设计与结果分析

### 2.1 数据集

本研究基于 Wildfire 森林火灾检测图像数据集<sup>[26]</sup>进行研究实验,该数据集共 5 662 张图像,包含烟和火两类目标,覆盖多种场景(如山地、林地)、光照条件(白天、夜晚)及火焰阶段(初燃、蔓延),使该数据集能够有效增强模型在真实复杂环境中的泛化能力,提升火灾识别的准确性。数据集图像示例如图 10 所示。



图 10 数据集火灾图像示例

为验证本文方法的泛化性能,进一步采用 Forest Fire 数据集<sup>[27]</sup>进行实验。该数据集包括 3 304 张图像,依据烟雾的密度、颜色和扩散形态等细粒度特征,进一步划分为 5 个烟雾子类别,包含大量火焰与烟雾相互交织的图像,对模型在复杂视觉干扰下分离并识别不同目标提出了更高要求。

### 2.2 实验环境及参数配置

实验基本环境见表 1,模型在 GPU 上训练 200 轮,图像大小 640×640,批次大小为 16,初始学习率为 0.01,优化器为 SGD,模型的性能通过平均精度均值  $mAP$ 、精度  $P$ 、召回率  $R$  以及推理速度  $FPS$  作为评估指标。

表 1 实验环境配置表

参数	实验配置
操作系统	Linux
GPU	NVIDIA GeForce RTX 3090
显存	24 GB
CUDA	12.1
深度学习框架	PyTorch 2.3.0
开发工具	PyCharm
开发语言	Python 3.11

平均精度均值 ( $mAP$ ) 是每个类的平均精度,它衡量了模型对所有类别的整体检测性能。 $mAP$  的计算公式如式 (11) 所示:

$$mAP = \frac{\sum_{m=1}^M AP(m)}{M} \quad (11)$$

其中,  $M$  为检测到的类总数,  $AP(m)$  为模型在第  $m$  类上的  $AP$  值。

精度 ( $P$ ) 和召回率 ( $R$ ) 的计算分别如式 (12) 和式 (13) 所示:

$$P = \frac{TP}{TP+FP} \quad (12)$$

$$R = \frac{TP}{TP+FN} \quad (13)$$

其中,  $TP$  为正确预测为正例的样本数量,  $FP$  为错误预测为正例的样本数量,  $FN$  为错误预测为负例的样本数量。

推理速度 ( $FPS$ ) 为每秒检测和识别的图像数量,数值越大,则检测速度越快,计算公式如式 (14) 所示:

$$FPS = \frac{N}{T_{total}} \quad (14)$$

### 2.3 实验结果与分析

#### 2.3.1 消融实验

为了验证本文算法的有效性,在实验环境、实验参数和数据集均相同的条件下进行了多个模块组合的 8 组消融实验,以 YOLO11 为基线实验,在此基础上依次添入 Strip Pooling 空间池化模块、SEAM 注意力机制、DySample 动态上采样算子,实验结果见表 2。

表 2 消融实验结果

Baseline	Strip Pooling	SEAM	DySample	$mAP$ (%)	$P$ (%)	$R$ (%)	$FPS$ (f/s)
√	—	—	—	82.8	82.9	74.8	64.05
√	√	—	—	85.9	84.9	75.7	55.84
√	—	√	—	84.5	83.6	77.6	69.60
√	—	—	√	85.3	84.5	76.2	71.52
√	√	√	—	85.3	84.2	76.4	61.76
√	√	—	√	84.9	83.8	74.7	62.40
√	—	√	√	84.2	83.7	72.8	76.48
√	√	√	√	86.5	85.1	78.2	68.86

由表 2 可知,在独立基线模型测试中,模型的平均精度均值  $mAP$ 、精度  $P$  以及召回率  $R$  分别为 82.8%、82.9% 和 74.8%。通过模块化改进策略的逐项验证,各优化模块均展现出显著的性能优势。通过集成 Strip Pooling 空间池化模块,捕获了烟雾扩散的长程特征,有效增强了模型对长距离空间特征的提取能力,使  $mAP$  提升 3.1%,精度提升 2.0%;引入重新设计的空间

增强注意力检测头 Detect-SEAM, 使  $R$  达到 77.6%, 较基线模型提升了 2.8%; 采用 DySample 动态上采样算子替代传统插值方法, 不仅使  $mAP$  值、 $P$  值分别提升了 2.5%、1.6%, 而且使单模型达到 71.52 f/s 的最佳推理速度。

当全部改进均融入基线模型后, 综合性能显著提升: 平均精度均值  $mAP$  最终提升至 86.5%, 较基线模型提升了 3.7%, 精度提升了 2.2%, 召回率提升了 3.4%, 表明模型在保持较高准确性的同时显著增强了对正样本的识别能力, 构建了更优的精度-召回平衡. 推理速度提高了 4.81 f/s, 在显著提高检测精度的基础上仍满足实时处理需求. 为了更直观地反映改进后模型性能的提升, 图 11 展示了改进前后模型的平均精度均值对比情况。

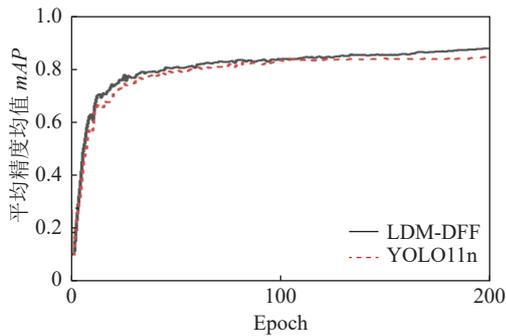


图 11 模型改进效果对比

### 2.3.2 不同注意力机制比较

为了验证本文注意力模块的有效性, 保持实验环境与参数不变, 进行通道注意机制 SE、混合注意力机制 CBAM、无参数注意力机制 SimAM 和 SEAM 注意力机制的对比研究. 结果见表 3, SEAM 在  $mAP$ 、 $P$ 、 $R$  以及  $FPS$  指标上均展现出明显优势. 结果表明, SEAM 注意力机制在增强特征表达能力与兼顾计算效率方面具备综合优势, 具有较强的实用性和先进性。

表 3 不同注意力机制比较

注意力机制	$mAP$ (%)	$P$ (%)	$R$ (%)	$FPS$ (f/s)
SE	82.9	82.1	74.2	60.02
CBAM	82.7	82.3	72.8	45.28
SimAM	82.2	81.5	73.3	35.84
SEAM	84.5	83.6	77.6	69.60

### 2.3.3 对比实验

为了验证本文算法的有效性, 保持实验环境和实验参数不变, 在此数据集上选取了多种代表性经典目

标检测算法作为对比基准. 所选算法涵盖了单阶段、两阶段及基于 Transformer 的多种先进架构, 包括轻量级网络 YOLOv7-tiny、YOLOv8n 与 YOLO11n, 两阶段检测器 Faster R-CNN, 单阶段检测器 SSD, 基于 Transformer 的 DETR 算法, 以及兼具效率与精度的 EfficientNetv2 算法, 实验结果见表 4。

表 4 对比实验结果

算法	$mAP$ (%)	$P$ (%)	$R$ (%)	$FPS$ (f/s)
YOLOv7-tiny	82.3	81.2	74.5	65.02
YOLOv8n	82.6	82.1	74.2	79.36
YOLO11n	82.8	82.9	74.8	64.05
Faster R-CNN	82.6	80.9	74.7	21.46
SSD	82.4	81.2	74.6	22.08
DETR	83.2	83.3	76.3	45.33
EfficientNetv2	83.8	83.7	76.8	31.05
LDM-DFP	86.5	85.1	78.2	68.86

表 4 列出了不同算法的  $mAP$ 、 $P$ 、 $R$  和  $FPS$  比较结果. 相较于其他算法, 本文算法在核心指标平均精度均值和召回率等均高于其他算法, 提升了模型的检测性能. 本文算法的平均精度均值达到了 86.5%, 高出基线模型 3.7 个百分点, 精度达到了 85.1%, 较基线模型提升了 2.2%, 召回率达到了 78.2%, 较基线模型提升了 3.4%, 推理速度  $FPS$  达到了 68.86 f/s。

为了更直观地显示本文算法性能, 图 12 展示了在 Wildfire 数据集中不同模型场景下火灾检测的视觉比较. 相较于基线模型, 本文提出的基于长程依赖建模与动态特征融合的 LDM-DFP 检测算法, 不仅在保持低误报率前提下实现了烟雾边缘特征的精准提取, 成功检测出与复杂背景混淆的烟雾, 提升了检测性能, 而且在对同一小目标检测时保持了更高的精度, 展现出了本文算法的优势所在。

### 2.3.4 泛化实验

为了验证本文算法的泛化性与适用性, 保持基本实验环境和实验参数不变, 选取 Forest Fire 森林火灾数据集进行泛化实验. Forest Fire 数据集包括 3304 张火灾烟雾图像, 可分为 5 种不同类别. 该数据集与本文研究范畴相同, 各算法在训练过程中均触发 Early Stopping 机制, 本文仅截取 150 轮数据作为训练结果. 实验结果见表 5, 在 Forest Fire 数据集上, 本文算法在关键指标上均取得明显提升: 平均精度均值  $mAP$  达到 44.3%, 相较于基线模型提升 1.8%; 检测精度达到 53.1%, 较基准模型 5.2% 的增幅尤为突出, 表明算法

在降低误报率方面取得重要进展;召回率 48.7% 比基线模型提升 1.1%, 验证了模型对真实火灾特征的捕

捉能力. 因此, 本文算法具有良好的泛化性, 对实际应用研究具有重要意义.

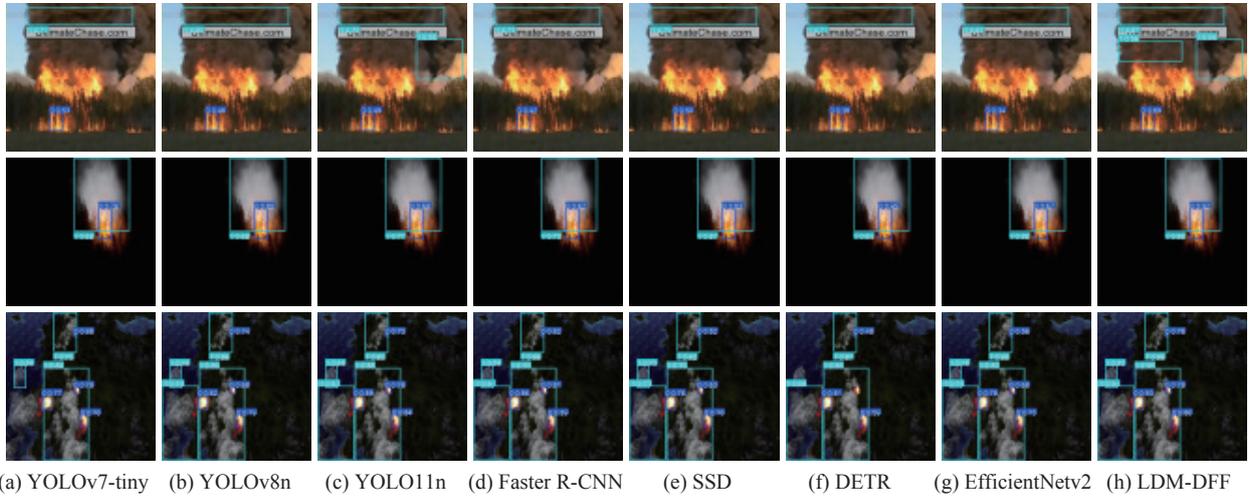


图 12 Wildfire 数据集上不同算法检测效果对比

表 5 泛化实验结果

算法	$mAP$ (%)	$P$ (%)	$R$ (%)	$FPS$ (f/s)
YOLOv7-tiny	42.1	39.7	46.7	65.02
YOLOv8n	40.8	39.4	47.7	79.36
YOLO11n	42.5	47.9	47.6	64.05
Faster R-CNN	42.1	41.2	45.6	21.46
SSD	42.4	40.8	46.6	22.08
DETR	43.2	47.6	46.9	45.33
EfficientNetv2	42.7	42.1	44.8	31.05
LDM-DFF	44.3	53.1	48.7	68.86

### 3 结论与展望

针对森林场景下火灾目标模糊、背景干扰及高精度检测需求, 提出了基于长程依赖建模与动态特征融合的森林火灾检测算法. 该算法构建了 Strip Pooling 双路径特征提取架构, 采用纵向条带池化与局部卷积的协同工作机制, 使  $mAP$  值达到 86.5%, 较基线模型提高了 3.7%; 重新设计了空间增强注意力检测头 Detect-SEAM, 通过特征通道重标定与空间位置聚焦机制实现特征优选, 有效增强烟雾、火焰等目标的显著性特征响应, 使  $R$  值达到 77.6%, 较基线模型提升了 2.8%; 在颈部网络引入动态上采样器 DySample, 替代传统插值方法的固定参数模式, 实现多尺度特征图的自适应融合与空间分辨率重建, 达到了 71.52 f/s 的推理速度.

然而, 模型在跨场景适应性和极端条件下的鲁棒性仍有提升空间, 需进一步优化以应对低光照、烟雾遮挡等复杂环境. 未来工作将聚焦轻量化网络架构与

多模态特征融合, 通过知识蒸馏与硬件感知量化压缩技术, 在进一步提高检测精度的同时降低计算量, 构建适用于机载边缘设备的森林火灾实时预警系统.

### 参考文献

- 王晶磊. 基于改进 YOLOv5 的火灾检测模型研究 [硕士学位论文]. 太原: 中北大学, 2023.
- 李泽琛, 李恒超, 胡文帅, 等. 多尺度注意力学习的 Faster R-CNN 口罩人脸检测模型. 西南交通大学学报, 2021, 56(5): 1002–1010. [doi: 10.3969/j.issn.0258-2724.20210017]
- 杨海川, 茹志鹏, 张诗雨. 基于无人机遥感图像的森林火灾检测. 宁夏大学学报 (自然科学版), 2022, 43(1): 17–21. [doi: 10.3969/j.issn.0253-2328.2022.01.005]
- Wang QJ, Lin JY, Zhou Y, *et al.* Indoor fire detection system based on multi-sensor information fusion algorithm. Proceedings of the 4th IEEE International Conference on Pattern Recognition and Machine Learning (PRML). Urumqi: IEEE, 2023. 637–641.
- Vijayakumar A, Vairavasundaram S. YOLO-based object detection models: A review and its applications. Multimedia Tools and Applications, 2024, 83(35): 83535–83574. [doi: 10.1007/s11042-024-18872-y]
- 马跃, 黄周睿, 周雯, 等. 基于感受野注意力的轻量化林火检测算法. 计算机工程, 2025, 51(9): 350–361. [doi: 10.19678/j.issn.1000-3428.0069454]
- 杨国为, 刘璇, 郜敏, 等. 基于多尺度特征融合的轻量级火灾检测算法. 计算机工程与应用, 2024, 60(23): 229–237. [doi: 10.3778/j.issn.1002-8331.2404-0262]

- 8 Park M, Ko BC. Two-step real-time night-time fire detection in an urban environment using static ELASTIC-YOLOv3 and temporal fire-tube. *Sensors*, 2020, 20(8): 2202. [doi: [10.3390/s20082202](https://doi.org/10.3390/s20082202)]
- 9 李宝民, 王小鹏, 孙茜容, 等. 基于概率两阶段 CenterNet2 的林火图像预警检测方法. *计算机工程与科学*, 2023, 45(10): 1884–1890. [doi: [10.3969/j.issn.1007-130X.2023.10.019](https://doi.org/10.3969/j.issn.1007-130X.2023.10.019)]
- 10 张立国, 张琦, 金梅, 等. 一种改进 YOLOv5s 的森林火灾烟雾检测算法. *计量学报*, 2024, 45(9): 1314–1323.
- 11 张彬彬, 帕孜来·马合木提. 基于 YOLOv3 改进的火焰目标检测算法. *激光与光电子学进展*, 2021, 58(24): 2411003. [doi: [10.3788/LOP202158.2411003](https://doi.org/10.3788/LOP202158.2411003)]
- 12 Rahmi KIN, Febrianti N, Prasasti I. Forest and land fire smoke detection using GCOM-C data (case study: Pulang Pisau, Central Kalimantan). *IOP Conference Series: Earth and Environmental Science*, 2021, 893: 012068. [doi: [10.1088/1755-1315/893/1/012068](https://doi.org/10.1088/1755-1315/893/1/012068)]
- 13 Fralenko PV. Neural network methods for detecting wild forest fires. *Scientific and Technical Information Processing*, 2025, 51(5): 497–505.
- 14 Hoang QQ, Hoang QL, Oh H. An efficient forest smoke detection approach using convolutional neural networks and attention mechanisms. *Journal of Imaging*, 2025, 11(2): 67. [doi: [10.3390/jimaging11020067](https://doi.org/10.3390/jimaging11020067)]
- 15 Hou QB, Zhang L, Cheng MM, *et al.* Strip pooling: Rethinking spatial pooling for scene parsing. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 4002–4011.
- 16 杨豪, 方承志, 罗佳辉, 等. 基于增大感受野的场景文本检测算法. *智能计算机与应用*, 2023, 13(8): 30–36. [doi: [10.3969/j.issn.2095-2163.2023.08.006](https://doi.org/10.3969/j.issn.2095-2163.2023.08.006)]
- 17 Yu ZP, Huang HB, Chen WJ, *et al.* YOLO-FaceV2: A scale and occlusion aware face detector. *arXiv:2208.02019*, 2022.
- 18 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 19 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision (ECCV 2018)*. Munich: Springer, 2018. 3–19.
- 20 Qin XY, Li N, Weng C, *et al.* Simple attention module based speaker verification with iterative noisy label detection. *Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2022)*. Singapore: IEEE, 2022. 6722–6726.
- 21 田青, 王颖, 张正, 等. 改进 YOLOv8n 的选通图像目标检测算法. *计算机工程与应用*, 2025, 61(2): 124–134.
- 22 郝荣荣, 马巧梅, 谭亚军, 等. 基于可变形卷积与注意力机制的复杂火灾检测. *计算机与现代化*, 2025(6): 79–85, 119.
- 23 Liu WZ, Lu H, Fu HT, *et al.* Learning to upsample by learning to sample. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 6004–6014.
- 24 高家源, 张新. DMH-YOLO: 针对屏幕外观缺陷检测与分割的高效算法. *计算机工程与应用*, 2025, 61(13): 151–164. [doi: [10.3778/j.issn.1002-8331.2411-0399](https://doi.org/10.3778/j.issn.1002-8331.2411-0399)]
- 25 王俊霖, 方睿, 刘金智, 等. 基于 RAD-YOLO 的织物疵点检测算法. *棉纺织技术*, 2025, 53(6): 58–65. [doi: [10.3969/j.issn.1001-7415.2025.06.009](https://doi.org/10.3969/j.issn.1001-7415.2025.06.009)]
- 26 Wildfire. <https://universe.roboflow.com/wildfire-q3rm1/wildfire-ajbuc/dataset/1>. [2025-02-15].
- 27 Forest fire. <https://universe.roboflow.com/object-detection-n6mxu/forest-fire-tyoo7/dataset/10>. [2025-02-15].

(校对责编: 张重毅)